**MARIUSZ BARCZAK**

# APPLICATION OF BIG DATA IN TOURISM DEMAND RESEARCH

Издателство на Българската академия на науките
"Проф. Марин Дринов"
Professor Marin Drinov Publishing House
of Bulgarian Academy of Sciences

Sofia 2025

**Reviewers:**

**Marcin Kęsy –** Professor, Doctor of Sciences, Pomeranian Higher School in Starograd Gdanski, Poland

**Igor Britchenko** – Professor, Doctor of Sciences, University of the National Education Commission in Kraków

**Maksym Bezpartochnyi -** Professor, Doctor of Sciences, State University of Intelligent Technologies and Telecommunications, Ukraine

© Copyright by Mariusz Barczak, Economics and finance
Mariusz Barczak - ORCID: 0000-0001-5270-6222

The Bulgarian Academy of Sciences (Българска академия на науките, БАН)
Sofia, Bulgaria 2025

**ISBN: 978-619-245-528-6**

*It is with immense gratitude that I dedicate this book to my wife.*
*Thank you for believing in me, even when I had doubts myself, and the countless moments you sacrificed to enable me to work on this project.*
*It is thanks to you and your presence that this book has become a reality*

Table of contents

INTRODUCTION ..... 4
CHAPTER I: CONCEPTUALISING THE FUNCTIONING OF ..... 9
THE TOURISM SECTOR ..... 9
1.1. Basic terminology used to analyse phenomena in the tourism sector ..... 9
1.2. Tourism as an economic sector ..... 17
1.3. Determinants of competitiveness of tourist regions (tourist reception areas) ..... 27
1.4. Trends in the tourism sector nationally and internationally ..... 36
CHAPTER II: TOURIST DEMAND ..... 47
2.1 Determinants of tourism demand ..... 47
2.2 Tourism demand determinants ..... 61
2.3 Tourist decision-making models ..... 67
2.4 Theoretical basis for market segmentation in tourism ..... 76
CHAPTER III: TECHNOLOGIES, METHODS AND TOOLS USED IN BIG DATA ANALYSIS ..... 85
3.1. The importance of big data in today's economy ..... 85
3.2. Impact of AI and machine learning on data analytics ..... 100
3.3. Technologies, methods, tools and sources of big data in the tourism sector ..... 109
3.4. Issues surrounding the use of big data (ethical and legal considerations) ..... 118
CHAPTER IV: OPPORTUNITIES TO ANALYSE THE TOURISM SECTOR ..... 128
USING BIG DATA ..... 128
CHAPTER V: APPLICATION OF BIG DATA IN ANALYZING TOURISM DEMAND RESEARCH IN THE KUJAWSKO-POMORSKIE VOIVODESHIP ..... 142
5.1. Diagnosis of internal and external factors affecting the tourism sector in the Kujawsko-Pomorskie Voivodeship ..... 142
5.2. Overview of tourism sector research in the Kujawsko-Pomorskie Voivodeship ..... 149
5.3. Exploring technologies, methods and tools for tourism demand research ..... 153
5.4. A working model for big data analytics in the tourism sector ..... 156
5.5. Challenges and future directions for the application of big data in the tourism sector 183
Completion ..... 189
Bibliography ..... 192

# INTRODUCTION

The interest in tourism demand research has important practical implications (Dwyer & Forsyth, 2010). Firstly, tourism demand is the foundation on which tourism-related business decisions are based. Given the key role of demand as a determinant of business profitability, estimates of expected future demand are a very important element in the planning of activities. Accurate forecasts of tourism demand are essential for effective planning of tourism-related activities, especially given the perishable nature of the tourism product. Secondly, investments in tourism, especially investments in tourism infrastructure such as airports, motorways and railways, require long-term financial commitments and the costs of the investments undertaken can be very high if the investment projects do not meet their design capacity. Therefore, forecasting long-term demand for tourism-related infrastructure is often an important part of project appraisal. Thirdly, the macro-economic policies of governments or the micro-economic policies of regions largely depend on the relative importance of the different sectors within the tourism reception area. Therefore, accurate forecasts of tourism sector demand help to formulate and implement appropriate tourism strategies in the medium to long term.

The challenge of determining the size and structure of tourism demand (especially in a statistical context) is the need to constantly seek new sources and methods of obtaining information on it (Saluveer & Raun, 2020). Therefore, the use of primary information becomes essential. In addition to empirical research (partial, based on survey techniques and primary data acquisition), which has played an important role in the Polish market in recent years in filling the information gap on tourism demand mechanisms, there are now increasing attempts in both academic and business circles to use information from the 'data cloud' (big data) generated by Internet and mobile device users. Developments in information and communication technology (ICT) and computing are creating such opportunities. Rapid advances in ICT technology, digitisation, widespread access to devices that generate and record data, and their connection to the global web, are creating vast information resources that offer unprecedented analytical opportunities (Wozniczka, 2018). The potential of big-data analysis (BDA), including for tourism demand analysis, has been recognised relatively recently, but its use is growing rapidly. One of the first areas of application of big data became marketing (including tourism), and the fast and financially measurable effects of data analysis were the driving factor behind this trend (Barczak, Brudnicki & Wierzbiańska, 2023).

Data on tourism traffic collected by public statistics cover only the registered accommodation base. Information characterising the structure of the participants in this traffic, due to shortcomings in public statistics, has to be obtained by means of partial surveys, which significantly affects their quality

due to the problem of representativeness. Other ancillary statistics, information and data from various institutions or facilities, as well as those obtained during events or surveys of tourist attractions, are fragmentary in nature and only provide a limited representation of the nature, scale and structure of the phenomenon under study. Among the specific objectives of these studies, the most frequently identified were (Szafrańska, Włodarczyk & Dziedzic, 2015):

- determining the volume of tourist traffic, i.e. the number of visitors to the region, broken down into domestic visitors (with a breakdown by voivodeship) and foreign visitors, as well as into day visitors and tourists;
- to determine the dynamics of change in the volume of tourist traffic in the region, based on data from previous years;
- to determine the structure of tourism participants by place of residence, age, gender, education, occupational and material status;
- creating a socio-demographic profile of the different categories of visitors, both domestic and international;
- to obtain characteristics of arrivals in the region in terms of the purpose of the visit, length of stay, persons accompanying and means of transport used;
- determine the average level of expenditure incurred during a stay in the region, converted per person and/or per day;
- identifying visitor behaviour, ways of spending their time, levels of satisfaction with the tourism offer and opinions on safety, cleanliness and value for money.

The group of tourism parameters most commonly recognised and described in regional tourism studies (including those conducted in Poland) includes (Matczak 1992, Pociecha 2002, Brudnicki 2006, Włodarczyk 2011, Brudnicki & Barczak 2017, Skalska 2020, Barczak, Brudnicki & Wierzbiańska, 2023)**:**

- size, seasonality and length of stay, frequency of arrivals, by which it can be determined how many tourists arrive in the region/city and when, how long they stay, whether there are periods of the year when the volume of traffic is significantly higher, what proportion of traffic is made up by returning tourists;
- the structure of the tourist flow, which determines who comes to the region/location (who the visitors are) - demographic characteristics (age, gender, family status), social characteristics (education, occupation, social background), nationality;

- spatial characteristics, showing both the geographical range of tourist traffic (defined by the place of origin of visitors) and the places and areas of concentration of traffic in the region (also by basic units - communes, districts);
- the method of organisation, indicating the form of the trip (individual, group), the organiser (independent, organised organiser) and the means of getting to the region/location, as well as the means of transport used;
- the purpose of arrival, allowing the predominant forms of tourism in the region/location under study to be identified;
- motives for choosing the region/reception area as a tourist destination, which make it possible to determine what factors determined the choice of the studied region;
- ways of spending time during a stay, making it possible to determine which elements of the tourist potential of a region/location are of greatest interest (e.g. by preparing a ranking of tourist attractions);
- expenditure during the stay (trip), allowing the estimation of income from tourism and tourism-related activities;
- opinions about the region or locality (tourist reception area), regarding the preparation of the space for tourism (assessment of tourist development elements and level of service, e.g. professionalism of tourist staff, hospitality of residents ftp.) and enabling appropriate measures to be taken to improve unsatisfactory elements and develop elements perceived positively.

The aim of this study is to present a model for utilizing Big Data in the process of forecasting tourism demand, taking into account methods of analyzing large data sets in the context of market trends and tourist behavior. This model is based on the application of web scraping techniques to collect data from a booking platform, enabling the identification of patterns and the prediction of future demand for tourism services.

The study conducted opens up the possibility of seeking answers to key questions regarding the application of modern analytical tools in tourism, including:

- How can Big Data obtained through web scraping be used to forecast tourist traffic in the area of tourist reception?
- How can an analysis of the availability of accommodation help to predict the future volume of tourism in an area?

- What are the projections for future hotel occupancy in the study area based on accommodation availability?
- How can forecasts of accommodation availability indicate changes in tourists' preferences in terms of location and type of facilities?

The answers to these questions not only allow for a better understanding of tourism market dynamics, but also provide valuable information for tourism managers, accommodation owners and local authorities, supporting them in making more informed strategic decisions.

In an effort to find answers to the research questions posed, the author developed a model called Tourism Demand Horizon, which serves as a reference for his empirical research. The model is based on a detailed analysis of tourism demand data and takes into account key factors affecting the dynamics of the tourism sector.

In addition, in order to deepen the analysis and expand the theoretical perspective, the author proposed Tourism Demand Horizon+, which is based on the theoretical assumptions of the new model. This concept combines existing analytical frameworks with new methodological approaches, enabling a better understanding of the mechanisms that shape tourism demand.

The introduction of both models allows for a complementary view of the problem under study - on the one hand through the analysis of empirical data, and on the other through the exploration of theoretical assumptions and their potential implications for future research and management practice in tourism.

The research undertaken and the related research questions represent an important step towards a better understanding of the analytical potential of modern technology in the tourism sector. The use of web scraping and advanced business intelligence tools such as Power BI opens up new possibilities for forecasting and managing tourism in tourism reception areas. The publication and the planned provision of an interactive dashboard allow for the practical application of these analyses and the further development of knowledge in this area.

The structure of the publication has been designed to enable the reader to gradually learn about particular aspects of the tourism sector. It is divided into five chapters that form a logical and coherent narrative. The first chapter introduces the reader to the basic concepts and terminology used in the analysis of tourism phenomena. It also discusses the importance of tourism as a sector of the economy, the determinants of competitiveness of tourist regions and key trends in tourism, both at national and international levels. The second chapter, 'Tourism demand', focuses on the analysis of factors influencing tourism demand. It presents the determinants, prospects for development, models of the

tourist decision-making process and the theoretical basis of market segmentation, which provide the foundation for effective demand management. The following third chapter, 'Technologies, methods and tools used in big data analysis', highlights the growing role of advanced technologies in data analysis. The chapter discusses the importance of big data, the impact of artificial intelligence and machine learning on analytics and the specifics of using these technologies in the tourism sector. Attention is also given to the ethical and legal challenges associated with their use. The fourth chapter, 'Opportunities for tourism sector analytics using big data', presents the analytical potential of big data technologies in the context of tourism. Particular emphasis is placed on the possibilities of integrating data from different sources and on the practical applications of big data analysis in tourism research. The fifth chapter, entitled "Applying big data to analyse tourism trends and segmentation of tourism demand in the Kujawsko-Pomorskie Voivodeship", takes the considerations to the regional level. The chapter includes an analysis of the factors influencing the tourism sector in the Kujawsko-Pomorskie Voivodeship, a review of research to date, and a proposed model for the use of big data in regional research. It also describes the challenges and future directions of advanced analytics applications in tourism. The whole is rounded off with a conclusion that summarises the most important findings from the analyses, as well as a bibliography and appendices that can serve as a source of inspiration and additional material for readers.

In the research part of this publication, the web scraping technique (the process of automatically extracting data from websites) was used to obtain data on the size of the accommodation base in the Kujawsko-Pomorskie voivodeship. The data was collected from the booking.com portal, which made it possible to obtain detailed information on the available accommodation facilities. The data collection process was conducted both in real time and by formulating queries regarding the availability of offers on future dates. This approach allowed not only the ongoing monitoring of the state of the accommodation facilities, but also the forecasting of tourist demand in the region, taking into account variables related to seasonality and planned travel. The research was conducted in November/December 2024. As a result, a database containing more than 40,000 records on 802 accommodation facilities in the Kujawsko-Pomorskie Voivodeship was created. The data made it possible to analyse the availability and structure of accommodation in the region.

# CHAPTER I:
# CONCEPTUALISING THE FUNCTIONING OF THE TOURISM SECTOR

## 1.1. Basic terminology used to analyse phenomena in the tourism sector

Tourism is an extremely complex phenomenon, not only socially but also economically. The term encompasses a wide range of activities related to travel, leisure and also cultural exchange. Understanding the basic terminology used in the analysis of this phenomenon is essential to properly study this sector.

Tourism is defined as an activity involving travel to destinations outside a fixed place of residence, for a period of no more than 12 months, mainly for leisure, work or health purposes. The concept is complex as it encompasses not only travel but also the entire system of services provided to tourists, including transport, accommodation, catering and other accompanying services (Patel, 2021).

Motivations for travel and the types of tourists have changed over the centuries. Prior to the 18th century, few people travelled for leisure - most trips were for trade, pilgrimage, religious or educational purposes. The demand for travel was mainly driven by economic requirements and obligations rather than the need for leisure. In ancient and medieval times, leisure travel was limited to short trips to fairs, festivals or other events. In the 18th century, when the 'Grand Tour' gained popularity among the young European elite, voyages of an educational or commercial nature continued to dominate. Exploratory expeditions to Africa, America and Australia were funded by governments and companies in order to develop commercial contacts. It was not until the 18th and 19th centuries that travel to satisfy personal needs, so-called 'final demand' tourism, began to play a greater role. Health trips, such as to Epsom in England, had existed before, but it was only at this time that they became popular among the aristocracy, and later the fashion was extended to sea bathing. Mass tourism only developed with industrial progress, when the development of transport - especially railways and steamships - made travel more accessible and cheaper, which increased interest in leisure tourism. Working in cities and the increased mobility of workers led to the separation of family members, generating the need for family gatherings. On a larger scale, colonisation and migration to the New World led to the development of long-distance VFR (visiting family and friends) tourism, which still plays an important role in international tourism today (Bull, 1998).

Travelling has a long history, but in the 20th century the growth of tourism was slowed down by two world wars, but since the 1950s tourism has taken on its modern form and the World Tourism

Organisation (UNWTO) has introduced statistical definitions of it (Fletcher, Fayall, Gilbert & Wanhill, 2018).

Generally, two categories of travel can be distinguished, i.e. inbound visitors (inbund tourism) and other inbound travellers (other inbound travellers). The first group includes tourists and day visitors. Tourists are those who arrive in a country for a short period of time and stay overnight, while day visitors are those who do not stay overnight. The main destinations, according to the UNWTO classification, are (Glossary of tourism terms, UNWTO, 2024):

1. Business and professional matters, i.e. travel related to professional duties, attending conferences, business meetings, training courses or trade fairs. Those travelling for this purpose are not primarily for recreation, but to fulfil professional tasks.
2. Holidays, leisure and recreation are trips for relaxation, rest, sightseeing and other recreational activities. It is the most common reason for tourist trips.
3. Visiting family and friends with the aim of meeting relatives, visiting friends or family members. This destination is particularly popular in regions with high population migration.
4. Education and training is travel related to learning, studying, taking courses or other forms of education. It refers to both short-term and long-term stays in another country or region.
5. Health and medical care generally involves the use of medical services, treatment, rehabilitation, surgery or specialised health procedures. This type of tourism includes both planned and urgent trips.
6. Religion and pilgrimage, i.e. journeys undertaken for religious reasons, such as attending pilgrimages, visiting religious sites or participating in religious ceremonies.
7. Other destinations include trips that do not fit into the above categories, such as shopping, cross-country transit or attending special events. This can include a variety of less popular trip motivations.

*Figure 1 . Destinations according to UNWTO*




Source: own development using DALL-E

Each of these objectives defines the motivation for the trip and influences the choice of destination, length of stay and the nature of the trip.

The second group, other inbound travellers, covers a broad spectrum of people who arrive in the country but not necessarily for tourism purposes. This group includes, among others: cross-border workers, seasonal workers, short-term workers, long-term workers, as well as nomads and refugees. In addition, this category includes transit passengers who do not enter the economic and legal territory of the country, public transport crews, people coming to settle in the country, as well as long-term students and patients and their families. The final subcategory is other travellers e.g. diplomats, military personnel and members of the armed forces on manoeuvres (Fletcher, Fayall, Gilbert & Wanhill, 2018).

The UNWTO distinguishes three main forms of tourism for each country. The first is domestic tourism, which includes the activities of residents within the borders of their country of residence, whether travelling domestically or abroad trips. This is the dominant form of tourism, accounting for about 80% of all tourism activity. The second form is outbound tourism, comprising residents' activities outside their country, whether the trip is domestic or international. The third category is inbound tourism, which refers to the activities of non-residents within the borders of a country during their inbound trip.

During the creation of the TSA (Tourism Satellite Account, 2008), the UNWTO decided that tourism should be recognised as a sector consisting of multiple industries, according to international standards. These industries have been listed to show what areas of activity are directly related to tourism (more extensively in chapter 1.2)

Tourism is a complex socio-economic phenomenon that encompasses a wide range of motivations and forms of travel. The diversity of tourism results not only from geographical and cultural conditions, but also from the needs and interests of travellers. Contemporary classifications of tourism are based on criteria such as purpose of travel, nature of participation, type of activity or expected level of involvement. Table 1 provides an overview of the main forms of tourism, grouped according to key tourism motivations. Each of these groups includes a variety of travel types that meet the specific needs and expectations of tourists. For example, cognitive tourism includes forms of travel aimed at gaining knowledge and experiences related to nature, culture or language. Leisure tourism, on the other hand, focuses on relaxation and entertainment, and health tourism focuses on improving physical and mental fitness.

An introduction to this classification provides a better understanding of the diverse needs of today's tourists and how different forms of tourism contribute to the economic and social development of tourist regions.

*Table 1 . Overview of types and forms of tourism*

| Motivation group | Form of tourism |
|---|---|
| Cognitive tourism | nature tourism, geotourism, ornithological tourism, polar tourism, tourist safari, ecotourism, cultural tourism, ethnic tourism, festival tourism, fanotourism, culinary tourism, sightseeing tourism, linguistic tourism, volunteer tourism |
| Leisure tourism | agrotourism, marine tourism, leisure tourism, commercial tourism, fun tourism, erotic tourism |
| Qualified tourism | hiking, cycling, equestrian tourism, skiing, speleological tourism, sailing, canoeing, diving, adventure tourism |
| Health tourism | spa tourism, spa and wellness tourism, medical tourism |
| Pilgrimage tourism | religious tourism |
| Business tourism | individual business travel, exhibitions, corporate tourism, congress tourism, incentive tourism |

Source: Kurek, W. (Ed.). (2008). *Turystyka [Tourism]*. PWN Scientific Publishers.

Tourist types can be classified in many different ways, depending on their preferences, personality traits and mode of travel. Cohen (1972), analysing issues related to types of tourist experience, proposed a classification based on dividing them into four distinctive types.

*Table 2 . Typology of tourist distribution*

| | |
|---|---|
| Typology of Tourist Distribution | •Organised Mass Tourist |
| | •Individual Mass Tourist |
| | •Explorer |
| | •Drifter |

Source: Cohen E. (1972) *Towards a sociology of international tourism*, Social Research, 39, 164-182.

Personalised mass tourists are those who prefer trips fully organised by travel agencies. Their trips are usually carefully planned, include standard tourist attractions, and there is limited contact with the local culture. Such tourists usually choose well-known and safe destinations, expecting comfort and security. Their entire trip is predictable and without major surprises. Individual mass tourists are also people who use travel agencies, but to a lesser extent than organised mass tourists. Although their trips are partly organised, they have more freedom and can make their own decisions about, for example, additional attractions or leisure activities. Although they are still concerned with comfort and

predictability, they are more open to spontaneous experiences. Explorers are tourists who strive for individual and authentic experiences, often avoiding mass tourist attractions. They plan trips on their own, seeking to learn about local culture and the daily lives of local people. Although explorers value authenticity, they usually do not sacrifice total comfort and safety, for example by choosing reliable means of transport or accommodation. Drifters are the most independent and spontaneous group of tourists who aim for total immersion in the local culture, avoiding any tourist infrastructure. They often travel without a fixed itinerary, moving from place to place according to circumstances. Their aim is to experience local traditions, lifestyles and nature to the maximum, even at the expense of comfort and safety (Cohen, 1972).

Table 3 showing the categories of tourists according to P.L. Pearce and their characteristic types of behaviour is also based on a complex system of motivations and patterns of tourist action. Contemporary tourism encompasses a variety of travel modes in which participants make different decisions depending on their experiences, goals and values. From those who prefer luxury and prestige to those who seek contact with nature and environmental protection, each type of tourist reflects a different approach to travel. This classification illustrates how tourists' behaviour varies with regard to interaction with local culture, risk-taking, self-reliance in organising trips, and interest in influencing the local community. This juxtaposition provides a better understanding of the psychological and social aspects of tourism, which is crucial for both researchers and tourism practitioners, helping them to tailor their offerings to meet the needs of different groups of tourists.

*Table 3 . Categories of tourists according to P.L. Pearce and characteristic types of behaviour*

| Tourist category | Characteristic types of behaviour |
|---|---|
| Tourist | The traveller takes photos, collects souvenirs, visits popular tourist attractions, but does not interact deeply with or try to understand the local culture. |
| Traveller | He moves quickly between places, tries local food, visits well-known attractions and prefers to explore on his own. |
| Holidaymaker | He takes photographs of the places he visits, visits popular tourist attractions and buys local souvenirs while supporting the local economy. |
| Jet-setters | He lives a life of luxury, oriented towards social position, seeks sensual pleasures, keeps in touch with his ilk, visits famous places |
| Businessman | He lives a life of comfort, focused on social prestige, seeks sensual experiences, maintains relationships with people of similar status, and visits popular and prestigious places when travelling. |
| Migrant | He has difficulty communicating in the local language, keeps in touch mainly with people with similar experiences, does not understand the local population and does not live a luxurious lifestyle. |
| Coservationist | He is interested in preserving the environment, avoids buying souvenirs, does not exploit the local community, visits and takes photos himself. |

| | |
|---|---|
| Explorer | He discovers new places on his own, is passionate about the local community, focuses on environmental issues and avoids buying souvenirs. |
| Missionary | He avoids buying souvenirs, seeks a deeper meaning of life, lives a modest lifestyle without luxuries and is not interested in sensual pleasures, while keeping a keen eye on the local community. |
| Overseas student | He samples local food, documents his journey with photographs and closely observes the local people. |
| Anthropologist | She carefully researches local communities, explores on her own, focuses on environmental issues and avoids consumerism by documenting her experiences. |
| Hippie | He avoids buying souvenirs, leads a modest lifestyle, does not focus on gaining social prestige and does not financially support the local community. |
| International Athlete | Travels to international sporting competitions with a focus on physical activity, often with limited time to explore local culture |
| Overseas journalist | He documents journeys and events in the places he visits, while also analysing and observing local communities in the context of his professional work. |
| Religious pilgrim | He travels for spiritual development, visiting sacred sites, often choosing modest forms of travel, focusing on prayer and contemplation. |

Source: Kurek (2008), s. 134.

Plog (1972), on the other hand, developed a model that divided tourists according to their degree of risk aversion. This model classifies tourists into three levels: allocentric (willing to take risks), mid-centric (who like to explore but value comfort) and psychocentric (who avoid risk and the unknown). Each of these tourist types has its own specific shopping preferences. Allocentric tourists usually organise their own trips, visit remote, less popular destinations, stay overnight with local people and try to explore the culture or natural environment of the visited region. They rarely return to the same places. Mid-centric tourists, on the other hand, prefer organised travel in the form of 'packages' but with elements of self-selection; they visit familiar destinations, balancing novelty with comfort. Psycho-centric tourists, on the other hand, opt for full-service mass tours, usually travelling to destinations culturally similar to their own, choosing mass accommodation and often returning to the same places. Each of these types of tourists has different demands on tourism products and responds differently to economic changes, which affects the way they contribute to the economy of tourist destinations (Bull, 1998).

Another important concept that is essential to understand in order to analyse the tourism sector is the tourism product, which comprises the totality of services and experiences offered to tourists. This product must be tailored to tourists' needs and be of high quality and uniqueness.

*Table 4 . Overview of selected tourism product definitions*

| Author | Definition of the tourism product |
|---|---|
| Wodejko (2006) | The tourism product in a broad sense is a composite of what tourists do and the assets, facilities and services they use in doing so. |
| Lazarek (1999) | The unity of tourism assets in the travel destination and the complementary goods and services that enable a temporary change to the permanent environment and make the stay at the travel destination possible and enjoyable. |
| Hołderna-Mielcarek (1998) | A tourism product is a complex product, shaped by many different actors, consisting of many individual elements - partly similar, partly heterogeneous, but always complementary. |
| Gołembski (1998) | A tourism product is all the goods and services created and purchased in connection with a trip away from the place of permanent residence, whether before the trip, during the trip or during the stay away from the hometown. |
| Middleton (1996) | A comprehensive offer that includes all the goods and services that tourists purchase and consume while travelling to and staying at a destination. The tourism product consists of five main elements:<br>• Tourist attractions - natural or man-made attractions that attract tourists (e.g. monuments, national parks, festivals).<br>• Tourism infrastructure - facilities and amenities that support tourists, such as hotels, restaurants, transport and other services.<br>• Accessibility - transport to the destination and between attractions.<br>• Brand image and awareness - tourists' perceptions of a place, image and reputation.<br>• Price and value - costs associated with travel, accommodation, food and other expenses. |
| Rapacz (1994) | In a narrow sense it includes tourist assets, i.e. attractions of various kinds in the form of natural environmental and sightseeing assets attracting tourists and advertised by companies. The broader meaning of the word includes, in addition to tourist attractions, the whole set of goods and services enabling the tourist to arrive at the place where the tourist attractions occur, to stay and to use them. |
| Stanton (1975) | A complex of tangible and intangible qualities, including quality, standard, class, prestige, price, trademark, commercial brand and others, which the purchaser can accept if they are compatible with his needs and he feels personal satisfaction through their purchase. |
| Kotler (1967) | A tourism product is a bundle of physical , service and symbolic elements and characteristics that contribute to the achievement of both satisfaction for the buyer and economic benefits for the seller. |

Source: own compilation based on the literature.

In order to interpret the structure of an area's tourism product, it is necessary to refer to Middleton's (1994) concept from the 1990s, which distinguishes five key elements of a tourism product: attractions and environment, infrastructure and services, accessibility, image and price. However, as Panasiuk (2019) rightly points out, this concept is not fully in line with economic principles. A tourism product should be considered as a transacted offer. The image of the destination,

although important, is not part of this commercial offer. It is a subjective image created by the media and promotional activities, which does not necessarily reflect reality. The image influences tourists' decisions and supports the branding of the destination, but is not in itself an object of sale. It is also worth emphasising that the image is dynamic, can change over time and vary according to groups of tourists or forms of tourism. Similar concerns apply to price as an element of the tourism product structure. The price reflects the supply structure and utility values of the good in question, but is not in itself an element of the transaction, but only the basis for its realisation. The singling out of accessibility as a separate element should also be approached critically, as it is derived from the infrastructure and transport services already singled out earlier.

The multidimensionality of the tourism product is one of the important concepts in the tourism management literature. It is important to note that the tourism product is not a one-dimensional offer consisting only of tourist attractions, but a complex set of elements that together create a holistic tourism experience. Within this concept, the tourism product is often divided into three main components: the core product, the extended product and the potential product.

The core product represents the essence of the tourism offer and the main factor attracting tourists to a particular destination. As Kotler, Bowen and Makens (2010) point out, the core product in tourism is most often a tourist attraction. Complementing this definition is Buhalis' (2000) assertion that the core product is the main promise that a place makes to its visitors, i.e. what the tourist hopes to see and experience in the chosen destination.

The second category is the augmented product, which includes any additional services that enhance the tourist's experience. According to Holloway (2009), an augmented product is a collection of services and facilities that contribute to the convenience and satisfaction of tourists during their stay. These can include items such as accommodation, food and beverage, local transport, guide services, and ancillary attractions such as optional tours, spas and theme parks. As Smith and Xiao (2008) point out, the augmented product is designed to enhance the perceptual value of the tourism offering, helping tourists to enjoy not only the primary attraction, but also all the accompanying amenities. Not infrequently, it is the augmented product that determines the ultimate customer satisfaction with the trip.

The final element of the tourism product is the potential product, which is the future services or attractions that can be developed in a location to attract more tourists. Kotler (2003) points out that the potential product includes innovations that can influence the further development of tourism in a region, such as new attractions, expansion of infrastructure, or new forms of recreation. This means that the tourism product is dynamic and subject to constant modification, depending on the needs of

the market and the development possibilities of the destination. Buhalis (2000) adds that the development of potential products is crucial to ensure the competitiveness of tourist destinations in the global market, as destinations constantly need to offer something new to attract returning and new tourists.

The multidimensionality of the tourism product plays an important role in tourism management, as it allows for a holistic approach to the creation of tourism offers. As Middleton (1994) notes, this approach helps destinations to better respond to the needs of different market segments - from individual tourists to family groups to business tourism. Each of these segments requires the right balance of core, extended and potential product to achieve maximum customer satisfaction.

## 1.2. Tourism as an economic sector

Tourism is defined as a set of relationships and phenomena arising from the travel and stay of visitors, provided that it does not involve permanent residence or engagement in remunerated work (Hunziker, 1951), is considered a natural part of human activity.

In addition, international travel has become more accessible as more and more people with lower social and economic status are finding ways to realise international travel (Martins, Gan & Ferreira-Lopes, 2017).

The tourism economy is a specific form of economic activity that combines elements of both the service and goods sectors, focusing on meeting the needs of tourists. The scientific literature emphasises that tourism is an important factor in the socio-economic development of many regions, especially those with rich natural and cultural assets. It is a tool for income generation, job creation and the development of technical and social infrastructure, including catering and hospitality. The tourism economy is based on cooperation between different actors - from tourism businesses to local authorities to cultural and natural institutions. It is these elements that together contribute to the creation of an attractive tourism product that attracts customers from all over the world. Definitions of the tourism economy in various studies point to its complexity and multidimensionality, emphasising that its development has an impact on the wider national economy, including gross domestic product, investment, tax revenues and job creation.

*Table 5 . Overview of definitions tourism economy*

| Author | Definition |
|---|---|
| UNWTO (2017) | The tourism economy is a complex of activities involving tourism and its impact on socio-economic development through income and employment. |
| Tribe (2016) | The tourism economy encompasses both direct and indirect effects of tourist spending on the economy, including impacts on various service sectors. |
| Kruczek (2011) | The tourism economy includes businesses providing tourism services and other tourism-related industries such as transport, catering, hospitality. |
| Kaczmarek, Stasiak, Włodarczyk (2010) | The tourism economy encompasses a variety of actors that work to meet the needs of tourists, while also influencing the development of the region. |
| Page (2009) | Tourism economics is the study of the economic effects of tourism travel and its impact on regional development. |
| Meyer (2008) | The tourism economy is a system of economic activities related to serving tourists, including both the production and distribution of goods and services. |
| Rapacz (2007) | The tourism economy covers all sectors directly and indirectly related to tourism, including transport, accommodation, catering and cultural services. |
| Gołembski (2002) | The tourism economy is the totality of economic processes involved in the creation, provision and consumption of tourism services in a specific region. |
| Domanski (1997) | The tourism economy is a set of economic activities carried out in relation to tourism services, affecting various sectors of the economy. |

Source: own compilation based on the literature.

Most of the definitions provided emphasise the structure of the services and products that are offered to tourists, which applies to both indirect and direct activities in the sector. In addition, the definitions often refer to the importance of in an economic context, describing it as a source of income and also as a part of the economy with an impact on other sectors.

Undoubtedly, the development of tourism is fostered by other phenomena of a social nature such as the increase in leisure time (including the prospect of discussions on reducing labour force

participation to 4 days per week), increasing life expectancy - according to the WHO, in 1900 the average life expectancy was 32 years, whereas in 2019 was 72.6 years (WHO, 2024), an increase in the size of the world's population (1.65 billion in 1900 to about 8 billion in 2024) (United Nations, 2024), ease of travel, and strong computerisation, including the development of social media indirectly encouraging travel.

Until the pandemic, the tourism economy was one of the most dynamic economies in terms of sectoral importance, with its impact usually measured in terms of its share of gross domestic product or employment in the industry. The impact of the pandemic on the global tourism sector is undeniable. UNWTO data shows that, globally, declines were -72% (2020 compared to 2019) and, despite a gradual recovery, the tourism economy in 2023 still did not reach pre-pandemic levels, remaining 20% below the 2019 figure. This represented a loss of 2.1 billion international arrivals (UNWTO, 2024). International tourism export receipts fell by 62% in 2020 and by 59% in 2021, representing a loss of around US$2.6 trillion over the two years. One of the main challenges remains adapting to new economic and geopolitical realities that may affect the pace of recovery. UNWTO stresses the need for innovative solutions and sustainable development strategies in order for the tourism sector to continue to grow in a way that is resilient to future crises. In this context, access to current and predictive data on the situation of the sector takes on particular importance .

Tourism sector encompasses various forms of economic activity in many sectors t. e.g. hotel, catering, tourism organisers, pilots and guides, but also transport or commerce. Nevertheless, there is no doubt that we are dealing with a market in which there is a sphere of demand and supply seeking a price equilibrium, although in reality globalisation has made the tourism market transnational, it should be noted that market representatives are exposed to constant distortions through external and internal factors. The complexity of the tourism product structure, as well as its real and virtual nature, poses challenges in both the research sphere of tourism supply and demand.

The result of such a situation is, among other things, the information asymmetry operating in the market. The contemporary market is characterised both by an excess of information concerning market participants, producers' offers and market processes, including competition, and by the incompleteness of information, which is its immanent feature (Panasiuk, 2016a). On top of this, an important part of the supply offer in the sector is directed not only to tourists and visitors, but also to local communities, where, on the one hand, one can look for elements favouring an increase in the quality of life of the inhabitants, while on the other hand, through increased traffic, increased cost of living, environmental degradation or noise and pollution, it also negatively affects tourist reception areas and their inhabitants.

As Panasiuk (2016a) points out, a more appropriate term for the processes taking place at the interface of demand and supply in tourism is precisely the tourism market, and not, as often encountered in the literature on the subject, the tourism services market. For interpreting the phenomena occurring in a market in which not only services are the subject of exchange is not adequate to its scope. For the market for tourism services is only an element of the tourism market.

While developing the Tourism Satellite Account (TSA), an internationally recognized statistical system created by the UN, UNWTO, and OECD to measure the economic impact of tourism on a country's economy, the UNWTO decided that tourism should be recognized as a sector encompassing various industries, in accordance with international standards. These industries were listed to clearly indicate the economic areas that are directly related to tourism, highlighting the multi-faceted nature of the sector (UNWTO, 2008). Within accommodation, different types of establishments are distinguished, such as hotels and similar accommodation (code 55100), youth hostels (55202), campgrounds and sites for recreational vehicles and trailers (55300), resorts and holiday villages (55201) and other collective forms of accommodation (55900). The restaurants and related services category lists licensed restaurants (56101), unlicensed restaurants and cafes (56102), takeaways and mobile food outlets (56103), other catering services (56290), event catering activities (56210) and licensed clubs (56301) and pubs and bars (56302). The rail passenger transport industry (code 49100) includes intercity passenger rail transport. The road passenger transport sector includes taxi services (49320) and other land passenger transport (49390). For water transport, marine and coastal passenger transport (50100) and inland waterway passenger transport (50300) are distinguished. Passenger transport by air includes both scheduled passenger transport (51101) and non-scheduled passenger transport (51102). The transport equipment rental industry includes the rental and leasing of cars and light motor vehicles (77110), as well as the rental of watercraft and aviation equipment (77341, 77351). Travel, tour and tour guide services include the activities of travel agents (79110), tour operators (79120), tour guide services (79901) and other booking services (79909). The cultural activities area includes performing arts (90010), performing arts support activities (90020), arts creation (90030) and arts facility management (90040). Industries related to museums and historic sites include museum operations (91020) and historic site management (91030), as well as botanical gardens and nature reserves (91040). Sports and leisure activities include gambling and betting (92000), management of sports facilities (93110), other sports activities (93199), amusement and theme parks (93210) and other forms of recreation (93290). Finally, country-specific tourism industries include activities of exhibition and fair organisers (82301) and activities of conference organisers (82302), as well as rental and operation of conference and exhibition centres (68202).

The tourism sector encompasses a wide range of services, from passenger transport (rail, air, sea, etc.) to accommodation, catering and services related to culture, entertainment and recreation. It is worth noting that, in addition to direct tourism services, tourism influences many other economic sectors that are complementary, such as booking services and the activities of travel agencies.

A list of sections and PKD codes assigned to the tourism sector in Poland is presented below (Biznes.gov.pl, 2024):

Section H: Transport and storage

- 49.1: Intercity passenger rail transport
- 49.10.11: Passenger transport by railway for excursion and tourism
- 49.31.1: Passenger rail transport, urban and suburban
- 49.32: Passenger taxi services
- 49.39.2: Passenger transport by aerial cableways, funiculars and ski-lifts
- 49.39.3: Non-scheduled passenger land transport
- 50.1: Sea and coastal passenger transport, including short-sea shipping
- 50.3: Inland passenger waterway transport
- 51.1: Air passenger transport

Section I: accommodation and food services

- 55.1: Hotel and similar accommodation services
- 55.2: Holiday and other short-stay accommodation services
- 55.3: Camping grounds services (including caravan or trailer park services) and camping services
- 55.90.13: Sleeping accommodation services in railway carriages and other transport equipment
- 56.1: Restaurant and other food service services
- 56.3: Beverage preparation and serving services

Section N: Administrative and support services

- 79.1: Services provided by travel agents, tour operators and retailers
- 79.9: Other reservation services and related services

Section R: cultural, entertainment, sporting and recreational services

- 91.02: Services provided by museums
- 91.03: Services connected with the operation of historical sites and buildings and similar visitor attractions

- 91.04: Services connected with the operation of botanical and zoological gardens and nature conservation areas and sites
- 93: Sport, entertainment and recreation services

It is noteworthy that in the sections presented, the tourism sector is not assigned to a single division, but is found at multiple points in the classification. Such an observation reinforces the idea that such a wide range of tourism-related activities can affect the entire economy. At the same time, the sections mentioned above lack many other elements that are indirectly related to the tourism economy, such as (after Panasiuk 2008) agriculture (agro-tourism), production activities (production of equipment for tourist needs), trade (food or souvenirs), construction (creation of tourist infrastructure), warehouse management and communications, financial intermediation (banking and insurance services), research (scientific research), public administration (units supervising the functioning of the sector, issuing passports and visas), education (tourist education), health care (spa tourism), municipal activities (waste management) and social activities.

*Table 6 . Direct and indirect tourism economy*

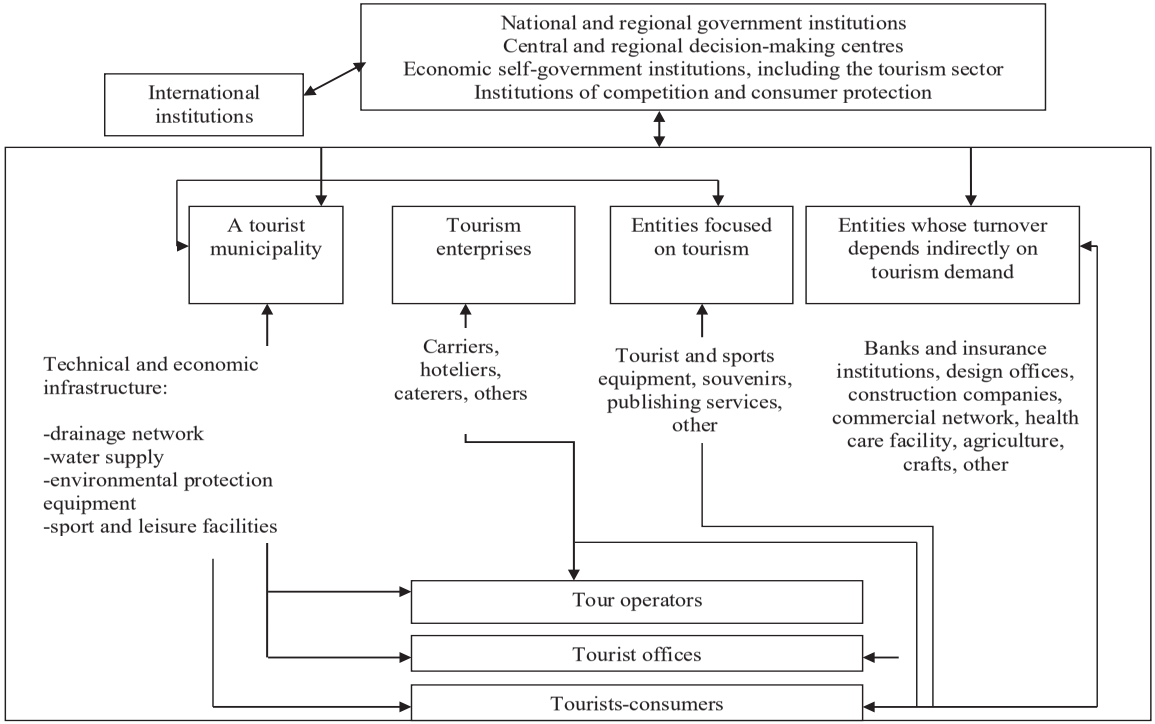
| Direct | | Intermediate |
|---|---|---|
| **Typical** | **Supplementary** | |
| Hospitality | Production of maps and guidebooks, production of souvenirs | Manufacture of sports goods, clothing, fitness equipment |
| Catering | Travel insurance | Protection of the natural environment |
| Passenger transport | Cultural and entertainment facilities | Telecommunications services |
| Travel agencies | Sports and leisure facilities | Utilities |
| Infrastructure for tourism assets | Local government tourism development activities | Production of multimedia for the tourism sector |
| Activities of tourism associations and organisations | Activities of tourism promotion organisations | Cosmetics production |

Source: own elaboration based on Panasiuk, A. (2008). *Gospodarka turystyczna [Tourism economy]*. Warszawa: PWN

As Panasiuk notes (see Table 6), the tourism economy is one of the elements of the structure of the national economy, interpenetrating and supported in its functions by other economic sectors, and the main part of the processes in the tourism economy are realised by the direct tourism economy and the tourism entrepreneurs representing it.

Kornak and Rapacz defined the tourism economy as a set of diverse functions, developed both directly and indirectly, to satisfy people's needs related to their desire to participate in various forms of tourism and leisure (Kornak & Rapacz, 2001). The entity structure of the tourism economy is shown in Figure 2, which illustrates a comprehensive picture of the entire tourism sector. The key criterion when considering whether an entity belongs to the tourism economy is the consumer criterion. All goods and services that are created and offered for the needs of, or purchased by, tourists can therefore be included in the tourism economy sector.

*Figure 2 . Tourism economy diagram*




Source: *Turystyka [Tourism]*, red. W. Kurek, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 341.

An increase in the number of travellers brings a number of benefits to local communities. These include an increase in the number of jobs in destination countries, an increase in the GDP of the host economy, the development of infrastructure used by residents and tourists, and many others. Tourism development is also stimulated by the tourism multiplier effect (Barczak, 2009). Tourism, due to its interdisciplinary nature, affects many sectors of the economy, becoming a catalyst for development at the regional level. Its importance in the international economy is also reflected in the balance of payments, which compiles revenues and costs resulting from tourism exports and imports, as illustrated in Table 7. At the national level, revenues from foreign arrivals are called 'invisible exports', while citizens' expenditure on foreign travel is referred to as 'invisible imports'. Countries where invisible exports are a key part of the economy are so-called tourism monocultures, which include the Maldives, Seychelles, Antigua and Barbuda, Aruba, British Virgin Islands and Vanuatu (Barczak, 2014).

*Table 7 . Tourism balance*

| Receipts (tourism exports) | Expenditure (tourism imports) |
|---|---|
| Direct receipts from international visitor arrivals. | Expenditure from residents' own travel abroad. |
| Export of goods, consumer goods, equipment, facilities for tourism. | Imports of goods, consumer goods, equipment, facilities for tourism. |
| Proceeds from communication services provided to international visitors. | Communication expenses of own residents related to foreign tourist travel. |
| Foreign tourism investment. | Tourism investment abroad |
| Proceeds from tourism investments abroad. | Interest expenses on foreign investments and loans. |
| Proceeds from employment of own residents abroad in the tourism sector. | Expenditure on employment of foreigners in the tourism sector in the country. |
| Tourism promotion expenditure. | Tourism promotion expenditure. |
| Miscellaneous services (e.g. vocational training, licences). | Miscellaneous services (e.g. vocational training of employees abroad). |

Source: *Turystyka (Tourism),* red. W. Kurek, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 349

In the Polish legal system, the Minister for Sport and Tourism is responsible for tourism issues, carrying out tasks through the Department of Tourism in the Ministry of Sport and Tourism (Gov.pl, 2024). This department deals with a wide range of activities, which include the development and promotion of tourism, the creation of tourism programmes and the development of the country for tourism. It is also responsible for improving tourist safety and regulating the tourism market, supporting entrepreneurship, improving the quality of services and protecting consumer rights. In

addition, the Department of Tourism deals with legal issues related to hotel services, the activities of tour managers and tour guides and the organisation of tourist events. An important area of its activity is also the recognition of qualifications obtained in EU countries and cooperation in the development of EU tourism law. The Department benefits from EU structural funds and actively participates in international tourism organisations. The Department is also responsible for education and training of staff in tourism, conducting statistical research and cooperating with educational and research and development institutions. In addition, it supervises the activities of the Polish Tourist Organisation and manages the Polish Tourist Voucher and the System of Public Registers in Tourism.

It is worth noting that the history of public institutions responsible for tourism in Poland is complex and includes numerous structural changes. Over the years, responsibility for tourism has passed through various ministries, including the Ministry of Communications (1945), the Central Committee for Tourism (1978), the Ministry of Industry (1990), and the Ministry of Transport and Maritime Economy (2000), among others. In 2007, these competences were assigned to the Ministry of Sport and Tourism and later taken over by the Ministry of Development (2019) and the Ministry of Development, Labour and Technology (2020). Such institutional instability has influenced the changing role of tourism, depending on how it is perceived - as part of the economy or as a form of recreation. The lack of constancy in tourism governance has not been conducive to long-term public investment, which is key to the development of the sector. The only permanent institution remains the Polish Tourism Organisation (PTO), whose functions remain unchanged.

In order to increase the effectiveness of Poland's promotion in tourism, both at the national and international level, the Polish Tourism Organisation was established in 1999 (PTO Act, 1999). Its main objective is to support cooperation between state administration bodies, local governments and organisations representing tourism entrepreneurs, including industry associations. The PTO, which is a legal entity supervised by the minister responsible for tourism, deals with, inter alia, the promotion of Poland as an attractive tourist destination, the development of a tourist information system and support for the modernisation of tourist infrastructure. The organisation actively cooperates with regional and local tourism organisations, inspiring their formation and development.

The system includes the creation of Regional Tourism Organisations (ROTs), which usually operate within provinces, and Local Tourism Organisations (LOTs), which operate on a smaller, local scale. A key element of this system is close cooperation between local authorities and other relevant tourism stakeholders to develop joint action strategies. ROTs act as a platform to coordinate the activities of local governments, industry organisations and tourism enterprises. They were established in accordance with the Act on the Polish Tourism Organisation of 25 June 1999 and the regulations on

associations. Their task is to support the tourist development of the region by integrating various stakeholders and promoting the region on a wider scale. This allows for more efficient use of promotional resources and optimisation of activities without additional expenditure. Although ROTs are not direct branches of the Polish Tourist Organisation, they are an important tool for local governments in promoting the region both nationally and internationally (Zawadzka, 2014).

## 1.3. Determinants of competitiveness of tourist regions (tourist reception areas)

The approach to competition in economic terms has evolved over the years. The classical period of economic thought was dominated by the views of Smith, Ricardo and Malthus (Gorynia & Łaźniewska, 2009), who saw competition as the key to optimal resource allocation and maximisation of social welfare. In Smith's Inquiries into the Nature and Causes of the Wealth of Nations (1776), the search for prosperity boiled down to understanding the sources of economic growth. Smith argued that the pursuit of individual benefits in a competitive environment results in social benefits, with the 'invisible hand' of the market steering the economy towards the optimum. Ricardo, on the other hand, insisted that openness to international trade promoted competitiveness, which he explained in the theory of comparative costs, promoting international specialisation as the path to prosperity.

In contrast, the neoclassical school of economics (Martin, 2003), which gained dominance in the later period, promotes a model of perfect competition and minimal state interference. In this theory, perfect competition is the ideal state in which the economy operates most efficiently. International trade is the key driver of economic growth here, and monopoly is seen as an exception to the competitive rule. However, this approach is criticised for being static, as it assumes ideal market conditions, which cannot always be achieved in reality.

In contrast, Keynes (2003) challenged the assumptions of the neoclassical school, arguing that self-regulation of the market through price elasticity does not ensure equilibrium. In his theory, it is important for the state to stimulate consumption and investment, which is particularly important in conditions of economic crisis, as in the 1930s. Keynesian theory promotes state intervention to correct market failures, especially in difficult times like the Great Depression.

Another approach, development economics (Martin, 2003), which emerged in the mid-twentieth century, focuses on long-term economic development and analyses the factors that support competitiveness at regional and national levels. Key aspects include openness to international trade, structural change in the economy and foreign investment. The theory emphasises the movement of

resources to higher value-added sectors, which has a positive impact on countries' growth and competitiveness.

New growth theory, also known as endogenous growth theory, explores the importance of human capital, technological progress and innovation as sources of growth. It introduces the concept of economies of scale and increasing benefits from investments in knowledge and technology, which promotes economic growth that is generated within the economic system. Endogenous technological progress in this theory assumes imperfect competition, in which spending on research, technological development and innovation is the determining factor for competitiveness.

The slice of approaches to competing and competition presented illustrates the variability of approaches from the perspective of place and time, among others. Ultimately, contemporary approaches also suggest that the pursuit of sustainable competitive advantage may be illusory. McGrath (2013), analysing contemporary developments, argues that in an era of digitalisation and globalisation, companies need flexible strategies based on short-term advantages that can change rapidly with the market.

It is worth noting in this connection that the concept of competitiveness is defined differently in the literature, which is mainly due to the various manifestations of competition in the context of socio-economic development. The development of competition theory has contributed to the emergence of an almost uniform definition of the term in modern economic literature (see Table 8 for details).

*Table 8 . Overview of competition definitions*

| Author | Assumptions |
|---|---|
| Brown, Heskett (2010) | Competition is a system of pressure and incentives in the economy in which different actors compete for resources and consumers, leading to improved services, lower prices and higher quality. |
| Penc (2008) | Competition is the process of different actors competing to achieve analogous goals and carrying out actions that make it difficult or impossible for rivals to achieve them |
| Krugman, Obstfeld (2006) | Competition is a process of interaction in the market in which different producers seek to increase their efficiency and innovate to meet consumer needs and stabilise the market. |
| Liebmann (2003) | Competition is a form of rivalry between economic actors in which companies vie for the favour of customers by offering more attractive transaction terms, including better prices and quality. |
| Porter (2001) | Competition in a given sector depends on five basic forces: rivalry for position among existing competitors, bargaining power of customers, bargaining power of suppliers, threat of new entries, threat of substitute products and services |

| | |
|---|---|
| Drucker (1998) | Competition is a mechanism that regulates the activities of firms by evaluating the efficiency of each other's economic activities in order to gain an advantage under conditions of limited resources and demand. |
| Blaszczynski, Stygares (1995) | Competition is a market situation, the interaction of sellers or buyers with each other. In the case of competition between sellers, each tries to win the customer by making the most favourable offer to the seller. Buyers can also compete with each other, trying to get the desired product by making the most advantageous offer to the seller. |
| Błaszczyński, Stygares, Stygares (1995) | Competition is the rivalry between companies seeking to increase their market share, leading to improved product quality and to innovation, which ultimately means lower prices for customers |
| Bernard, Coli (1994) | Competition is the free, total and genuine confrontation of all economic actors in the sphere of both demand and supply of goods and services, means of production and means of capital |
| Syrek (1994) | Competition in the colloquial sense is the struggle of entrepreneurs for the market. Choosing the most favourable exchange conditions to ensure the highest possible income |
| Bremond, Salrot (1994) | Competition is rivalry, competition to increase market share and profits |
| Majewska-Jurczyk, Jurczyk (1993) | Competition is a market situation in which firms or sellers independently seek to patronise buyers (gain the support of buyers) as a condition for them to achieve a given economic objective, which may be an increase in profit, sales or market share |
| Kamerschen, McKenzie, Nardineli (1990) | Competition is the process by which market participants, in pursuit of their interests, attempt to make offers that are more favourable than others in terms of price, quality or other characteristics that influence the decision to enter into a transaction |

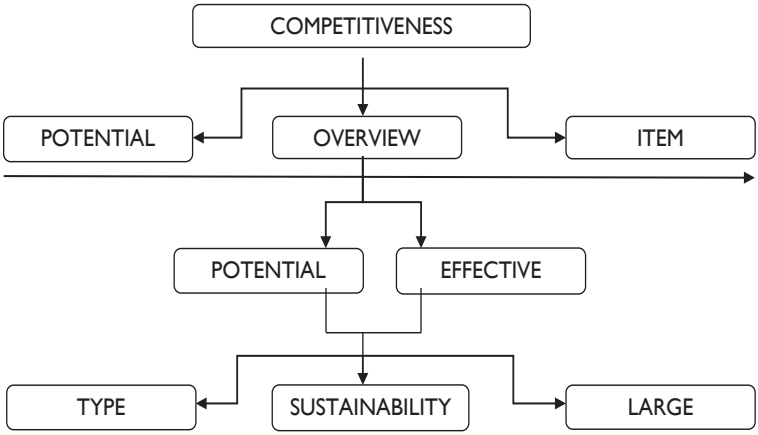
Source: own compilation based on the literature.

Defining the competitiveness of enterprises is less controversial than establishing a definition of the competitiveness of spatial units. This difference is due to the different characteristics and objectives that characterise enterprises and spatial units, as well as the different nature of their competition. Enterprises focus primarily on maximising profit in the long term, while spatial units focus mainly on improving the standard of living of residents. Bienkowski (2008) depicts this relationship on the "competitiveness pyramid" model." The differences between enterprises and spatial units are apparent in the market in which they compete, their priorities for action and their objectives. For enterprises, the competitive market includes products, services and factors of production, and the overriding rationale is economic, which translates into the pursuit of increased efficiency of operations for increased profits. Spatial units, on the other hand, compete within their area, prioritising social needs, with the aim of meeting the needs of the community that inhabits the area and increasing the efficiency of the national economy. In addition, the competitiveness of enterprises is embedded in

market mechanisms, while the competitiveness of spatial units is not based on the operation of a pure market to prevent the marginalisation of the weakest regions.

The conceptualisation of the notion of competitiveness makes it possible to distinguish three elements that make up the competitive model, i.e. potential, advantage and competitive position. A diagram of the concept of competitiveness is shown in Figure 3 below.

*Figure 3 . Conceptualising competitiveness*




Source: Barczak, M. (2014). *Koopetycja jako źródło przewagi konkurencyjnej obszaru recepcji turystycznej na przykładzie Bydgoskiego Węzła Wodnego [Co-opetition as a source of competitive advantage of a tourist reception area on the example of the Bydgoszcz Waterway]*.

In the context of a tourism region, competitiveness can be analysed using all the elements shown in the figure, such as potential, advantage, position, as well as the division into potential, effective, type, sustainability and size aspects. The potential of a tourism region refers to its natural resources, cultural heritage, infrastructure, access to transport and the quality of tourism services. This potential provides the basis for attracting tourists, shaping the unique character of the region. It encompasses elements that can attract different audiences, from those seeking relaxation to adventure enthusiasts or business tourists. High potentials range from attractive landscapes, historical monuments to entertainment and recreational offerings, which are the cornerstones of the region's competitiveness.

Competitive advantage is the unique characteristics that distinguish a region from other tourist destinations. Advantage can stem from natural assets such as mountains, sea or lakes, but also from unique cultural events, local cuisine or a tourist-friendly climate. A region's advantage can also be

based on innovative solutions, such as advanced tourism infrastructure, digitisation of services or attention to sustainability, which is increasingly important for environmentally conscious tourists. A region that is able to maintain and develop its advantages increases its attractiveness and attracts more visitors. An overview of selected definitions of competitive advantage is presented in Table 9.

*Table 9 . Competitive advantage in the literature*

| Author | Definition of competitive advantage |
|---|---|
| Pierścionek (2013) | A company can only sustain itself in the market if it has an adequate competitive advantage. |
| Gorynia (2009) | Competitive advantage in terms of competitive position is the result of a set of competitive instruments, which are the components of a competitive strategy. |
| Cetinski, Milohnic (2008) | The essence of competitive advantage boils down to a company doing something better compared to its rivals. |
| Szymura-Tyc (2002) | The ability to create and deliver value for the customer and the company, greater than its competitors. |
| Strategor (2001) | A collection of assets valued by the market that the company possesses. |
| Porter (2001) | Competitive advantage even at its best is only temporary. |
| Ma (1999) | The asymmetry or differentiation of each characteristic or factor that allows a company to satisfy customers' needs to a higher degree than its competitors, thereby achieving higher added value and performance mastery. |
| Rokita (1995) | The ability to utilise resources and competitive conditions in product supply and sales markets in such a way as to generate greater added value for the company and its actual and potential customers than the value generated by competitors |
| Barney (1991) | Competitive advantage can be achieved if the current strategy enables value creation for the customer and is not the strategy used by current or future competitors. |

Source: own elaboration.

Today, where big data is very often becoming currency, it can also be argued that the combination of historical data, real-time analysis and predictive analytics enables competitive advantage. Supporting the process of gaining advantage through big data makes organisations, and ultimately regions, not only responsive to market changes, but also strategically prepared for them. Certainly, the possibilities of using datasets allow for a competitive advantage over competing regions by better adapting products and services to the needs and preferences of the target group, and thus a better focus on the customer.

A region's position refers to its place in comparison with other tourist regions. This positioning is the result of both resources and advantages, as well as the region's reputation in the eyes of tourists

and marketing activities. A region's positioning in the tourism market is the result of long-term promotional activities, cooperation with the tourism industry and efforts to develop a good reputation among visitors. Regions that have a higher position attract larger groups of tourists and can compete more effectively for financial resources and investment. An interesting example of research into the competitive position of regions (and it should be made clear that administrative regions do not always correspond to tourist reception areas) is the competitiveness ranking of European Union regions, the so-called RCI - Regional Competitiveness Index). The RCI is an indicator developed by the European Commission since 2010 that measures the ability of regions to offer attractive living and working conditions. The ranking is based on a number of indicators, such as innovation, infrastructure, education and labour market accessibility, and its results are published every three years, with the latest edition in 2022 (EU Regional Competitiveness Index 2.0, 2022). Although tourism is not singled out, some of these indicators can indirectly affect tourism competitiveness as they relate to the attractiveness of the environment and the quality of life of the population, which is key to the development of the tourism industry. Regions with high competitiveness, such as capital city regions or those with large cities, also attract more visitors, which is conducive to tourism. Such regions tend to have a well-developed infrastructure and a higher standard of living, which can have a positive impact on their attractiveness as a tourism destination. Indirectly, the development of infrastructure and the efficiency of the labour market, supported by cohesion policy, contribute to the sustainable development of tourism in EU regions, especially those with lower competitiveness.

For example, Kujawsko-Pomorskie is in the category of regions in transition in the Regional Competitiveness Index (RCI) 2022, which means that its competitiveness is not at the level of developed EU regions, but is catching up with more developed areas. The RCI indicators cover various pillars such as institutions, macroeconomic stability, infrastructure, health, education, labour market and innovation. Transition regions, such as Kujawsko-Pomorskie, perform better in basic ('Basic') pillars such as basic education and stability of institutions, but still lag behind in more advanced indicators such as technological readiness and innovation, which are in the 'Innovation' sub-index. The goal for transition regions is to improve competitiveness through the development of advanced technologies and the improvement of innovation infrastructure, which could help to attract investment and create jobs (EU Regional Competitiveness Index 2.0 - 2022).

When considering potential and effective aspects of competitiveness, the region needs to assess which resources are still under-utilised (potential) and which resources are effective in increasing the number of tourists and improving the quality of life of residents. Potential aspects of competitiveness can be developed to attract new groups of tourists, while effective aspects indicate which resources are

of greatest benefit and can be expanded. The type of competitiveness of a region can include price competition (e.g. cheaper accommodation and services) or quality competition (exceptional quality of offer, unique attractions). Different regions choose different strategies, depending on their offer and target group of tourists. Sustainability, on the other hand, is a key aspect, as it refers to a region's ability to maintain its attractiveness in the long term, regardless of changing tourism trends. Sustainable competitiveness requires investment in infrastructure, care for natural resources and adaptation of the offer to market expectations. A region that cares for its resources and operates in a sustainable manner is more resilient to crises and better able to attract loyal tourists. However, it should be noted that sustainability is always analysed in hindsight and there is no sustainable competitive advantage. Therefore, it is important to seek answers to the question of the sources of stable competitive advantage. Pierścionek (2007) argues that sustainable advantage results from the ability to generate short-term competitive advantages. Porter (2010), on the other hand, suggests that the longevity of an advantage depends, among other things, on the number of its sources and the difficulty of imitating them. In this context, access to knowledge and information, especially those related to the structure and volume of tourism, is of particular importance. Through the use of modern analytical technologies, such as big data analytics, regions can significantly improve their ability to understand tourist needs, identify market trends and respond effectively to changes in traveller preferences.

Big data allows the collection and processing of huge amounts of data, which include not only demographic or geographical data, but also real-time user behaviour data, such as from online bookings, social media activity, or traffic analysis in urban areas. This enables tourism regions to create personalised offers, allocate resources more precisely, and make better strategic decisions, improving their concreteness at the level of all stakeholders. Access to such data is an important source of competitive advantage, as it enables them to better predict tourist behaviour and respond more quickly to changing market conditions. At the same time, the difficulty of obtaining and effectively using such data - both because of its complex analysis and the need for an appropriate technological infrastructure - makes it extremely challenging for competitors to imitate these activities.

As a result, regions that effectively use big data in tourism management can not only achieve short-term competitive advantages, but also build a sustainable advantage based on unique analytical and strategic capabilities. This approach can become a key element of their long-term development and the basis for building a positive image and attracting new tourist segments.

The last element, size, refers to the scale of a region's competitiveness - it includes the number of tourists, the geographical scope of influence and the ability to serve a variety of needs. Regions with

greater competitive size are able to attract larger groups of tourists and offer a variety of leisure activities that cater for a wide range of audiences.

Satisfying tourism needs takes place in travel destination areas that attract tourist traffic and are referred to as tourist destinations (Niezgoda 2006). Each such area, and in particular the stakeholders operating in it, significantly influence its development and, as a result, the development of tourism nationally. According to definition of the World Tourism Organisation (UNWTO), a tourism reception area can be a town, city or region. The UNWTO restricts this concept to the mesoscale level to exclude single resorts as well as countries, states or international metropolises from this category. The area thus refers to places that offer differentiated tourism products to tourists and where a specific tourism policy is implemented (Majewska, 2008).

Local authorities often look to tourism as a means of improving the quality of life of local residents, as a result of the fact that much of the infrastructure serves both tourists and local communities. An increase in the number of visitors, on the one hand, supports the development of tourist reception areas and, on the other, benefits local residents who can benefit from revitalised public spaces.

The revitalisation of space by local authorities not only improves the aesthetics and functionality of the surroundings, but also creates opportunities for local tourism initiatives. Thus, local governments often stimulate the creativity and innovation of local entrepreneurs, which is in line with Schumpeter's (1991) concept of seeing innovation as a key driver for regional economic development.

The relationship between the creation of new tourism products and an increase in the quality of life of residents is usually a positive one. The expansion of the infrastructure and service network, driven by the development of tourism, makes urban space more accessible and welcoming for both visitors and residents. However, it is difficult to capture the full spectrum of motives that drive people to visit particular regions, as these can be as varied as the purposes - from business to leisure.

Tourist reception areas can bring economic benefits to a region due to unique characteristics that distinguish them from other places. The creation of competitive strategies should take into account the development of values that are valued higher by tourists than the offer of the competition, thus maintaining market advantage (Niezgoda, 2012). Co-operation between actors in these areas is crucial, as tourists benefit from the whole package of services rather than individual offers. A coopetition strategy, combining cooperation and competition, allows for increased competitiveness without eliminating rivals, which is beneficial for all stakeholders (Meyer, 2006).

As Ritchie and Crouch (2000) note, the competitiveness of a tourist reception area is the sustained benefits of tourism development for all actors in a given socio-economic system. From a

tourism perspective, the competitiveness of a region means the ability to create and present an offer that would differentiate it from other destinations, attracting tourists and increasing the socio-economic benefits of tourism development (Derek, 2008). Maintaining the competitiveness of a tourist area requires the creation of values that will be valued higher by tourists than those offered in competing regions. When shaping a region's unique product, it is worth avoiding a homogeneous offer, as standardisation and copying reduces regional diversity while increasing the possibility of substitution and competition between tourist areas (Kornak and Rapacz, 2001). In the age of globalisation, products devoid of uniqueness lose their attractiveness to tourists due to their universality (Gołembski, 2000).

A competitive perspective does not exclude looking at the tourism sector also from the perspective of cooperation, which, occurring simultaneously, creates space for coopetition. The term was first used in a publication by Noord (1993) and developed by Brandenburger and Nalebuff (1996). It defines a situation in which two or more business partners simultaneously compete and cooperate, assuming repeated interactions. In classical competition, the goal was to eliminate a rival or limit its access to the market, whereas coopetition transforms competition into a game in which all participants can achieve higher profits. In this game, it is about maximising the value produced and competing for the largest share of it (Rapacz, Gryszel, Jaremen, 2010; Czakon, 2009).

Definitions of coopetition vary from author to author: Lado, Boyd and Hanlon (1997) describe it as a specific way of dealing with competitors, Fjeldstadt (2004) treats it as a competitive activity, while Dagnino and co-authors (2009) see it as a system based on partial compatibility of interests. Niemczyk and Stańczyk-Hugiet (2014) define coopetition as a synergy strategy that results from the interaction of firms in a network and their appropriate positioning relative to other market participants.

Co-opetition gives companies, and also regions, the opportunity to gain a competitive advantage by integrating resources and competences, and interacting with competitors, suppliers and customers. In the tourism industry, this cooperation is particularly important because of shared resources such as nature and cultural heritage.

The joint use of modern technologies, such as big data analytics, provides the opportunity to more accurately predict visitor behaviour, which supports the development of personalised services and the improvement of their quality. Such an approach builds not only the competitive advantage of individual stakeholders, but also positively influences the image of the entire region as an attractive tourist destination. Coopetition enables synergies to be created between different stakeholders, leading to more sustainable development and increasing the attractiveness of the area for both tourists and investors. Sharing knowledge and resources is therefore becoming a key element in the struggle for a sustainable competitive advantage that is difficult for other regions to imitate.

## 1.4. Trends in the tourism sector nationally and internationally

The modern world is seeing dynamic changes in the structure and consumption preferences of society, which has a direct impact on the development of tourism demand. One of the key factors shaping this market is the increasing affluence of societies. Rising incomes translate into a greater willingness to spend on travel and related experiences. People are increasingly opting for more frequent, albeit shorter, holidays as a result of changing lifestyles and attitudes to work and leisure.

Globalisation and the development of transport, especially low-cost airlines, are making travel significantly easier. The availability of cheap airfares and the developing transport infrastructure means that more and more people can afford to travel abroad. This is a phenomenon that not only increases the number of travellers, but also diversifies destinations. The development of social networks is also having a not inconsiderable impact on the travel industry. Platforms such as Facebook, Instagram and X allow people to share their travel experiences in real time, which in turn inspires others to travel. In addition, access to a wealth of information, user reviews and opinions helps with travel planning, destination selection and tourist attractions.

It is also worth mentioning the growing role of technology in the form of translators and mobile applications that make it easier to communicate and navigate in a foreign environment. They make language and cultural barriers less perceptible, which further encourages the decision to travel to unknown parts of the world.

All these factors paint a picture of a booming tourism market that is becoming increasingly accessible and attractive to a wide audience. The following sections of the chapter will discuss in detail the trends and forecasts for tourism demand in the context of these developments.

The impact of the pandemic on the tourism sector worldwide is indisputable. UNWTO data (see Table 10) shows that the global decline was -72% (2020 compared to 2019) and the tourism economy, despite a gradual recovery, has still not reached pre-pandemic levels (-20%, 2023 compared to 2019). Of course, individual countries reacted differently to the pandemic, and the data below show that the tourism sector in Poland fared much better than the global and Central and Eastern European average. Additionally, the situation in Poland in recent years has been much better than in Germany, the Czech Republic and Slovakia (Barczak, 2023a).

In most cases, such a significant decrease was due to the total border closure, which in May 2020 covered almost all regions of the world (see Table 4). However, with increased knowledge and understanding of the situation and the introduction of a vaccination regime, restrictions were gradually

relaxed or lifted. Noticeably, destinations in Europe and the Americas were quicker to reopen to tourists, while destinations in Asia and the Pacific were more cautious (Barczak, 2023b).

In most cases, such a significant decrease was due to the total border closure, which in May 2020 covered almost all regions of the world. However, with increased knowledge and understanding of the situation and the introduction of a vaccination regime, restrictions were gradually relaxed or lifted. Noticeably, destinations in Europe and the Americas were quicker to reopen to tourists, while destinations in Asia and the Pacific were more cautious. For example, 156 countries closed their borders in April 2020, and by November 2021, this number had fallen to 46 (UNWTO, 2023).

*Table 10 . Impact of Covid-19 on international tourist arrivals (YTD rate - %)*

| YTD comparison period | World | Central Europe Eastern | Poland | Germany | Czech Republic | Slovakia |
|---|---|---|---|---|---|---|
| 2020 vs. 2019 | -72% | -70% | -60% | -69% | -74% | -65% |
| 2021 vs. 2019 | -70% | -63% | -54% | -70% | -76% | -77% |
| 2022 vs. 2019 | -35% | -41% | - 25% | -28% | -33% | -36% |
| 2023 vs. 2019 | -20% | -28% | n.d. | -17% | -17% | -14% |

YTD: year-to-date.
Source: own elaboration based on: www.unwto.org

In the first three quarters of 2020, international arrivals to EU countries fell by 67.5% compared to the same period in 2019. Depending on the specific member state, this decrease oscillated between 46% and 84% (European Court of Auditors, 2021). Restrictions on travel have resulted in a drastic decrease in the number of tourists, which in turn has significantly reduced the contribution of the tourism industry to the economy. In 2020, the sector lost nearly two million jobs in the European Union and its share of GDP halved compared to 2019 - from approximately 10% to 5% of GDP (WTTC, 2021).

The end of the pandemic was officially announced at the Emergency Committee International Health Regulations (IHR) meeting on 4 May 2023. In line with this decision, the World Health Organization (WHO) recommended that Covid-19 cease to be treated as a public health emergency of international concern. As a result of this meeting, travel recommendations were also made, indicating that restrictions should be phased out and that proof of vaccination should no longer be required as a prerequisite for international travel (UNWTO, 2023).

According to a report by the World Economic Forum (WEF, 2024), international tourist arrivals and the contribution of the tourism sector to global GDP are back to pre-pandemic levels in 2024. Similarly, notes Mastercard (Mastercard Economics Institute, 2024), which, analysing aggregated and anonymised transactional data, finds that many records are being broken in the tourism economy in 2024 as, for example, it turns out that 9 out of 10 record-breaking days in terms of spending in the global airline ticketing sector fall in that year. And McKinsey & Company's research (2024) finds that 60% of travellers surveyed said they were more interested in travelling than before the pandemic, and this result holds across all groups and nationalities surveyed. It appears that travel is becoming a priority and, after spending on food at home and spending on food in restaurants, is in third place in consumer opinion (McKinsey Consumer Wise Global Sentiment Survey, 2024). Younger generations seem to be driving the increase in interest in travel. Millennials and Generation Z made an average of nearly five trips in 2023, surpassing the score of Generation X and baby boomers, who travelled less than four times (see Table 11). Millennials and Generation Z also report spending an average of 29 per cent of their income on travel, while Generation X spends 26 per cent and baby boomers 25 per cent.

*Table 11 . Average number of trips by age group*

| Age group | Domestic trips (average) | International travel (average) | Difference (%) |
|---|---|---|---|
| Gen Z (18-25 years) | 2.3 | 2.1 | -9 |
| Millenials (26-41 years) | 2.7 | 2.3 | -16 |
| Gen X (42-57 years) | 2.2 | 1.7 | -26 |
| Baby boomers (58+ years) | 2.3 | 1.1 | -71 |
| Total | 2.4 | 1.9 | -23 |

Generation Z, n = 396; millennials, n = 2,037; generation X, n = 1,660; baby boomers, n = 968.
Source: McKinsey & Company. (2024, May). *The state of tourism and hospitality 2024*. Travel, Logistics & Infrastructure Practice, p. 17.

Unfortunately, the tourism sector should grow more evenly, as the ten most popular destination countries in the world currently receive 45 per cent of all tourist spending, including domestic travel (UNWTO, 2024). However, new destinations such as Thailand, the Philippines, Vietnam, Peru and Rwanda are starting to gain popularity, attracting more tourists and attention to the tourism market. There is also a noticeable increase in residential rentals and a marked rebound in hotels in the post-pandemic period (McKinsey & Company, 2024).

*Table 12 . Value of accommodation bookings by type of stay*

| Year | Hotels (billion $) | Rental of flats (billion $) | Other (billion $) | Total value ($ billion) |
|---|---|---|---|---|
| 2017 | 749 | 94 | 135 | 977 |
| 2019 | 842 | 118 | 144 | 1104 |
| 2021 | 508 | 92 | 119 | 719 |
| 2023 | 792 | 156 | 128 | 1076 |
| 2025* | 1048 | 195 | 163 | 1406 |
| 2027* | 1235 | 232 | 190 | 1657 |

*forecast

Source: own compilation based on McKinsey & Company. (2024, May). *The state of tourism and hospitality 2024*. Travel, Logistics & Infrastructure Practice, p. 56.

The World Economic Forum also maintains an index of destinations that are most conducive to travel and tourism development (Travel & Tourism Development Index, 2024). It lists countries ranked according to a set of factors and policies that enable sustainable and resilient growth in the tourism sector. The United States tops the list with a score of 5.24, followed closely by Spain with 5.18, followed by Japan (5.09), France (5.07) and Australia and Germany (5.00 each). A score of 4.96 was achieved by the UK, followed by China (4.94), Italy (4.90) and Switzerland (4.81). The global average for all countries is 3.96. Poland is ranked 27 in this index with a score of 4.40.

*Table13 . Accommodation and travel expenditure 2023-2030 (forecast)*

| Criterion | 2023 | 2030 (forecast) |
|---|---|---|
| Domestic overnight stays (billion) | 16.1 | 19.8 |
| Intra-regional overnight stays (billion) | 6.8 | 11.6 |
| Interregional accommodation (billion) | 4.5 | 7.3 |
| Domestic travel expenditure ($ billion) | 4.1 | 6.0 |
| Intra-regional travel expenditure ($ billion) | 0.9 | 1.7 |
| Interregional travel expenditure ($ billion) | 0.5 | 0.9 |

Source: own compilation based on McKinsey & Company. (2024, May). *The state of tourism and hospitality 2024*. Travel, Logistics & Infrastructure Practice, p. 6.

Projections for the development of the tourism sector are also indicated by Booking.com (Travel Predictions, 2024), which has identified seven key predictions based on a survey of 27,000 travellers from 33 countries.

*Table 14 . 7 major travel trends for 2024*

| | |
|---|---|
| Travel Trends | •Anonymity and identity change |
| | •Water Journeys (cruises) |
| | •Healt and well-being |
| | •Spontaneity |
| | •Sustainable development |
| | •Local cuisine |
| | •Budget luxury |

Source: Travel Predictions 2024, Booking.com

As Table 14 shows, travel will become a tool for transformation and exploration of new identities, driven by the growing need for anonymity. There will be a growing interest in water-based expeditions, such as cruises, and wellness travel, which focuses on recovery and health. The trend for spontaneity will gain momentum, with more travellers choosing unplanned trips, and environmental awareness will drive demand for sustainable travel. Local cuisine will become an important factor in destination selection and at the same time luxury experiences combined with affordable prices will attract the attention of those looking for unique offers.

Another important international factor affecting the tourism sector at a regional level is the war in Ukraine, which has caused both an increase in energy prices and changes in the structure of visitors to Poland. According to a report by the Polish Economic Institute, the war had a significant impact (Polish Economic Institute, 2023):

- increase in the price of supply and investment goods (44%),
- increasing the prices of products and services offered by companies (45%),
- higher operating costs (42%),
- increased business risk (35%).

*Table 15 . Perspective on changes in travel plans as a result of the war in Ukraine*

| Declared answer to the question "How has the ongoing war between Russia and Ukraine affected your travel decisions?" | Value [%] | |
|---|---|---|
| I was not planning any overnight stays in the next six months | 8,7% | |
| This did not affect me at all. | 45,4% | |
| I will wait and see how the war situation unfolds before I book a trip. | 8,8% | 45,9% |
| I made the trip with some changes from the original plan (accommodation/transport). | 7,4% | |
| I changed my journey from international to domestic. | 7,5% | |
| I changed my itinerary to avoid countries neighbouring the conflict area. | 12,8% | |
| I rescheduled it for another date. | 5,0% | |
| I completely cancelled my trip. | 4,4% | |

Source: *Monitoring sentiment for domestic and intra-European travel,* European Travel Commission 2023, p. 29, [in:] https://historicalcity.eu/wp-content/uploads/2023/12/European-Travel-Commission-Report-2023-2024.pdf

According to a survey by the European Travel Commission, 13% of Europeans indicated (see Table 15) that the conflict negatively influences their decision to choose this destination and they will avoid coming to close conflict zone countries (+3% than in 2022) (European Travel Commission, 2023). For nearly 46% of respondents, the war appeared to be a factor influencing their decisions, with as many as nearly 13% indicating that they had decided to change their holiday destination to a country that is not adjacent to a conflict area.

Table 16 shows selected trends in the tourism market that are shaping tourist preferences and behaviour around the world. These changes reflect both the changing expectations of travellers and the impact of global events such as the COVID-19 pandemic and the war in Ukraine. It is worth noting the diversity of trends, from philanthourism to microtourism to workation and digital nomads. These phenomena influence how travel is planned, what destinations are chosen and what experiences are most important to tourists.

*Table 16 . Trends in the tourism market*

| Name of the trend | Description |
|---|---|
| Philanthurism | Philanthropic travellers are a new group of tourists who want to contribute to the development of the places they visit by consciously choosing their holiday destination. Most aim to support the local economy, including small and medium-sized enterprises. Their aim is to make a lasting, positive impact on local communities, even after their stay has ended (Global Travel Trends Report, 2023). |
| Micro-tourism | In the post-covid era, consumers and tourists are more willing to consider alternative ways to meet their needs, including choosing new, closer destinations (Top 10 Global Consumer Trends, 2022). |
| Back to nature | Among Polish tourists, natural attractiveness and the cultural richness and values of local communities appear to be key criteria when choosing a holiday destination. Interest in forests, landscape parks is also steadily increasing (Nationwide Survey of Travellers, 2021). |
| Slow travel | The slow living philosophy is becoming an increasingly comprehensive and well-established approach, encompassing aspects such as a slower pace of life, mindful eating, quiet movement and travelling with attention to the present. It is a lifestyle rooted in mindfulness and attentiveness, especially when travelling (Travel + Leisure Magazine, 2024). |
| Culinary tourism | In tourism, eating and drinking has moved from being seen solely as a basic necessity of life to being the main attraction of travel. Research indicates that for 60% to 74% of travellers, culinary aspects are a key factor in choosing a tourist destination (Global Travel Trends, 2022). |
| In search of happiness | Travellers will be increasingly driven by the need for self-acceptance, which will become a priority for them above 'show holidays' designed to impress others. They will seek experiences tailored to their individual needs. During the pandemic, many people re-evaluated their life priorities. Travellers now aspire to live and travel more fully, making tourism increasingly a means of fulfilling individual aspirations (Top Global Consumer Trends, 2022). |
| Hunger to travel | After almost two years of travel restrictions, people around the world are expressing a stronger desire than ever to travel, book holidays and explore new places. The motivation to travel is now the sheer joy of moving around, experiencing other cultures and experiencing novelty, as seen in the growing enthusiasm to plan and execute trips (Top Global Consumer Trends, 2022). |
| Workation, bleisure, and digital nomads | Modern lifestyles are increasingly combining elements of work with relaxation and travel, leading to the popularity of models such as workation (a combination of work and holiday) and bleisure (a mix of business travel with elements of leisure). The COVID-19 pandemic has significantly accelerated the remote working trend, changing the relationship between employers and employees and making it possible to work outside the traditional office. Many professions and people from different age ranges |

| | |
|---|---|
| | discovered the possibility of working remotely. This, in turn, has opened up new possibilities for choosing leisure locations that can also serve as temporary offices. |
| Digital seniors - the silver tsunami | The growing number of older people each year will increasingly use online booking options, including travel services. |
| Metaversum and virtual worlds | In 2021, the transfer of activities to the virtual space has become a widespread phenomenon (Top Global Consumer Trends, 2022). Trends such as the personalisation of offers, online payments and hybrid events are gaining ground, enabling a more personalised and interactive approach to travelling and discovering the region. |

Source: own compilation based on Barczak M. (2023). *Diagnoza - załącznik do Strategii budowy i promocji produktów turystycznych województwa kujawsko-pomorskiego (Diagnosis - Annex to the Strategy for construction and promotion of tourism products in the Kujawsko-Pomorskie Voivodeship),* Toruń, Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.

In addition to the trends listed in Table 16, awareness of the environmental impact of tourism, including CO2 emissions and plastic pollution, is growing all the time. Travellers will increasingly consider the carbon footprint of transport, the energy efficiency of hotels, zero waste practices in restaurants and the issue of over-tourism and its effects during the high season.

All the trends selected, fit into the most important trend - Sustainability in tourism, which is a kind of closing bracket. Tourism is becoming more environmentally conscious. Tourists' expectations of transparency in environmental measures and motivations such as eco-certification for accommodation facilities are also increasing (Top Global Consumer Trends, 2022).

Tourism that harmonises with the sustainable development goals set out in the UN Agenda 2030 (United Nations, 2015) is becoming increasingly important. There is a growing number of travellers who make tourism choices in full awareness of their impact on sustainable development, taking into account aspects such as carbon footprint (including carbon footprint certificates) or the impact on the degradation of terrestrial and aquatic ecosystems. Thus, tourism can contribute to a more sustainable development of regions, offering decent working conditions and promoting even economic development (Barczak, 2023a).

The consequence of technological change in the tourism sector is primarily the integration and internationalisation of activities. This means the integration of local and regional service providers into global booking and sales systems, which leads to the creation of new links. On the other hand, the increased use of technology in the offering and selling of services enables faster and more accurate consumer analyses, thanks to the use of methods and tools based on so-called big data. Adapting to these changes not only helps to increase the competitive advantage of companies operating in the tourism industry, but also supports local governments in managing tourism in the region more effectively (Barczak, 2023a).

Many service providers in the travel industry, present on various websites and portals, offer features that allow users to create, share and express opinions, advise others and share their experiences with travel. Content generated by other Internet users is considered more valuable by consumers than advertisements created by companies, especially among millenials. Research has shown that reading travel reviews written by other tourists helps internet users to gain information about destinations and travel services, evaluate different alternatives, avoid destinations and services they might be dissatisfied with, confirm decisions to go to a particular destination, imagine the trip, reduce risks and uncertainty, and plan trips effectively (Dabija, Brandusa, & Tipi 2018, Gretzel, Yoo & Purifoy 2007).

In this context, it is worth highlighting what is known as content marketing, which involves the creation and distribution of valuable content to attract, acquire and engage specific target groups in order to trigger intended and beneficial buyer actions (Content Marketing Institute, 2024).

As part of the changing trends in tourism, changes related to search and analysis of offers are also noticeable. For example, both AR and VR have great potential for application in the tourism sector, but their use should depend on the target group and the type of tourism product offered to potential customers. For example, VR makes it possible to visit places that may be difficult to access, expensive or protected due to their historical value. This application also works well in the effective promotion of tourist reception areas and in the training process of tourism staff, which can improve the quality of the tourist experience. AR, on the other hand, allows live information to be added to the real environment, for example enriching a tour with directions along the route or additional information about local services. AR may be more appealing to travellers who want to keep in touch with the real environment while receiving enriched information, while VR may be a better option for those seeking an intense virtual experience (Barczak, 2023a).

*Table 17 . Interpretive differences between VR and AR*

| Scope | VR - Virtual Reality | AR - Augmented Reality |
|---|---|---|
| User environment | The user enters a completely virtual environment that is separate from the real environment. Using special VR goggles, the user is completely immersed in an artificially created world. | AR superimposes virtual elements on the user's real environment. The user is still aware of the real environment, but with additional information or objects superimposed on it. |
| Immersion | It provides full immersion, isolating the user from the surrounding world. The user experiences full virtual reality | The immersion in AR is less deep, as the user is still aware of the real environment. AR aims to enrich reality with additional information. |
| Interaction | Interaction in VR often takes place using motion controllers, gestures or other special devices. The user can interact with the virtual environment in ways that do not necessarily correspond to real gestures. | Interaction in AR can take place using smartphones, tablets, AR glasses or special controllers. Virtual elements are superimposed on the real environment, making it possible to interact with them. |
| Examples of applications | Virtual reality games, simulations, virtual training, 360-degree videos. | Mobile apps with information overlaid on an image of the real environment (e.g. Pokemon GO), enhanced travel guides, apps to help with repairs or design. |
| Tools | VR goggles that lock the user into a virtual environment. | AR glasses, smartphones, tablets that allow the user to see both real and virtual elements. |

Source: Barczak M. (2023). *Diagnoza - załącznik do Strategii budowy i promocji produktów turystycznych województwa kujawsko-pomorskiego (Diagnosis - Annex to the Strategy for construction and promotion of tourism products in the Kujawsko-Pomorskie Voivodeship),* Toruń, Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.

# CHAPTER II:

# TOURIST DEMAND

## 2.1 Determinants of tourism demand

Demand is one of the key concepts in economics, referring to the quantity of goods and services that consumers are willing to purchase at a given price over a given period of time (Onoua & Smith, 2020). In the context of scientific analysis, the study of demand is an important part of understanding the dynamics of markets and consumer behaviour. Demand can be determined by many factors, such as price, consumer income, preferences, as well as technological change and innovation. In the tourism sector in particular, demand plays a key role in business planning and forecasting, influencing investment decisions, pricing policies and resource management. Understanding how demand is shaped in different contexts is essential for predicting market trends and optimising decision-making in many industries.

Tourism demand is one of the elements of the tourism sector and, therefore, tourism economics research, being an important element in the search for answers on the dynamics and behavioural patterns of consumers of tourism services. The definition refers to the quantity of tourism goods and services that consumers are willing to purchase at a given price over a given period of time (an overview of definitions is presented in Table 18). In the literature, tourism demand is analysed in terms of a number of variables, such as consumer income, prices of tourism services, transport costs and tourists' preferences and motivations. Tourism demand models often take into account price and income elasticities, which allows for more precise forecasting and planning of marketing activities and public policies.

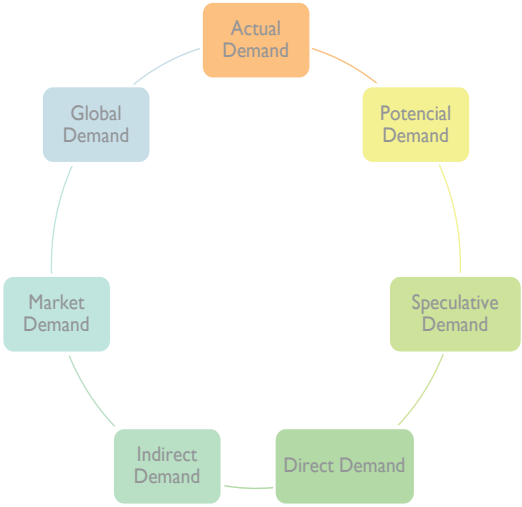
Analyzing the definitions presented in the table, it can be observed that the authors agree on the fundamental understanding of tourism demand as the quantity of goods and services that consumers are willing to purchase. However, there are differences in the approach to the study and modelling of demand. Dwyer & Forsyth (2006) emphasise a broad economic view, taking into account the various factors that influence demand. Lim (1997) focuses on detailed analysis of economic variables such as income and prices, while Song, Romilly & Liu (2000) and Turner focus on demand forecasting using advanced econometric techniques.

*Table 18 . Overview of definitions of tourism demand*

| Author | Definition |
|---|---|
| Dwyer (2020) | Tourist demand is the number of visits by tourists from the country of origin to the destination country or the spending by tourists from the country of origin in the destination country. |
| UNWTO (2019) | Tourism demand refers to the total demand for tourism services and products generated by tourists over a specified period of time. This demand includes not only tourist spending on goods and services, but also tourism-related capital investment and government spending on tourism services and infrastructure. |
| Troisi (2019) | Tourism demand is the amount of goods and services that, at a given point in time, a buyer is willing to purchase at existing prices in a given country (domestic tourism) or at a given exchange rate (foreign tourism). |
| Yazdi & Khanalizadeh (2017) | Tourism demand is the demand for international tourism as measured by the number of tourist arrivals and their spending . |
| Discord (2012) | Tourism demand is the amount of demand for a tourism product at a given price and over a given period. |
| Stabler, Papatheodorou & Sinclair (2010) | Tourist demand is the number of tourists and their spending, which can be forecast on the basis of various macroeconomic factors . |
| Wodejko (1998) | The volume of demand for a tourism product at a given price and over a given period. |
| Hunziker (1961) | Tourism demand is the sum of tourism goods, services and commodities that tourists are willing to purchase at a given price level and over a given period of time. |

Source: own compilation based on the literature.

*Figure 4 . Types of demand*




Source: own study.

Actual demand refers to the quantity of goods and services that consumers actually purchase within a given period. It primarily represents realized demand, reflecting consumers' actual purchases in the market. Actual demand is an essential indicator for market analysis because it allows us to understand what products are desired and in what quantities. It takes into account all current economic conditions, such as price levels, consumer income and product availability (Glover, 2009).

Potential demand is the anticipated or expected amount of demand that can be realised when certain conditions are met, such as an increase in consumer income or an improvement in product quality. It is demand that has not yet materialised, but may become actual in the future. Potential demand analysis is crucial for companies that want to assess growth opportunities and expansion into new markets (Uysal, 1998). Potential demand analysis can enter a new dimension with the Torusim Demand Horizon model (more in Chapter 5), which allows potential demand to be forecast with greater accuracy.

Speculative demand occurs when consumers purchase goods with a view to the future value of those goods rather than their current use. This is most often the case in financial markets (e.g. shares, currencies) and real estate markets, where buyers anticipate future price increases. Speculative demand is relevant to risk analysis as it often contributes to price bubbles that can destabilise markets. Speculative demand in the context of tourism takes on a specific meaning, especially in markets such

as investments in tourist properties, the purchase of holiday homes, participation in holiday clubs or bookings for all-inclusive trips. In this case, consumers invest in tourism goods with a view to their future value and not necessarily their current use. An example would be the purchase of property in popular tourist destinations, where the value is expected to increase due to the increasing attractiveness of the region, planned infrastructure or increased tourist influx.

Direct demand refers to the demand for products that are purchased directly by final consumers to meet their needs. Examples include the purchase of food, clothing, electronics or cars. Primary demand reflects consumers' preferences and needs for the products they consume directly and forms the basis for most market analysis in the consumer goods sector. Indirect demand, also known as derived demand, refers to the demand for goods and services that results from the demand for other products. An example is the demand for raw materials that are needed to produce other goods. For example, the demand for steel is derived from the demand for cars and buildings. This type of demand is crucial in the industrial sector, where purchasing decisions are often determined by the level of demand for final products. In the context of tourism, direct demand refers to the demand for products and services that are directly consumed by tourists, such as airline tickets, accommodation, food or guide services. Indirect demand, on the other hand, refers to the demand for goods and services that support these direct needs of tourists, such as raw materials for hotels, transport vehicles or tourism infrastructure. The interplay between direct and indirect demand is particularly evident in the context of seasonality in tourism. During peak demand periods, such as summer holidays or holiday periods, tourism businesses see an increase in demand for their services, which directly translates into larger orders of goods and services from suppliers in other sectors. For example, an increase in flight bookings during the peak season can increase demand for fuel, which in turn has a direct impact on the energy sector. In the long term, this interaction leads to infrastructure development and economic growth in tourist regions, which become more attractive to investors and tourists.

Market demand is the sum of all demand for a product or service in a particular market at a given time. It is an aggregation of the demand of all consumers and allows an assessment of how large the market is for a particular good. Market demand provides the basis for analysing market dynamics and for making decisions about development and marketing strategies (Gancarczyk, 2003).

Global demand refers to the demand for products or services in the world market. In an era of globalisation and international trade, global demand is becoming increasingly important, especially for companies operating in international markets. Understanding global demand allows companies to better adjust their sales and marketing strategies, as well as to assess the impact of global economic trends, such as changes in exchange rates or trade policies.

Tourism demand is characterised by specific properties that have a significant impact on the structure and conditions of the supply side of the market. Several key aspects are most commonly identified in the literature (Dziedzic & Skalska 2012):

- Complementarity: involves combining different elements of supply into a single tourism product. Tourist demand includes many components, such as accommodation, transport, food, attractions and other services purchased by visitors.
- Substitutability: refers to the possibility of easily substituting one product for another of similar utility value. There are two types of substitution: internal, which increases competitiveness in the sector, and external, which is a threat to tourism services.
- Mobility: indicates the need for tourists to move, while there is no spatial mobility of supply, at least in the short to medium term.
- Restitutability: assumes that the need to travel is so strong that changes in the consumer's economic situation do not immediately lead to the abandonment of tourism. This may distort the price and income elasticity of demand, and price increases or income decreases will not necessarily reduce tourism use.
- Seasonality: refers to the temporary concentration of tourism demand, which in the short term can lead to price increases and exacerbate inflationary phenomena.

In the tourism sector, demand is specific in that, among other things, it is not always satisfied at the moment of disclosure, due to the fact that tourists realise their intention to purchase after a certain period of time or are limited by factors such as income (Niezgoda, 2012) or other conditions described later in this chapter.

The volume and structure of tourism demand is influenced by many variables of an economic, demographic, social, infrastructural or marketing nature, among others

In economic terms, research by Song, Qiu and Park (2023) indicates that GDP, exchange rates and prices are key determinants of tourism demand. The GDP context can be analysed both from the perspective of the tourists' country of origin, which relates to their income, and the destination country, which reflects the development of infrastructure to attract tourists. Meanwhile, exchange rates significantly influence tourists' purchasing power, as they directly affect the attractiveness of a tourist destination through changes in travel and accommodation costs. When the value of a destination country's currency falls, the cost of travel to that country for tourists from other countries decreases. This, in turn, increases the demand for tourism as tourists can get more for their money in the destination country. Conversely, when the value of the destination country's currency increases, the cost of travel for foreign tourists increases. As a result, the demand for tourism may fall as travel

becomes less profitable. A study in India by Sharma and Pal (2020) found that exchange rate volatility has an asymmetric effect on tourism demand. This means that the response to an increase and decrease in exchange rates is not equal.

Research by Martins, Gan and Ferreira-Lopes (2017) provided an understanding of how macroeconomic variables such as exchange rates, prices and GDP affect tourism demand. The study used data from 1995 to 2012 from 218 countries and tourism demand was measured by the number of tourist arrivals and on-site spending. The study used three econometric models to analyse the panel data. The aforementioned studies also confirmed that the depreciation of the national currency increases tourism demand, making the country more affordable for foreign tourists. The impact of exchange rates on tourism demand is positive.

In contrast, a study by Muryani, Permatasari and Esquvias (2020) analysed the determinants affecting tourism demand in Indonesia using a panel model and the generalised method of moments (GMM) to analyse data for 2000-2014. They confirmed that the higher the per capita income, the higher the propensity to spend on tourism. Similar to other studies, the depreciation of the Indonesian rupiah increases tourism demand, making Indonesia more affordable for foreign tourists. The aforementioned study also describes how the terrorist attacks in 2002 and 2005, the tsunami in 2004 and the financial crisis in 2008 showing how these events had a negative impact on tourism demand in Indonesia, and how the introduction of free tourist visas in 2003 had a positive impact.

Research by Dwyer and Forsyth (2006) to analyse various econometric models for analysing the impact of exchange rates and prices on tourism demand. These models included long-run equilibrium analysis (cointegration) and error correction models to identify both long-run and short-run dynamics in the data. To obtain more reliable results, the authors used triangulation of methodologies, combining both quantitative and qualitative methods. Destructive case studies and descriptive techniques were used alongside formal econometric analyses to provide a more complete picture of the impact of exchange rates and prices on tourism demand. As the authors point out, a decrease in relative prices in the destination country, compared to prices in the tourist's home country, leads to an increase in tourism demand. Prices therefore have a direct impact on travel and accommodation costs, which is a key determinant of tourist decisions. Furthermore, the authors proved the thesis that the depreciation of the destination country's currency increases its price attractiveness, which in turn leads to an increase in tourist arrivals. The authors emphasise that tourists often react to changes in exchange rates, which can significantly influence their decisions regarding tourist destinations

The literature review also showed that Tovmasyan's (2023) study, which aimed to identify and assess factors influencing tourism demand in Armenia poiterted that a 1% increase in GDP leads to a 0.22% increase in domestic tourists while a 1% increase in CPI leads to a 0.12% decrease in domestic tourists. The study used least squares regression (OLS) and weighted least squares regression (WLS) to analyse the data. Heteroskedasticity (Breusch-Pagan test) and autocorrelation (Breusch-Godfrey test) tests were performed. The Cochrane-Orcutt methodology was applied to eliminate autocorrelation in the model.

An interesting review of data analysing demand from 32 countries sending tourists to Turkey over the period 2000-2006 was made by Gormus and Gocer (2010). These countries cover different regions of the world, allowing for a broad analysis of the determinants of tourism demand for inbound tourism to Turkey. The study used the real GDP of the sending countries as a measure of income which showed that a 1% increase in real GDP leads to a 0.19%-0.23% increase in the number of tourists. This result indicates an income elasticity of tourism demand, meaning that as income increases in the sending countries, the number of tourists travelling to Turkey increases.

In social terms, the determinants affecting tourism demand were reviewed by Sobko and Kolonowska-Matynia (2019), who used the chi-square independence test and the Chuprov convergence test to verify the hypotheses. Empirical data, on the other hand, were collected through a diagnostic survey conducted in the local tourism market. The cited studies showed that gender is an important factor influencing tourism demand, but is not the most important determinant. Contrary to some studies that emphasise the key role of gender in consumer behaviour, the results of this study indicate that gender per se is not the main determinant of tourism choices. An interesting conclusion concerns the degree of urbanisation of the place of residence. Residents of urban areas may prefer different forms of tourism compared to those living in urban areas. Urbanisation influences accessibility to tourism infrastructure and lifestyle, which in turn influences tourism choices. Similarly, family situation as a demographic determinant has a significant impact on demand, as families with children may prefer different destinations and forms of tourism compared to single travellers or couples without children. Family situation influences the choice of tourism product, including the choice of accommodation and tourist attractions. The last described element is whether the tourist is a migrant, as migratory movements and the place of origin of tourists can influence their tourism preferences. Migrants may have different tourism preferences compared to the local population due to cultural differences and life experiences.

The social aspects described by Sobko and Kolonowska-Matynia (2019), relate to culture, consumer motivations or the amount of leisure time consumers have. The culture in which a consumer

grows up or functions has a significant impact on his or her tourism choices. A consumer's cultural background influences preferences for destinations, tourism activities and leisure activities. Motivations for tourism decisions are a key factor influencing tourism demand. Motivations can be related to the need for relaxation, discovering new places, education, health or other personal goals. Undoubtedly, the increase in leisure time, resulting from increased life expectancy and improved health conditions, is encouraging the development of tourism

As Dwyer and Forsyth (2006) point out, investment in tourism infrastructure is key to long-term growth in tourism demand. The authors suggest that policies that support infrastructure development, such as the construction of new airports and improved public transport, can contribute to the attractiveness of tourist destinations. Transport is one of the most important elements of infrastructure influencing tourism demand. The availability of various modes of transport, such as airports, motorways, railways, and the quality and efficiency of these services significantly influence tourists' decisions. As Martins, Gan and Ferreira-Lopes (2017) note, the development of the transport sector allows consumers to choose from a wider range of destinations, which increases tourism demand. Accommodation and catering facilities are another key aspect of tourism infrastructure. The quality, availability and standardisation of these services are important determinants of tourism demand. A study by Borrego-Domínguez, Isla-Castillo and Rodríguez-Fernández (2022) found that the number of available beds positively correlates with tourist demand in Spain, indicating the importance of investment in accommodation to attract tourists. It is also noticeable that the expansion of accommodation tends to lead to an increase in the supply of other services, which may in turn favour existing socio economic development (Machado, Contreiras & Duarte 2019; Chau & Yan 2021). In addition, tourist attractions, such as museums, national parks, historical monuments and other places of cultural and natural interest, are key factors in attracting tourists. Recreational services such as amusement parks, sports centres and wellness facilities also play an important role in driving tourism demand. Takahashi's (2020) article highlights that regional differences in tourism infrastructure lead to differences in tourism demand, which can have significant implications for regional policy

The information obtained on tourism demand can be used by local tourism operators to better adapt their offer and to develop strategies to attract tourists with different preferences. The results of the determinant analysis can find wide application, for example in tourism marketing (Sobko & Kolonowska-Matynia, 2019).

In marketing terms, it should be noted that effective marketing in tourism has a significant impact on creating demand. An important aspect of marketing efforts is the study of content created by both professionals and consumers, which influences perceptions of tourism destinations. Increasingly,

analysis of online data, such as blog posts and online searches, is being used to predict sales of tourism packages. UGC, i.e. customer-generated content in the form of reviews, blog posts, photos and videos, is not only a form of destination promotion, but also a tool for researching market preferences and needs. This type of data is invaluable as it comes directly from people using tourism services, giving a reliable picture of their expectations and satisfaction levels. On the other hand, marketer-generated content (MGC), such as advertisements, social media posts or advertising campaigns, shapes the image of the tourism offer and builds its attractiveness.

The analysis of this data, especially in the context of new technologies such as big data, plays a key role in predicting future demand. Online data analytics tools make it possible to monitor online search trends, social media interactions, and review analysis to better tailor tourism offerings to changing consumer preferences. For example, an increase in the number of searches related to a specific destination can signal a growing interest in that destination, which provides an opportunity to adjust marketing activities accordingly, such as promoting selected travel packages.

What's more, by analysing data from the internet, tourism organisations can better understand what elements of an offer are most desired by tourists - from tourist attractions to additional services to the quality of accommodation. Based on this data, it is not only possible to better target advertising campaigns, but also to personalise the offer, which is becoming a key element of modern marketing strategies.

The literature therefore cites an extensive list of factors influencing tourism demand and one of the chosen forms of presenting research results in this area is the word cloud. Authors of studies in this area (e.g. Song, Qiu & Park, 2022) have published results presenting a word cloud of the variables that have been most frequently analysed in empirical studies related to tourism demand. Larger words in the cloud (e.g. 'GDP', 'income', 'exchange rate') indicate the variables that are most frequently used and most relevant in tourism demand studies. The popularity of these variables is due to their availability, ease of data collection and their key theoretical and empirical importance in tourism research. However, the authors emphasise that while the popularity of these variables is justified by their strong links to tourism demand, future research should carefully justify and discuss the choice of variables.

*Figure 5 . Word cloud showing factors relevant to tourism demand*




Source: Song, H., Qiu, R. T. R., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management*, *94*, 104655. DOI: 10.1016/j.tourman.2022.104655.

Due to the multiplicity of tourism forms and services, as well as the complexity of the tourism product, it can be assumed that different types of tourism demand may have different types of variables explaining the volume and structure of demand. The growing list of variables, or rather the search for their impact on tourism demand, even includes elements such as corruption or the rule of law in the destination country (Bayram et al. 2024).

Another example of searching for answers as to what factors can influence tourism demand is the system used by Amazon, which relates to the sale of books in an online bookstore, but precisely indicates how the data collected can help increase sales. Namely, this system is based not on looking for patterns and demand factors, but on analysing data and relationships between products, indicating similar recommendations in real time (Mayer-Schönberger & Cukier, 2013). In practice of research, the challenge is not only to ensure high data quality, but also to be able to produce datasets that have real relevance among others that may be irrelevant or unnecessary (Lee & Sohn, 2016).

As Panasiuk (2016b) notes, tourism demand cannot be analysed without including consumption at the place of permanent residence in the analysis and without linking one type of demand to the other. Tourist demand differs in both size and structure from residence demand. Tourists tend to be more willing to spend more during tourist trips than on a daily basis at their place of residence. Thus, the issue of tourism demand research can be analysed in a multifaceted way. Certainly, the innovative

research conducted by Mastercard (Mastercard Economics Institute, 2024), in the context of the analysis of payment card use, will make it possible to fill this research gap by using so-called big data.

*Table 19 . Research issues of tourism demand*

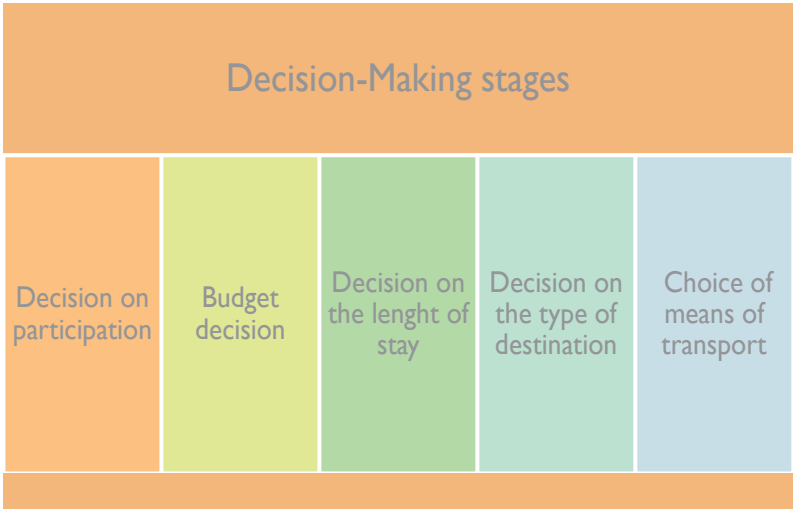
| Perspective | Examples of test criteria |
|---|---|
| Tourist demand of residents (e.g. city, region, country, social group) | Destination |
| | Departure time |
| | Form of accommodation |
| | Mode of transport |
| | Form and type of tourism |
| | Expenditure |
| | Seasonality |
| Tourist demand at the destination | Volume of tourist traffic |
| | Traffic structure |
| | Length of stay |
| | Satisfaction with your stay |
| | Impact of events |
| | Type of attraction visited |
| | Tourist spending on site |
| | Assessment of service availability |
| Demand structure processes and trends | Effectiveness of marketing campaigns |
| | Impact on other economic sectors |
| | Impact of global crises (health, military, climate, natural disasters) |
| | Changes in tourist preferences |
| | The role of technology and the digitalisation of services |

Source: own elaboration based on Panasiuk, A. (2016). *Polityka Unii Europejskiej w oddziaływaniu na wielkość i strukturę popytu turystycznego. Współczesne Problemy Ekonomiczne [European Union policy in influencing the volume and structure of tourism demand. Contemporary Economic Problems],* (13), 116. Uniwersytet Szczeciński.

Whether the decision to choose a tourism product is planned or spontaneous, big data analytics can provide information to better tailor offers to current customer needs, predict trends and optimise marketing activities in real time.

In this sense, five stages can be considered: the decision to participate; the tourism budget decision; decisions on the frequency and length of stay; the decision on the type of destination; and the final decision on the destination and choice of mode of transport (see Figure 6).

*Figure 6 . Decision-making stages*




Source: own elaboration

A broader view of the phases consumers go through may look different, as the first phase is usually the awareness of the need for leisure or travel, which is driven by internal factors such as fatigue or the need to relax, but also external factors such as advertising or recommendations from friends. The next phase is the search for information, where social media and the opinions of other consumers become important. At this point, the consumer compares different offers and alternative destinations, taking into account price, quality, availability, safety, but also personal preferences or experiences. In the next phase, the consumer already decides on a specific tourism offer and makes a purchase, a moment sometimes influenced by other incentives such as promotions or discounts, for example. Once the purchase has been made, the consumer experiences the tourism service, which influences their satisfaction. Certainly, the phases that consumers go through allow the definition of key decision-making stages (see Table 20), each of which is crucial and at the same time critical in the context of the choice of a specific tourism offer. The different decision-making stages of tourism

demand are influenced by many different factors, including age, income, education, as well as other demographic or psychological factors.

Referring to generational diversity, it should be pointed out that modern tourists can be divided into four generations (Seweryn 2017 & Rojewska, Dobrowolska, Kreft 2023), i.e:

- The Baby Boomers (BB) generation (1946-1964): Born after World War II, they grew up during the post-war economic boom. In the context of tourism, they prefer more traditional forms of travel, such as package tours, cruises and stays at established resorts as well as trips related to improving health and family relationships. They value comfort and quality of service, as well as travel safety. Although they may have difficulties with modern technology, they are willing to use it to a basic degree. People of this generation, due to the end of their careers, have more time to travel.
- Generation X (1965-1979): They seek authentic experiences in tourism, but often prefer tried and tested destinations and forms of travel. They are open to new technologies, although they often opt for solutions that combine traditional approaches with modern conveniences, e.g. using travel management apps while planning trips in a more classic way
- Generation Y (1980-1994): Generation Y, also known as Millenials, prioritise mobility and diversity of experience. In the context of tourism, they prefer independent travel, often seeking unique and 'off-the-beaten-path' destinations. They value authenticity and are willing to make spontaneous travel decisions. Raised around modern technology, they use it extensively to plan, book and share their experiences on social media.
- Generation Z (born after 1995): Familiar with technology from an early age, Generation Z sees travel as an integral part of their lives. For them, tourism is an opportunity to learn new skills and seek challenges. They use the latest apps, social media platforms and VR technology to plan and experience a trip before it even starts. They are willing to seek out the latest trends, such as eco-tourism and activity-based travel.

Each generation has its own distinctive characteristics and experiences that influence their attitude during the various decision-making stages, and the alpha generation (2010-2024) and beta generation (those born in 2025 and beyond) are already queuing up to plan their journey.

*Table 20 . Differences in decision-making by different generations of tourists*

| Generation/ The decision-making stage of the tourist | Decision to participate | Budgetary decision | Decision on length of stay | Decision on the type of destination | Choice of mode of transport |
|---|---|---|---|---|---|
| BB (1946-1964) | They prefer familiar and tried-and-tested places, often inspired by previous experiences. | They are less price-sensitive, preferring convenience and quality. | They often choose longer trips | They choose venues that offer comfort and culture and often use reviews and informative content. | They most often use cars or planes, preferring convenient modes of transport. |
| X (1965-1979) | Decisions are often motivated by family needs, the desire to provide activities for all family members. | They usually have a set budget, but are open to promotions and last-minute offers. | Trips tend to be shorter than for Boomers, due to family and work commitments. | They choose family destinations and often rely on reviews and information from the internet. | They use cars or planes, they choose reliable means of transport |
| Y (1980-1994) | They often choose destinations influenced by social media and online inspiration. | Very price-sensitive, they often look for deals and promotions. | They prefer shorter but more frequent trips throughout the year. | They like variety - from active leisure to culinary travel, and often choose less popular destinations. | They are keen to use low-cost airlines and public transport |
| Z (1995-2012) | Decisions are heavily inspired by social media and online trends. | Budget is important, many rely on family support. | They prefer shorter stays and often make decisions shortly before departure. | They choose destinations that offer unique experiences and are keen to explore lesser-known places. | They use low-cost airlines, preferring environmentally friendly transport options. |

Source: Seweryn, R. (2017). *Pokolenie Baby Boomers na rynku turystycznym na przykładzie odwiedzających Kraków*. Marketing i Zarządzanie, 2(48), 105–116. https://doi.org/10.18276/miz.2017.48-10. Rojewska, K., Dobrowolska, K., & Kreft, P. (2023). *Analiza pokoleń – jak zrozumieć współczesnego turystę?* Qtravel.ai. https://www.qtravel.ai/wp-content/uploads/2023/06/Analiza-pokole-jak-zrozumie-wspczesnego-turyst-1.pdf

## 2.2 Tourism demand determinants

There are many models in the literature outlining what determinants influence the volume and structure of tourism demand (Eugenio-Martin 2004, Assaf & Josiassen 2012, Song & Li 2008, Divisekera 2013, Ibragimov, Perles-Ribes & Ramon-Rodriquez 2017), and one of the most popular is the division into economic, demographic, geographical, legal and political, media influence and social factors, which was proposed by Middleton and described by Niezgoda (2012), among others.

Each of these elements can be further subdivided into smaller elements such as economic variables can be divided into economy-wide, income or price variables (Wodejko, 1998).

Factor influencing tourism demand is household income, calculated on a per capita basis. Tourism needs are categorised as higher-order, meaning that they are financed from a free consumption fund, arising after basic needs have been met. As a result, tourism becomes possible once an adequate level of income is reached. Some forms of tourism, such as health tourism or family travel, can be considered as primary goods, which lowers the income threshold.

*Table 21 . Economic sources of impact on tourism demand*

| Group of factors | Economic variables |
|---|---|
| Occurring in the area generating tourist traffic | the level of individual income (and especially the free consumption fund), the distribution of income (this appears to be mainly a division between income earmarked for tourism and goods substitutable to it), entitlement to leave, the length of paid leave and the extent to which it is taken, purchasing power of the currency, tax policy and control of tourism expenditure |
| Occurring in the tourist reception area | price level,competition in the market, quality of tourism products, economic regulations in the tourism sector |
| Other | the relationship between prices in the two areas (especially the prices of tourism services), promotional activities undertaken by tourist reception areas in the sending area, exchange rates, relationship of travel time to cost |

Source: Bull, A. (2005). *The economics of travel and tourism* (2nd ed.). Oxford: Longman

It can be seen from Table 21 that the operation of tourism is strongly influenced by a number of economic variables in both the tourism generating area and the tourist reception area. Factors such as income levels, distribution of funds, purchasing power of currency, prices of tourism services, and exchange rates influence tourism demand, but economic regulations and promotional activities undertaken by tourist destinations are also important and can stimulate interest in travel.

However, it should be stressed that other factors that are not directly related to the rationality of human behaviour in an economic context also play a key role in tourism. These factors include, among others, fashion, i.e. the popularity of certain destinations at a given time, which often goes beyond economic justifications, and tourists choose destinations because of their momentary appeal in media and culture. Other such factors include social trends or cultural preferences, as, for example, some groups of tourists travel to certain places for cultural or religious reasons, which are not always based on purely economic calculations. Emotional motivations or climatic and environmental factors are also important.

Leisure provisions vary widely around the world, both in terms of annual and weekly days off. Nonetheless, there is a global trend towards longer leisure periods, which encourages increased travel opportunities. Bull (2005) points out that when the main constraint to the growth of tourism is scarcity of time, rather than insufficient income (as is the case, for example, in Japan). Thus, any liberalisation of labour laws, leading to longer holidays or weekends, results in a high elasticity of tourism demand. This manifests itself in a greater interest in longer distance trips and longer lengths of stay, which translates into higher spending by tourists.

Thus, consumers' purchasing decisions in the tourism market are much more complex than just economic choices, and their behaviour is strongly influenced by a variety of non-economic factors. Social, cultural, psychological and emotional influences are important in shaping tourism demand, making the study of these determinants challenging. While economic aspects such as income or prices can be easily measured and understood, factors such as membership of social groups, the need for self-fulfilment or concern for image are difficult to quantify, yet have a huge impact on consumer decisions. For example, fashion and the desire to belong to a certain reference group often shape tourists' choices, regardless of the financial cost. Consumers choose destinations that are currently popular among their community, which may not have an economic rationale, but are driven by a desire to be up-to-date or follow trends. An example would be travelling to well-known destinations that are trending on social media. In such a context, even if other destinations offer better financial or logistical conditions, tourists will still choose the trendy ones to mark their belonging to a particular cultural group.

Psychological theories such as Maslow's hierarchy of needs or Veblen's theory of motivation (Scott, 2017) show how important emotional and status factors are in the decision-making process. Tourism often fulfils needs for self-fulfilment and prestige, as manifested, for example, in ostentatious consumption. Trips to luxury destinations or the use of expensive tourist services can be a manifestation of social status and a way of building one's image. This is the so-called demonstration effect, where consumers choose tourism products not only for their functionality, but also because

others around them choose similar services. In short, the Veblen effect in tourism shows how people are often driven not only by the value of the experience itself, but also by its social perception, leading to an increase in demand for luxury travel and services. This dynamic process can also vary from culture to culture, where different societies view ostentatious consumption and luxury experiences differently.

Maslow's hierarchy of needs (Osei, 2022), on the other hand, is widely used in tourism as a tool to understand travellers' motivations. Maslow distinguishes five levels of needs that can be satisfied through tourism. In Maslow's hierarchy of needs, tourists at the first level seek satisfaction of basic physiological needs such as food, shelter and rest. The provision of adequate accommodation and food is crucial. Next, safety needs refer to a sense of protection from threats, so tourists choose destinations that they perceive as safe, avoiding places with conflicts or health risks. At the level of needs for belonging and love, tourists seek social interaction and bonding with loved ones, making group travel or visiting family popular. The needs for recognition and prestige, on the other hand, manifest themselves in the desire to travel to prestigious destinations and enjoy luxury services, which can emphasise social status. In this case, the Veblen effect makes the higher price of tourism services more attractive. At the highest level, self-fulfilment needs relate to personal development, discovering new cultures and realising one's full potential, which is often associated with adventure and educational tourism.

Demographic changes such as an ageing population also affect tourism demand, as older people are increasingly participating in the tourism market. The industry needs to adapt its offerings to their specific needs, such as the availability of infrastructure or recreational programmes. Senior tourism is gaining importance, especially in European countries where the ageing process is progressing rapidly. Equally important is the sense of security, which significantly influences tourism decisions. Threats, such as terrorism or natural disasters, temporarily reduce demand, but data show that once the threat ceases, interest quickly returns. In turn, the individualisation of consumer needs is leading to a growing demand for travel offers tailored to unique preferences, making flexible, personalised offers more popular than traditional packages. Understanding these motivations is key to effectively designing travel offers that can meet the different needs of travellers at different stages of their lives.

Prices are widely recognised in the literature as a key determinant shaping the demand for tourism services. According to basic economic principles, there is an inverse relationship between prices and demand: an increase in prices leads to a decrease in demand, while a decrease in prices triggers an increase in interest in a particular product or service. In the case of tourism, price changes affect not only demand for specific tourism services, but also consumer behaviour in relation to

substitute and complementary goods. It is also important to note that price changes induce two important phenomena. The first is the effect on real incomes - a reduction in the price of tourism services increases consumers' available resources, which translates into an increase in demand as tourists can afford more services or a higher standard. The second effect is the so-called substitution effect, which means that a decrease in the price of tourism services makes them more attractive compared to substitute goods (e.g. alternative leisure activities), which become more expensive in relative terms.

Price is a key indicator in tourism demand analysis, as it helps to understand how sensitive tourists are to changes in the price of services offered, such as accommodation, transport or tourist attractions. Price elasticity of demand is measured by the ratio of the percentage change in demand volume to the percentage change in price. If the coefficient of price elasticity of demand is greater than 1, it means that the demand for a tourism product is elastic, i.e. tourists respond significantly to price changes. If the coefficient is less than 1, we speak of inelastic demand, meaning that price changes have little impact on the number of tourists or the number of bookings. The literature, including a study by Biryukova et al. (2019), highlights that the price elasticity of tourism demand can vary significantly by region, type of tourism and target group. For example, demand for luxury tourism tends to be less elastic, as customers from higher income segments are less sensitive to price changes. In contrast, for mass tourism, especially during the low season, the price elasticity of demand can be much higher, with the result that tour operators often introduce discounts to attract more customers and fill vacancies in hotels or transport. Price management in the tourism sector is a key element of marketing strategy, and considering price elasticity of demand allows companies to optimise revenues and maximise profits. During periods of economic crisis, price elasticity of demand plays a particular role as tourists become more sensitive to price changes, which can lead to a shift in demand to cheaper forms of tourism, such as domestic tourism or last minute offers.

Thus, prices not only affect the level of tourism demand directly, but also indirectly, by changing the structure of consumers' spending and their decisions to choose between different leisure options (Wodejko, 1998) .

Analysing the impact of prices on the tourism market requires a segment-specific approach, as the response to price changes can vary depending on the type of tourism. For example, business tourism tends to have lower price elasticity, meaning that price changes have little impact on demand in this segment. Shopping-related tourism, on the other hand, may show greater sensitivity to price changes. In addition to price factors, there are other determinants affecting demand and price elasticity in the tourism sector. One of these factors is urgency, which makes consumers willing to incur higher costs

in order to satisfy their need immediately, regardless of price. Another important factor is the share of tourism expenditure in the consumer's budget - the higher the percentage of the budget spent on tourism trips, the lower the willingness to accept high prices. Another important aspect is the possibility of subsidies, for example in the framework of social tourism, where the effect of shared expenditure reduces price sensitivity. The principle of complementarity, or the so-called sunk money effect, also shapes consumer behaviour - if the purchase of one product conditions the correct functioning of another, consumers are willing to pay more for complementary products. In addition, the unique value of a product, or the unique value effect, can result in an increased willingness to pay a higher price. It is also worth considering the effect of awareness of the existence of substitutes - the easier it is to substitute a product or service for another of similar utility, the greater the price elasticity. Ultimately, the quality-price ratio plays an important role in shaping consumer decisions in the tourism market - the more attractive the quality-price ratio, the greater the willingness to use a given tourism offer, which affects the price elasticity of demand (Dziedzic & Skalska, 2012).

In the context of tourism demand analysis, reference should also be made to income elasticity of demand, which is one of the key concepts in economics. It refers to how demand for a product changes in response to changes in consumer income. In simple terms, it is a measure of the sensitivity of demand to changes in income. There are various theories and empirical studies in the economic literature that attempt to describe and explain this mechanism. Based on the available research, a number of factors influencing income elasticity of demand can be identified, as well as applications of the concept in different contexts, including international tourism.

Income elasticity of demand is defined as the ratio of the percentage change in demand to the percentage change in consumer income. The formula for calculating income elasticity is as follows:

*Figure 7 . Formula for income elasticity of demand*

$$E_Y = \frac{\%\Delta Q}{\%\Delta Y}$$

Where:

- E_Y is the income elasticity of demand,
- %Delta Q represents the percentage change in the quantity of goods purchased,
- %Delta Y is the percentage change in consumer income.

Source: Panasiuk, A. (2020). Marketing w turystyce i rekreacji [Marketing in tourism and recreation]. PWN.

Depending on the value of elasticity, we can distinguish between three main product categories (Crouch, 1992):

- Normal goods - when income elasticity is positive. An increase in income leads to an increase in demand for these goods. An example is overseas tourism, where higher incomes encourage more frequent travel.
- Luxury goods - if the income elasticity exceeds 1, it means that an increase in income leads to a more than proportional increase in demand. International tourism often falls into this category, especially in developed countries.
- Lower-end goods - when income elasticity is negative. When income increases, demand for these goods decreases as consumers switch to higher quality products. In tourism, this can be a choice of more exclusive destinations instead of local ones.

Income elasticity of demand, or the sensitivity of demand to changes in income, as already mentioned, is a key concept in economics, and its value can vary depending on a number of factors. For example, luxury products tend to have a higher income elasticity than basic products. In the tourism sector, trips to exotic, remote destinations, which are perceived as luxuries, show greater sensitivity to changes in consumer income than local trips, which can be treated as a basic need (Crouch, 1992). A second important factor affecting income elasticity of demand is the level of consumer income. In higher income countries, demand for luxury goods, including international travel, has a higher elasticity. Consumers with higher incomes are more willing to spend their surplus income on goods that offer prestige, comfort and uniqueness. In lower-income countries, a greater proportion of the budget is spent on basic needs, making the income elasticity of demand for luxury goods, such as international travel, lower (Var, Mohammad & Icoz, 1990).

The income structure of society also plays an important role in shaping the income elasticity of demand. An increase in income inequality leads to differences in consumer responses to changes in income. Higher income groups may be less sensitive to further increases in income because their consumption needs are already largely satisfied. In contrast, middle- and low-income groups may respond much more intensely to increases in income, as manifested by an increased demand for luxury goods to which they previously had no access (Uysal & Crompton, 1984). According to Crouch's (1992) research, the average income elasticity of demand for international tourism is around 1.76, suggesting that it is more sensitive to changes in income than many other sectors. However, these values can vary by region and market specificity. For example, in developing countries, where

international travel is still a luxury for most of the population, the income elasticity may be higher than in developed countries.

Tourism demand is shaped not only by factors related to general macroeconomic conditions, but also by policy interventions targeting specific aspects of the tourism market. Policy actors can use a variety of instruments of an economic, legal and administrative nature to influence both demand and supply in the tourism sector. These interactions are often reciprocal and difficult to separate, as incentives directed at one area (e.g. demand) can also affect the other (e.g. supply). The main policy tools affecting tourism demand include regulated prices and taxes and tax exemptions, especially for services targeting marginalised groups. Supporting demand can also be done through subsidies for services of a social nature, such as support from social funds of workplaces, which increases the accessibility of tourism for wider social groups. Equally important are instruments related to the financing of specific services, such as spa tourism, where public funds, such as the National Health Fund, play a significant role. In addition, the availability of consumer credit and currency regulations have a direct impact on consumer decisions to spend on international tourism. Other demand-supporting mechanisms include visa and customs fees, which can restrict, on the one hand, and encourage international travel on the other. Attention should also be paid to the role of tourism promotion, which aims not only to increase awareness of the tourism offer, but also to create demand, especially in segments of new markets. From this point of view, economic policy, through a variety of regulatory and financial tools, has the capacity to shape tourism demand in a multidimensional way. In the context of global challenges, such as climate change or rising transport costs, these interventions can play a key role in stabilising and developing tourism markets, especially in regions with a greater dependence on tourism.

## 2.3 Tourist decision-making models

There are many publications in the literature on models of the tourist decision-making process. As described by Eugenio-Martin (2003), the choice of tourist destination is determined by four other decisions and the decision-making process includes five stages: the decision to participate, the decision on the tourism budget, the frequency and length of stay, the type of destination and the final choice of place and mode of transport.

The model presented by Barczak (2014) illustrates the tourist's decision-making process in the context of choosing a tourist offer. This process consists of several stages, which are influenced by various factors such as regional and business competition. The first stage is to decide on the type of

tourism the tourist wants to practice. This can be leisure tourism, cultural tourism, adventure tourism, health tourism, etc. The choice of the form of tourism depends on the preferences, interests and needs of the tourist. This choice can be influenced by various factors such as tourist trends, recommendations from friends, availability of information and offers presented by travel agencies. The tourist then decides on a specific region to travel to. This can be domestic or foreign, urban or rural, mountainous, coastal, etc. This decision-making stage is influenced by the marketing activities of the tourist regions, their competitiveness, as well as the opinions of other tourists. Competing regions are also important in this stage as they try to attract as many tourists as possible through attractive offers, promotions and unique attractions. The final stage is the selection of a specific tourism offer, including accommodation, attractions, food and other travel-related services. This choice is influenced by competing tourism businesses such as travel agents, hotels, carriers and other travel service providers. These companies seek to attract customers through competitive prices, high quality services and extras in the form of special offers and promotions.

Another popular TNS travel model - the Travel Model (Department of Industry, Tourism and Investment) outlines the seven main stages tourists go through when planning and executing their trips. These stages include:

- Dreaming: in this phase, travel planners begin to consider different destinations and holiday ideas.
- Planning: during this stage, they gather information and seek recommendations on where to go.
- Booking: refers to the physical process of booking tickets, accommodation and other travel elements.
- Anticipating: preparing to leave, associated with excitement, excitement and sometimes some apprehension.
- On the road: The moment when travellers hit the road and the holiday officially begins.
- Destination: time spent on holiday, at a destination.
- Post-Holiday Sharing: returning to everyday life and sharing memories, often on social media.

In contrast, the tourist decision-making process, as described by Smallman and Moore (2010), emphasises the complexity and multi-layered nature of the decisions made by tourists. Traditional models of tourist decision-making often focus on a 'choice set' approach, where independent variables such as destination attractiveness or price explain tourists' choices. However, Smallman and Moore point out that this approach does not take into account decision-making processes as continuous and dynamic activities that evolve as travel progresses and tourists interact with the environment.

According to Smallman and Moore (2010), tourists' decision-making process should be considered in a more naturalistic way, including both emotions and unexpected factors that influence tourists' choices in real time. Traditional models also fail to take into account the complex social and contextual relationships that often shape tourists' decisions as a group. Rather than treating decisions as single-minded and based solely on rational considerations, the authors suggest that many decisions are made intuitively or spontaneously, and that their ultimate shape depends on the environment in which the tourist is located. Smallman and Moore point to the need for a broader study of tourists' decision-making processes based on narratives that can better capture the changing and emergent actions that lead to the choice of tourist destination or activity.

McCabe, Li and Chen (2015) argue that tourism decisions are dynamic and involve two cognitive systems:

- Level one, intuitive, fast, automatic and less effortful, based on heuristics. This is an emotional and simplified system that can be used when tourists have a low level of involvement in the decision-making process. For example, tourists may choose destinations that are familiar to them based on recognition or popularity of the destination among friends (social heuristics).
- The second, rational, level is slower, more analytical and requires more effort. This system involves more complex reasoning based on logic, and is used when tourists have a high level of involvement in the decision, e.g. carefully analysing the available information and comparing alternatives.

Niezgoda and Zmyślony (2003) presented a models in which they emphasised the complexity of tourists' purchasing decisions. In their view, the process is a multi-stage one and tourist decisions are shaped by a number of factors that are not limited to economic aspects alone. The authors pointed out that a tourism service is supposed to satisfy a tourist's complex of needs, which means that the choice of a tourism offer depends on a number of variables, including economic, emotional, social and situational variables. In addition, an extremely important element of this process is the tourist's previous experience, knowledge and skills, which influence the final choice of tourist offer.

One of the key elements highlighted by Niezgoda and Zmyślony (2003) is the role of the tourist's experience, which plays an important role in their subsequent choices. This experience shapes the tourist's expectations, as well as their ability to evaluate the different offers available on the market. Tourists who have travelled before are more likely to analyse different options and to search for information about the destination. In this context, information search and analysis skills become extremely important. These tourists are able to assess which sources of information are reliable and make travel planning decisions based on this information. Nowadays, with the development of

technology and easy access to the internet, the role of these skills has increased as tourists have access to a wide range of information that can influence their choices. The modern tourist no longer relies solely on traditional forms of information, such as travel guides or travel agencies. Instead, tourists are increasingly planning their trips themselves, relying on information available online. This independent travel planning process gives them more control over their decisions and allows them to tailor their offerings to their individual needs. Another important aspect is the ability to analyse information, which allows tourists to better assess the offers available and their value. Tourists are able to compare different offers in terms of price, quality, location or additional amenities. This enables them to choose the offer that best suits their needs and budget. This ability also avoids potential problems associated with choosing an offer that might not meet their expectations. In this context, tourists are becoming more informed consumers who make considered decisions rather than choosing offers on impulse.

Other aspects influencing the comfort of a holiday in a destination include knowledge of the local language or the ability to get around using local means of transport. Knowledge of the local language can make communication with locals much easier, which has a positive impact on the overall travel experience. Tourists who know the basics of the language are more likely to get helpful information from locals, which can enrich their tourism experience. Additionally, being able to use local communications makes tourists more mobile, allowing them to explore places outside of the main tourist attractions on their own. These types of skills can greatly enhance the enjoyment of a trip, as tourists feel more independent and confident in an unfamiliar environment.

The emotional and social aspects mentioned by Niezgoda and Zmyślony (2003) also play an important role in tourists' decision-making. Tourism is often a way of escaping everyday stresses and problems, as well as an opportunity to experience new emotions and relationships. Tourists may be motivated by the desire to experience new cultures, to make social connections, or simply to relax and unwind. The emotional and social aspects associated with travel influence which destinations and forms of tourism are chosen. For example, tourists seeking tranquillity and contact with nature may prefer rural areas, while those interested in nightlife and urban attractions will choose big cities.

Tourist decisions often involve both systems at different stages of the process, e.g. when choosing a destination, tourists may start with an intuitive choice and then move on to analyse the available options, especially when faced with a high information load or time constraint. The model also assumes that individual factors such as cognitive ability, level of engagement, or access to information can influence which decision-making system is used at any given time. In the publication cited above, McCabe, Li and Chen (2015) outlined a process diagram that begins with the tourist's decision to choose a destination, which is influenced by individual differences such as the need for

cognitive engagement, belief in intuition and demographic factors. Next, the level of involvement of the tourist in the decision-making process is a key factor. When the level of involvement is low, level one is triggered, which is emotional, intuitive, automatic and requires less cognitive effort. In this case, the tourist may be guided by heuristics such as recognition heuristics (choosing a destination based on familiarity with it) or social heuristics (choosing a popular destination visited by friends or celebrities). At a high level of engagement, level two is triggered, which requires more cognitive effort and is more rational and complex. At this stage, the tourist undertakes an information search that depends on their cognitive abilities and demographic factors. If the tourist experiences cognitive overload during this process, he or she may revert to simpler decision-making strategies in level one, using heuristics such as lexicographic heuristics (selecting destinations based on a single most important criterion, such as price) or trade-off heuristics (balancing different characteristics). However, if cognitive overload is low, the tourist continues the decision in level two by moving to a complex evaluation of the available options, such as utility maximisation (choosing the destination offering the greatest benefit). Ultimately, the choice of destination may be the result of an interaction between both systems, depending on the level of engagement and complexity of the situation. The model also takes into account the principles of systems involvement that influence decision-making, such as the accuracy-effort principle, the avoidance of negative emotions and the need to justify the outcome of a decision.

Moore et al (2012) investigated the decisions made by tourists on the ground, i.e. while travelling, and their analysis focuses on understanding how different factors influence the decision-making process in the context of travel to New Zealand. The model emerged from an intensive qualitative analysis based on 140 interviews with international tourists. The authors proposed a four-dimensional 'cascade model' that describes the decision-making process as the result of the interaction of several underlying dimensions: (in)flexibility, place and time of decision, social group composition and stage of travel. The dimension of (in)flexibility describes the degree of openness to modification of plans. At one extreme of this dimension are decisions that are rigidly defined before the start of the trip, such as the choice of routes and accommodation, which are 'locked in' before arrival. These types of decisions have less flexibility and are more difficult to change. An example of this is when a tourist books accommodation in advance for their entire stay. On the other hand, there are flexible decisions that can be modified on the fly, such as the choice of daily activities or places that the tourist visits spontaneously. An example of a flexible decision could be a tourist who decides to visit places recommended by other travellers or local people. The flexibility of the decision also depends on the type of trip - tourists travelling for a shorter period of time are more likely to plan in advance, while those staying longer are more likely to make decisions spontaneously. The second dimension relates

to where and when decisions are made - before the trip (off-site) or already on-site. Pre-trip decisions tend to be more rigid and related to key elements of the trip, such as the itinerary or main attractions. On-site decisions, on the other hand, mainly concern day-to-day activities or minor aspects of the trip. The timing of decision-making can also change during the trip. At the beginning of a trip, tourists tend to plan more, and as the trip progresses they become more open to spontaneity. In addition, accommodation decisions may be more flexible in the middle phase of the trip, when the tourist has more freedom to choose accommodation. Another key dimension is the composition of the social group with which the tourist is travelling. Decisions made by tourists are often the result of interactions with other members of the travel group, as well as with people met during the trip. For example, the decisions of families travelling with children may require more planning and some arrangements before the trip to meet the needs of all group members. However, the composition of the group may change during the trip - people travelling to visit friends or family may spend part of the trip together and part independently. In this case, on-site decisions may still be influenced by the advice and suggestions of friends, even if they are not travelling together. The last dimension, trip stage, refers to changes in decision-making style at different stages of the trip. Tourists in the initial phase of a trip tend to have more fixed plans, while in the middle phase they make more flexible decisions. In the final phase of the trip, decisions may change depending on priorities, such as choosing more comfortable accommodation before returning home.

The decision-making model presented by Dwityas & Briandana (2017), on the other hand, describes the impact of social media on tourism decision-making. The authors of the study point out that in the era of Web 2.0 and evolving information and communication technology, the way tourists gather information and make travel decisions has changed significantly. Social media has become a key source of information, influencing all stages of the process, from planning to the post-visit experience. The decision-making process, according to the authors, in travel is divided into three main phases: before the trip (pre-trip), during the trip (during-trip) and after the trip (post-trip). In each of these phases, social media plays an important role by providing information and enabling interaction between users (see Table 22).

*Table 22 . Phrases of the tourist decision-making process*

| Name of the decision phase | Description |
|---|---|
| Pre-travel phase (pre-trip) | In the pre-travel phase, the tourist identifies the need to travel, starting from the awareness stage. Users, by reviewing content posted by others, identify potential destinations and available tourism products. |
| | The tourist then moves on to the information-gathering stage. At this point, social media acts as an invaluable source of knowledge about destinations, available attractions, transport and accommodation. Tourists benefit from the experiences of other users by reading reviews, viewing photos and videos and comparing different options. |
| | The alternatives evaluation stage, in which tourists evaluate the information collected, based on User Created Content (UCC). UCC, in the form of opinions, reviews and ratings, allows for a better understanding of the available options and their actual value. |
| | In the pre-travel phase, the tourist moves on to the decision to travel, i.e. finalising plans and travel-related purchases such as airline tickets and hotel bookings. |
| During-trip phase | During the additional preparation phase, the tourist may seek additional information about the places they plan to visit. |
| | Tourists also use social media to share their experiences in real time. By posting photos, status updates and tagging places, tourists create content that can be useful to other travellers. Comments and reactions from other users in real time become part of the process of sharing information and enhancing the tourist experience. |
| Post-trip phase | Once the trip is complete, tourists move into the trip evaluation phase, which includes reflection on the completed trip, evaluation of the tourism products and services and the information resources they used to make their decisions. During this phase, tourists share their experiences by posting reviews, photos and videos of their trip on social media platforms. |
| | One of the most important elements of this phase is the assessment of travel satisfaction. Satisfaction, or lack thereof, influences the future travel decisions of the tourist and the decisions of other users who use the content provided. |

Source: own elaboration based on Dwityas, N. A., & Briandana, R. (2017). Social media in travel decision making process. *International Journal of Humanities and Social Science, 7*(7), 193-201.

In the post-pandemic era and warfare in Ukraine or the Middle East, the analysis of risk factors is becoming more and more prevalent in research, which is also considered in relation to tourists' decision-making model. Risk and uncertainty play a key role in tourists' decision-making. In the tourism literature, risk is understood as an assessment of the possibility that undesirable events, such as accidents, health problems, crime or political instability, will happen during the trip. Uncertainty, on the other hand, refers to the incomplete knowledge that tourists have about the destination, which can cause difficulties in predicting how the trip will go. Karl's (2016) findings show that tourists differ in their attitudes to risk. Some are more willing to make risky decisions, such as travelling to less developed or politically unstable destinations. Other tourists, on the other hand, are more risk-averse and prefer safe, tried-and-tested destinations with developed tourism infrastructure where the level of

risk is low. One of the key elements of the model is the development of a typology of tourists based on their attitudes to risk and uncertainty in the decision-making process. The authors identified five main types of tourists:

- Risk and uncertainty avoiders - Tourists belonging to this group prefer well-known destinations with a developed tourist infrastructure. They tend to avoid places that may be dangerous and prefer organised guided tours. They often return to places they have visited before and avoid experimenting with new destinations.
- Risk-averse - Tourists in this group are less willing to take risks but, compared to the first group, are more open to travelling to new places, provided these places are safe and well organised. They plan their trips themselves, although they prefer them to take place in predictable conditions. They are reluctant to take safety risks, but may try new foods or engage in less risky activities.
- Newcomers looking for safety - These tourists like to explore new places, but in a safe way. They prefer destinations that are new, but at the same time well organised and free of major risks. They often use tours organised by travel agencies to minimise risk and uncertainty, but value the opportunity to explore new regions.
- Novelty and adventure seekers - Tourists belonging to this group actively seek out new and less familiar destinations. They are keen to try local food and engage in spontaneous, less organised forms of travel, which involves a higher level of risk. They organise their trips independently and the safety and infrastructure of the destination is not a priority for them. It is more important for them to experience adventure and discover new cultures.
- Risk-takers - This is the riskiest group of tourists, who deliberately choose destinations with low security and minimal tourist infrastructure. They make decisions spontaneously, often without a specific plan for their trip, which involves a high level of uncertainty. These tourists value intense experiences and are willing to take risks to get them.

The analysed model also takes into account the influence of other factors, such as socio-demographic variables, on the decision-making process. Research indicates that age, education level, profession, and travel experience play a significant role in shaping tourists' attitudes towards risk.

- Age - older tourists, especially those over 60, are more likely to be risk-averse. They are more likely to choose safe, well-known destinations. Younger tourists, especially those aged 20-29, are more open to risky travel and experimenting with less popular destinations.

- Education - higher levels of education are associated with a greater willingness to take risks. Tourists with higher education are more likely to choose less familiar destinations, even if this involves a higher level of uncertainty.
- Travel experience - the greater the travel experience, the greater the propensity to take risks. Tourists who travel frequently are more confident in situations that require dealing with uncertainty and are therefore more likely to choose more exotic and risky destinations.

The model comprehensively depicts how different types of tourists make destination choice decisions, taking into account risk and uncertainty. It also shows that decision-making in the context of travel is a dynamic process in which both the individual characteristics of tourists and the characteristics of destinations influence each other. Risk and uncertainty are key factors that shape tourists' final decisions.

*Table 23 . Group decision-making process*

| Name of the decision process | Description |
|---|---|
| Information gathering | Groups of tourists seek information on a variety of topics, such as transport, tourist attractions, accommodation or food. An important element of this stage is the variety of information formats, including texts, photos, reviews or videos. Participants in the study showed that they use a variety of sources, such as social media (Facebook, Weibo), search engines (Google, Baidu), maps (Google Maps) and travel portals (TripAdvisor, C-trip) |
| Sharing information within the group | Once the information has been collected, there is a sharing stage within the group. Group members use various communication tools such as instant messaging (WhatsApp, WeChat), email, phone calls or online collaboration tools (Google Docs). Depending on the size of the group and the degree of intimacy between its members, the preference for tools may vary. |
| Knowledge transfer | The next step is the transfer of knowledge between group members. In this model, knowledge is understood as processed information that group members transform using previous experiences and newly acquired information. This process is based on the theory of encoding and decoding - senders encode information according to their meaning structures and receivers decode it based on their own interpretations. Differences in perception between senders and receivers can lead to misunderstandings, which in turn affect the group's decisions. |
| Decision-making | The final stage is decision-making**.** Based on the knowledge gathered, group members make choices about the itinerary. The most commonly used tool for decision-making is instant messaging, which facilitates quick and efficient sharing of information. However, there are also some disadvantages, such as lack of clarity or overlooking important messages |

in the flurry of communication. Members of large groups often have to sift through lengthy discussions to understand how final decisions were arrived at, which can be time-consuming and annoying.

Source: own elaboration based on Zhang, L., Sun, X., & Wagner, C. (2018). Exploring the group holiday decision-making process with the support of technology. *Information Processing & Management, 54*(3), 397-409.

The decision-making model described in Zhang, Sun and Wagner (2018) focuses on the decision-making process of tourism groups using technology. The authors analyse how groups of tourists gather, share information and make decisions when planning a holiday together. Participants in the study showed that the use of online tools such as WhatsApp, WeChat, Google maps and travel portals significantly facilitates the collection and sharing of information and joint decision-making. Groups with more travel experience rely less on detailed information about accommodation and tourist attractions, while less experienced groups seek more detailed, official information. Zhang, Sun and Wagner's model points out that the decision-making process varies according to factors such as group size, intimacy between group members, travel experience and cultural differences. For example, Chinese tourists showed a greater propensity to use travel portals offering comprehensive information about destinations, while tourists from Western countries are more likely to use maps and photos of places. In addition, larger groups made up of people with weaker ties are more likely to use social media to share information, while smaller groups prefer more direct methods of communication, such as phone calls.

## 2.4 Theoretical basis for market segmentation in tourism

One of the forerunners of segmentation theory was Smith (1956), who proposed dividing the market into smaller, homogeneous groups of consumers with common characteristics or needs. This approach was a response to economic theories of perfect competition, with Saunders (1980) pointing out that organisations should identify differences in the consumer population to better respond to demand, as the population is heterogeneous. In the literature, the most common types of segmentation are demographic segmentation, geographical segmentation, psychographic segmentation, behavioural segmentation.

Developments in technology and big data analytics allow for a more dynamic and personalized approach to segmentation, taking into account not only static demographic characteristics, but also changing consumer preferences and behavior in real time. Some researchers (Kotler et al., 2017) point to the growing role of predictive segmentation based on artificial intelligence algorithms, which may

challenge traditional approaches based on rigid categories. Critics of classical segmentation also point out that in reality, consumers rarely belong to a single, well-defined group, and their preferences can be fluid and change depending on situational context and cultural influences.

Consequently, while segmentation remains a key tool in tourism marketing and management, its effectiveness depends on its ability to integrate new methods of data analysis and adapt to dynamic changes in consumer behavior.

Demographic segmentation is based on characteristics such as age, gender, occupational status, income and education, facilitating the identification of consumer preferences. Geographical segmentation divides the market by location, taking into account the influence of climate, culture and infrastructure. Psychographic segmentation, on the other hand, analyses values, lifestyles and interests, providing a deeper understanding of consumer motivations. The last one, behavioural segmentation, focuses on actual behaviours such as purchasing patterns and brand loyalty, particularly useful for digital marketing.

According to the research review presented by Dutta and Bhattacharya (2015), most of the studies using the types of segmentation described above covered specific tourism reception areas (see Figure 8-9), with the pattern of the segmentation process in the tourism sector generally being similar.

From the perspective of a tourism region, the segmentation process begins with defining the target markets, i.e. identifying which groups of tourists will be analysed with a view to attracting them to the area. Appropriate segmentation variables are then selected, such as demographic, psychographic or behavioural characteristics, which will serve as the foundation for further analysis of tourist preferences.

The next step is to conduct research, usually in the form of surveys or other methods of collecting data from visitors or potential tourists. The aim of this stage is to gather information about their expectations, motivations and behavioural patterns. The collected data are then analysed using methods such as factor analysis or clustering techniques, which allow segments of tourists with similar preferences to be distinguished. This enables the tourist region to better understand the diversity of the visiting groups and tailor its offer to their unique needs. This type of segmentation enables tourism organisations and the region itself to develop marketing strategies more effectively, which in turn helps to attract the right groups of tourists and improve their level of satisfaction with their visit.

*Figure 8 . Market segmentation process in the tourism sector*




Source: own compilation based on Dutta, S., & Bhattacharya, S. (2015). Segmentation research in tourism: a brief literature review. *TSM Business Review*, *3*(1), 69.

Due to the diversity of needs, preferences, attitudes and the different life stages and lifestyles of consumers, there is no single variable that can be used effectively to segment tourism markets (Andereck & Caldwell, 1994). In tourism, it is recommended to use multivariate segmentation or a combination of different variables to obtain a more precise analysis than with only one variable (Middleton, 1994; Havitz & Dimanche, 1990; Morrison, 1994).

*Figure 9 . Overview of research on tourism demand segmentation*

| **Basis for segmentation** | **Details of the study** |
|---|---|
| Demographic | Beh & Bruyere, 2007 (Kenya, 465*), Carmichael & Smith, 2004 (Canada, 333428), Dolnicar, 2005 (Australia, 430), Frochot, 2005 (Scotland, 734), Hsieh et al, 1992 (Hong Kong, 807), Jang et al., 2004 (France, 496), Johns & Gyimothy, 2002 (Denmark, 1099), Kastenholz et al., 1999 (Portugal, 187), Kim & Jogaratnam, 2003 (USA, 514), Koh et al, 2010 (USA, 214), Mohsin & Ryan, 2007 (India, 1026), Molera & Albaladejo, 2007 (Spain, 335), Loker-Murphy, 1997 (Australia, 690), Oh et al, 1995 (Australia, 1030), Park & Yoon, 2009 (Korea, 252), Pesonen et al., 2011 (Finland, 195), Rid et al., 2014 (Gambia, 430), Ryan & Glendon, 1998 (UK, 1127), Sarigollu & Huang, 2005 (Latin America, 265) |
| Geographical | Beh & Bruyere, 2007 (Kenya, 465), Carmichael & Smith, 2004 (Canada, 333428), Dolnicar, 2005 (Australia, 430), Frochot, 2005 (Scotland, 734), Hsieh et al., 1992 (Hong Kong, 807), Jang et al., 2004 (France, 496), Johns & Gyimothy, 2002 (Denmark, 1099), Kim & Jogaratnam, 2003 (USA, 514), Koh et al, 2010 (USA, 214), Mohsin & Ryan, 2007 (India, 1026), Molera & Albaladejo, 2007 (Spain, 335), Loker-Murphy, 1997 (Australia, 690), Oh et al, 1995 (Australia, 1030), Park & Yoon, 2009 (Korea, 252), Pesonen et al., 2011 (Finland, 195), Rid et al., 2014 (Gambia, 430), Ryan & Glendon, 1998 (UK, 1127), Sarigollu & Huang, 2005 (Latin America, 265) |
| Psychographic | Beh & Bruyere, 2007 (Kenya, 465), Carmichael & Smith, 2004 (Canada, 333428), Dolnicar, 2005 (Australia, 430), Frochot, 2005 (Scotland, 734), Hsieh et al., 1992 (Hong Kong, 807), Jang et al., 2004 (France, 496), Johns & Gyimothy, 2002 (Denmark, 1099), Kim & Jogaratnam, 2003 (USA, 514), Koh et al, 2010 (USA, 214), Mohsin & Ryan, 2007 (India, 1026), Molera & Albaladejo, 2007 (Spain, 335), Loker-Murphy, 1997 (Australia, 690), Oh et al, 1995 (Australia, 1030), Park & Yoon, 2009 (Korea, 252), Pesonen et al., 2011 (Finland, 195), Rid et al., 2014 (Gambia, 430), Ryan & Glendon, 1998 (UK, 1127), Sarigollu & Huang, 2005 (Latin America, 265) |

| | |
|---|---|
| Behavioural | Carmichael & Smith, 2004 (Canada, 333428), Hsieh et al., 1992 (Hong Kong, 807), Johns & Gyimothy, 2002 (Denmark, 1099), Kim & Jogaratnam, 2003 (USA, 514), Koh et al, 2010 (USA, 214), Mohsin & Ryan, 2007 (India, 1026), Molera & Albaladejo, 2007 (Spain, 335), Loker-Murphy, 1997 (Australia, 690), Oh et al., 1995 (Australia, 1030), Pesonen et al., 2011 (Finland, 195), Ryan & Glendon, 1998 (UK, 1127), Sarigollu & Huang, 2005 (Latin America, 265) |

*numbers in brackets indicate sample size

Source: own compilation based on Dutta, S., & Bhattacharya, S. (2015). Segmentation research in tourism: a brief literature review. *TSM Business Review*, *3*(1), 69.

Market segmentation brings strategic and tactical benefits. Strategically, it allows companies to assess customer profitability and focus resources on key segments, while helping to analyse the competition and build lasting relationships. Tactically segmentation promotes a deeper understanding of customers, enabling offers to be better tailored to their needs, increasing satisfaction and brand loyalty.

From the perspective of market segmentation analysis, not only its definition, which implies the development of the demand side of the market and implies the rational and precise adaptation of product and marketing activities to the requirements of the consumer or user, but also the objectives and benefits that segmentation brings, is important. The purpose of segmentation is to group buyers on the basis of similar needs and behaviour and to introduce marketing programmes dedicated to them. Segmentation criteria can relate to both the consumer and the product, as described in Table 24.

*Table 24 . Segmentation criteria*

| Criterion | | Description |
|---|---|---|
| Criteria relating to the consumer | Socio-economic | Income, occupation, education, place of residence |
| | Demographic | Age, gender, family size, phase of the family life cycle, nationality |
| | Psychographic | Lifestyle, activities, interests, opinions |
| Product-related criteria | Consumption patterns | Frequency of use, ownership of or access to other products, brand loyalty |
| | Terms and conditions of purchase | Type of shop, time of purchase, volume of one-off purchase, frequency of purchase |
| | Benefits offered | Consumer product knowledge, perception of benefits |

Source: Żuchowski, I. (2007). *Segmentacja rynku - ujęcie teoretyczne [Segmentation of the market - theoretical approach]. Zeszyty Naukowe WSES*, (4), 119-132.

Nowadays, tourism stakeholders are increasingly defining the behavioural profiles of tourists in order to better adapt marketing tools to specific groups. The conducted assessment of tourism products assumes three levels of detail: area products, understood as constellations of attractive places, tourist routes and forms of tourism. The behavioural profile of a tourist is a comprehensive description of his/her behaviour, preferences, activities and decisions made during the trip. The analysis of this

profile focuses on understanding how the tourist reacts to different stimuli, what choices he or she makes and what his or her preferences are, so that the tourism offer can be better tailored to individual needs and expectations. The basic criteria defining a tourist's behavioural profile are presented in Figure 10.

*Figure 10 . Elements in defining a tourist's behavioural profile*




Source: own study.

Example of defining tourist segments is the study for the capital city of Warsaw (Walas, 2021). It was based on analyses of data, reports and publications, surveys among Warsaw tourism entrepreneurs and Warsaw tourist information points, workshops with representatives of the city hall, members of the WOT and expert analysis. Based on the aforementioned analyses, the following behavioural profiles were defined (Walas, 2021):

- domestic tourists: the occasional visitor, the economic newcomer, the conference taster of the city, the school excursionist, the enthusiastic explorer;
- foreign tourists: individual cultural explorer, package minimalist, memory traveller, sentimental polonus, lifestyle traveller;
- residents: local micro-tourist, opinionated temporary resident.

*Figure 11 . Package minimalist - example of a behavioural profile for Warsaw*

*They arrive by coach or plane with a travel agency or another organiser for the first time. They represent the 40+ age group. According to the programme, they stay for two to three nights, often as part of a sightseeing tour around Poland. They stay in a hotel. They accept attractions from the "Top 10" list included in the package but show little interest in attractions not included in the itinerary. They are curious about what the city and its daily life look like, and the modernity of the place often surprises them. Following the programme, in the evening, they "disappear," depending on their national temperament, to go, for example, to a restaurant, pub, or just wander around the city. The coach takes them from one attraction to another. They listen attentively to the guide. They are not particularly interested in Polish cuisine, although they want to try it at least once. Being part of a group provides them with a sense of safety. While shopping, they look for handicrafts and local souvenirs.*

Source: Walas, B. (2021). *Segmenty turystów, obszary produktowe oraz rynki emisyjne - weryfikacja [Tourist segments, product areas and issue markets – verification]*. Warszawa: Wydawnictwo WSTiE.

An important study is the segmentation of domestic tourists conducted by Kantar for the Polish Tourist Organisation between 28 June and 8 July 2019. The main aim of the survey was to understand the tourism behaviour of Poles, defined as those who have taken at least three tourist trips in the last 24 months, each lasting a minimum of one night, including at least one paid trip carried out in Poland. A total of 2,650 Poles aged 18 and over took part in the survey. Of these, 12% were those who had not taken a tourist trip in the past two years. The majority of respondents, 88%, said they had travelled as tourists, of which 42% were domestic tourists. The survey focused on 1,000 domestic tourists, outlining the preferences and behaviours of Polish travellers (Kantar, 2019). The survey in question divided domestic tourists into five segments:

- Undemanding Kowalscy (Smiths) (25%) - the largest group, who are likely to value simple and accessible forms of tourism with no special requirements for luxury or unique attractions.
- Spontaneous athletes (23%) - a group that values physical activity and spontaneity during trips, likely to prefer active leisure and outdoor sports.
- Travel agency customers (20%) - people who use travel agencies, suggesting that they value convenience and organised leisure activities.
- Freebirds (18%) - relaxed tourists, probably preferring the freedom to plan and execute their trips.
- Holiday enthusiasts (14%) - the smallest group for whom tourism is a passion and travel is carefully planned with a variety of experiences in mind.

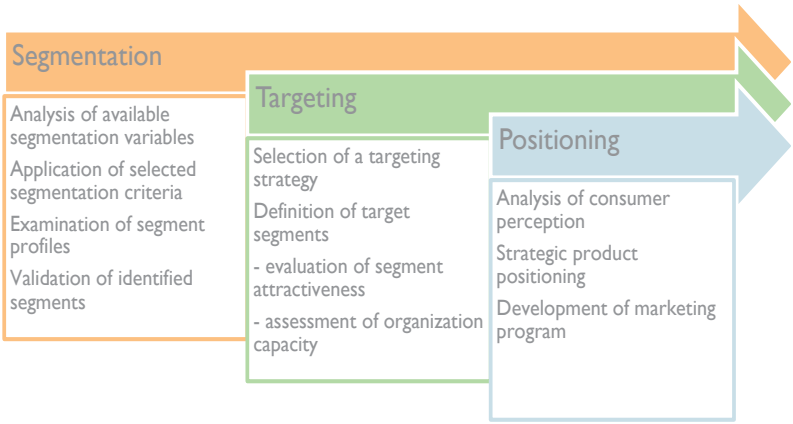
The use of big data and AI in the customer segmentation process can bring many advantages over traditional segmentation methods. Mandapuram et al. (2020) pointed out among the benefits of such a solution that algorithms work without human bias and discover new patterns in the data that

humans may not notice. With such tools, it is also possible to achieve a high level of personalisation while being able to scale to the entire market. The use of AI in the marketing process, thanks to precise segmentation, makes it possible to offer customers a more digital and individually tailored shopping experience (Florez-Lopez & Ramon-Jeronimo, 2009). Content is better tailored to customers' needs, allowing for more effective segmentation and relationship building according to consumer preferences.

Huang and Rust (2021), referring to the possibilities of segmentation using AI tools, indicated a division into mechanical, thinking and feeling AI in relation to the unique application at different stages of marketing planning. They defined mechanical AI as remote sensors, machine translation, classification and clustering algorithms that allow for systemisation and standardisation of activities. They defined thinking AI as machine learning algorithms and neural networks that recognise patterns in texts, voice or images. Sentient AI, on the other hand, is described by the authors as a tool for understanding emotional customers and for personalising relationships through sentiment analysis. The division indicated seems reasonable, although it needs clarification that it does not refer to AI as sentient, thinking and mechanical but refers to its use. The authors have provided a detailed review of research in the scopes discussed, including from the perspective of marketing strategies in terms of STP (segmentation, targeting, positioning).

Research shows that data analytics can be used to identify patterns that are not always obvious to marketers. For example, text processing and machine learning techniques can automatically analyse loan applications, classifying borrowers as 'reliable' (those who will repay) and 'unreliable' (those who will not repay) (Netzer et al., 2019). On the other hand, automated text analysis together with correspondence analysis allows for psychographic segmentation of customers in the art market (Pitt et al., 2020). Data mining can also be used to segment tourists based on their perceptions of the importance of different destinations, with better results than traditional grouping methods (Valls et al., 2018).

*Figure 12 . Segmentation Targeting Positioning process*




Source: Dibb, S., & Simkin, L. (2008). *Market segmentation success: Making it happen!* Routledge.

Marketing (STP) is crucial not only for businesses, but also for tourism regions that aim to attract specific groups of tourists. In the context of tourism region management, the STP model enables tourism offerings to be tailored to the specific needs of different market segments, which increases a region's attractiveness and allows it to compete effectively with other destinations. Through segmentation, regions can better understand their target tourist groups and then tailor infrastructure, services and promotional campaigns to meet their expectations. Targeting, on the other hand, allows a region to focus on those segments that are most profitable and have the greatest potential to attract longer-stay or higher-spending tourists. Positioning, in turn, allows the region to build a consistent image that differentiates it from competing tourist destinations. The STP process can bring many benefits to both the tourist region and the company, but it also comes with some disadvantages as well as advantages (see Table 25).

Table 25 . Advantages and disadvantages of the STP model from an organisational and regional perspective

| Aspect | Advantages for the region | Disadvantages for the region | Advantages for organisations | Disadvantages for organisations |
|---|---|---|---|---|
| Segmentation | Better understanding of the needs of different groups of tourists, ability to adapt infrastructure | High costs and time-consuming market analysis | Allows you to focus on specific customer groups | The need for accurate data, which generates costs |
| Targeting | More effective attraction of tourist segments with high economic potential | Possibility of missing out on less profitable but culturally or locally relevant segments | Efficient allocation of resources to the most profitable segments | Risk of missing out on other potential customer groups |
| Positioning | Opportunity to build a strong image for the region, differentiating it from its competitors | Difficulties in changing the image once it has been defined | Strengthening the brand and its recognition among target customers | Costly marketing activities and the need for ongoing investment in maintaining brand positioning |

Source: own elaboration

In summary, the STP model allows, at regional level, a more personalised approach to tourists and a fine-tuning of the offer, which is conducive to building a competitive image. However, it is important to bear in mind the potential limitations, such as the high cost of analysis and the risk of missing out on less popular but important groups of tourists.

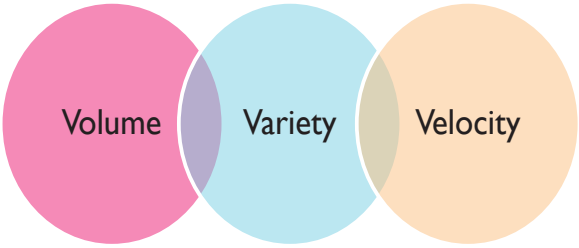
# CHAPTER III:

# TECHNOLOGIES, METHODS AND TOOLS USED IN BIG DATA ANALYSIS

## 3.1. The importance of big data in today's economy

With the development of the digital economy, the quantity and quality of digitally created information is steadily increasing, forming a variety of data sets called big data. Their analysis and processing are valuable because they can lead to new knowledge (Kachniewska, 2014). Big Data analytics are used wherever large volumes of digital data require the acquisition of new information or knowledge (Douglas, 2011). This phenomenon is intensifying with the increasing availability of the internet and e-services, which are naturally adapted to the use of databases. As a rule, scientific publications point to three main characteristics of such collections, i.e. volume, variety and velocity - see Figure 13. This definition refers, firstly, to the large volumes of data, which are most often measured in terabytes or petabytes, making it possible to conduct analyses with a much greater degree of detail and accuracy than ever before (for an overview of big data definitions, see Table 26). Secondly, the variety of data types that come from many sources (e.g. social media) makes it necessary to look for models that allow data to be analysed in such a way that they can complement rather than exclude each other. On top of this, thirdly, the speed with which new data is generated and delivered becomes crucial in the process of monitoring and responding quickly to changing needs of, for example, consumers.

*Figure 13 . Attributes of big data (traditional view)*




Source: own compilation based on P. Russom, Big data Analytics, TDWI Best Practices Report, The Data Warehousing Institute, 2011, p. 6.

Some authors (Gandomi & Haider 2015, Fan & Bifet 2013, Borne 20114) extend this concept to 5V, 7V and even 10V (see Figure 14). In these approaches, there is also reference to 'veracity', which emphasises that data can be inconsistent, incomplete and inaccurate and also 'value', which shows the

value hidden in the data. The broadest concept also includes 'variability' and 'virtual data' (Kisiolek, 2018).

In the context of Big Data, volume as already mentioned refers to the huge amount of data generated and collected every day. Data comes from a variety of sources, such as social media, sensors or online transactions, leading to the creation of huge collections. Velocity, on the other hand, refers to the speed at which data is processed and analysed. Today, many systems require real-time analysis for immediate results. Another aspect is variety, i.e. the variety of formats and types of data that are analysed - this can be structured, semi-structured and unstructured data such as text, images or video. An important attribute is also reliability (veracity), which refers to the quality and accuracy of the data. Not all data are equally reliable, which can lead to erroneous conclusions if not properly validated. The validity of the data (validity), i.e. the relevance and correctness of the data in a given context, is crucial when making decisions based on analysis. The ultimate goal of data analysis is to derive value (worth), i.e. the benefits that can be derived from the data. Businesses need to consider how the data can deliver profits or other benefits. The variability of data, i.e. its dynamic changes over time, is also a challenge, so Big Data analytics must be flexible and ready to process changing information. Another aspect is venue, which refers to the location of data generation and processing. On a global scale, data can come from different systems, countries and platforms, which affects how it is managed. It is also important to understand the specific terminology and standards used in data analysis, which relates to vocabulary. Different industries may use different conventions, which requires an appropriate alignment of analytical tools and methods. The last important attribute is vagueness, which refers to the uncertainty and ambiguity of the data, which can make it difficult to interpret.

*Figure 14 . Attributes of big data (contemporary view)*

| | |
|---|---|
| **Attributes of big data** | volume - refers to the vast amount of data generated and stored in systems, often exceeding traditional processing capabilities |
| | •velocity - represents the speed at which data is generated, transmitted, and processed, requiring technologies capable of real-time analysis. |
| | •variety - refers to the diversity of data formats, including structured (e.g., databases) and unstructured (e.g., multimedia, text). |
| | •veracity - concerns the quality and reliability of data, which may be incomplete, inconsistent, or erroneous, affecting its analytical usefulness. |
| | •validity - refers to the accuracy and correctness of data in the context of its application, which is crucial for obtaining reliable analytical results. |
| | •value - highlights the hidden value within data, which can be extracted through proper processing and analysis for informed decision-making. |
| | •varibility - refers to changes in data patterns over time, indicating that trends and relationships may evolve dynamically. |
| | •venue - relates to the location and sources of data, which may originate from various platforms, devices, and geographically distributed systems. |
| | •vocabulary - encompasses the semantics and terminology used in data analysis, which is essential for accurate interpretation of information |
| | •vagueness - refers to the ambiguity and imprecision of data, which can lead to misinterpretations if not properly processed and categorized. |

Source: Borne K. Top 10 big data challenges: a serious look at 10 big data vs, Tech. Rep. 2014 [Online]. Available: https://www.mapr.com/blog/top-10-big-data-challenges-look-10-big-data-v

Of course, although big data makes it possible to review and analyse larger data sets this does not at all mean that all of this data represents useful information. From the perspective of, for example, Google Advertaising ID (GAID) or other social media-generated data, it is important to note that one user may have multiple accounts, or, for example, one account may be used by multiple users. This raises significant issues related to fault tolerance and data quality (Katal, Wazid & Goudar, 2013). Undoubtedly, the use of big data collections offers a number of advantages (see - Fig. no. 15).

*Figure 15 . List of the benefits of big data analytics.*

| | |
|---|---|
| List of the benefits of big data analysis | •More effcient targeting of marketing activities |
| | •Precise segmentation of customer base |
| | •Better recognition of market opportunities |
| | •Improvement of business decisions through faster business decision-making |
| | •Automation of decision-making in real-time processes |
| | •Trend analysis |
| | •Market forecasting |
| | •Innovation design |
| | •Optimisation of marketing and production activities |
| | •Relationship management |
| | •Understanding consumer behaviour |

Source: own compilation based on 1) P. Russom, Big data Analytics, TDWI Best Practices Report, The Data Warehousing Institute, 2011, p. 11. 2) Vriens, M. & Kidd, P. (2014, November/December). The big data shift: what every marketer needs to know about advanced analytics. *Marketing Insights*, 23-29. 3) Bean, R. (2017). How companies say they're using big data. *Harvard Business Review*. https://hbr.org/2017/04_4) Kalkowski, J. (2016). *Handling a flood of data*. https://www.brandpackaging.com/articles/85628

Figure 15 shows the various benefits of Big Data analytics, which are crucial for effective business operations. Big Data analytics enables more accurate targeting of marketing activities and precise segmentation of the customer base to better tailor offers to consumer needs. Rapid business decisions based on Big Data analysis enable better identification of market opportunities and improved strategic decisions. The automation of real-time decision-making processes and the analysis of market trends support the prediction of future market behaviour, which is essential for designing innovations and optimising marketing and production activities. As a result, organisations can better manage customer relationships and understand customer behaviour, resulting in increased business efficiency.

*Table 26 . Overview of big data definitions*

| Author | Definition |
|---|---|
| Cierpiał-Wolan, Stateva (2023) | A way of acquiring knowledge and learning about reality that is made possible by new technologies for creating and processing large data sets. New sources of data, combined with innovative methods of processing them (particularly machine learning), make it possible, in many cases, to publish information on socio-economic phenomena and processes in real time, as well as to improve the quality of forecasts. |
| Iwasiński (2022) | Big data are large data sets (at least peta- or even exabytes), characterised by diversity (different types, formats) and high flow dynamics (data are produced and recorded at a rapid pace) |
| Kuzmar S. (2019) after Cox and Ellsworth, (1997) | Big data is large datasets with analytical potential, the amount of which needs to be maximised in order to extract informational value. |
| Rojszczak (2019) | Big data is a technology that allows the prediction of future events based on the analysis of large sets of data from different sources, often used to predict user behaviour. |
| McKinsey Global Instutute (2016) | Big data is defined by three characteristics, i.e. 1) Volume: Data is extremely voluminous, ranging from terabytes to zettabytes 2) Variety: Big Data includes structured, semi-structured and unstructured multimedia data. 3) Velocity: Data streams must be processed and analysed in real time. |
| Fasel, Meier (2016) | Big Data is data that, due to its size, variety of structures and speed of change and availability, cannot be processed in traditional database systems and requires the use of NoSQL technology. |
| Gartner Group | Additional characteristics according to the Gartner Group are the additional two Vs, i.e. value: Big Data applications should increase the value of a company, and veracity: data is often ambiguous, requiring specific algorithms to assess its reliability. |
| Maślankowski (2015) | Big data are tools that allow the storage and processing of large data sets, which would not be possible using traditional methods such as relational databases. The definition includes characteristics such as large volume, high variability (Velocity) and high diversity (Variety). In addition, emphasises reliability (Veracity) and data value. |
| Kachniewska (2014) | Big data are data sets characterised by significant size, variability and diversity. Their processing and analysis are valuable as they can lead to new knowledge. These characteristics are referred to as the 4Vs: Volume, Velocity, Variety, Value. |
| Sagiroglu, Sinanc (2013) | Big data are data sets of large volume, variety and complexity that are difficult to store, analyse and visualise. The process of examining large amounts of data to uncover hidden patterns and correlations is called big data analytics. |
| Russom (2011) | The definition of big data is based on three Vs: Volume, Variety and Velocity (speed). These three characteristics are key to understanding big data and its implications for analytics. |

Source: own study based on the cited authors.

The scale of data creation in the digital economy is difficult to estimate, but it is nevertheless estimated that the amount of data collected is increasing by 35-50% every year. This is due, for example, to the number of payment card transactions made every second, the amount of content published on social networks and many other sources. This data is subject to analysis as it provides a wealth of valuable information about customers, which translates into personalising marketing campaigns, estimating sales, or strengthening customer loyalty. Many business models of companies such as Google, Meta, Amazon, LinkedIn or Polish Allegro are created on the analysis of these data sets (Kachniewska, 2014).

The information shared is a valuable source of data collected and used by different actors according to their needs. This includes not only the data that the user uploads independently and consciously, but also those collected based on the user's activities in cyberspace, such as our preferences, the shops we visit and the goods and services we view or choose (Stępień, 2018).

The internet is a major source of information retrieval and processing, creating huge datasets analysed as required. Contrary to popular belief, big data is not anonymous and unimportant. From a data protection point of view, it is valuable information, often having the status of personal data, including sensitive data. In an age of technological advances, the use of big data can bring numerous benefits and support development in various social, economic and political areas, but it also poses a serious threat to individuals' privacy (Stępień, 2018).

The optimism with which some analysts and many companies, such as those in the retail and insurance industries, are approaching big data analytics indicates the viability of such investments, even though their full potential is only just beginning to be recognised (Smith, 2015). Surveys of Fortune 1000 companies between 2012 and 2017 show that more and more companies are seeing the benefits of investing in big data analytics systems. In 2017, executives at more than 80 per cent of the largest US companies considered these investments successful, with almost 50 per cent of them pointing to tangible, measurable results (Bean, 2017). According to Hopkins, vice-president of Forrester Research, about 75% of US companies consider big data to be crucial to their business, although less than 30% of them admit to being able to make full use of the knowledge gained, indicating a skills problem in advanced data analytics (Woźniczka, 2018). However, nowadays, correlating data is becoming more important than understanding shopper behaviour.

Big data analytics is contrasted with so-called traditional marketing analytics, pointing out that, in contrast, it focuses on the processing of large streams of incoming data in real or near-real time instead of using specific metrics to indirectly and with some delay evaluate individual marketing investments (Xu, Frankwick & Ramirez, 2016). Quite quickly, it has become apparent that the role of

research based on big data sets in at least some of the areas of marketing interest is becoming primary over research conducted by other methods. At the same time, it is pointed out that the value of marketing research results increases when big data analysis is combined with other research methods in the research process (Woźniczka, 2018).

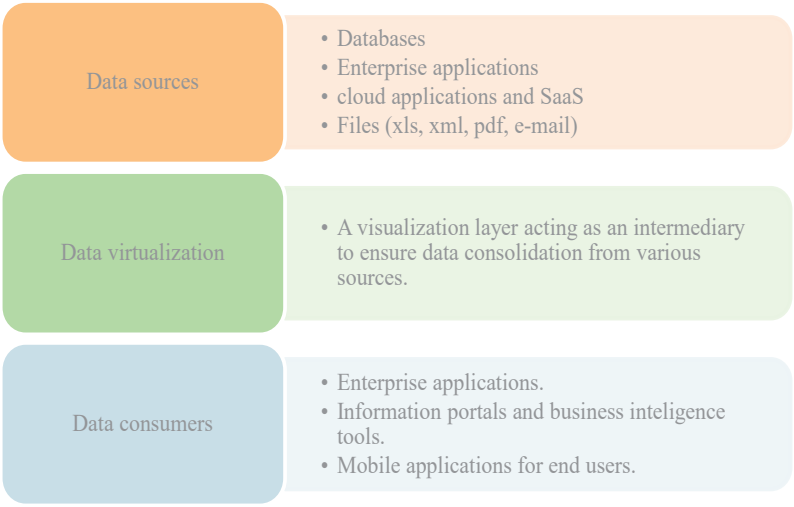
Research shows the beneficial impact of big data on product development. With access to multiple data sources in real time, companies using big data analytics can continuously observe consumer behaviour and competitors' actions. They can also monitor customers' reactions to new products and services and obtain detailed information about their opinions on product features, pricing and sales strategies (Xu, Frankwick & Ramirez, 2016). Big data sources can be used in public statistics in several ways (Kitchin, 2015):

- Partial replacement of existing statistical sources (current data).
- Provide additional statistical information in the same area but from a different perspective (additional data).
- Revision of estimates from existing statistical sources (revised data).
- Provision of completely new statistical information in the field (new alternative data source).

Big data is stored in what is known as a data lake, which is a place to store large amounts of raw data in its orginal form. The data lake concept was first introduced by CTO Pentaho Dixon (EMC², 2011)

An important element of the use of datasets is their visualisation capabilities. This process is the link between data sources and their consumers. In the case of tourism sector data, it can provide a pathway from the data zipped about tourists to their consumers, for example, the management units of a tourism reception area in a region. This process is illustrated in Figure 16.

*Figure 16 . The process of linking data sources with consumers*




Source: own compilation based on Singh, A. (2021). Methods for data lakes implementation. Intercathedra 3(48), p. 154. http://dx.doi.org/10.17306/J.INTERCATHEDRA.2021.00130

Figure 16 depicts the process of linking different data sources to data consumers in an organisation. This process has three main components: data sources, data virtualisation and data consumers (Data Consumers). In the model presented in chapter five, i.e. the Tourism Demand Horizon, the 'data virtualisation' and 'data consumers' stages have been combined into one (third) element, while, in second place, there is an AI module to integrate data and analyse behavioural patterns. Data sources include various types of data, such as databases, enterprise applications, cloud applications and SaaS services, as well as files in various formats such as xls, xml, pdf or emails. These are the primary sources from which data is collected for analysis.

Data virtualisation is an intermediary layer that acts as a tool to consolidate data from many different sources. It provides a unified representation of data, facilitating further processing and analysis without the need to physically move or copy data to a single location. Data consumers, on the other hand, are the ultimate recipients of the processed data, such as enterprise applications, information portals and business intelligence tools, as well as end-user mobile applications. Data consumers use the provided information for business decision-making, reporting or further analysis.

*Table 27 . Advantages and disadvantages of using big data*

| Disadvantages | Advantages |
|---|---|
| SOCIAL | |
| Invasion of privacy is a serious problem associated with big data. Data is often collected without users' knowledge and consent, raising ethical and legal questions | Big data can improve social services by better understanding user needs and behaviour and responding more quickly to changing conditions |
| Big data has the potential to reinforce existing inequalities and create new ones (only those who provide data are considered in decision-making) | The use of big data in social media analysis can provide valuable information on public opinion and sentiment, which can help to better manage social initiatives and policies |
| ECONOMIC | |
| The high costs associated with the implementation of big data technology, including hardware, software and specialist employment, can be a barrier for many businesses | Big data analytics can lead to more precise and faster business decisions, increasing operational efficiency and optimising business processes |
| Lack of adequate analytical skills in companies can lead to the full potential of big data not being realised | Companies can better anticipate market changes and identify trends, allowing for better strategic planning and increased competitiveness |
| Companies may ignore or underestimate the preferences and behaviours of consumers who do not shop in ways that big data tools can easily capture and analyse (e.g. certain promotions may not be offered to them, etc.). | |
| POLITICAL | |
| Risk of abuse by governments or organisations that may use the data for surveillance and control of the public, which may lead to violations of civil rights | Big data can support public policymaking by providing accurate and timely data to better understand social and economic issues |
| | Big data analytics can help predict and prevent crime and manage crises, helping to improve public safety |
| TECHNOLOGICAL | |
| Technological complexity and the need for specialised staff can pose significant challenges for companies implementing big data | The development of big data contributes to technological advances through the development of new analytical tools and algorithms, allowing for more sophisticated data analysis |
| Problems with integrating different systems and data sources can lead to inefficiencies and errors in analysis | The ability to process massive amounts of data in real time enables faster and more precise responses to changes in the business environment |
| ENVIRONMENTAL | |
| Processing and storing huge amounts of data requires a significant amount of energy, | Big data can support environmental action by analysing data on emissions, natural resource management and climate change monitoring |

| | |
|---|---|
| which can lead to an increased carbon footprint and negative environmental impact | |
| LEGAL | |
| Existing legislation often fails to keep up with the rapid development of big data technology, leading to loopholes and potential abuse | The introduction of appropriate regulations can protect users' privacy and ensure the safe and ethical use of data |
| Regulatory complexity and the diversity of regulations in different countries can hinder the global implementation and management of big data | Big data can support law enforcement by analysing criminal data and supporting investigations |

Source: own compilation based on Harford, T. (2014, December). Big data: are we making a big mistake? *The Financial Times.* Retrieved from https://www.ft.com

Table 27 shows the advantages and disadvantages of using big data technology in different areas. The table outlines the advantages and disadvantages of using Big Data technology in several key areas: social, economic, political, technological, environmental and legal. Each of these areas offers unique benefits, but also poses significant challenges for Big Data users that require detailed consideration if the technology is to be used effectively and responsibly.

In the social context, big data brings both benefits and significant risks. On the positive side, big data can improve social services and facilitate a better understanding of user behaviour in a dynamic changing environment. Analysing big data sets enables more accurate predictions, which in turn can lead to more efficient management of resources and support actions to improve quality of life. This enables social institutions and organisations to better tailor their services to the needs of citizens. However, with these benefits also come serious problems. Violations of user privacy are one of the biggest challenges in the context of big data. Often, data is collected without the full knowledge or consent of users, raising ethical and civil rights concerns. In addition, there is a risk that big data can be used to reinforce existing social inequalities, as resources and technologies are most often available to larger and wealthier actors. Maintaining a balance between the potential benefits and the risks of privacy and ethics violations is therefore crucial.

From an economic perspective, big data offers companies the opportunity to optimise business processes, increase operational efficiency and make decisions faster. Analysing big data sets allows new market trends to be identified, which in turn enables companies to do better strategic planning and increase competitiveness. The data can also support the anticipation of market changes, allowing companies to better prepare for future challenges and opportunities.

However, the implementation of big data comes at a high cost, both in terms of hardware and skilled personnel, which can be a barrier for smaller companies. A lack of appropriate analytical tools

and knowledge can lead to the full potential of big data not being realised. Companies may also ignore data on consumer preferences and behaviour, which can negatively impact their bottom line if they fail to use it appropriately.

Big Data also has a significant impact on the political sphere. On the positive side, this technology can support the creation of better public policies, enabling the prediction of crises, the analysis of demographic and economic data, which can ultimately contribute to improving the quality of life of citizens. Big Data can help to better manage crises, as well as identify potential security risks. There is a risk of abuse by government bodies that may use the data for excessive control and surveillance of citizens. These abuses can lead to violations of civil rights, especially in countries with authoritarian regimes where access to data can be used to maintain power.

Big Data contributes to technological advances by supporting the development of analytical tools and enabling more advanced data analysis. The ability to process big data in real time is becoming a key tool for companies and institutions that want to be competitive in the global market. However, the technological complexity of big data and the need to work with new tools is challenging. Implementing new systems and integrating them with existing structures can be complicated, leading to delays and operational problems. Another challenge is the compatibility of data from different sources, which can lead to difficulties in analysing it.

In the environmental context, Big Data can contribute to environmental protection by enabling the monitoring of climate change and the analysis of data on energy consumption, water resources and other natural resources. As a result, Big Data technologies can support environmental initiatives and sustainable development. However, processing and storing huge amounts of data requires significant amounts of energy, which can have a negative impact on the environment. The increase in energy consumption and IT infrastructure can contribute to greenhouse gas emissions, which contradicts sustainability efforts.

The final area that the table covers is the legal context. The benefits of Big Data, such as supporting the fight against crime and identifying threats, are significant, but at the same time require the creation of appropriate legal regulations. Current legal regulations do not always keep up with rapid technological developments, leading to ambiguities in privacy and data protection issues. The lack of adequate regulations and differences in international laws can lead to violations of user privacy and the use of data in an unethical manner. Consequently, there is a need to develop coherent legal provisions to adequately manage data and protect users' privacy. The implementation of these laws should go hand in hand with the development of Big Data technologies to ensure that data is used responsibly and in accordance with the applicable legal standards.

A review of the literature has shown that users often accept the terms of use of online services, including social media, without first reading the terms and conditions and their assumptions. After agreeing, users upload huge amounts of data, which are later collected and used for big data analysis (Urban, 2014). As shown in a study conducted among 272 respondents at the Wrocław University of Economics, as many as 57% of respondents did not carefully read the terms and conditions of a social media site and 73% did not read the privacy policy, with 73% of them being aware, as they indicated that they knew how cookies worked (Chomiak-Orsa & Mrozek, 2017).

*Figure 17 . The concept of data matching*

| | |
|---|---|
| Data Matching Concepts | Deterministic record matching |
| | Probabilitic record matching |
| | Data fusion |
| | Statistical matching |
| Identity Resolution Techniques | Text matching algorithms |
| | Tokenization and vectorization |
| Duplitace Detection Techniques | Fuzzy Matching |
| | Machine learning algorithms |
| | Image utilisation |

Source: Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer.

The figure number 17 presents the concepts of data matching, which refer to the process of identifying and linking records relating to the same object in different datasets. Data matching methods can be divided into four main groups (Cierpiał-Wolan & Stateva, 2023):

- deterministic record matching: Simple logical rules match key variables such as an identification number or address. Deviations lead to Type I (false match) or Type II (no record match) errors,
- probabilistic record matching: Estimating the probability of random matches and mismatches takes into account uncertainty and lack of complete data, allowing for a flexible approach to record matching,
- data fusion: Merges two datasets to create a new dataset containing all unique records, and missing data is filled in using interpolation methods,

- statistical matching: A procedure involving micro (unit-level pooling) and macro (parameter estimation taking into account the covariance matrix of the variables) approaches.

Identity resolution is the process of identifying and integrating records relating to the same entity. Techniques include text matching algorithms such as Levenshtein or Jaro-Winkler, tokenisation and vectorisation, which allows the meaning of text to be expressed in numerical form. Duplicate detection involves identifying and eliminating duplicate records. Techniques include fuzzy matching, K-NN algorithms, clustering methods and neural networks that find similarities between data. Algorithms that compare images, such as colour histograms and feature descriptors, can be used to deduplicate data on objects advertised on web portals, allowing data from different sources to be combined efficiently. Nowadays, it is becoming a huge challenge for statistics to provide good quality information, which should be available almost in real time and a prerequisite is the integration of data from multiple sources (in particular big data) (Cierpiał-Wolan & Stateva, 2023).

Some authors, such as Maślankowski (2015), define key Big Data quality metrics that are essential for assessing and improving the quality of data used in analysis and processing. Each measure is described by a name, a method of representation and a measure, which is a percentage.

- unambiguity measures how many observations (records) are not unambiguous and can be interpreted in different ways,
- objectivity assesses what percentage of records may be mapped incorrectly,
- the timestamp containment meter checks how many observations can be assigned to a specific timestamp,
- granularity or degree of detail measures how many observations can be assigned to higher levels of the hierarchy, such as the provincial or district scale,
- the Duplicate Data Detection Meter assesses what percentage of records were identified as duplicates during data processing,
- completeness checks what percentage of observations, key-value records were completed in the processing,
- data provenance assesses what percentage of observations come from sources without data and were provided for data integration and analysis,
- pattern conformity measures how many observations cannot be mapped as not conforming to a pattern,

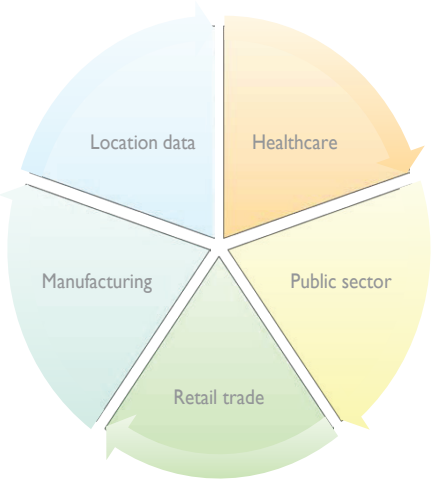
- classification and categorisation assesses what percentage of observations cannot be assigned to any category and classification,
- precision checks the percentage of observations that lose precision when matching to a pattern,
- interpretability assesses how many records can be misinterpreted,
- integrity checks how many observations are inconsistent in the processing,
- consistency measures what percentage of observations have been transformed or duplicated during data processing.

The use of such a tool, according to the author, should help monitor and assess data quality in the context of Big Data, enabling more precise and reliable analyses and decision-making based on this data.

The McKinsey Global Institute has identified the potential of big data in five main themes (Manyika et al., 2011):

- healthcare: clinical decision support systems, individual analytics applied to patient profile, personalised medicine, cost base modelling for staff, analysis of disease patterns, improving public health;
- public sector: creating transparency through available linked data, discovering needs, improving efficiency, aligning activities with relevant products and services, making decisions through automated systems to reduce risks, introducing new products and services;
- retail: in-store behaviour analysis, variety and price optimisation, product location design, operational efficiency improvement, distribution optimisation, logistics optimisation, website-based marketing;
- production: demand forecasting, improving delivery planning, sales support, product development, website optimisation;
- personal location data: intelligent route planning, geo-targeted advertising or emergency response, urban planning, new business models.

*Figure 18 . The potential of big data according to McKinsey*




Source: own compilation based on Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The Next Frontier for Innovation.* McKinsey Global Institute.

Of course, McKinsey identified only some of the potentials for implementing big data analytics. Many others can also be found in the literature, such as in the tourism sector, opportunities are pointed out for creating tourism experiences, strengthening relationships with service recipients, and increasing their participation in the promotion and creation of the tourism offer (Kachniewska, 2014). In terms of marketing, this refers to the possibility of performing market forecasts, creating buyer profiles or determining purchase needs and motives (Kachniewska, 2020). In any case, the use of big data analyses can be seen not only on the demand forecasting side, but also when optimising production processes (Iwasiński, 2022).

The development of big data analytics has its foundations in the idea of collecting a lot of details about the past in order to, with the appropriate analytical tools, find connections and correlations to predict the future. With this approach, it is assumed that many deficiencies or errors can be found in such collections. However, the most serious of these, due to access to the technology in which we leave most of our big data footprints, is the systematic omission of people who are on the 'margins' of big data due to poverty, geography or living culture. Such a remark once again submits the author of this paper to confirm the assumption that the model should consider a combination of multiple data sources that are complementary and at the same time exclude duplication of the same information (Lerman, 2013). A similar conclusion indicating that forecasting can be improved when different

sources and methods are combined has also been pointed out by other researchers (e.g. Bangwayo-Skeete & Skeete 2015; Choi & Varian 2009; Gunter and Önder 2016; Jungnher and Jürgens 2013; Pan and Yang 2016).

Summarising the importance of the impact of big data on the economy, it is also important to introduce the emerging concept of 'business intelligence' in the literature and in practice, which refers to the technology of data warehousing and integration, as well as query, reporting and analysis tools that allow users (most often business users) to access information themselves (Popeangă & Lungu 2012). One of the most popular tools in this area is Power BI developed by Microsoft, which allows users to transform data into interactive visualisations and reports. It is part of the Microsoft Power Platform ecosystem, integrating with various data sources such as Excel, SQL databases, servers, text files and many others. The tool also supports advanced data mining, automatic generation of visualisations and the ability to ask questions in natural language, making it easy even for those without advanced analytical knowledge.

## 3.2. Impact of AI and machine learning on data analytics

Data science is the most common term used to describe an interdisciplinary approach to analysing large data sets in order to draw valuable conclusions and predict trends. It is through data science that companies and organisations can better understand data, make more accurate decisions and predict future customer or market behaviour.

Blum, Hopcroft & Kannan (2020) defined data science as analysing data directly, often in large volumes, without reason or other justification, which requires a significant amount of software engineering skills.

It is worth noting that, in terms of identifying patterns in data, humans are able to perceive up to three attributes or characteristics of an object (such as a product, service or community). These attributes are also referred to as features or variables (Saura & Bennett, 2019). In contrast, with patterns from data science, it is possible to detect hundreds or even thousands of such attributes simultaneously (Berry & Linoff, 2004).

Patterns discovered using data science techniques allow for practical inferences, i.e. what researchers or data analysts intend to extract from the detected patterns (Kelleher, Mac Namee & D'arcy, 2015). In this context, 'inference' refers to the ability of patterns to provide relevant information that can help solve the problem under analysis.

Analysing large data sets is still a challenge, and as a result, methods are constantly being sought to analyse them efficiently. These include sampling, data condensation, density-based

approaches, grid-based approaches, divide-and-conquer methods, incremental learning or distributed computing (Tsai et al., 2015).

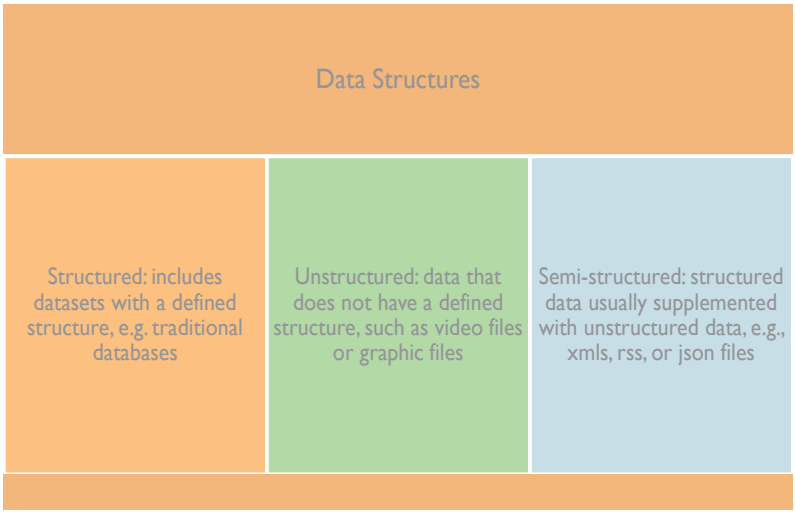
The amount of data generated by internet users (e.g. in social networks) and society in general (e.g. payment card data) is constantly increasing from a quantitative perspective. Although quantity does not always mean the size of the data in one source, but also the number of new sources available to acquire it. As a result, it is nowadays possible to accurately explore this data and also make inferences based on the many different sources available. From an economic perspective, this usually means getting to know consumers even more accurately and tailoring products and services to their expectations.

As already mentioned, in the overview of the definition of big data, this concept applies to data sets with high volume, variety, streaming, as well as variability and complexity (Sagiroglu & Sinanc, 2013) and also characterised by different structures (see Figure no. 19).

In view of the limitations pointed out and the problems of analysis and interpretation, which can be considered a weakness of big data, it is reasonable to conclude that for some research problems, the analysis of single cases (so-called small data) may be of greater value.

Table 30 provides an overview of selected tools used in big data analytics, covering various aspects of data processing, storage and analysis. These tools, such as Hadoop or Cassandra, enable the efficient processing and storage of huge data sets in distributed systems. Others, such as Mahout or MLPACK, support advanced analytics and the implementation of machine learning algorithms on large datasets. Nvidia CUDA and Twitter Storm, on the other hand, are used for fast data processing, especially in the context of data streams and parallel computing. The table shows that big data tools allow the comprehensive management of different types of data, enabling organisations to make better use of the information they have and make more accurate decisions based on real-time data analysis.

*Figure 19 . Breakdown of big data*




Source: Russom P. (2011), *Big Data Analytics*, TDWI best practices report, The Data Warehousing Institute (TDWI) Research.

Regardless of the type of data structure from a research perspective, a variety of AI and machine learning (ML) techniques can be used for processing, interpretation and prediction. The first method is supervised learning, where a model is trained on labelled input data. An example of the use of such a model would be the labelling of messages in a mailbox as spam or non-spam. In this approach, results are predicted from the input data. In unsupervised learning, the model analyses the data without pre-assigned labels. In this way, the model is allowed to discover hidden structures and patterns by creating, for example, groups of similar observations. A third method called reinforcement learning (RL) is a technique in which an algorithm makes decisions in a specific environment and works on a trial-and-error basis observing the results of its decisions on that basis, so it learns which actions produce better results. The most common applications of RL include computer games (e.g. playing chess), robot control (e.g. moving around a room), business optimisation (e.g. inventory management) (Jordan & Mitchell, 2015).

*Table 28 . Differences between ML and DL*

| Scope of comparison | Machine learning | Deep learning |
|---|---|---|
| Scope | A term for methods to learn from data. | A sub-domain of ML focused on networks with multiple layers. |
| Structure of models | Simple models like decision tree, linear regression. | Multiple processing layers, which allows for a hierarchical structure. |
| Processing | Can be run on standard computers or servers (smaller data sets) | They require significant computing resources and also specialised hardware such as GPUs (large data sets). |
| Feature detection | The expert must manually extract the relevant features from the data so that the model can learn from them. | The model can automatically detect features and patterns from the raw data. |
| Applications | Sales forecasting, analysis of financial data. | Analysis of complex and unstructured data. |

Source: own study.

It can be presented that ML is just one of the key methods of AI, as it allows machines to learn and adapt based on variables. DL, on the other hand, is a subset of ML that uses neural networks to model complex patterns in large data sets.

*Figure 20 . Relationship between AI, ML and DL*



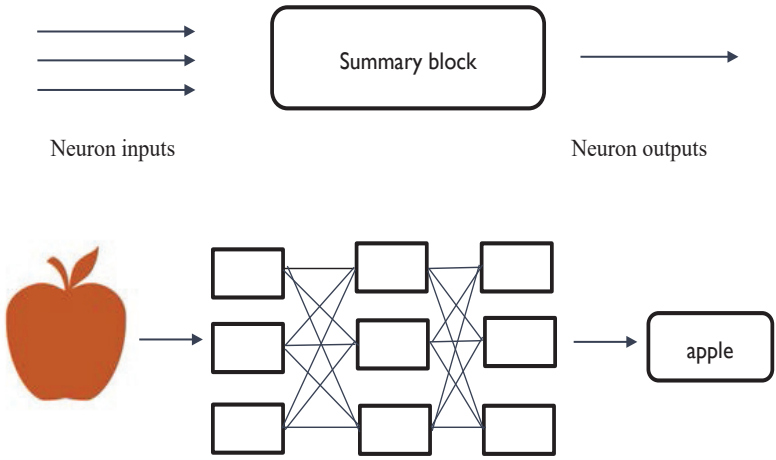

Source: own study.

In order to illustrate the difference between modern ML or DL-based systems compared to previous solutions, an example can be used where previously a programmer had to anticipate all possible user combinations by creating appropriate source code to modern solutions, where it is sufficient to provide access to huge amounts of varied data. However, as Sokolowski et al (2024) note,

contemporary limitations of genAI include uncritical analysis and often biased conclusions, copyright infringements, computer memory requirements and high power consumption.

In DL analysis, the biggest role is played by so-called Artificial Neural Networks (ANNs), which are mathematical models inspired by the structure and operation of natural neural networks in the human brain.

*Figure 21 . Artificial neuron application diagram.*




Source. www.sztucznainteligencja.org.pl (accessed 11.11.2024)

Some deep networks are capable of learning on their own, without the need for data tagging or human supervision. The artificial neurons in such networks are arranged in multiple layers. The outer layer receives stimuli, which are then passed on to the next layer (see Fig. no. 21). As in humans, after passing through several layers, the information reaches the 'brain' of the machine. In the human brain, information such as 'pain', 'heat' or a 'familiar face' is received, while in the machine a number appears. Based on its value, the machine takes the appropriate action, such as recognising an object or performing a specific task. The most common ML models are presented in Table 29.

*Table 29 . Main machine learning models*

| Type | Description |
|---|---|
| EM - Ensemble models | Used to make predictions using a model source. |
| DL - Deep Learning | Deep learning neural networks that have different layers and discover patterns . |
| MV - Machine Vision | As part of deep learning neural networks, MV enables visual identification and recognition of objects, people, products, etc. |
| NLP - Natural Language Processing | Based on text analysis and defined patterns, NLP works with language and text to identify inferences that are invisible to humans. |

Source: Saura, J.R. Using Data Sciences in Digital Marketing: framework, methods, and performance metrics. Journal of Innovation & Knowledge (2020). https://doi.org/10.1016/j.jik.2020.08.001

In many cases, providing massive amounts of data to AI-based systems will allow knowledge to be gained and transformed into actionable insights. The potential of big data can be harnessed through integration with AI automating a wide range of processes and creating challenges to complex challenges. In the tourism sector, this could include, for example, demand forecasting to better prepare for busy periods and allow for better resource management, including the implementation of sustainable tourism strategies and better management of natural resources as well.

*Figure22 . Integration of big data and AI*




Source: own study.

ChatGPT is by far the most recognisable generative artificial intelligence product, with around 50% of the online population in the six countries surveyed having heard of it. It is also by far the most used generative AI tool in the six countries surveyed. Despite this, frequent use of ChatGPT is rare, with only 1% using it daily in Japan, rising to 2% in France and the UK and 7% in the US. Many people who claim to have used generative AI have only used it once or twice, and it has not yet become part of routine internet use (Fletcher & Nielsen, 2024) (data was collected via an online survey conducted between 28 March and 30 April 2024 in six countries: Argentina, Denmark, France, Japan,

the UK and the US with a sample of at least 2,000 people in each country). The launch of ChatGPT by OpenAI in November 2022 (version GPT-3.5) and subsequent developments have generated tremendous interest in generative AI. Both the underlying technologies and the variety of applications and products incorporating elements of generative AI have developed rapidly (albeit unevenly), especially since the publication of the landmark paper on 'transformers', which contributed significantly to advances in foundation models and large language models (LLMs) (Vaswani et al. 2017).

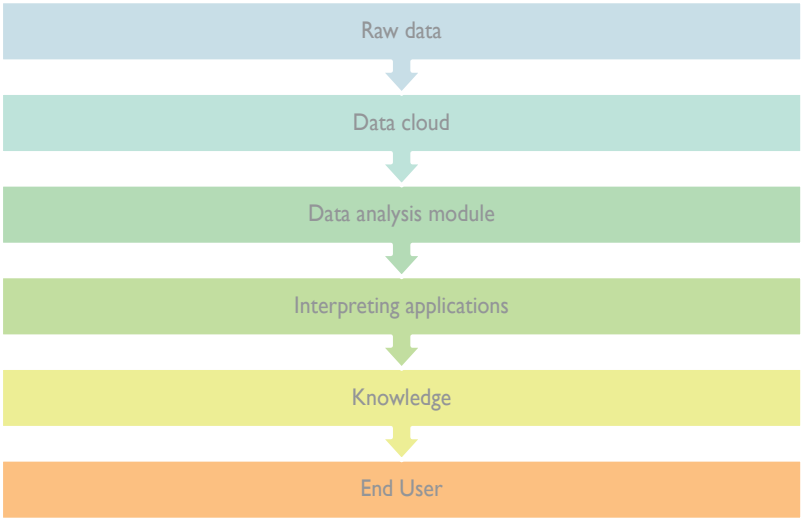
*Table 30 . Overview of selected tools for big data analysis*

| Name of the tool | Objective | Characteristics |
|---|---|---|
| Big Table | Storage | Distributed database system, built on the Google file system |
| Cassandra | Storage | An open-code database management system designed to handle massive amounts of data in a distributed system |
| Google File System | Storage | Proprietary distributed file system developed by Google - part of the inspiration for Hadoop |
| Hadoop | Processing, calculation | An open source software platform for processing huge datasets on certain problems in a distributed system |
| Hbase | Storage | A distributed, non-relational open source database, modelled on Google's Big Table, used to store and manage large volumes of unstructured data. |
| Mahout | Analytics | An open source library to help build scalable machine learning algorithms for Hadoop clusters. Facilitates analysis and prediction on big data. |
| MapReduce | Processing, calculation | A software platform introduced by Google to process huge datasets of specific problems in a distributed system |
| Mashup | Integration | An application that combines data or functionality from two or more sources to create a new service. Used to quickly integrate data from different sources into one place. |
| MLPACK | Analytics | A high-performance machine learning library for analysing large data sets. Allows the implementation of advanced algorithms. |
| Nvidia CUDA | Processing, calculation | A tool developed by Nvidia for parallel processing using graphics processing units (GPUs) to accelerate computing in big data. |
| Titan | Storage | Distributed graph database for storing and analysing large complex networks (e.g. social networks). Scalable and optimised for large datasets. |
| Twitter Storm | Processing, calculation | Open-source system for real-time data processing. Used to analyse data streams such as social media data. |

Source: Kisiołek, A. (2018). Analysis of Twitter posts using Big Data tools included in the R package. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (362), 309 and Pospiech, M., & Felden, C. (2012). Big data - A state-of-the-art. In *Proceedings of the Americas Conference on Information Systems*, 1-23. [Online]. Available: http://aisel.aisnet.org/amcis2012/proceedings/DecisionSupport/22

Research in the tourism sector, taking into account the variety of services available and the diversity of demand, including the creation of behavioural profiles of customers, indicates a favourable environment for the development of artificial intelligence in this industry. The complexity and multiplicity of services offered, as well as the need to accurately understand tourists' preferences, create an ideal environment for the application of advanced AI technologies that can significantly improve the delivery of personalised tourism experiences, and to be able to do this they will first use user data.

*Figure 23 . The idea of big data analytics in the cloud*




Source: own study.

In the context of AI, which is considered one of the revolutionary yet most invasive technologies (Ivanov & Webster, 2019), it is important to note that most researchers in, for example, the hospitality industry, focus their research on analysing the benefits and risks of introducing this technology (Buhalis et al., 2019; Ivanov and Webster, 2019; Murphy et al., 2017; Samara et al., 2020; Tung and Au, 2018; Tung and Law, 2017; Tussyadiah et al., 2020). However, as Filieri (Filierii et. Al. 2021) writes, the tourism sector is in greater need of empirical research to find out how AI technologies will affect the sector in the near future. The author of this paper notes that the development of AI technologies in the sector will not only have a significant impact on the ability to acquire more datasets, but will also significantly facilitate their analysis.

This revolution was very quickly recognised by the European Commission, which developed the world's first regulations on artificial intelligence, the so-called AI Act (EUR-lex, 2024). The aim of the developed solution is to ensure the safe and ethical use of AI technologies, taking into account citizens' rights and supporting technological innovation. The adopted solutions refer, inter alia, to transparency of algorithms as well as labelling of AI-generated content. Among the key elements of the act are regulations concerning AI systems for critical infrstructure, education, law enforcement or health care. The regulations indicate in which areas the risk is unacceptable such as social assessment systems used by the government, where it is subject to high risk and where the risk of use is minimal such as in games. Relating these regulations to the tourism sector and its role in people's lives, it can be assumed that it is in the area of minimal risk. The AI Act enters into force 20 days after publication in the Official Journal of the European Union (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024) and violations are subject to fines, calculated on the basis of the company's global annual turnover or a certain amount, whichever is higher.

The development of AI technologies for data analysis, including in real time, may provide an answer to the problems of tourism load in many cities, where this phenomenon causes traffic problems, environmental pollution and a decrease in the quality of life for residents. Intelligent systems will certainly help to channel tourist traffic so as not to overburden places or regions where tourist capacity has been temporarily exceeded. AI cameras, for example, can help with this, as they can count the number of tourists staying in a given tourist destination at a given time, and this may apply not just to a single place or space, but several places or even dozens of spaces or rooms. This concept is close to the idea of the so-called smart city, which is based on the use of modern information and communication technologies (ICT) to improve the quality of life of citizens, the efficiency of urban services and the promotion of sustainable development. Smart cities integrate a variety of systems and data to manage urban resources more efficiently, which typically includes areas such as transport, energy management, water, wastewater, public safety, healthcare and education.

Smart cities aim to create more integrated, efficient and sustainable urban environments, using modern technology to improve the lives of their residents.

However, applying smart city ideas to smart tourism is still a challenge. On the one hand, in the era of Industry 4.0, long-term observations with high frequency can be collected from transport, energy, environment or opinion and are comprehensive in nature. On the other hand, the challenge lies in understanding the implications of the data for decision-making.

## 3.3. Technologies, methods, tools and sources of big data in the tourism sector

Where an in-depth understanding or analysis of a phenomenon of tourism is needed, a scientific approach requires a precise definition of its subject matter and the development of methods to measure its characteristics and dynamics over time. The complexity of tourism is not only due to the number of elements it comprises, but above all to the nature of the interrelationships between these elements and their specific characteristics. Tourism can be interpreted in a variety of ways, whether as a form of activity, an industrial sector, or a social or economic phenomenon. Consequently, the challenges associated with its measurement arise primarily from the difficulty of defining the concept precisely.

Due to the significant economic importance of tourism, it is justified to use economic terminology and to analyse this phenomenon in terms of supply and demand. As already mentioned, tourism demand is determined by the number of people travelling to a specific destination (tourists and destinations). The measurement of this phenomenon can be based, for example, on four basic elements: number of people (tourists), economic value (expenditure and revenue), time (length of stay and trip) and space (distances and length of trip) (Song et al., 2010). The first two categories of measurement are most commonly used. Despite the efforts of various national and international organisations, there are significant differences in the methods and sources of collecting tourism demand data in different countries. Data often come from counters at borders (police, immigration services) and are supplemented by surveys conducted at points of entry, such as airports and ports. In other cases, measurements are taken at tourist accommodation facilities. It also happens that the data are taken from selected regions and the results extrapolated; in other cases, the measurements cover the full activity of all accommodation facilities. In addition, there is a risk of counting the same tourists more than once if they travel around the country and stay in different places (Bagio, 2019). Errors in measurement can be at the regional, national and international level. For this reason, it is particularly important to integrate data from different data, hierchise them and eliminate repetition.

Several methods for forecasting tourism demand have been developed. Today, we can identify three main groups (Peng et al. 2014) - see Fig. no. 24.

*Figure 24 . Tourism demand forecasting methods*

| Time series methods | Econometric methods | Artificial intelligence methods |
|---|---|---|
| use historical data to predict future trends | look for relationships between demand and certain explanatory variables, such as GDP, price or other variables, in order to make useful predictions about the development of the tourism market | use the latest developments in computer science and computing methods |

Source: own elaboration based on Peng,B.,Song,H.,&Crouch,G.I.(2014).Ameta-analysisofinternationaltourismdemandforecastingandimplicationsforpractice.TourismManagement,45,181-193.

The simplest group of time series models are linear models, which assume that changes in the data occur in a linear fashion over time. They use a linear regression approach, where dependent values are predicted from previous values of the series. Although these models are easy to interpret and implement, their use is limited to cases where only simple patterns are present.

One of the more advanced, but still relatively simple, methods is the Exponential Smoothing (ETS) model. ETS models allow a more flexible approach to modelling data, allowing both trends and seasonality to be taken into account in time series. These techniques are particularly useful in cases where data exhibit short-term changes and forecasts aim to reflect both stable and changing patterns over time (Wołkowicz, 2015). One of the most popular and versatile models for time series analysis is ARIMA (AutoRegressive Integrated Moving Average). The ARIMA model combines three elements: autoregression (AR), differentiation (I - Integrated) and moving average (MA). Autoregression refers to the prediction of future values based on linear combinations of past values of a time series. Differentiation removes trends from the data, making the series stationary, and moving average eliminates the impact of short-term fluctuations by smoothing the data. ARIMA is particularly effective in analysing data where there are long-term relationships or non-stationary trends (Nelson, 1998).

Variants such as SARIMA (Seasonal ARIMA), allow seasonal effects to be taken into account. In tourism, where seasonal fluctuations are often observed (e.g. changes in the number of tourists depending on the season), SARIMA is particularly useful for modelling such phenomena. Incorporating additional seasonality elements into ARIMA allows for a better fit of the model to the data and more precise forecasts (Chang & Liao, 2010).

Econometric models are based on identifying the relationship between a time series and other economic variables. Often used in tourism demand analyses, these models take into account variables such as GDP, prices, the unemployment rate or consumer spending, allowing for a better understanding

of the impact of macroeconomic factors on the phenomena under study. Among the most commonly used econometric models are VECM - Vector Error Correction Models (Pokharel, Poudel & Grala, 2018) and VAR (Vector Autoregression), which allow the analysis of interdependencies between multiple variables simultaneously (Cao, Song, 2017).

With the development of technology and wide access to advanced analytical tools, statistical techniques based on artificial intelligence are becoming increasingly important. Methods such as neural networks, fuzzy time series and genetic algorithms allow the modelling of more complex patterns and relationships that are difficult to capture using traditional methods. These algorithms learn from the available data and adjust their parameters as new information becomes available, making them particularly useful for forecasting in rapidly changing markets such as the tourism sector (Baggio, 2019).

Although nowadays every tourist and visitor leaves behind many digital footprints, there is still a gap in the form of groups of tourists who do not appear in any traditional measurements and capturing this phenomenon is not easy, although attempts are made through the analysis of electricity consumption, municipal waste volumes, etc. (De Cantis et al. 2015; Ezeah et al. 2015; Petrosillo et al. 2006).

Consumers' daily struggle with information overload means that they are unable to accurately process and evaluate all available options. As a result, instead of committing time and resources to analysing every option, consumers often base their decisions on trusting the opinions of others. This is a quick and cost-effective way of acquiring information in an era of information overload (Kachniewska, 2020a). On the other hand, tourism supply creators also face an overabundance of information about consumers, looking for clues to better meet their needs. The data comes from different sources, has different formats and is stored in a variety of forms, such as numbers, text, images or sounds, and is created in many ways. The data collected has great analytical potential, as it usually includes both structured transactional data and unstructured behavioural or preference data.

As Kachniewska (2020a) writes, the most important opportunities in big data analytics include: gaining a competitive advantage in the market by creating personalised predictive models in which the customer is offered perfectly tailored products and services, and optimising business management (Zhan et al, 2017; Erevelles et al, 2015; Wedel & Kannan, 2016; Xu et al, 2016).

*Figure 25 . Data acquisition methods and sources*

| | |
|---|---|
| Methods of data acqusition | Text mining |
| | Web scraping |
| | Crowdsourcing |
| | Exploration of ERP and CRM systems |
| Sources of data acqusitionSystem logs | Clickstream data |
| | IoT sensors |
| | Mobile data |
| | Commercial and financial transactions |
| | Geolocation |
| | Open data |
| | Market research data |
| | Streaming data |

Source: own study.

The physical and temporal remoteness of tourism supply and demand, as well as the complex structure of the tourism product, make the availability and visualisation of information crucial to the tourist decision. Furthermore, the distribution of tourism services is more susceptible to digitalisation than the distribution of food or clothing (Kachniewska, 2014b). Every airline, hotel or car rental booking leaves a digital footprint, which is insignificant on its own, but when combined with the entire data set becomes valuable information. From it, conclusions can be formulated and decisions made (Davenport, Barth & Bean, 2012). Analysing the data left by users of booking portals also allows predictive models to be created. Combining this unique information, as Kachniewska (2014a) has already rightly pointed out, can create the basis for creating competitive advantage. After all, knowledge has become extremely valuable, with the speed of obtaining it also being important.

Web scraping is nowadays defined as a key method of collecting big data for public statistics. This method allows the automatic extraction of data from the Internet using programmes called web robots. These robots are tasked with retrieving relevant information, including, for example, the tourism sector. Web robots analyse the content of websites in real time, eliminating the need to download and store web pages. This is a practical solution for collecting data at regular intervals (Maślankowski, 2019). The data obtained by this method is increasingly being used to measure

inflation (Lunnemann & Wintr, 2006). The Billion Process Project is an MIT initiative to collect and analyse prices from websites on a daily basis, making it possible to monitor inflation and other indicators almost in real time (Cavallo & Rigobon, 2016). Similar methods are also used by Eurostat, the UK Office of National Statistics, the Belgian Statbel (Juszczak, 2021) or the Polish CSO. Of course, many researchers like Radzikowski, Smietanka (2016) or Juszczak (2021) rightly point out that many goods and services that are traditionally included in the inflation basket do not appear at all in the case of online shopping. For this reason, existing methods need to be modified to produce reliable data.

There are many programmes designed for web-scraping, but due to the variety of website codes, it is still considered the most effective method to manually create scraping code for a specific site. The most commonly used development environments for product price scraping are R and Python, which offer many specialised scripts and libraries. In Python, it is possible to use the website API, simulate user actions using the Selenium package, or download the website code and work offline using the Beautiful Soup package (Juszczak, 2021).

In Poland, the National Bank of Poland is investigating the use of data acquired via web-scraping as part of the e-cpi project, which focuses on forecasting inflation based on actual data. Thanks to the rapid data acquisition by web-scraping, methods based on nowcasting prove to have 11% less error compared to the best ARMA models (Macias & Stelmasiak, 2018). The Centre for Social and Economic Analysis (CASE) also conducts research in this area in Poland. It publishes the Online CASE CPI, collecting data from around 50 online shops, which covers around 87% of the inflation basket (Radzikowski & Śmietanka, 2016). The indicated benefits related to automation and immediate use certainly outweigh the disadvantages related to obtaining a complete picture of the market (e.g. such a solution does not include smaller shops). It is also impossible to agree with Juszczak, who points out as a disadvantage that many shops present offer prices rather than transaction prices (Juszczak, 2021), as nowadays prices presented on websites usually reflect current market offers, and in many countries there are regulations requiring that prices presented to consumers must be true and in line with actual transaction prices. More challenging may be the analysis of data from social media sites, where much of the content may be inaccurate or untrue, with very often deliberate and large-scale input (Rojszczak, 2019). Nevertheless, the hypothesis indicating that predictions generated by big data will faithfully reflect users' preferences and expectations has been confirmed by, among other things, the revelation of Cambridge Analytica's activities (more in section 3.4).

Scrapping has proven to be an effective research method with a clear objective of developing databases and testing hypotheses. The literature presents the results of various research projects using big data collected from web portals and other online sources.

For example, The Burning Glass Research revealed the most feasible transitions and required skills by analysing and clustering job vacancies. Data was collected by scraping from more than 40,000 online sources, covering more than 50 million vacancies from 2016-2017, and these skills were grouped into 550 clusters and 958 occupation types according to the O*NET classification, the standard taxonomy of occupations in the US. In contrast, the Where the Works.org project in the UK, a labour market supply and demand analysis model for middle-skill occupations, revealed discrepancies between candidates' qualifications and employers' requirements. However, the analysis was conducted in the context of occupations, not competences (Vankevich & Kalinouskaya, 2021).

Kureková et al. (2015) pioneered the use of a European publicly managed EURES platform. They conducted a content analysis and comparison of employers' demand for skills in several European countries, finding that the mix of skills required varies considerably from country to country, suggesting that there is no universal set of requirements and that national institutions and structures strongly influence the demand for workers.

*Table 31 . Advantages and disadvantages of webscraping*

| Disadvantages | Advantages |
|---|---|
| Data instability | Rapid data acquisition |
| Data quality problem | Collection of data from various sources |
| Regulations | Process automation |
| Technical requirements | Enrichment of statistical data |
| Problems with data duplication | |

Source: own study.

The analysis of data extracted using web scraping is done using text mining, which is a method of text mining and processing to look for patterns, trends and relationships in textual data (Maslańkowski, 2019).

An important element of webscraping is the analysis of robots.txt files, which is used to manage web robot traffic to websites. This is a simple text file that resides in the site's root directory and contains instructions on which pages can be indexed by search engine robots and which cannot. Verification of the file's existence indicates whether the site allows indexing and retrieval. Of course, in the absence of a robots.txt file, web scraping can still be performed, but it would be good practice to inform the site owner of your intention to scrape the data.

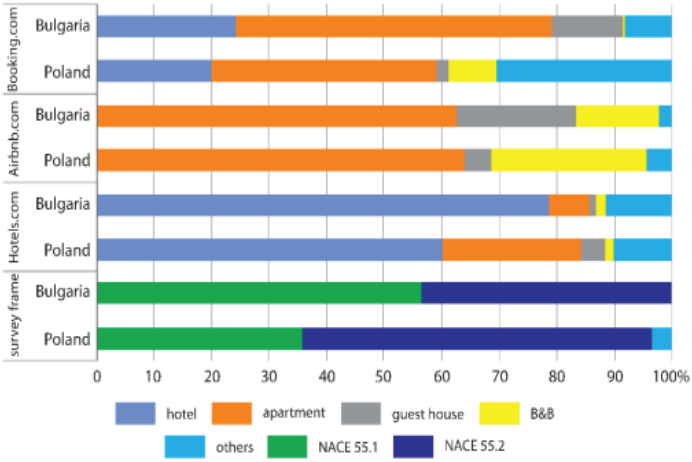
Big data in urban bicycle system research was applied in Poznan, where data was downloaded using the Phyton script, which downloaded the locations of all bikes every 10 minutes from May to October 2019. In this study, the data was initially downloaded up to every two minutes, but due to the volume of the data, it was decided to extend this interval which allowed the collection of more than 19 million urban bicycle locations. An element that was not discarded from the survey was the relocation of bikes by the operator. Such a decision was conditioned by the analysis that displacements represent no more than 10% of the total traffic. The results showed that short trips (up to 2.5 km in a straight line) predominated, with 90% of the trips not exceeding a distance of 5 km. Such a research procedure can be implied in the analysis of other cities and its elements can be transferred to other modes of transport used in tourism. However, it is worth considering how to combine the results of such data and analysis with socio-economic data (Dzięcielski, Radzimski & Woźniak 2020).

Another example is the Booking.com webscraping platform, where data was obtained from 8884 facilities in Poland and 3873 facilities in Bulgaria, collecting 33 variables, i.e. the web address of the ad, the name of the object, its address (street, number, postal code, city), type of accommodation, room name, maximum number of guests, rental price, area of the object, quality rating, number of nights, number of adults, number of children, number of rooms, number of double beds number of single beds, number of sofas, number of reviews, ability to communicate in English, availability of parking, restaurant, bar, breakfast on site, availability of internet, television, cleaning service, air conditioning, laundry, spa, gym, swimming pool and facilities for the disabled (Cierpiał-Wolan & Stateva 2023).

Similar tools were also used for Hotels.com for 4620 facilities in Poland and 532 in Bulgaria. The analysis collected twelve variables such as the web address of the Hotels.com ad, the abbreviated web address on the portal, the name of the facility, its address, its type, the number of rooms, the availability of parking, restaurant, bar, the availability of breakfast on site and the availability of laundry and cleaning services In contrast to the data collected from Booking.com, only 11 accommodations (0.2%) did not meet any classification criteria (Cierpiał-Wolan & Stateva 2023).

Using the Airbnb.com web scraping method, information was also collected on 12,556 accommodations in Poland and 4,174 in Bulgaria. Eight variables were extracted: the web address of the ad, the shortened address on the portal, the name of the property, its address and type, the maximum number of guests, the rental price and the number of beds. Compared to Booking.com, almost all properties on Airbnb.com (99.9%) were classified by a specific type. However, this classification uses generic names for the types of facilities, which are not consistent with the classifications used in official statistics (Cierpiał-Wolan & Stateva 2023).

*Figure 26 . Structure of tourist accommodation in Bulgaria and Poland*




Source: Cierpiał-Wolan, M., & Stateva, G. (2023). *Ocena metod integracji danych dotyczących turystyki z uwzględnieniem big data [Evaluation of tourism data integration methods with big data]*. Wiadomości Statystyczne. The Polish Statistician, 68(12), s. 36.

Trust in the source of information plays a key role. The assessment of trustworthiness underpins the decision to use it. User-generated content, created by consumers themselves, has gained a reputation as independent and trustworthy sources in the decision-making process (Sweeney et al., 2008). K-H. Yoo and U. Gretzel (2007) also noted that reviews by other travellers are often considered by readers to be more up-to-date, interesting and reliable than information provided by travel providers (Wąsowicz-Zaborek, 2019).

Research by Dwivedi et al (2024) indicated that the integration of AI can bring transformative change, but also highlights challenges from the perspective of companies, customers and regulators. The aforementioned publication developed a research agenda that was intended to help understand new technologies from the perspective of the scientific world. It also pointed out the potential problems that AI-based systems may encounter when delivering services. Many companies are exploring ways to harness the potential of ChatGPT and similar tools in their operations. Microsoft has introduced ChatGPT into its Bing search engine and Google, Microsoft, Meta and Open AI have pioneered the implementation of ChatGPT in customer interactions. In the hospitality and travel industry, the technology is only just emerging, with growth projections of 3% between 2022 and 2027, reaching US$9 trillion by 2027 (Tourism and Hospitality Market Forecast, 2023). The authors describe how ChatGPT can help tourists make more informed decisions about choosing places to visit, restaurants

and travel dates. ChatGPT also provides personalised suggestions, which can increase the number of tourists in specific locations, thereby stimulating tourism demand

*Table 32 . Overview of big data analysis tools*

| Name of the tool | Description |
|---|---|
| Apache Flume | Apache Flume is a distributed, reliable and accessible system for efficiently collecting, aggregating and moving large volumes of log data from many different sources into a centralised data store. |
| Apache Sqoop | Apache Sqoop is a CLI tool designed to transfer data between Hadoop and relational databases. Sqoop can import data from an RDBMS, such as MySQL or Oracle Database, into HDFS and then export the data back after transforming it using MapReduce |
| Apache Kafka | Apache Kafka is a distributed streaming platform designed to handle large amounts of real-time data. It enables streaming, storing and processing of streaming data from various sources to analytical systems. |
| Apache Spark | Apache Spark is a distributed computing platform that enables fast processing of large data sets. Spark offers the Spark SQL library for data analysis, as well as the Spark MLlib for machine learning, enabling complex big data analysis. |
| Hadoop MapReduce | MapReduce is a programming model and mechanism for processing large data sets in Hadoop clusters. It works by dividing tasks into smaller parts (mapping) and merging the results (reducing), enabling efficient distributed analysis. |
| Elasticsearch | Elasticsearch is a search engine and analysis engine that enables fast and scalable searching of large data sets. It is mainly used for analysing log data and searching textual content. |
| Apache Hive | Apache Hive is a tool that enables data processing and analysis in Hadoop using a SQL-like language, enabling data analysts to use SQL query language to work on large datasets stored in HDFS. |

Source: own elaboration based on 1) Guller, M., & Mutlu, G. (2020). *Big Data Analytics: Apache Spark and Hadoop*. Elsevier. 2) Noghabi, S., Jeyaraman, J., & Chambers, C. (2020). *Kafka: The Definitive Guide*. O'Reilly Media. 3) White, T. (2015). *Hadoop: The Definitive Guide*. O'Reilly Media. 4) Shmueli, G., Bruce, P. C., & Patel, N. R. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. Wiley. 5) Grolinger, K., Hayes, M., Higashino, W. A., L'Heureux, A., Allison, D. S., & Capretz, M. A. M. (2013). *Challenges for MapReduce in Big Data Analytics*. *Concurrency and Computation: Practice and Experience*, 25(12), 1798-1818.

## 3.4. Issues surrounding the use of big data (ethical and legal considerations)

In the past, written forms, such as books, and works of art were the main medium of information. Over time, these were joined by photographs, sound recordings and video. Today, almost all human-generated information is digitised and stored electronically on personal computers, and increasingly shared on social networks. This type of data is usually poorly structured and rarely subject to external scrutiny.

One of the main data sources for big data analysis is publicly available information on the internet. E-service operators, including social networks, offer access to unclassified data published by users. Much of this information is available through software interfaces (APIs) to third-party content providers. Information about user activity, such as time of publication, geolocation, language of the post or interactions with other users, is supplemented with additional parameters to facilitate analysis. Big data analytics can thus make use of hundreds of millions of new posts from social networks on a daily basis. It should therefore be noted that the sources of big data can be not only data published by users themselves, but also public social, healthcare or tax records remaining in the IT systems of public authorities. Thus, although these collections are generally characterised by anonymity, once several anonymised databases have been collated, it creates a space for the re-identification of a given entity or person (Rojszczak, 2019).

In some cases, data on units also allow for the assessment of the quality of education as in the Economic Fate of Graduates (Ela.nauka.gov.pl, 2024). This is an innovative system developed by the Information Processing Centre - National Research Institute on behalf of the Ministry of Education and Science. The system analyses data from the Social Insurance Institution (ZUS) and the POL-on system (Integrated Information Network on Science and Higher Education). The main objective of the ELA system is to provide reliable information on the situation of graduates of Polish universities in the labour market. It analyses, among other things, the time taken to look for a job, the level of earnings, the level of unemployment and the professional activity of graduates from various fields of study. Thanks to this, future students can better plan their educational path, choosing faculties and universities that offer the best career prospects. The ELA system is recognised as one of the most innovative systems of its kind in Europe. Information is presented both nationwide and by individual universities and fields of study. The system not only monitors the fate of graduates from recent years, but also provides reports and analyses that help to improve the quality of education and adapt the educational offer to the requirements of the labour market.

The ability to combine multiple pieces of information from different databases allows the discovery of new information that is not present in the original data. When analysing anonymised datasets, such new information may be the identity of a person. By combining several anonymised datasets, it is possible to re-identify the subjects in question. The first studies of this type were carried out before the development of big data technology and showed that with three pieces of information about a person - postcode, date of birth and gender - 87% of the US population could be identified (Tene, 2011).

Results come from a study published in 2015, involving a sample of 1.1 million anonymised payment card transactions. It found that around 90% of the people making these payments could be identified. These examples suggest that the anonymisation of large data sets does not always effectively prevent identification. The data still contains linkages and logical identifiers that, when correlated with external sources, can enable identification of the user and reveal information from the original database. It is possible to remove such internal linkages, for example by removing unique user or transaction identifiers, but then the data becomes useless for further analysis. As noted by Mattioli (after Rojszczak, 2019), "context-free data can also be meaningless."

*Table 33 . Privacy risks in relation to big data*




Source: Nunan, D. and Di Domenico, M. L. (2013). Market research and the ethics of big data. International Journal of Market Research, 55(4), 505-520.

No matter how beneficial these activities may be to customers, citizens or other members of the community, they should be fully aware that their personal data is being used to carry them out, especially as the benefits achieved by companies or institutions are usually much greater. Ethical concerns are heightened when the processing of personal data involves vulnerable groups such as children, young people, the elderly, the sick or various social minorities. In addition, it is controversial to carry out attractive personalised marketing communications that promote harmful or restricted products, such as alcohol, tobacco products, pharmaceuticals or weapons (Lee et al. 2016). As early as 2010, the insurance company Aviva began commissioning data analyses of its customers' consumption habits, based, among other things, on information from their loyalty cards. The purpose of these analyses was to assess customers' physical fitness, predict their life expectancy and estimate health and life insurance risks (Scism & Maremont, 2010). In 2012, it was revealed that the US retail chain Target, based on an analysis of its female customers' purchases, assessed which of them were pregnant and which trimester they were in. Based on this, delivery dates were predicted and tailored offers were sent accordingly.

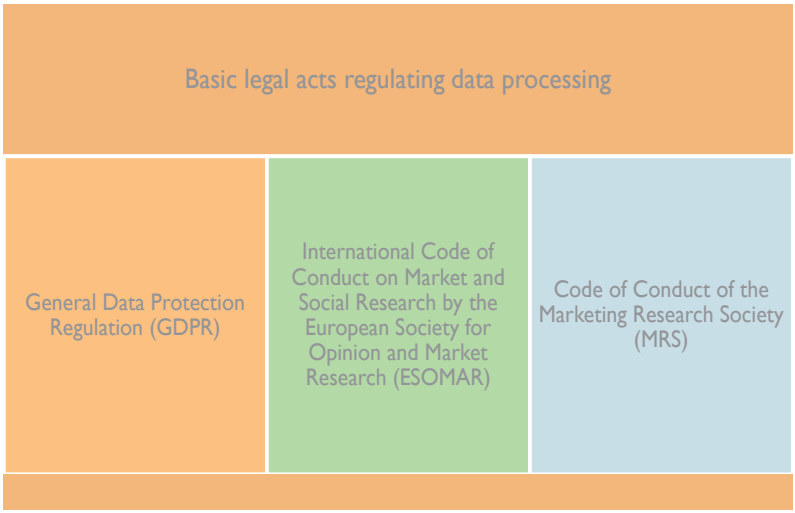
The need to develop appropriate legal regulations and maintain high ethical standards for the processing of big data is becoming increasingly apparent. Despite efforts to increase the security of web users and protect their privacy, big data processing is still controversial. On the one hand, users

of big data benefit in the form of better products, more convenient communication and more efficient service. On the other hand, some audiences see it as surveillance and manipulation. There is also an asymmetry of access to big data, where large market players and institutions that can afford to access and process this data have an advantage, which widens the divide between larger and smaller companies and institutions, putting the former in a privileged position.

Richards and King (2019) note that large data sets are used to generate predictions affecting different areas of life, such as shopping, health, education, or security. Despite this, people are often unaware of what data is being collected and shared with third parties. The authors argue that current privacy laws are insufficient because secondary uses of big data can reveal past, present and future privacy violations. Big Data collects a lot of personal data, such as call history or location, which is already in the hands of governments and companies, and its collection is accelerating. Richards and King point out that the growing role of big data requires the creation of an ethic of this technology to protect the rights of individuals. In addition, as Culnan and Williams (2009) note, the growth of big data databases allows for breaches of privacy by the employees who manage these datasets, and allows for identity theft with even more detailed information.

Nunan and Di Domenico (2017) noted that the advent of networked sensors on various devices and the speed of their analysis means that this data is very often analysed automatically without taking into account the consent of individuals. Even if those managing the datasets decided to remove the identifiable data they would certainly judge this action as impractical. Furthermore, even if users are aware of and theoretically able to control their data, it is undeniable that this data becomes a commodity and that data brokers collect, buy and sell this information without the knowledge and consent of users.

*Figure 27 . Basic legal acts governing data processing*




Source: Woźniczka, J. (2018). *Big Data w marketingu: szanse i zagrożenia [Big Data in marketing: opportunities and threats]. Studia Oeconomica Posnaniensia*, 6(6). https://doi.org/10.18559/SOEP.2018.6.3, s. 43-44.

In July 2024, an important document on harmonised rules for artificial intelligence was also published in the Official Journal of the European Union. The so-called AI Act (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024) establishes uniform rules for the development, marketing and use of artificial intelligence systems within the European Union. The purpose of the act is to ensure that AI-based technologies are developed in line with EU values.

The AI Act classifies AI systems according to their level of risk and defines the relevant requirements for them:

- systems with unacceptable risks are those that may lead to violations of fundamental rights (e.g. 'social scoring' for governments) - according to the act, they are completely prohibited,
- high-risk systems are those used in areas such as healthcare, education, transport, which must meet special requirements in terms of security, transparency and accountability,
- limited-risk systems that require specific transparency and user information on the use of AI,
- minimum-risk systems are systems that do not present serious risks and can be used without additional requirements.

The AI Act introduces important regulations for the use and processing of big data in the context of artificial intelligence systems. This document significantly affects the way big data is collected, processed and analysed. The requirement to comply with the provisions of the RODO is a

key element, forcing companies to take special care to protect privacy and personal data. Due to the processing of sensitive information, organisations have to comply with strict transparency and accountability requirements - big data-based systems require accurate documentation and constant monitoring, which often necessitates the allocation of additional resources. Operators of these systems are required to meticulously maintain documentation of models and analyses, allowing for more complete control over the results and how they are obtained. At the same time, the regulations aim to eliminate bias and counter discrimination. The AI Act places specific requirements that AI and big data systems must be free of bias and not lead to discrimination based on age, gender or ethnicity, meaning that models must be thoroughly tested and calibrated. Alignment with security requirements is also an important aspect of regulation, particularly where these systems may pose a risk to the health or safety of users. Any system analysing big data that potentially puts users at risk will have to meet particularly stringent standards. The AI Act forces organisations to implement measures to ensure compliance with EU requirements, which not only contributes to security and confidence in analytics results, but also poses operational challenges. Companies will have to adapt their existing data processes, as well as incur additional costs to implement new procedures and monitoring tools. These regulations create a more responsible environment for the development of artificial intelligence and big data analytics, although their implementation may present significant challenges.

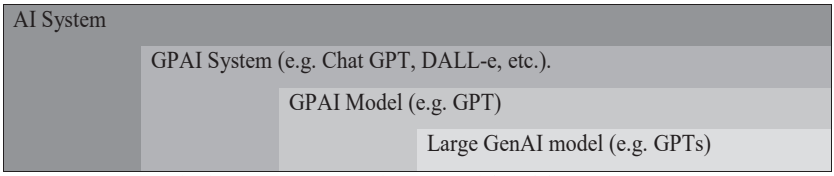
In many cases, AI models are considered so-called 'black box' models, meaning that they are too large or too complex to be understood by humans, or more broadly by systems whose internal decision-making processes are opaque. The increasing reliance on AI models whose internal decision-making mechanisms remain opaque, so-called 'black box' models, has led to increased interest and research in this area. These models are characterised by high complexity and difficulty to fully understand, which poses a challenge in terms of assessing their reliability, identifying potential errors and preventing recurrent negative effects of their operation. Indeed, such AI models may be vulnerable to third-party attacks, assimilate harmful biases present in the data, and may not cope well with real-world changes in data distribution (Outeda, 2024).

In order to understand the relationship between models and systems, researchers (Quintais 2024) also identify structures to understand the relationship between different elements of AI technology (see Fig. no. 28):

- AI system is the largest area in the figure representing the entire AI system, which includes various components including GPAI systems and models. This is a broad category that includes all AI systems, regardless of their purpose or mode of operation,

- inside the AI system area is the GPAI system, which includes specific general-purpose artificial intelligence applications such as ChatGPT or DALL-E. The GPAI system refers to a larger structure that includes the AI model and the additional components needed to make it work, such as user interfaces, input collection systems and ways to generate results,
- inside the GPAI system is the GPAI model, which is a key element that determines the intelligence and functions of the system. Examples of such models include the GPT (Generative Pre-trained Transformer) and other models that may or may not generate content. The GPAI model is responsible for processing data and generating responses within the GPAI system,
- the innermost part of the diagram shows large generative AI models, such as GPTs, which have been trained on large datasets and can generate new content based on given inputs. This area indicates specialised models that are generative (i.e. create new content) and are trained intensively on huge datasets to act as the main decision engine of the system.

*Figure 28 . Structure of AI systems and models*




Source: Quintais, J. P. (2024). Generative AI, copyright and the AI Act. *Available at SSRN* p. 9.

Also, ethics in the context of Big Data analytics is becoming an increasingly important topic, especially with the increasing amount of data being collected, processed and analysed by various institutions and organisations. In the context of big data, ethics refers to broad issues concerning the moral implications of collecting, processing and using vast amounts of data. In a slightly different sense, ethics also refers to ethical principles, which are specific guidelines or standards that derive from ethics and define how to act (Internetowy Słownik Języka Polskiego, 2024). This raises the need for the development of ethical principles to balance the benefits of Big Data analytics with the protection of individual rights and the preservation of social justice (see Table 34).

*Table 34 . Big data analysis from an ethical perspective*

| Name of ethical theory | Main assumptions | Relationship to big data analysis |
|---|---|---|
| Kantian ethics<br><br>Representative:<br>I. Kant (1724-1804), | It is a theory focused not on what an individual does, but on what they should do. Everyone is subject to the same standards and appropriate behaviour has clearly defined guidelines. In Kantianism, it is not the result of an action that is most important, but the principle behind that action. | In the context of Big Data, the implementation of this principle is problematic, as data is collected and analysed without the consent of individuals, meaning that their autonomy is not respected. When using data, organisations often disregard the informed consent of users, treating their information as a tool to pursue their own interests. |
| Utlylitarianism<br><br>Representatives:<br>J. Bentham<br>(1748-1832)<br>J. Stuart-Milla<br>(1806-1873)] | It judges what is right or wrong on the basis of the consequences of an action or principle. In contrast to the Kantian perspective, where the will that motivates action is most important, in utilitarianism what matters most is 'happiness' or the maximum good for the whole | Utilitarianism as a method of evaluating Big Data faces a number of difficulties related to accurate measurement of impact, biases and a lack of widespread knowledge of Big Data, which makes it difficult to transparently analyse its impact on society. |
| Theory of the social contract<br><br>Representatives:<br>T. Hobbes<br>(1588-1679)<br>J. Locke<br>(1632-1704)<br>J.J. Rousseau<br>(1712-1778) | In this theory, people are seen as rational beings who understand that in order to create and maintain a society, they must cooperate and agree to follow certain guidelines in order to benefit from society. To do this, they must choose rationality over their natural, selfish instincts | Big Data challenges this theory because people want the benefits of technology while also wanting to control its impact on privacy. Therefore, societies need to establish new rules that balance access to technology with the protection of individuals' rights to determine the morally acceptable aspects of Big Data |
| Virtue ethics<br><br>Representatives:<br>Aristotle<br>(384 BC - 322 BC)<br>Plato (428 BC - 347 BC) | This theory defines a virtue as a character trait or disposition that is deeply rooted in its possessor and makes that person good. There are two types of virtues: intellectual and moral | In the context of Big Data, this means examining the intentions of those using these technologies and determining whether their actions are consistent with virtues such as honesty and responsibility. For example, if Big Data is used in a way that violates privacy, this may indicate moral failings on the part of the users. However, when Big Data serves a social good, as in medicine, it reflects their virtuous intentions. |

Source: own compilation based on Herschel, R., & Miori, V. M. (2017). Ethics & big data. *Technology in Society*, *49*, 31-36

From the perspective of research in the big data era, there is also the theme of the impact of researchers (Markham, Tiidenberg, & Herman, 2018) on other people's lives on a daily basis in the context of the possibility of revealing a lot of sensitive data, so the scientific world should also look for ways to limit the negative effects of their actions. In traditional ethical theories such as Kantianism or utilitarianism, moral decisions were based on clearly defined principles, values and predictable calculations of right and wrong. However, in the context of big data and artificial intelligence, such an approach becomes difficult. Algorithms and machine learning models are often too complex to fully understand how they make decisions. The 'black box' problem makes the operations of these technologies opaque and not easy to explain - their internal mechanisms are hidden from the eye of researchers. As a result, it is difficult to determine clearly what effects the decisions made by the algorithms will have, and accountability begins to blur.

Another issue of ethics in big data analysis is the exploration of the role of 'nudging' (election steering) through advertising and social media. The study of the role of social media and its impact on society and politics shows how manipulation techniques can influence election outcomes, democracy and consumer behaviour. In 2018, it came to light that Cambridge Analytica, a UK company, accessed the data of millions of Facebook users without their explicit consent. This information was used to create targeted social media advertising campaigns, mainly for political purposes. The data was collected through an app created by researcher Alexander Kogan, which offered users to take psychometric quizzes. Although participation in the quiz was voluntary, the app gained access not only to participants' data, but also to information about their Facebook friends (Gardian, 2018). In this way, the data of around 87 million users was collected, often without their full knowledge or consent. The information extracted was used to build psychological profiles of voters and target them with precisely tailored messages designed to influence their choices. Cambridge Analytica used this strategy in, among others, Donald Trump's 2016 presidential campaign (The New York Times, 2018) and during the campaign for Brexit.

In addition to Cambridge Analytica, cases such as Clearview AI, Facebook, Google, Uber, TikTok and the PRISM programme demonstrate the serious consequences of unconscious or uncontrolled management of personal data. Clearview AI used facial recognition to collect photos without users' consent, raising privacy concerns. Facebook experimented with manipulating users' emotions. Google and Uber were accused of misusing location data, and TikTok faced allegations of user data security. The PRISM programme, in turn, revealed the scale of monitoring carried out by the NSA in collaboration with technology companies. These examples highlight the importance of a

responsible approach to privacy and data protection in the face of new technologies and big data analytics.

Similar practices have taken place in the travel sector. Some booking platforms and airlines used the practice of dynamic pricing, where prices increased based on a user's search history (BlueWhite.co.uk, 2024). For example, if a customer checked the price of a flight several times, the algorithm could increase the price, assuming that the customer had a strong interest in the flight. Such practices were controversial because customers were not aware that their online behaviour was affecting the cost of travel services.

# CHAPTER IV:
# OPPORTUNITIES TO ANALYSE THE TOURISM SECTOR USING BIG DATA

In attempting to analyse the tourism sector from the perspective of the use of big data, it should first be noted that there is potential for generating multiple datasets in this sector, but at the same time, the use of such data, tools and methods represents an innovation gap, as there is an apparent distance between the tourism sector and innovative standards and practices in other industries. As Kachniewska (2019) points out, the opportunities arising from big data analytics in tourism relate to five main dimensions (see Table 35), and focusing all attention on consumers and understanding their purchasing patterns helps the tourism industry which is conducive to offering personalised solutions and better understanding of market trends.

*Table 35 . Dimensions of Big Data Analysis in Tourism*

| **Big Data for the tourism sector** | | | | |
|---|---|---|---|---|
| Customer/Consumer | Product | Time | Location | Channel |

Source: Kachniewska, M. (2019). Big Data Analysis as a Tool for Predictive Intelligence and Experience Personalization in Tourism. In *Entrepreneurship and Management*, 20(2), 39-52.

As research indicates, many datasets are useless without the use of appropriate data processing techniques (Han & Kaber 2006, Kachniewska 2019). Extracting valuable knowledge from them that supports decision-making and trend forecasting is only possible after using a variety of methods applied independently or complementarily to each other. Otherwise, such collections are just collections of unrelated information with no practical benefit. Basic processing techniques include:

- characteristics, which is used to generalise, summarise and possibly differentiate data characteristics;
- classification, understood as the assignment of data to specific categories;
- regression, which is used to predict continuous values from data;
- association, i.e. discovering relationships between different databases and attributes of a single database;

- clustering commonly used to divide a large data set into smaller groups, which simplifies the design and implementation process and reduces similarity between classes;
- detection of changes and deviations;
- connection analysis, i.e. tracing connections between objects to develop models based on patterns of relationships using graph theory techniques;
- sequential pattern analysis, which is used to identify patterns in sequential data.

Interpreting big data from devices in accommodation facilities could prove to be an interesting research challenge. After all, hotels that have already implemented the IoT (internet of things) in their daily service activities have at their disposal all the real-time data on a guest's presence or absence in a room, how they use light, water, heating or air conditioning, as well as electronic devices, various TV channels or music. A study in 25 Polish 4- and 5-star hotels revealed that 67% of guests are not interested in access to TV and satellite radio - while the need for access to the internet and also VoD and other information platforms emerged among the main needs (Kachniewska, 2019).

With the growing importance of data and its analytics in the tourism sector, it has been noted (Amadeus, 2018) that 90% of travellers share their photos and travel experiences on platforms such as Instagram, Facebook or X. As is widely known, every day millions of new reviews related to places, hotels or restaurants visited appear online and this allows to assess whether the reviews are mostly positive, neutral or negative allowing to identify trends in customer opinions and behaviour.

The use of big data from social media can help to understand user behaviour based on three elements (Kachniewska, 2019):

- early warning (to respond quickly to a crisis if irregularities in the use of electronic devices and services are detected);
- real-time monitoring (to provide accurate information for formulating policies and plans by monitoring users' current behaviour, emotions and preferences);
- real-time feedback (getting groups' opinions on certain social activities based on real-time monitoring).

Big data analytics can improve the effectiveness of online marketing as it focuses on conversion and the ability to sell products with minimal exposure. With commercial tools offered by leading social media platforms such as Facebook, YouTube and SnapChat, travel providers can increase conversion rates. Simply attracting users to a page is not enough if they do not make a purchase. Modern predictive algorithms can help to estimate conversions, better understand users' values, identify destinations more precisely and optimise advertising campaigns. Among the 54 hotels that tested the new conversion

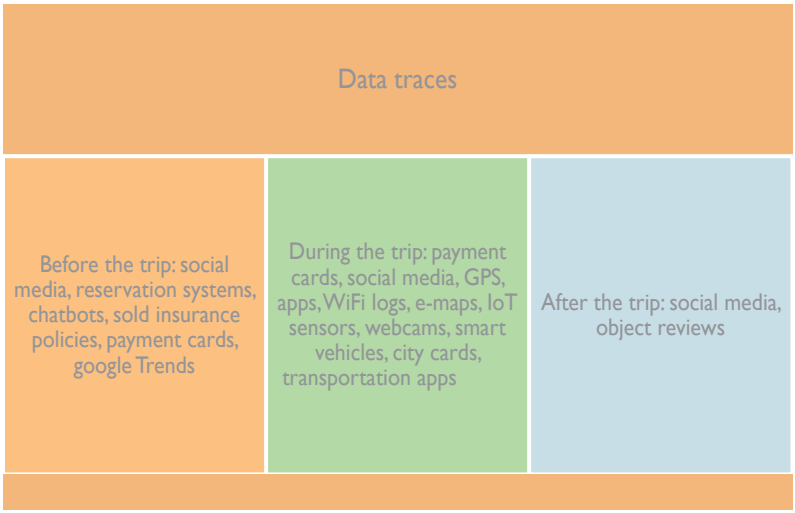
optimisation system, the programme managed to reduce the cost of online campaigns by 15 per cent and in other cases by 5-9 per cent (Travolution, 2018). With such solutions, hotels can increase their revenue while reducing their marketing spend.

In the context of the ever-increasing use of big data analytics, it is important to ask whether, as Kachniewska (2019) argues, the shaping of a tourist offer should be based on access to information about tourists' expectations and preferences or whether it should highlight attractions in such a way as to highlight local culture and tradition and encourage spontaneous discoveries and unexpected experiences. It can be a challenge for tourism organisers to strike a balance between using analytical tools and highlighting unique assets that very often go unnoticed in large datasets. The search for such a balance point will not only allow, in the author's opinion, to attract more discerning tourists, but will also help to support the sustainable development of the tourist reception area, highlighting its authenticity and uniqueness.

Many authors (Kachniewska 2020, Frederiksen, 2012; Hendrik & Perdana, 2014; Manyika et al, 2011) emphasise that the datasets available on the tourism sector, i.e. online transactions, web searches or social networks, are irrelevant when isolated from their context and only become meaningful when analysed together. A similar model of operation was adopted in the Tourism Barometer for the Kujawsko-Pomorskie Voivodeship, which is scheduled for implementation in 2024-2030 (Barczak, Brudnicki & Wierzbiańska, 2023). Extensive data analysis not only allows for a better understanding of the needs of tourists, but also helps to counteract the negative effects of the concentration of tourist traffic and promotes the activation of mechanisms to channel it to less busy destinations. Just as Google maps show us the road with the least amount of car traffic, implementing tourist management systems allows us to achieve such benefits, which in turn brings improved economic efficiency in economic terms and a better tourist experience in social terms (Kachniewska, 2020b).

Swan identified that every tourist and visitor leaves a trail of data behind them, which creates opportunities for business use (Swan, 2013). Researchers emphasise (Meker & Hong, 2014, Hendrik & Perdana 2011, Bollier & Firestone 2010) that big data is more reliable, as it provides information on the actual activities of tourists and not just the declarations usually derived from surveys, which also promotes an understanding of spending patterns. In addition, they make it possible to analyse consumer demand for various tourism products and services and allow information to be analysed even in real time. In the same way that "cookies" enable subsequent suggestions of products and services to consumers, developed applications based on user profile analyses could also suggest itineraries, trips and destinations.

*Figure 29 . The data trail and its potential use in tourism*




Source: own study.

In the future, such a role will be fulfilled by AGI (Artificial General Intelligence), which can significantly improve the customisation of tourist offers through advanced analysis and personalisation[1] . With its ability to process vast amounts of data in real time, AGI will suggest the best itineraries, tours, shops and attractions based on the tourist's preferences, previous travel history and current conditions (e.g. weather, current promotions, tourist traffic). For example, AGI will analyse data from social media, GPS location and previous purchases to suggest personalised itineraries that take into account less crowded routes and unique local attractions that may suit the tourist's interests. It will also be sure to send 'push' notifications of the latest offers, events or promotions in the area, which in turn will increase tourist satisfaction and engagement. As a result, such advanced solutions will not only help improve economic efficiency for local businesses, but also improve the overall tourism experience, making it more personalised and rewarding (McKinsey, 2024).

The application of big data analytics to tourism appears to have benefits on both the tourism demand and supply side, but raises significant concerns about privacy and control over data and its use.

[1] Artificial general intelligence (AGI) is a theoretical AI system with capabilities that rival those of humans, including reasoning, problem solving, perception, learning and language understanding. Although currently a theoretical concept, AGI has the potential to revolutionise many aspects of life and business in the future. AIs such as ChatGPT are, for now, only steps towards AGI.

Research conducted by Kachniewska (2020b) indicated that respondents, both from the Polish tourism and high-tech sectors, show moderate scepticism towards the future of smart tourism in Poland. These concerns relate not only to economic and political aspects, such as insufficient support from the government and local authorities, but also to socio-cultural challenges, including the digital gap that affects Polish society and small and medium-sized enterprises, making it difficult for them to implement innovative business models. The greatest opportunities for growth lie in changing consumption patterns and the growing volume of big data, which has significant market value in an environment of increasing competition. The success of smart tourism and the engagement of smart tourists largely depends on the ability to collect extensive data, adequate computing power and the effective use of algorithms to deliver personalised information in real time. While innovation in big data is key, it creates challenges around privacy, access to technology and business models. Smart tourism, although at an early stage of development, has promising prospects. It is important that local authorities and companies think carefully about data management and privacy issues.

In the case of Wąsowicz-Zaborek's (2019) research on web-generated content, content generated by users of the TripAdvisor, Zoover (a tourism platform founded in 2015 in the Netherlands) and Twitter (contemporaneous with X) platforms about the city of Kolobrzeg was analysed, and data was collected using an indexing robot (web crawler). The web crawler is recognised as a key element of web scraping, but there are differences between the two. A web crawler usually searches multiple websites for content, whereas web scraping extracts specific data from specific websites. A scraper usually focuses on extracting data from one or a few pages in a very precise way. The results of the research described in the cited publication indicate the usefulness of the content analysis method for assessing factors shaping the image of a travel destination. The most important factors in assessing the image of the city of Kołobrzeg turned out to be tourist attractions, events, catering offer and accommodation facilities. The study included only content published on open Internet platforms that did not require logging in. Only textual content was analysed, without photos and graphics, and the results referred to one town and three social media sites, which obviously limits the possibility of generalising the results to other travel destinations. However, the study may be useful for local government units and tourism organisations responsible for promotion in this particular case study or, when replicated in other regions, such results could form the basis for the development of promotional materials by tourism entrepreneurs (Wąsowicz-Zaborek, 2019).

The comments extracted from all three sites were subjected to a preliminary analysis with the TextMining module in the Statistica package to determine the frequency of words used. In the course of data reduction, conjunctions, prepositions and pronouns, as well as some numerals, were removed

from the list of generated words, and synonyms and close words were grouped into a single variable. The cleaned and reduced data were further analysed using SPSS software and the data were presented using a semantic profile (see Table 36).

The key factors influencing the perception of Kołobrzeg include the variety of tourist activities available, the cultural offer and the catering facilities. An analysis of the comments and ratings awarded made it possible to define the image of the city as seen in the surveyed websites. In the chart, negative ratings correspond to low or unfavourable opinions, while positive ratings reflect favourable impressions of users.

The study indicates the key areas important in creating the image of Kołobrzeg as a tourist destination. The results can support the promotional activities of institutions, organisations and entrepreneurs, helping to adapt the offer to the expectations of the audience. The use of user-generated content analysis also allows monitoring o shaping the perception of the city, constituting an effective promotional tool. The content can also reveal poorly rated areas of the offer, i.e. primarily those in need of improvement

*Table 36 . Importance of image attributes of Kolobrzeg*

| Attribute | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Possibility to practice various tourist activities | | | | | | | |
| Landscape, surroundings | | | | | | | |
| Nature | | | | | | | |
| Cultural attractions | | | | | | | |
| Nightlife and entertainment | | | | | | | |
| Commercial base | | | | | | | |
| Tourist information | | | | | | | |
| Sports facilities | | | | | | | |
| Public transport | | | | | | | |
| Accommodation | | | | | | | |
| Catering facilities | | | | | | | |
| Price level, costs | | | | | | | |
| Climate | | | | | | | |
| Possibility to relax | | | | | | | |
| Accessibility (openness) | | | | | | | |
| Security | | | | | | | |
| Social interaction | | | | | | | |
| Attitudes of residents | | | | | | | |
| Originality | | | | | | | |
| Level of service quality | | | | | | | |

Source: own elaboration based on opracowanie własne na podstawie Wąsowicz-Zaborek, E. (2019). *Treści generowane przez użytkowników Internetu jako źródło wiedzy o wizerunku miejsca docelowego podróży na przykładzie miasta Kołobrzeg [Content generated by Internet users as a source of knowledge about the image of a travel destination on the example of the city of Kołobrzeg]. Folia Turistica*, nr 53. DOI: 10.5604/01.3001.0013.7512

The large number of data sources is a natural feature of the tourism sector and poses a significant challenge in aligning supply with tourism demand and fostering the search for solutions through AI capabilities. Buhalis et al. (2019) cite four types of AI described by Hintze (2016) in their study. The first type is reactive AI, which has no memory or use of prior data, such as Deep Blue. The second type is memory-limited AI, which has a selective memory. An example is the way autonomous cars recognise and react to objects around them. The third type is theory of mind, referring to machines capable of understanding the mental states of other individuals, such as beliefs, desires and intentions, which influence the decisions they make.

In previous years, the tourist generally used a search engine and advanced filters to find the offer that best suited his or her needs and expectations. Nowadays, this process is increasingly taking place through 'black box AI', where an algorithm set by the creator will match the 'best offer'. This approach and the lack of transparency in the creation of algorithms will certainly allow for a lot of abuse. Of the AI systems that are likely to be implemented most rapidly in the tourism sector, translation systems stand out. As Cohen (2004) stated, language is one of the greatest sources of

discomfort and anxiety. Language barriers have always hindered tourism exploration and automating this process will greatly facilitate destination exploration.

Although technological innovations are penetrating the tourism sector faster and more frequently, many tourists still prefer traditional methods of purchasing and using services. As with any other technology or innovation, tourists can be divided into the categories listed by Rogers (2010):

- innovators,
- early adopters,
- early majority,
- late majority
- marauders.

Traditional approaches are often driven by the need for security, attachment to tried and tested methods and innovative solutions often raise concerns about complexity or the need to learn new tools. For innovators and early adopters, technological novelties present an attractive challenge and opportunity to improve the travel experience, while the late majority and marauders tend to adopt new solutions only when they become commonplace and easily accessible. In the context of tourism, this means that some tourists prefer to avoid modern apps or service automation, instead relying on face-to-face contact with staff, printed guides or traditional booking methods, which they find more understandable and predictable.

Considering the risks and benefits of AI, Tussyadiah and Miller (2019) distinguished three types of users:

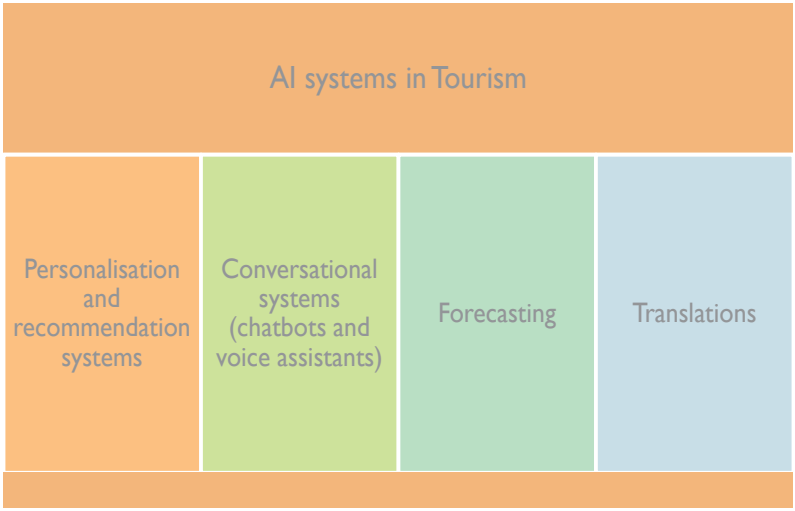
- marauders (who see high risks and low benefits associated with AI),
- enthusiasts (who perceive high benefits and low risks associated with AI)
- realists (who are aware of both the likely benefits and risks of AI).

These authors found that those with negative feelings towards AI were those who had not used such technologies before (Tussyadiah and Miller 2019). Considering these typologies, it can be seen that attitudes towards AI in tourism are strongly influenced by users' previous experiences with new technologies. Marauders are often concerned about loss of privacy, dehumanisation of services or potential technological glitches that may affect their travel experience. Enthusiasts, on the other hand, see AI as a tool that streamlines processes, offers greater personalisation of services and reduces the effort involved in organising travel. Realists are characterised by a pragmatic approach that combines openness to innovation with a cautious approach to its implementation.

Another risk associated with the widespread use of AI is bias in AI systems. All humans exhibit bias (including, of course, those programming AI algorithms) and biases are natural in humans, such as biases related to race, gender, age and economic status. The problem is that AI is much more powerful than humans and can potentially increase the biases ingrained in algorithms (Smith, 2019). In practice, this could lead to AI systems disproportionately promoting luxury hotels or expensive attractions in areas preferred by wealthier customer groups, ignoring the needs of more frugal travellers. In addition, algorithms may inadvertently exclude specific destinations or regions that do not fit the stereotypical expectations of tourists, thus limiting tourism diversity and promoting homogeneity of offerings.

The problem of bias in AI can be particularly relevant in rapidly evolving technologies such as chatbots that answer customer questions. If the training data on which AI is based contains cultural biases, chatbots may give inappropriate or biased answers, negatively affecting the customer experience. An overview of AI systems is presented in Figure 30.

*Figure 30 . AI systems for tourism*




Source: Bulchand-Gidumal, J. (2020). Impact of artificial intelligence in travel, tourism, and hospitality. In Z. Xiang et al. (Eds.), Handbook of e-Tourism. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-05324-6_110-1

As AI becomes more versatile, capable of solving a wide variety of problems and learning autonomously, its functions may extend beyond its current boundaries. Reaching the level of AGI would mean the ability of machines to understand and perform any task at a level comparable to humans, regardless of the domain or situation.

However, the shift from narrow AI to AGI also brings serious challenges and ethical questions. AGI could radically change the way industries such as tourism operate, offering extremely personalised experiences, but at the same time increasing the potential risks of bias, lack of control and autonomous decisions. As such, the development of AI into AGI will require not only advanced technology, but also strict oversight, ethical approaches and regulation to ensure that these systems are implemented responsibly and safely.

*Table 37 . Factors transforming AI into AGI and implications for tourism*

| Factor | AI | AGI |
|---|---|---|
| Visual perception | AI systems are a long way from achieving human sensory perception. | The use of AGI in recognising tourist sites, analysing users' photos to personalise recommendations or detecting threats in real time (e.g. in adventure tourism). |
| Sound perception | Humans use sound to identify spatial features of the environment. AI has limited capabilities in this area and difficulties in interpreting sounds. | Improving voice interpreters and real-time speech recognition systems, which would make tourism services more accessible to people from different language groups. |
| Natural language processing | To match human cognisance, AGI would need to understand and process information from various sources with full comprehension. | Development of advanced chatbots capable of conducting natural conversations in different languages, which would facilitate bookings, answering tourists' questions or handling complaints in a more personalised way. |
| Problem solving | The AGI system should diagnose and solve problems using general knowledge and adaptability. | Automatically diagnose and resolve problems, such as accommodation, flight rescheduling or emergency assistance (e.g. lost luggage), without the need for human intervention. |
| Motor skills | AI robots have not yet attained the precision movement skills for complex tasks such as braiding hair. | Introducing robots in hotels and tourist attractions to perform precision tasks such as delivering meals, cleaning or operating sports equipment. |
| Navigation | Navigation systems such as GPS have made progress, but years of work are still needed to create robots that navigate autonomously. | Improving autonomous transport vehicles in tourism, such as buses and tour boats, which would increase the efficiency and safety of tourism transport. |
| Adaptation to new situations | AGI should be able to learn from experience and adapt to new situations without the need for human intervention. | The ability to adjust offers and itineraries in real time, e.g. in the event of adverse weather conditions, attraction closures or other unexpected changes. |
| Social and emotional competences | AGI should understand and respond to emotions and social behaviour in a human-like way. | Better personalisation of services by understanding tourists' emotions, e.g. automatic adaptation of service in hotels to the guest's mood (more private space, playful interactions, support in stressful situations). |
| Creativity | AI systems will have to rewrite their own code, which requires understanding huge amounts of code written by humans and finding new ways to improve it. | Creating unique tourism experiences, such as generating new tours, designing customised attractions or personalised marketing campaigns. |
| Social and emotional involvement | For robots and AI systems to be truly effective, humans should want to interact with them. The robot needs to interpret facial expressions and changes in tone of voice that reveal emotions. Some AI systems can do this, but to a limited extent. | Building deeper relationships with customers through systems that understand and respond to individual needs and preferences, strengthening customer loyalty to destinations and tourism services. |

Source: own compilation based on McKinsey & Company. (n.d.). What is Artificial General Intelligence (AGI)? McKinsey Explainers. Retrieved June 29, 2024, from https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-artificial-general-intelligence-agi

An intersting review of research methodologies in the tourism and hospitality sector, including for data modelling for demand forecasting, using AI was provided by the work of Doborejh (2021) - see Table 38

*Table 38 . Overview of research methodologies for the tourism sector*

| Author/year | Context of the study | Data | Methodology |
|---|---|---|---|
| Machine learning | | | |
| Zhang (2014) | Identifying the main factors influencing Chinese citizens' intentions to travel abroad | The questionnaire included demographic and personal values variables collected in Nanjing, March-May 2011. | - C4.5 Rule-PANE - Ensemble neural network - Logistic regression |
| Donaire (2014) | Understanding tourist behaviour with different perspectives of the same place. | 1,786 photos taken by tourists in the valley (Spain) and published on Flickr. | Ward's method |
| Giglio et al. (2019) | Identification of clusters surrounding points of interest. | 26,392 photos taken in 6 cities in Italy between 2014 and 2016 and published on Flickr. | - DBSCAN - k nearest neighbours - Euclidean distance |
| Höpken et al. (2020) | Tourist traffic survey and point of interest. | 13,545 photos taken by tourists in 2015 and published in the media. | - k-mean - (DBSCAN) |
| Ramos-Henriquez et al. (2021) | Identification of key features of the hotel value proposition. | 250 features from 5136 lines collected in the Canary Islands region. | - Boruta for feature selection - (SVM) for classification |
| Xie et al. (2021) | Forecasting demand for Chinese tourism. | Search engine data. | - LSSVR model (with LSSVR-GSA) |
| Work on AI methods | | | |
| Silva et al (2019) | Forecasting tourism demand | Tourism demand of ten European countries. | - NNAR - Denoised neural networks |
| Liu (2011) | Forecasting the number of tourists | Time series data on the number of visitors from 1998 to 2009 in Weifang. | - (BPNN) - Gradient method |
| Golmohammadi et al (2011) | Predicting tourist satisfaction with travel | 1,870 surveys collected from tourists at the airport over three months. | Hybrid neural network |

| | | | |
|---|---|---|---|
| Zhu et al (2018) | Forecasting tourist flows | Tourism demand in Singapore's six major regions, collected from 1995 to 2013. | Copula-based approach |
| Claveria et al (2015 | Forecasting tourism demand and comparing the performance of three different ANN models | Tourist arrivals, collected between 2001 and 2012. | - (MLP) - (RBF) - Elman recursive neural network |
| Sanchez (2020) | Identification of intentions leading consumers to cancel hotel bookings | 10,000 booking records of a four-star hotel in Spain, collected between 2016 and 2018. | - (SVM) - (ANN) - (GA |
| Kwon (2020) | Identification of underlying factors influencing consumer preferences for restaurants | 4,799,240 restaurant reviews from 2004-2014, collected from Yelp.com | Linear mixed-effects model |
| Big data deep learning methods | | | |
| Li (2018) | Forecasting tourist flows | Beijing tourism data from 2011 to 2015, collected by the Beijing Tourism Association (BTA). | - Principal component analysis (PCA) - Index based on a complex dynamic factor model (GDFM) |
| Zhang B et al (2020) | Forecasting tourist flows | Tourist demand in Jiuzhaigou, China, collected from 2013 to 2017 | (LSTM) - Contextual long-term memory (C-LSTM) - Deep belief network - (BPNN) |
| Lado-Sestayo (2019) | Forecasting hotel profitability | Data on hotels and tourist destinations, collected between 2005 and 2011 | Multilayer neural network |
| Chang Y-C (2020) | Forecasting hotel reviews and responses | 113,685 hotel reviews | NLP |
| Zhang K (2019) | Understanding tourist behaviour and perception | 35,356 photos of tourists from Flickr | Deep learning technologies |
| Al Shehhi (2020) | Forecasting hotel room rates in selected Gulf Cooperation Council (GCC) cities | 2800 observations of daily room sales, collected between 2010 and 2017 | - (SARIMA) - Deep belief network - (ANFIS) - Polynomial smooth support vector machine (PSSVM) |
| Wu (2021) | Forecasting tourist arrivals in Macau, China | The data covered the period from 1 January 2017 to 30 April 2019, resulting | (SARIMA) - (LSTM) |

| | | | |
|---|---|---|---|
| | | in a total of 850 daily observations for each group | |
| He (2021) | Improving the accuracy of tourism demand forecasting | The data covered the period from 1 January 2017 to 31 October 2019, resulting in a total of 1,033 observations for each country/region. Data from places such as China, Hong Kong, Taiwan, the Philippines, South Korea and Singapore were included in the analysis. | - (SARIMA) - Convolutional Neural Network (CNN) - (LSTM) |

Source: Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N. (2021). Artificial intelligence: a systematic review of methods and applications in hospitality and tourism. International Journal of Contemporary Hospitality Management, 1-23. Advance online publication. https://doi.org/10.1108/IJCHM-06-2021-0767

The research underscores the growing importance of big data analytics in analyzing travel trends. Data from social media, search engines or booking systems are making it possible to forecast traveler behavior with increasing accuracy. The use of machine learning methods, such as DBSCAN, SVM, ANN, LSTM or SARIMA, shows their effectiveness in analyzing a variety of data types - from images to texts to tourist flow figures. A particularly important area of research is forecasting tourist demand. Much of the analysis focuses on predicting the number of tourists, which is crucial for the tourism industry, especially in terms of managing infrastructure, optimizing the supply of hotel services and developing effective marketing strategies. Equally important is the analysis of tourists' preferences and satisfaction, where artificial intelligence methods make it possible to identify key factors influencing customer satisfaction. The use of social media feedback and tourist reviews makes it possible to better tailor offers to consumer expectations and more effectively personalize services. In conclusion, the use of advanced data analytics in tourism research contributes to a better understanding of market trends and travelers' preferences. The use of big data and artificial intelligence makes it possible not only to forecast tourism traffic, but also to adjust business strategies based on actual customer behavior. This trend will intensify with the increasing availability of data and the development of analytical tools.

# CHAPTER V:
# APPLICATION OF BIG DATA IN ANALYZING TOURISM DEMAND RESEARCH IN THE KUJAWSKO-POMORSKIE VOIVODESHIP

## 5.1. Diagnosis of internal and external factors affecting the tourism sector in the Kujawsko-Pomorskie Voivodeship

In proceeding with the diagnosis of the internal and external factors affecting the tourism sector in the Kujawsko-Pomorskie Voivodeship, the following methodological breakdown can be used:

- Impact of internal factors
    - Natural resources
    - Cultural resources
    - Historical resources
    - Infrastructure resources
- Impact of external factors
    - Socio-cultural conditions (including trends described in previous chapters)
    - Political and legal conditions
    - International conditions
    - Technological considerations
    - Economic conditions

Thus, starting from the diagnosis in terms of natural resources, it should be emphasised that the natural system of the Kujawsko-Pomorskie Voivodeship, located in the north-central part of Poland, is a key resource for the development of tourism. The region covers an area of 17,972 km² and is distinguished by its varied relief, including upland moraines, the Vistula, Noteć and Drwęca proglacial valleys and numerous lakes and forests. Terminal moraines, kames and drumlins add variety to the landscape, while areas such as the Toruńska and Grudziądz Basins add to the physiographic diversity. The highest point in the region is Czarna Góra (188.8 m above sea level) in the Krajeńskie Lake District and the lowest is the floodplain of the Vistula River at 12.2 m above sea level (Draft Spatial Plan, 2015).

The voivodeship is rich in mineral resources, such as rock salt, whose resources account for 37.7% of the national resources, as well as limestone, marl, peat, peat bog, quartz sands and medicinal waters. Mineral waters, chloride and sulphide, are found in places such as Ciechocinek, Inowrocław and Wieniec-Zdrój, and thermal waters have also been documented in Ciechocinek and Toruń. The resources of

therapeutic waters constitute approximately 12.2% of the national resources of this type, which makes the region attractive for spa tourism. It should be noted that the Kujawsko-Pomorskie Voivodeship ranks second in Poland (after the Zachodniopomorskie Voivodeship) in the number of spa tourism visitors (Kaszuba, 2023) - 153,170 inpatients compared to 192,813 in the Zachodniopomorskie Voivodeship.

A large part of the voivodeship is made up of forested areas and lakes, which occupy about 1.4% of its area. Forests cover 23.5% of the geodetic area of the region, with pine monocultures dominating in poor boreal habitats. Nature reserves, such as the Tuchola Forests, protect rare species of flora and fauna, including plants that prefer poor habitats. The province is also home to numerous animal species, including protected aquatic and migratory birds, as well as mammals such as raccoons and wolves. Ecological corridors of European and national importance that allow large mammals to migrate are an important part of the ecosystem. The province has 92 nature reserves, 10 landscape parks, 31 protected landscape areas and the Natura 2000 network, which includes areas of special bird and habitat protection. As part of UNESCO's Man and the Biosphere initiative, the province has the Bory Tucholskie biosphere reserve, which plays an important role in biodiversity conservation and environmental education (Draft Spatial Plan, 2015).

The natural attractiveness of the region is concentrated in areas such as the Tuchola Forest, the Brodnica Lake District, the Vistula River Valley, the area around Żnin with its set of lakes, and the Zalew Koronowski on the Brda River. The region has great tourist potential due to the presence of numerous lakes, river valleys and forests, as well as the possibility of developing water tourism on the Vistula, Brda, Wda and Drwęca rivers. The presence of mineral and thermal waters also favours the development of spa tourism, and the varied natural environment supports agrotourism. The location of the voivodeship in the lake district zone and the diversity of natural assets, such as gorges, river valleys, natural monuments and lakes, make the region an attractive place for tourists seeking contact with nature. The abundance of natural resources combined with a developed tourism infrastructure and nature protection offers great opportunities for the development of sustainable tourism (K-POT.pl, 2024).

The cultural heritage of the Kuyavian-Pomeranian region is extraordinarily rich and varied, reflecting a diverse range of tangible and intangible culture. Its development was influenced by numerous political, socio-economic and cultural transformations over the centuries. The region attracts tourists with its numerous architectural monuments, museums, cultural routes and artistic and musical events.

According to data from the Provincial Office for the Protection of Monuments in Toruń, the Kuyavian-Pomeranian voivodeship has 29595 historical architectural, construction, green and cemetery objects and 52330 immovable archaeological monuments. Among them are numerous Teutonic castles (e.g. the ruins of a castle in Venice and the Mysia Wieża in Kruszwica) and sacred buildings (the Collegiate Church of

Sts. Peter and Paul in Kruszwica, the oldest Benedictine monastery in Mogilno). Mennonite wooden architecture, represented by the arcaded house in Chrystkowo, is also an attraction. Industrial sites include the historic factories in Bydgoszcz, Toruń and Włocławek, a complex of sugar factories in the Chełmno region, the Bydgoszcz Canal, hydroelectric power plants in Żur and Gródek, and graduation towers in Ciechocinek. Technical monuments also include the 19th century salt mines in Inowrocław and the only lignite mine in northern Poland in Piła-Młyn. A curiosity is the Great Brda Canal with a unique aqueduct in Fojutowo. The region is also distinguished by park and palace layouts, such as the park in Lubostron or the palaces in Ostromecko, Lubraniec and Nawra. Wooden architecture includes a summer theatre in Ciechocinek and historic churches in Gąsawa and Włóki. Some of these objects have been moved to open-air museums, e.g. in Toruń or Kłóbka, to protect them from destruction (Diagnosis of the state and conditions of development of the Kuyavian-Pomeranian region, 2020)

Kujawsko-Pomorskie is also a region rich in museums and art galleries. The most important include the Archaeological Museum in Biskupin, the Museum of Pałuca Land in Żnin, the Museum of Soap and the History of Dirt in Bydgoszcz and the Wozownia Art Gallery in Toruń.

Prominent among the cultural routes are the Piast Route, linking key sites associated with Polish history, and the St James' Route, part of the European Camino de Santiago (Caminopolaco.pl, 2024).

The voivodeship organises numerous artistic events, including the Bydgoszcz Music Festival, the Camerimage Film Festival in Toruń, the International Festival of Romani Song and Culture in Ciechocinek, and the Knights Tournament at Golub Castle. Important cultural institutions include Opera Nova and the Pomeranian Philharmonic in Bydgoszcz, the Centre for Contemporary Art in Toruń and the Rothera Mills in Bydgoszcz.

The region has three UNESCO World Heritage sites: the medieval urban complex of Toruń (UNESCO.org, 2024), the Tuchola Forest Biosphere Reserve (K-POT.pl, 2024) and Bydgoszcz as a Creative City in the Field of Music (UNESCO.org, 2024)). Regional culinary products such as podpiwek Kujawski and plum jam from the Lower Vistula Valley, which are on the list of Protected Geographical Indication (MRiRW.gov.pl, 2024), are also an important part of the heritage.

Bydgoszcz and Toruń, the two capitals of the region, play a key role in cultural development. Bydgoszcz, known as an important music centre, organises numerous festivals, including the Bydgoszcz Opera Festival (Opera.Bydgoszcz.pl, 2024) and the Jazz Festival (BydgoszczJazzFestiwal.pl, 2024). Toruń, on the other hand, impresses with its wealth of historical monuments, including the Gothic Town Hall, the ruins of the Teutonic Castle and St John's Cathedral. Both cities also have a rich theatrical and museum offer, which enriches the tourist potential of the region.

Kujawsko-Pomorskie Voivodeship is distinguished by its rich historical heritage, which is an integral part of its cultural identity. The area includes historical regions such as Kujawy, Chełmno Land, Dobrzyń Land, Michałów Land, Pałuki, Krajna and Kociewie. The beginnings of settlement date back to the Stone Age, and in the Middle Ages the area of today's voivodeship developed intensively, resulting in the establishment of numerous towns. Many of them, such as Brodnica, Toruń, Chełmno or Włocławek, have preserved historic urban layouts entered in the register of monuments (Diagnoza stanu ..., 2020).

Rural heritage also plays a significant role. In Kujawy, Chełmno Land, Pałuki or Bory Tucholskie, elements of ancient rural layouts can be found. Among the oldest works of architecture are Romanesque churches, such as those in Kruszwica, Mogilno or Strzelno, which were built in the 11th-12th centuries. Gothic sacral buildings, found in large numbers in Toruń and Chełmno, are a testimony to the medieval heritage. In the modern period (16th-18th centuries) valuable monasteries and churches were built, e.g. in Bydgoszcz, Brodnica and Świecie. From later periods, orthodox churches, synagogues and churches have been preserved, including a Zaural-style church in Ciechocinek. The province is also famous for its defensive buildings. In the Middle Ages, Teutonic knights', bishops' and royal castles were erected here, of which those in Golub-Dobrzyń, Radzyń Chełmiński and Zamek Bierzgłowski, among others, have survived to this day. Cities such as Toruń, Grudziądz and Chełmno have fragments of city walls, towers and gates. Fortifications from the 17th-20th centuries, such as the Toruń Fortress and Grudziądz Fortress, are also valuable monuments. The region has witnessed numerous historical events, from battles with the Teutonic Knights, through the Swedish Wars and the January Uprising, to the September campaign of 1939. Memorials and places of martyrdom from World War II commemorate these dramatic events. The Kujawsko-Pomorskie Voivodeship was an important point on trade routes, such as the Amber Trail or medieval trade routes. Pilgrimage routes also led here, including the Way of St James to Santiago de Compostela.

The importance of historical heritage is indicated by the presence of eight sites on the list of Monuments of History. These include Biskupin, the Old Town in Toruń, the complex of graduation towers in Ciechocinek, the granaries in Grudziądz and the palace in Lubostron. These places, which are testimony to the past, represent exceptional tourism potential for the voivodeship (President.pl, 2024).

In recent years, the tourist attractiveness of regions, including the Kujawsko-Pomorskie Voivodeship, is no longer based solely on natural and cultural assets. A well-developed tourist infrastructure, including transport, accommodation, catering and a tailored tourist offer that meets tourists' changing expectations, plays a key role. Leisure comfort, a high standard of facilities and opportunities for active leisure and new tourist experiences are increasingly important.

Kujawsko-Pomorskie Voivodeship does not stand out in terms of the number of accommodation facilities compared to the rest of the country, but in recent years there has been a development of accommodation of a higher standard, especially hotels with conference facilities, catering for business, congress and sports and cultural events. As at 31 July 2022, there were 360 tourist accommodation establishments in the province, offering 31.0 thousand beds. Compared to the previous year, the number of facilities decreased by 0.3%, while the number of beds decreased by 0.3%. The accommodation density ratio places the voivodship in 8th place in the country. The largest number of beds is in Ciechocinek and Inowrocław - spa town municipalities. Ciechocinek, with 19 spa facilities offering 4,000 beds, accounts for more than half of the voivodeship's spa resources. Inowrocław, with 8 facilities, provides 1.9 thousand places. Wieniec-Zdrój in the municipality of Brześć Kujawski is also an important spa centre. Spa facilities, such as hospitals, sanatoriums and natural health care facilities, account for 18.1% of accommodation in Poland, making the voivodeship one of the leaders in this area (CSO, 2023).

Kuyavian-Pomeranian voivodeship is characterised by a well-developed cycling infrastructure. Both international routes, such as EuroVelo 2 and 9, and the Vistula Cycling Route (Eurovelo, 2024), which is the longest route in the voivodeship (467 km), run through the region. Local routes in Bory Tucholskie and the Vistula Valley are also popular, offering a variety of opportunities for cyclists.

The region's transport infrastructure, including roads, railways and the airport in Bydgoszcz, supports accessibility for tourists. The voivodeship is crossed by the A1 motorway and other important national roads, including the S5 and the planned S10, which connect the region with the largest centres in Poland. Good rail links and international waterways, which are part of the TEN-T network, increase the potential for external accessibility.

At the local government level, an important function is performed by the marshal of the voivodship, who supervises the activities of tourism organisers and the operation of hotel facilities. The marshal is responsible for maintaining provincial registers of tourism organisers and entrepreneurs facilitating the purchase of related services, as well as for supervising tour operators and tour guides (UFG Register, 2024).

An important element of tourism regulation is competition and consumer protection legislation. At central level, these activities are supervised by the Office of Competition and Consumer Protection and at local level by consumer ombudsmen. In the area of transport , the President of the Civil Aviation Authority, the Ombudsman for Passengers' Rights and the Railway Transport Authority play an important role.

In the context of technological considerations, there is no doubt that dynamic changes in tourism are leading to greater integration and internationalisation of activities. Local and regional players are

integrating into global booking and sales systems, which is creating new networks and organisations acting locally are competing globally. The use of technology enables faster and more accurate analysis of consumer data, allowing tourism businesses and local governments to manage tourism more effectively and increase competitiveness.

Modern data sources such as mobile advertising identifiers (GAIDs), data from booking portals, payment cards or AI cameras enable more sophisticated monitoring of tourism traffic. Social media, such as forums, blogs and social networks, play an important function in promoting tourism regions and products, while also enabling interaction with consumers. Opinions generated by internet users are seen as more credible than advertising content, especially among younger generations. These opinions influence decision-making, helping to select destinations, evaluate alternatives, avoid risks and plan trips effectively. Modern marketing efforts are increasingly based on content marketing strategies that focus on creating valuable content for potential customers. Instead of directly promoting products, content marketing provides information that helps consumers make decisions and reinforces brand loyalty.

Virtual reality (VR) and augmented reality (AR) are also gaining ground. VR enables virtual tours of places that are difficult to access or protected, which supports tourism promotion and staff training. AR enriches the tour with real-time information, such as directions along the route or additional data about attractions. Both technologies are tailored to different audiences: AR can appeal to tourists who prefer to interact with the real environment, while VR addresses the need for intense virtual experiences.

Digital tools that allow tourists to virtually experience tourism offers before they buy are growing in importance in modern marketing strategies. Virtual walk-throughs, videos and VR images help tourists to evaluate offers, which increases online bookings. Digital technologies also support the premium segment, where personalisation and a high standard of service are priorities.

The development of tourism in the Kujawsko-Pomorskie Voivodeship is also closely linked to the economic situation at both national and regional levels. Key factors influencing the sector are: Gross Domestic Product (GDP), the unemployment rate, wage levels and the purchasing power of the population.

An increase in GDP usually translates into an increase in consumer purchasing power and enables greater investment in tourism infrastructure. According to European Commission forecasts, the Polish economy will grow by 3% in 2024 and by 3.6% in 2025 . Such forecasts suggest a potential recovery of the tourism sector in the coming years (European Commmission, 2024).

A low unemployment rate undoubtedly favours the development of tourism, as it increases the financial stability of local residents. In Poland, according to the Central Statistical Office, the registered unemployment rate in December 2023 was 5.2% . In Kujawsko-Pomorskie Voivodeship, the rate was

higher at 7.1% . The highest unemployment was recorded in Radziejów County (14.3%) and the lowest in Bydgoszcz (2.3%) and Toruń (3.0%). Such variation can affect local tourism development, especially in areas with higher unemployment, where the tourism sector can be an important source of employment (GUS, 2024).

In turn, high inflation has a negative impact on the cost of travel, especially in terms of transport, accommodation and catering. According to the Central Statistical Office (GUS), inflation stood at 6.1% year-on-year in December 2023. Although this is a decrease compared to earlier months, it still represents a significant burden on household budgets, which may limit tourism spending (CSO, 2024).

The diversity of purchasing power in Europe creates an opportunity to attract tourists from countries with higher incomes. Countries such as Germany, the United Kingdom, Switzerland or Norway, where purchasing power is several times higher than in Poland, can be key target markets for the tourism offer of the Kujawsko-Pomorskie Voivodeship. Warsaw, as the leader in Poland with a purchasing power of EUR 17,980 per capita, also represents a significant internal market. It is worth noting that the average gross salary in Poland in 2023 was PLN 8033, while in the Kujawsko-Pomorskie Voivodeship it was PLN 6987 , which places the region below the national average (CSO, 2024). In an international context, the purchasing power per capita of Poland in 2023 was EUR 10.9 thousand, which is well below the European average of EUR 17.6 thousand (GfK Purchasing Power Europe, 2023).

The analysis of economic conditions indicates that the development of tourism in the Kujawsko-Pomorskie Voivodeship requires targeting promotional activities at international markets with high purchasing power. The development of tourism infrastructure and optimisation of the offer in terms of price can contribute to increasing the attractiveness of the region, even in the face of economic difficulties.

## 5.2. Overview of tourism sector research in the Kujawsko-Pomorskie Voivodeship

As mentioned earlier in this publication, tourism demand is directly linked to the supply of hotel accommodation, and an increase in the number of available beds contributes to an increase in tourist arrivals. Research indicates (Habibi, 2016) that destinations with well-developed hotel infrastructure are more likely to attract tourists, resulting in higher levels of tourist arrivals. This mechanism is based on the assumption that a greater supply of accommodation not only meets tourists' needs, but also influences the perception of the destination as more accessible and welcoming. When planning a trip, tourists pay attention to the availability and variety of accommodation options, which can influence their final destination decision. Additionally, the availability of accommodation influences price competitiveness, which is also a factor that attracts tourists, especially those who are price sensitive.

The main premise for the search for an answer to the question of how big data can be used to analyse tourism trends and segment tourism demand was the author's previous experience in monitoring tourism in the Kujawsko-Pomorskie region. Originally (e.g. Brudnicki & Barczak, 2012 and Brudnicki & Barczak 2015), research monitoring tourist traffic in the region was done in a traditional way.

Namely, the study of tourism, due to its specific nature and difficulties in fully recording data, the dynamics and mobility of tourists, and the variability of population structures, has been and still is a complex issue. These cognitive barriers exist irrespective of the research area - both in the regions generating tourism and in those receiving it. For this reason, and on the basis of an objective assessment of the solutions used in other regions, the 2012 research decided to use several complementary and verifying research methods. However, the most important element of the study was the use of a survey questionnaire. The survey questions addressed three main cognitive areas. Firstly, they were aimed at parameterising tourism, including an analysis of aspects such as the purposes of arrival, length of stay, frequency of visits, choice of accommodation, methods of obtaining information about the place of stay, accompanying persons, choice of means of transport and use of catering and other tourism services, as well as planned expenditure. Secondly, the questions concerned an assessment of the attractiveness of the destination and the level of quality of the infrastructure, such as the ability to recommend the city of Bydgoszcz as a tourist attraction. The questions also sought to assess the favourable and unfavourable aspects related to the development of tourism services. Thirdly, the survey was intended to identify the profile of tourists, including their leisure preferences and their

socio-demographic characteristics, such as their geographical origin (Brudnicki & Barczak, 2012). However, this approach had one major drawback, the collection of questionnaires took place over a long period of time (tourist season) and the results were available at the beginning of the following calendar year. This was due to the time-consuming nature of the research process, including the preparation of the final report, which had both quantitative and qualitative results.

Meanwhile, in the monitoring of tourism in the Kujawsko-Pomorskie Voivodeship in 2015, 5,000 interviews were decided upon, which met the requirement of obtaining a small estimation error (approximately 2%) and influenced the increase in the level of representativeness. In this case, the experience gained by Anszperger (2009), who, together with his team, defined the interview collection sites, was also used. The research tool was an interview questionnaire, developed after a detailed analysis of the 2009 and 2010 reports, and consisted of 20 questions and a section on the demographics of the respondents. Compared to previous editions, the number of questions was reduced. Most of the questions were closed, but three open-ended questions gave respondents the opportunity to speak freely. Some questions were province-specific, while others were targeted at a specific research region. The questionnaires were available in three language versions: Polish, English and German, and the electronic form was innovatively integrated with a mobile application that enabled online interviews to be conducted using tablets. The electronic version of the questionnaire eliminated the need to code the data and allowed for ongoing access to the results, including monitoring of the number of interviews conducted, analysis of responses and control of the research process. The questions included in the interview questionnaire allowed for a thorough study of tourist behaviour, making it possible, among other things: to create a detailed profile of visitors, both domestic and foreign, taking into account such parameters as country and place of residence, age, gender, education level, professional and material status. In addition, the questions allowed for an analysis of the specifics of the visits, taking into account the purposes of the trip, length of stay, accompanying persons and means of transport used. The questionnaire also made it possible to examine tourists' preferences and behaviour, assess satisfaction with the tourist offer, evaluate the attractiveness of a given place and estimate the average expenditure incurred during the stay (Brudnicki & Barczak, 2015).

Another study of tourist traffic in the Kujawsko-Pomorskie Voivodeship was conducted in 2021, and the data used in the study came from a data warehouse, which was obtained through queries and mobile ad impressions, as well as from publishers of specific applications. According to the contractor, it had a database of 24 million unique users in Poland during the study period. The survey and database used the DSP platform and the RTB mechanism. The DSP (Demand-Side Platform) is a technological tool that enables the automatic purchase of advertising space in digital media. It works

on the principle of optimising advertising campaigns, allowing advertisers to precisely plan their activities, select a target group and set budgets. The buying process takes place automatically and in real time, allowing ads to be dynamically tailored to users using different media, such as websites, mobile applications or online video. RTB (Real-Time Bidding) is a mechanism for buying advertising space in real time that works on an auction basis. Advertisers participate in auctions, competing to display their advertisements on a given space. The winner of the auction is the one with the highest bid per ad display. The RTB process takes place in fractions of a second, allowing ads to be instantly matched to website visitors or app users. The automation and speed of RTB mean that ads can be more precisely targeted and used effectively. The result of the report was a presentation of tourist traffic in the Kujawsko-Pomorskie Voivodeship including domestic and foreign tourists as well as visitors (Selectivv, 2021). This was a step towards a significant change in data acquisition and analysis, as it allowed for faster data acquisition and analysis. However, the results presented in the report differ significantly from the results obtained using other methods, i.e. the results presented in the big data analysis significantly underestimated the volume of tourist traffic, if only in comparison with the number of tickets sold or in relation to the degree of utilisation of accommodation according to the CSO. Nonetheless, as a result of this research, it was noted that the use of big data analyses to study tourism demand is a desirable direction in tourism research.

Research covering the entire Kuyavian-Pomeranian region was conducted four times: in 2009, 2010, 2015 and 2021. Their main objective was to identify and provide a descriptive overview of domestic and foreign tourist traffic, as well as to examine the characteristics and opinions of tourism participants, in particular regarding the level of satisfaction and assessment of the tourist attractiveness of the region. The research was mainly based on partial fieldwork, conducted through face-to-face interviews using a questionnaire, and on the analysis of data collected through the Desk Research method used in the preparatory phase. The analysis of reports from previous years (2009, 2010, 2015) indicates that the research was conducted in accordance with objective scientific standards and tourism monitoring practices, which ensured the high quality of the collected research material. In summary, the research carried out in the Kujawsko-Pomorskie region in recent years differed in terms of methodology, thematic scope and spatial coverage. Due to the high tourism potential and the impact of tourism on the economy, these studies were most often conducted in larger centres such as Toruń and Bydgoszcz. Less frequently they covered smaller tourist destinations such as Grudziądz and Chełmno. Despite being systematic, these studies differed in their methodology, making them difficult to compare (Barczak, Brudnicki & Wierzbiańska, 2023).

In order to carry out tourism research properly, it is necessary to apply a complex methodology that combines a variety of measurement techniques and uses data from many complementary sources. Achieving meaningful results requires both the experience of the researchers and a thorough knowledge of the specifics of the tourism market in question.

Therefore, Barczak, Brudnicki & Wierzbiańska (2023) developed the concept of a methodology for researching tourism in the Kujawsko-Pomorskie Voivodeship entitled the Tourism Barometer. According to the study, the Tourism Barometer was created to provide comprehensive information about the tourism sector in the Kujawsko-Pomorskie Voivodeship. This tool is to enable not only the estimation of the scale of tourist traffic in the region, but should also provide a wide range of data, the appropriate analysis of which may contribute to the increase of competitiveness of the voivodeship.

The Tourism Barometer concept is distinguished by its innovative combination of traditional research methods and modern technological solutions (see Figure 31). This combination allows for a detailed and holistic understanding of tourism and other related phenomena, and supports a more effective involvement of various stakeholders in the research process. Moreover, it allows the results of the current analyses to be compared with data from previous years, further strengthening its analytical value.

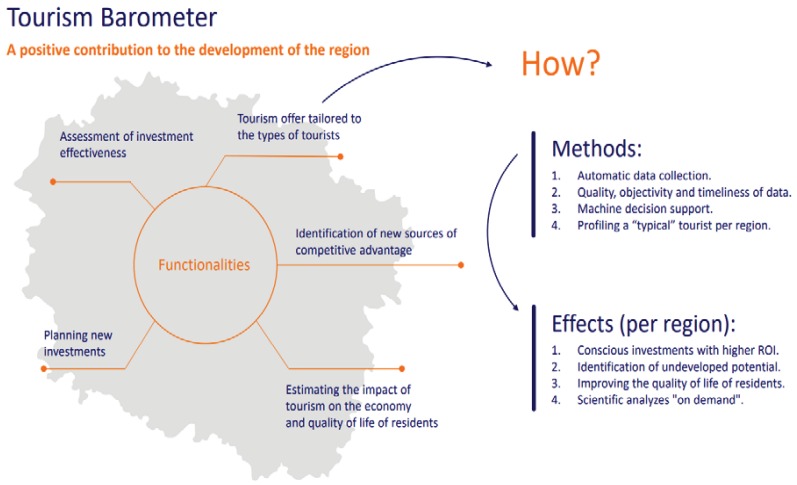
*Figure 31 . Data sources planned for use in the Tourism Barometer*




Source: Barczak M., Kulikowski S., Karczewski W., World Canals Coference 2024, Bydgoszcz (abstract).

Central to this approach is a system based on the diverse data sources that feed the tool. The proposed concept includes a dozen complementary and mutually validating research methods that can be used both in combination and independently. One of the key advantages of this solution is that some of the results can be obtained in real time via an interactive website, offering the user access to a dashboard. This will give report recipients the ability to independently select data sources of interest and generate individual analyses and reports.

*Figure 32 . The idea behind the Tourism Barometer*




Source: Barczak M., Kulikowski S., Karczewski W., World Canals Coference 2024, Bydgoszcz (abstract).

In the author's opinion, the Tourism Barometer is an important step towards the holistic management of the tourism sector in Kujawsko-Pomorskie Voivodeship, providing a detailed picture of its state and impact on the economy and society (see Fig. no. 32).

## 5.3. Exploring technologies, methods and tools for tourism demand research

The analysis to date shows that researching tourism in a given area requires a complex methodology, taking into account a variety of research methods and data from different, complementary sources. In Poland, such studies have already been conducted many times in different regions, but due to significant methodological differences the results were often difficult to compare. Most often, research has focused on analysing the profile of tourists, their travel behaviour and

estimating the scale of tourism. The specific objectives of the studies include estimating the number of tourists, dividing them into domestic and foreign tourists and same-day visitors, analysing the dynamics of tourist traffic over the years, and examining the demographic structure of visitors. In addition, the research included an analysis of the purpose of visits, length of stay, means of transport and level of spending, as well as an assessment of tourists' behaviour, satisfaction with tourism services and opinions on safety and value for money (Alejziak, 2010).

The overview presented in the figure 33, however, refers to traditional approaches in the esplanade of tourism research, while global experiences incorporate new technologies and thus access to data sources based on big data in their approaches.

For example, a study was carried out in the Czech market (Krkonošský národní park, 2024) in 2017-2019, for which the primary data source was mobile phone operator data. The SIM card data used in the tourism research was anonymised and then used to create geodemographic matrices. This type of analysis made it possible to determine tourists' mobility, origin and behavioural patterns in different locations. This also made it possible to track the nationality of tourists based on the country of origin of the SIM card and to analyse the behaviour of domestic tourists, e.g. movement patterns between different tourist spots in the country. A similar study was carried out in Estonia, where mobile network data was also used, but based only on roaming data. This method allowed estimation of the number of foreign tourists, but did not take into account domestic tourists (Ahas, Aasa, Silm & Tiru, 2007).

*Figure33 . Basic parameters of tourist traffic*

| | |
|---|---|
| Parameters of tourist traffic | •Volume, seasonality, lenght of stay, and frequency of arrivals |
| | •Structure of tourist traffic |
| | •Spatial characteristics |
| | •Organisation method |
| | •Purpose of travel |
| | •Reasons for choosing a tourist destination |
| | •Ways of spending leisure time |
| | •Expenses during the stay |
| | •Opinions about the region |

Source: Matczak, A., 1992, *Model badania ruchu turystycznego. Studium metodologiczne [Tourist Traffic Research Model: A Methodological Study],* Acta Universitatis Lodziensis, Wyd. UŁ, Łódź. M. Pociecha, 2002, *Metody statystyczne w zarządzaniu turystyką [Statistical methods in tourism management,]*, Albis, Kraków. R. Brudnicki, 2006, *Społeczno-ekonomiczne uwarunkowania zachowań turystycznych mieszkańców Torunia [Socio-economic determinants of tourist behaviour of Toruń],* UMK, Toruń. *Badania konsumentów usług turystycznych w regionach [Research on Consumers of Tourism Services in Regions],* 2015 (E. Dziedzic red.), POT, Warszawa. Włodarczyk, B. (red.), 2011, *Ruch turystyczny w Łodzi i w województwie łódzkim w 2010 roku [Tourist Traffic in Łódź and the Łódź Voivodeship in 2010],* ŁTN, Łódź. Brudnicki R., Barczak M., 2017, *Badania ruchu turystycznego – klucz do sukcesu oferty turystycznej [Tourism research - the key to the success of a tourism offer],* [w:] B. Sowińska, M. Barczak (red.), Zeszyt Naukowy WSG Tir nr 14, Bydgoszcz

An interesting approach was proposed by a team of researchers (Zheng, Liu, Yuan & Xie, 2012), which was based on an analysis of taxi rides, where the data came from taxi corporations. Another study conducted in a National Park within Alaska conducted an analysis of tourist traffic on popular hiking trails and data was collected using GPS information generated by guides of tours organised in the park (Stamberger, van Riper, Keller, Brownlee & Rose, 2018). The GPS tracking method was also used in a study of tourist traffic in Tenerife. In this research, GPS data were combined with information from social networks. The methodology was based on data from the photo-sharing website Flickr and GPS tracks obtained from the Wikiloc platform. With these tools, tourist traffic was mapped, analysing the spatial distribution of tourists, the most frequently visited attractions, trail networks and data on length of stay and countries of origin of tourists. In addition, a practical application of geotagged data has been proposed to plan optimal locations for new facilities such as information points (Barros, Moya-Gómez & Gutiérrez, 2019).

The search for holistic research methods combining different large data sets was also handled by the research team implementing the ESSnet Big Data project, i.e. the CSO (Statistics Poland), which was the leader of the WPJ, as well as NSI (Statistics Bulgaria), ELSTAT (Statistics Greece), HSL (Hesse Statistical Office, Germany), ISTAT (Statistics Italy), CBS (Statistics Netherlands), INE (Statistics Portugal) and NSI (Statistics Slovakia). Work on this project lasted from 2018 to 2020. The outcome of the project is the basis for the preparation of a prototype Tourism Integration and Monitoring System (TIMS) that would support the activities of European Union countries in the field of tourism statistics and assist in monitoring changes in a country's tourism sector (European Commission, 2020). The research team made a wide range of proposals, but a key element in the context of tourism surveys is their suggestion that, in addition to traditional statistical data from authorities and institutions such as the Border Guard, the surveys should also include information from online platforms related to travel planning, organisation and monitoring, using the 'web scraping' method. The WPJ team also proposes to include data from airports and analysis of credit and debit card transactions to get a more complete picture of tourist behaviour. As recommended by ESSnet Big Data II (Quaresma et aln, 2020), web scraping is the fastest and least costly way to extract statistical data for official statistics. However, in order to obtain high quality data, it is necessary to ensure that the raw data is readable and can be linked to other formats (Eurostat, 2020).

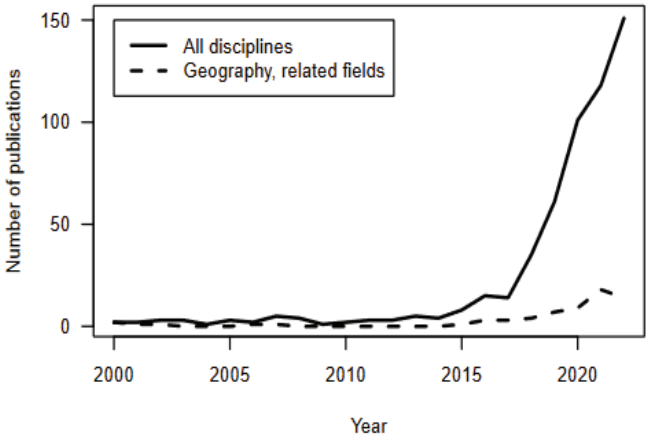
Web scraping is increasingly used by national statistical institutes (NSIs) to relieve respondent burden, speed up statistical processes, create new indicators and analyse background variables and (sub-)population characteristics (Quaresma et al., 2020).

## 5.4. A working model for big data analytics in the tourism sector

Since its inception in 1991, the World Wide Web ('Web') has significantly changed the way people get information. Today, the average user spends nearly seven hours a day on the Internet, accounting for more than 40% of his or her time when active (DataReportal, 2022). Despite these fundamental and rapid changes in personal and professional life, research methodologies do not yet fully reflect this radical transformation (Brenning & Henn, 2023). Web scraping is only just beginning to gain prominence in research - according to a search in the Web of Science Core Collection (WoS; search terms 'web scraping', 'web-scraping' and 'web-scraping' in all fields - see Fig. no. 34). Of course, the structure and content of websites can change at any time, often without notice, and service providers can introduce unexpected technological barriers. For this reason, researchers must be prepared to

dynamically modify scraping tools, which may require significant investment in software development.

*Figure 34 . Use of web scraping in academic literature since 2000*




Source: Brenning, A., & Henn, S. (2023). Web scraping: A promising tool for geographic data acquisition. arXiv. https://doi.org/10.48550/arxiv.2305.19893

Web scraping is used in research and commercial activities because it enables the efficient creation of large and fine-tuned data sets at relatively low cost. This allows companies to dispense with traditional, less efficient data collection methods, allowing them to respond in real time to rapidly changing market trends and consumer expectations (Hillen, 2019).

Web scraping has become a major tool for statistical and scientific research of all types (Snell & Menaldo, 2016). The increasing availability of open-source web scraping tools has greatly simplified the process for researchers, allowing them to create customised scrapers to meet their specific needs (Han & Anderson, 2021).

The source of data for web scraping is all public websites available on the Internet. Consequently, there are three basic types of web scraping tools:

- crawlers that directly access data from the web and extract information based on HTML/XHTML tags,
- robots that access structured or semi-structured data in various formats (e.g. JSON, CSV),
- applications that use APIs to transfer data directly from the host server to the local machine.

One promising possibility is to use an API, provided by the data provider. However, APIs are often difficult to obtain, only available for a limited time, or require upfront fees, even if researchers manage to access them. Moreover, APIs typically do not provide the full set of necessary data (Han & Anderson, 2021).

As computer performance advances, so does the potential for more sophisticated and intelligent web scraping algorithms (Asma, 2020). An important factor will be the increasing use of artificial intelligence and machine learning in these techniques. Currently, most scraping tools are based on simple methods that work for many sites, but are not universal. The implementation of machine learning could lead to the development of scraping techniques that will 'learn' from previous experience and be able to work effectively on a much wider range of websites (Suganya, 2019).

Figure 35 shows the process of generating information extracted through web scraping. It consists of several stages, which are visualised as successive steps from web access to data analysis. Below is a description of each step:

Stage 1 Internet access:

- URL searcher: This tool searches the internet for URLs that will be useful for data extraction.
- Scraper: A web scraping tool that collects web content, such as text, images or other information posted on a page.

Stage 2 Storage:

- URLs retrieved: URLs retrieved are stored in a database or other storage system, where they will be available for further analysis.
- Scraped content: Collected data from web pages (e.g. text, images) is saved and indexed so that it is ready for processing.

Stage 3 Data preparation:

- URL scorer: A tool for evaluating found URLs for relevance.
- Tokenisation: The process of dividing text into smaller units (tokens), such as words or phrases.
- Word filters (e.g. stop words): Elimination of so-called "stop words" (e.g. "and", "and", "but"), i.e. words that are not relevant to the analysis.
- Language-specific lemmatisation: The process of converting words to their base forms (lemmas) to aid text comprehension.

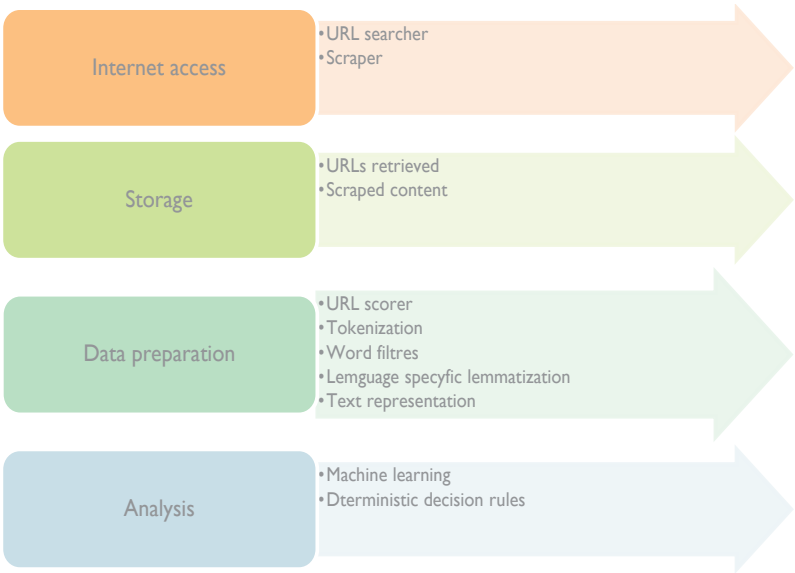
- Text Representation: Data is represented in formats such as 'Bag of Words' to facilitate analysis:
- Bag of Words: A simple representation of text where it counts how many times each word occurs in a document e.g. for sentiment analysis.
- Language encoding: The process of converting text into formats that computer systems can understand.
- Engineered Features: Creation of additional features based on the collected data that can help further analysis.

Stage 4 Analysis:

- Machine Learning: The process of using algorithms that learn from data and find patterns in the information collected.
- Build training & test sets: Data is split into learning and test sets to train the machine learning model and evaluate its effectiveness.
- Training classifier: Teaching a classification model so that it can recognise patterns in the data.
- Application classifier: Application of a trained model to classify new data.
- Deterministic decision rules: The creation and application of established decision rules to make decisions based on collected data.

The whole process illustrates how data extracted from websites is collected, stored, processed and analysed using advanced techniques such as machine learning.

*Figure 35 . The process of generating information extracted through web scraping*




Source: Quaresma, S., Maślankowski, J., Salgado, D., Ascari, G., Brancato, G., Di Consiglio, L., Righi, P., Tuoto, T., Daas, P., Six, M., & Kowarik, A. (2020). *Revised version of the quality guidelines for the acquisition and usage of big data (Deliverable K3: Work package K: Methodology and quality)*. ESSnet Big Data II, Eurostat. https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en

Web scraping studies for the hospitality industry were conducted by Han and Anderson (2021) and Doan et al. (2023), among others, the former presenting methods and tools (including code in Python) to populate traditional data sources (using Tripadvisor as an example), while the latter studies focused on data extraction from Booking.com and the results of these studies confirmed the feasibility and effectiveness of using web scraping techniques to extract big data in the tourism and hospitality sector. In this research, the focus was on hotels located in Ho Chi Minh City, Vietnam, to collect data from Booking.com. Web scraping techniques were applied using a Node.js-based crawler programme, which was developed using the Puppeteer library. The scraping programme was run in February 2023, collecting data on all visible hotels listed on Booking.com in Ho Chi Minh City at that time. The collected data was saved in CSV format for further analysis. The web scraping methodology used in this study enabled the effective and comprehensive collection of digital content from Booking.com, providing a large and diverse data set for analysis (Doan et al., 2023). The study included a lot of information, but most of it was about the past, as it referred, for example, to review ratings, review date, photos posted after a stay. Meanwhile, according to the author, the web scraping study of

Booking.com can also be used to analyse and forecast demand in tourist reception areas in the future. Following this approach, the author proposed Tourism Demand Horizon in his research.

In contrast, the study by Abdirazakov et al (2023) relied on quantitative data collected from Booking.com to examine the relationships between variables such as star ratings, prices, customer ratings, distance from the city centre and the number of hotel reviews in Almaty, Kazakhstan. The analysis showed that hotels with different star ratings differed significantly in terms of customer satisfaction. The results indicated a positive correlation between star ratings and overall customer reviews, meaning that hotels with higher star ratings tend to have higher customer reviews. Further analysis of the relationship between star rating and hotel price confirmed that hotels with higher star ratings have higher prices, which is consistent with the general perception that they offer better facilities and services. With regard to the distance of hotels from the city centre, the study found that hotels with higher customer ratings tended to be closer to the centre, although the relationship was weak and negative.

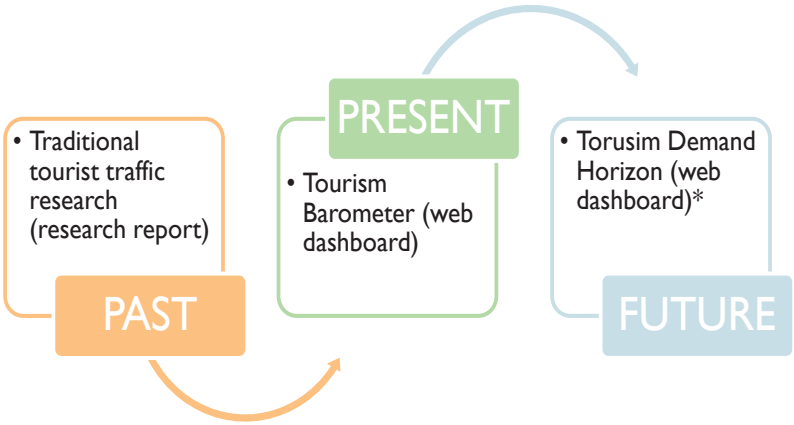
In contrast, Kamarazaman et al (2024) conducted a study using web scraping techniques on Tripadvisor for the Melaka region in Malaysia. Melaka was chosen as a research site because it is one of the areas recognised by UNESCO. The analysis covers 200 POIs (points of interest) in Melaka, which extend to the central areas and buffer zones of the city. These points of interest have been categorised into four heritage categories: built heritage, natural heritage, personal heritage and living heritage, some of which fall into more than one group. The TripAdvisor platform was used to collect the data, capturing a total of 14 different attributes. In addition, the study suggests that there is a weak positive correlation between price and distance from the city centre, meaning that a closer location to the city centre may lead to higher hotel prices. An analysis of the relationship between the number of reviews and the overall rating of a hotel revealed that more reviews may influence a slightly higher average rating, indicating that customers are more likely to give higher ratings to hotels with more reviews.

Online hotel review platforms such as Booking.com and Expedia play a key role in the travel industry, making it easier for travellers to book accommodation, flights and other travel services. Rather than generating purely social feedback, as with sites like Yelp or Tripadvisor, sites like Booking.com rely heavily on user transactions, making them transactional online travel agencies (OTAs). This allows customers to book accommodation and services at their convenience, using a global platform available in 43 languages. One of the unique aspects of Booking.com is its rating system, which is based on reviews from previous guests. After a stay, guests give ratings in various categories such as staff, cleanliness, location, comfort and value for money. These ratings are processed

and updated in real time, ensuring that the hotel's overall rating is a constant reflection of the current quality of service. With a feature to collect comments and reviews, the service allows guests to share their experiences, helping subsequent users to make informed decisions. Research, such as that conducted by Fleischer (2012), shows that factors such as location, hotel size and additional amenities can have a significant impact on room rates and customer decisions.

Traditional tourism surveys to date have been characterised by their time and cost intensity, which mainly resulted from the need to obtain information in the field through interviews with tourists and visitors. In addition, the processing of the data from such surveys usually took another few months. As a result, the recipient received a research report presenting information on what had happened in the tourism sector in the previous year, and therefore the data were historical.

*Figure 36 . The Toruism Demand Horizon idea*




Tourism Demand Horizon: a concept applied to the author's empirical research

Tourism Demand Horizon+: a concept applied to the theoretical assumptions of the new model

Source: own study.

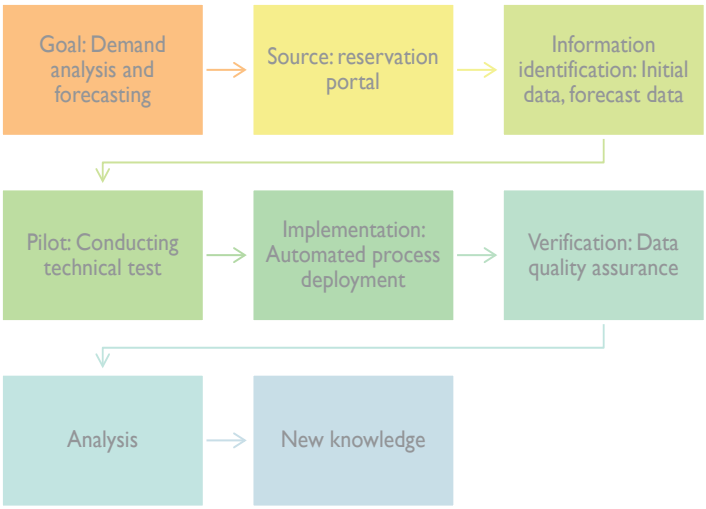
In a contemporary approach using, inter alia, big data, solutions have been sought (Barczak, Brudnicki, Wierzbiańska, 2023) to show in real time the state of the tourism sector, e.g. in a region (Tourism Barometer). To this end, tools were designed using, among other things.

- QR codes surveying NPS among tourists and visitors,
- mobile data in the form of Google GAID or Apple IDFA,
- data from AI cameras and curtain sensors located primarily in non-ticketed attractions.
- Web page scraping.

Furthermore, in such an approach it is possible to use transaction data from payment cards (Mastercard, 2024). Such solutions allow the use of the PowerBI software to present changes in the sector, e.g. via the online dashboard available on a daily basis. In this way, the research world was looking for solutions that would have higher application values from the point of view of the usefulness of the data for the tourism sector and its individual industries. The proposed concept was a unique combination of classic research methods and modern technological tools.

The experiments carried out to look for new methods and tools to study the tourism sector led the author of this publication to look for a solution that would allow him to see forecasts for the tourism sector at a regional level (although the solution could also be scaled up to other levels) in the future. The method was called the Toruism Demand Horizon and its assumptions are presented in Table 39 and Fig. 37 and Fig. 47.

*Figure 37 . Simplified diagram of how the Torusim Demand Horizon works, using big data from Booking.com as an example*




Source: own study.

*Table 39 . Process for capturing data for Tourism Demand Horizon using web scraping (methodological concept)*

| **Theoretical assumptions of the process according to. Brenning, A., & Henn, S. (2023)** | **Theoretical assumptions Tourism Demand Horizon** |
|---|---|
| Determine the purpose of the study and the data requirements - Identify what information is needed for the study. | The aim is to analyse and forecast tourist demand using real information from the booking portal about changes in the supply of facilities . [2]<br><br>Data necessary for the research process:<br>• A list of objects indicating their information for the date indicated.<br>• Creation of requests for the availability of facilities on specific dates in the future. |
| Selection of appropriate online sources, e.g. websites - Selection of sites to be used for data extraction. | For this publication, it was decided to download data from the Booking.com platform.<br><br>Ultimately, the use of other systems, i.e. Airbnb, Hotels.com or TripAdvisor, is suggested, with changes to their internal rules for users. All the portals indicated also transmit historical data directly to the European institutions. |
| Assessment of legal and ethical issues - Verify that web scraping on selected sites complies with applicable law and ethical principles. | Booking.com in its General Terms and Conditions (Booking.com, 2024) does not directly address the issue of scraping (automatic downloading of data) from its website. However, Booking.com reserves the right to restrict or block access to its platform in the event of suspected abuse[3] . As a result of scraping, it has been established, that it is possible to retrieve information on a maximum of 1,000 properties. [4] |

[2] Tourist demand refers to the number of tourists and their travel and service preferences, and analysis of the supply of accommodation provides information on the actual use of these services, which is one of the best measures of the level of demand. The use of tourist accommodation is largely determined by the demand generated by tourists. Higher levels of bookings and occupancy indicate higher demand, while a decline in the number of nights rented may signal a weakening of interest in a particular region or facility. Such data allows us to understand not only the overall number of tourists, but also their preferences, e.g. the type of facilities chosen (hotels, guesthouses, campsites) and the seasonality of demand.

[3] The Portal closely monitors each visit to the Platform and blocks anyone (and any automated system) if there is a suspicion that: an unreasonably large number of searches are being carried out; devices or software are being used to collect pricing or other data; an activity that unduly burdens our Platform

[4] Hotels.com: Hotels.com's terms and conditions include a provision prohibiting the use of automated means to access their systems, including data scraping - "You also agree not to: (...) access, monitor or copy any of the content of our Service using a robot, spider, scraper or other automated tool or any manual process - .

Airbnb: Airbnb also prohibits unauthorised access to its platform, which includes using bots or other automated methods to collect data without consent. - "Do not use bots, crawlers, scrapers or other automated tools to access, collect data or other content or otherwise interact with the Airbnb Platform."

Tripadvisor: Tripadvisor makes it clear in its terms of use that the use of automated means to access site content is prohibited. - "download any Content unless it is expressly made available for download by the Tripadvisor Companies." https://tripadvisor.mediaroom.com/PL-terms-of-use

| | |
|---|---|
| Identification of pages and information to be retrieved - Identify the specific pages and data to be retrieved. | Booking.com<br>• Output data: Name of object, url, city, province, date of scraping<br>• Predictive data: Name of facility, url, address, GPS coordinates, description of facility, raiting, voivodship, number of rooms, number of beds in rooms, date of data download, date of data verification (future dates indicated), number of vacancies. |
| Assessment of technical feasibility (pilot study) - Carry out technical tests to ensure that data acquisition is possible and effective. | The piloting was carried out several times and the results allowed for adjustments to the methodological assumptions made, e.g. in terms of the number of scraped objects or information about them. |
| Implementation of an automated web scraping process | Implemented 13.11.2024 |
| Testing, debugging and validation of the scraping tool - Validation of the tool and debugging. | Implemented from 13.11.2024- 4.12.2024 |
| Data quality assessment - Ensuring that the data acquired is as expected and of sufficient quality. | The substantive assessment of the data obtained and the verification of the data from the files obtained against the information available on the portal showed that there were no errors of a quantitative or qualitative nature. |
| Setting a timetable | Described in Table 40. |

Source: own study.

In line with the tool's described methodology, the research used Phyton software in the implementation phase of the automated process, which downloaded data from Booking.com using two scripts (see Appendix 1). In each iteration, the data were exported to xlsx format and then SQL operations were performed, including updates to the poi_list_view_clean table and cleaning of the poi_details_clean table using the TRUNCATE TABLE command. After each operation, data was transferred to the repository using Push/Pull commands. Webscraping was repeated four times, with the range of retrieval systematically decreasing. The main criterion for the selection of specific objects was the targeted study of objects for the Kujawsko-Pomorskie Voivodeship and the fact that they were available on the first day of web scraping. Thus, 802 facilities were defined in the list, which on 13.11.2024 had free rooms on offer for the indicated dates. The list of objects had information about the name of the object, web address, location (town and voivodeship), as well as the date of creation of the dataset.

*Table 40 . Data acquisition process in research*

| Date of collection* | Date from which data were extracted | | | |
|---|---|---|---|---|
| 13.11.2024 | 13-14.11.2024 | 20-21.11.2024 | 27-28.11.2024 | 4-5.12.2024 |
| 20.11.2024 | Nd. | 20-21.11.2024 | 27-28.11.2024 | 4-5.12.2024 |
| 27.11.2024 | Nd. | Nd. | 27-28.11.2024 | 4-5.12.2024 |
| 04.12.2024 | Nd. | Nd. | Nd. | 4-5.12.2024 |

*launching the script

Source: own study.

Table 41 presents a summary of the facilities available on the booking portal by location in the Kujawsko-Pomorskie Voivodeship, based on data as of 13.11.2024. The table includes the number of facilities for each location, which allows an analysis of the distribution of accommodation. The largest number of facilities is located in Toruń, where two categories are distinguished: Toruń, Old Town (241 facilities) and Toruń (113 facilities), making a total of 354 facilities in the whole of Toruń. This division indicates a difference in the way objects are described in the booking portal, where the Old Town is treated as a separate location. Bydgoszcz is second in terms of the number of sites, with 160 available locations. Other localities, such as Ciechocinek (34), Inowrocław (24), Grudziądz (23), Włocławek (22), Chełmno (17) and Brodnica (12), show a significantly lower number of facilities. The remaining localities with less than 10 sites offer a total of 156 sites

*Table41 . Overview of the facilities available on the booking portal by locality*

| Name of locality | Number of objects | |
|---|---|---|
| Toruń, Old Town | 241 | 354 |
| Toruń | 113 | |
| Bydgoszcz | 160 | |
| Ciechocinek | 34 | |
| Inowrocław | 24 | |
| Grudziądz | 23 | |
| Włocławek | 22 | |
| Chełmno | 17 | |
| Brodnica | 12 | |
| All others (less than 10 objects) | 156 | |
| Total | 802 | |

Source: own compilation based on 802 venues available on Booking.com as at 13.11.2024.

For comparison, in 2022, the Kujawsko-Pomorskie Office of Spatial and Regional Planning in Włocławek conducted an analysis of the accommodation facilities in the Kujawsko-Pomorskie voivodship (Kujawsko-Pomorskie Office of Spatial and Regional Planning, 2022), the result of which is presented in Table 42. The analysis shows that the number of accommodation facilities in the voivodship was 804. This was twice as high as the figure indicated by the CSO, as an important element affecting CSO statistics is the listing of facilities with only more than 10 beds. Therefore, if we take into account the facilities in the register of the marshal's office (161 facilities) and the municipal registers (643 facilities) the number of facilities is just mentioned 804. In addition, the number of facilities available for the Kujawsko-Pomorskie voivodship was verified in the quoted report and estimated at 752 facilities.

*Table 42 . Summary of the indicated number of facilities and accommodation places in the sources used in the Kujawsko-Pomorskie voivodship*

| Data source | Number of facilities (data for 2020) |
|---|---|
| Central Statistical Office | 374 |
| Register kept by the Marshal of Kujawsko-Pomorskie Voivodeship | 161 |
| Municipal registers | 643 |
| Booking.com | 752 |

Source: own elaboration based on Kujawsko-Pomorskie Biuro Planowania Przestrzennego i Regionalnego we Włocławku, Oddział w Bydgoszczy. (2022). *Baza noclegowa województwa kujawsko-pomorskiego [Accommodation base of the kujawsko-pomorskie voivodship]*. Opracowanie zespołu w składzie: M. Antkowiak-Dubińska, P. Biczkowski, K. Kaszuba. Kujawsko-Pomorskie Biuro Planowania Przestrzennego i Regionalnego.

In turn, according to the Central List of Hotel Facilities (CWOH) available on the official website of the Ministry of Sport and Tourism (data as of December 2024), 155 facilities have been defined in Kujawsko-Pomorskie Voivodeship.

Table 43 and Figure 38 show a summary of the number of tourist facilities registered in the Central Directory of Hotel Facilities (CWOH), their presence on Booking.com and a comparison with data from the Central Statistical Office (GUS). The data was broken down by category of facilities and their standard (number of stars). Hotel facilities constitute the largest group among the analysed categories. Most are 3-star hotels, of which 74.64% are available on Booking.com. In contrast, the percentage of occurrence of 1-4 star hotels on Booking.com ranges from 69.44% to 80% and one 5-star hotel is present in both the OWH data and the booking platform. Of the 155 hotels registered with CWOH, 112 (72.26%) are available on Booking.com. This indicates a significant visibility of venues on this platform, although there is still room to increase this number. The CSO data indicates the existence of additional facilities that are not included in the CWOH or on Booking.com, e.g. holiday homes, holiday resorts, colony centres, spa facilities or agritourism accommodation.

*Table 43 . Number of sites appearing on the CWOH and Booking portal*

| Type of facility | Category of facility | Number of facilities in the CWOH[5] | Appearance on Booking.com[6] | | CSO data[7] |
|---|---|---|---|---|---|
| Hotels | 1 star (*) | 5 | 4 | 80% | 138 |
| | 2-star (**) | 36 | 25 | 69,44% | |
| | 3-star (***) | 71 | 53 | 74,64% | |
| | 4-star (****) | 26 | 20 | 76,92% | |
| | 5-star (*****) | 1 | 1 | 100% | |
| Campsites | | 4 | 1 | 25% | 4 |
| Motels | | 4 | 2 | 50% | 5 |
| Pensions | | 8 | 6 | 75% | 7 |
| Other | | Nd. | Nd. | Nd. | 247 |
| Total | | 155 | 112 | Nd. | 401 |

Source: own compilation based on data from CSO (July 2024), CWOH (December 2024) and Booking.com (December 2024)

Table 44 shows the changes in the number of different types of accommodation facilities registered in the CSO databases between 2020 and 2024. Detailed data for 15 categories of facilities are shown, which allows an analysis of trends in the accommodation base in Poland. The number of accommodation facilities increased from 374 in 2020 to 401 in 2024, an increase of approximately 7.2%. The largest increase was recorded in 2024, suggesting a recovery of the tourism sector after years of stagnation due to the pandemic. Among accommodation establishments, hotels remain the dominant category, increasing from 128 in 2020 to 138 in 2024. This growth was fairly constant, except for a slight decrease in 2021. The most dynamic growth has been in less formal segments such as guest rooms, agritourism and other hotel facilities.

[5] Status as of December 2024.
[6] Status as of December 2024.
[7] Status as at 31 July 2024.

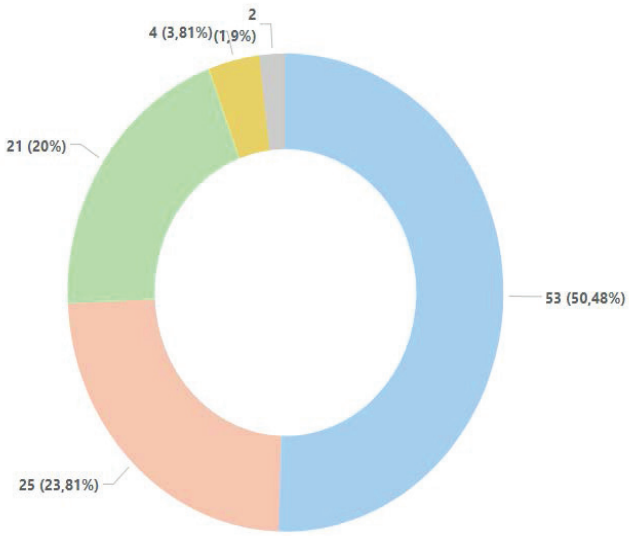
*Table 44 . Overview of kujawsko-pomorskie voivodeship accommodation facilities appearing in CSO registers*

| Type of facility | 2020 | 2021 | 2022 | 2023 | 2024 |
|---|---|---|---|---|---|
| hotels | 128 | 126 | 130 | 138 | 138 |
| motels | 7 | 6 | 6 | 6 | 5 |
| pensions | 7 | 8 | 8 | 8 | 7 |
| other hotel facilities | 48 | 44 | 49 | 47 | 54 |
| tour homes | 1 | 1 | 1 | 1 | 1 |
| shelters | 2 | 2 | 2 | 2 | 1 |
| school youth hostels | 4 | 5 | 5 | 5 | 5 |
| holiday centres | 35 | 34 | 33 | 30 | 30 |
| colony centres | 3 | 2 | 2 | 1 | 1 |
| training and recreation centres | 9 | 10 | 11 | 12 | 13 |
| tourist cottage complexes | 11 | 10 | 10 | 10 | 12 |
| campsites | 3 | 3 | 3 | 4 | 4 |
| camping grounds | 10 | 10 | 8 | 7 | 8 |
| hostels | 7 | 7 | 7 | 6 | 8 |
| spa facilities | 29 | 28 | 29 | 29 | 30 |
| guest rooms/private accommodation | 34 | 31 | 35 | 40 | 49 |
| agritourism accommodation | 20 | 18 | 18 | 18 | 21 |
| other tourist facilities | 16 | 16 | 15 | 16 | 14 |
| Total | 374 | 361 | 372 | 380 | 401 |

Source: own compilation based on CSO records (as at 31 July 2024).

Analysis of the data on hotel standard shows that the CWOH list sample is dominated by 3-star establishments, which account for more than half of all analysed hotels (50.48%). Such a predominance of this category suggests that the tourist offer is focused on the medium price segment, adapted to the needs of a wide group of tourists. The second most numerous group are 2-star hotels, which account for almost 24% of the establishments, indicating the considerable popularity of budget accommodation options. 4-star hotels, although less numerous, account for 20% of the sample, reflecting an important, albeit more exclusive, part of the tourism market. The results suggest that the hotel market in the Kujawsko-Pomorskie region is mainly focused on the medium standard of service, with limited offerings in the most luxurious segments.

*Figure 38 . Structure of WKP hotel facilities by standard in the CWOH*




1-star hotel (*): 3.81% (4 responses)
2-star hotel (**): 23.81% (25 responses)
3-star hotel (***): 50.48% (53 responses)
4-star hotel (****): 20% (21 responses)
5-star hotel (*****): 1.9% (2 responses)
Source: own study.

The analysis of the average ratings of tourist facilities on the Booking.com platform, presented in Table 45, makes it possible to indicate the best rated counties in three rating ranges: 9.00-10.00, 8.00-8.99 and 7.00-7.99. In the range of the highest ratings (9.00-10.00) were districts such as Radziejowski, Sępoleński, Tucholski, Golubsko-Dobrzyński, Chełmiński and Aleksandrowski. This data can serve as a guideline for local governments and entrepreneurs to analyse the quality of tourism services and identify areas for improvement. High ratings indicate the attractiveness of tourism and the high standard of facilities in the regions concerned, which can influence the attraction of more tourists.

*Table 45 . Counties best ranked in raiting*

| Average rating of facilities in the district | District |
|---|---|
| 9,00-10,00 | radziejowski, sępoleński, tucholski, golubsko-dobrzyński, chełmiński, aleksandrowski |
| 8,00-8,99 | brodnicki, wąbrzeski, bydgoski, świecki, Toruń, mogileński, Grudziądz, Bydgoszcz, rypiński, żniński, inowrocławski, lipnowski, toruński, nakielski, Włocławek |
| 7,00-7,99 | Grudziądz |

Source: own elaboration based on data from Booking.com

Fig. no. 39 shows the total number (7373) of beds available in all establishments in the Kujawsko-Pomorskie voivodship as at 4-5 December 2024 as at 13 November 2024. However, verification as at 4 December showed 5619 beds. Accordingly, 24% of the places were booked during the analysed period.

*Figure 39 . Occupancy rate of accommodation in the Kujawsko-Pomorskie voivodship*




Source: own elaboration

*Table 46 . Average number of available beds by county (as of 13.11.2024)*

| District | 13-14.11.24 | 20-21.11.24 | 27-28.11.24 | 4-5.12.24 | Average per county |
|---|---|---|---|---|---|
| Toruń | 2477 | 922 | 2467 | 2549 | 2103,75 |
| Bydgoszcz | 1087 | 1454 | 1567 | 1541 | 1412,25 |
| aleksandrowski | 413 | 553 | 572 | 530 | 517 |
| Włocławek | 252 | 264 | 370 | 373 | 314,75 |
| brodnicki | 336 | 265 | 310 | 306 | 304,25 |
| inowrocławski | 227 | 283 | 259 | 304 | 268,25 |
| toruński | 232 | 144 | 235 | 215 | 206,5 |
| bydgoski | 163 | 227 | 229 | 199 | 204,5 |
| świecki | 187 | 212 | 229 | 181 | 202,25 |
| Grudziądz | 174 | 159 | 239 | 225 | 199,25 |
| żniński | 189 | 46 | 175 | 301 | 177,75 |
| golubsko-dobrzyński | 89 | 174 | 122 | 118 | 125,75 |
| włocławski | 84 | 98 | 97 | 98 | 94,25 |
| tucholski | 71 | 104 | 98 | 72 | 86,25 |
| lipnowski | 71 | 75 | 75 | 71 | 73 |
| nakielski | 72 | 68 | 70 | 55 | 66,25 |
| chełmiński | 42 | 52 | 53 | 61 | 52 |
| radziejowski | 43 | 51 | 52 | 31 | 44,25 |
| mogileński | 35 | 36 | 39 | 37 | 36,75 |
| sępoleński | 40 | 21 | 47 | 39 | 36,75 |
| grudziądzki | 9 | 15 | 15 | 15 | 13,5 |
| rypiński | 11 | 13 | 15 | 0 | 9,75 |
| wąbrzeski | 7 | 8 | 8 | 8 | 7,75 |

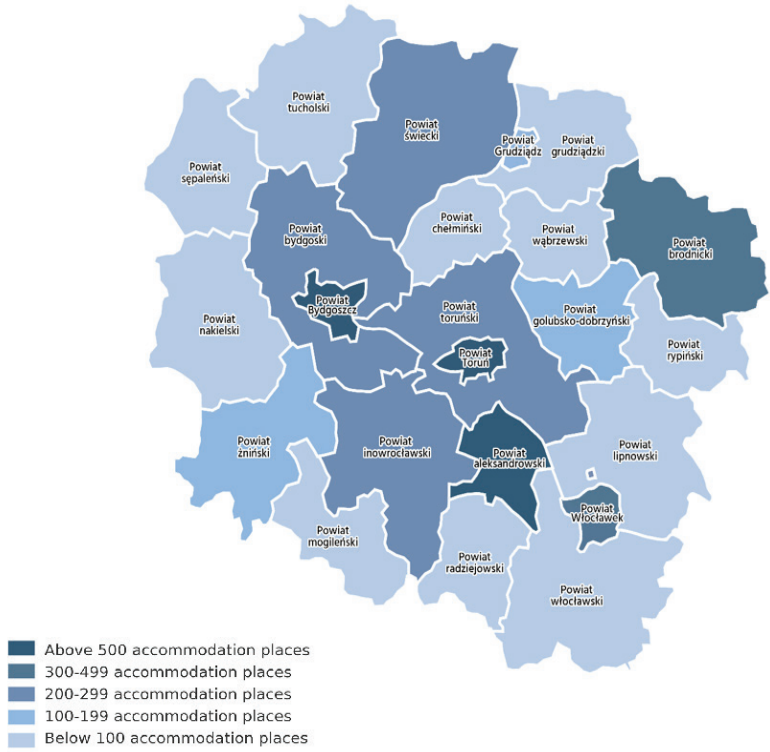
Source: own study.

Table 46 and Figures 40-44 show the average number of available beds in the counties of Kujawsko-Pomorskie Voivodeship for specific weeks from November to December 2024. The results show significant variation in the availability of accommodation between counties.

The highest average number of beds was recorded in cities with poviat rights, such as Toruń (2103.75) and Bydgoszcz (1412.25), reflecting their importance as major tourist and economic centres in the region. Among the remaining poviats, the highest values were achieved by Aleksandrowski county (517), while the lowest average was recorded in Wąbrzeski county (7.75) and Rypinski county (9.75). In the weeks analysed, fluctuations in the number of beds are visible, especially in the larger centres, which may indicate fluctuating seasonal demand or specific events in the region such as the 32nd International Film Festival of the Art of Cinematography Camerimage taking place from 16-23 November 2024 in Toruń. In this case, the increased demand for accommodation services available on the booking portal is particularly evident. Rural counties and those with less tourism potential, such as

Mogileński (36.75) and Sępoleński (36.75), are characterised by low availability of accommodation. This indicates a limited tourist offer or lack of infrastructure in these areas.

*Figure 40 . Average number of beds in the districts during the period under review*




Source: own study.

*Figure 41 . Number of beds on Booking.com in the counties as of 4-5.12.2024 as of 13.11.2024*

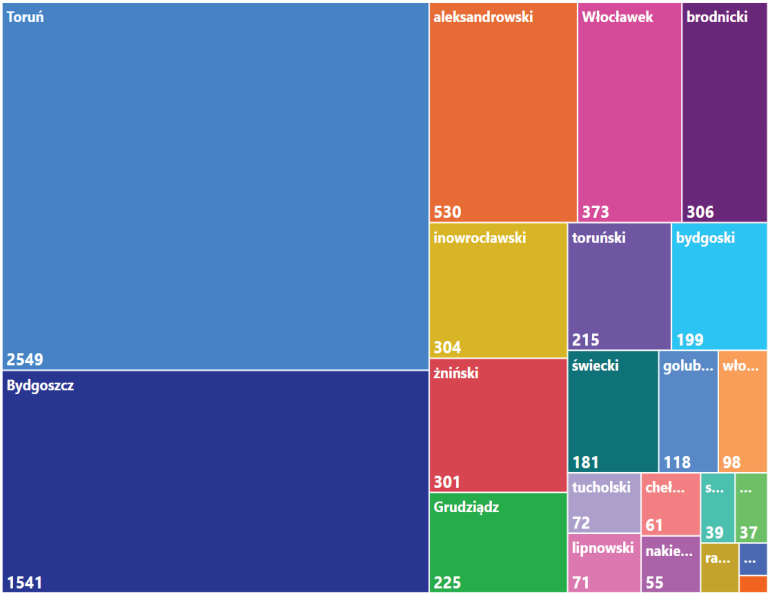
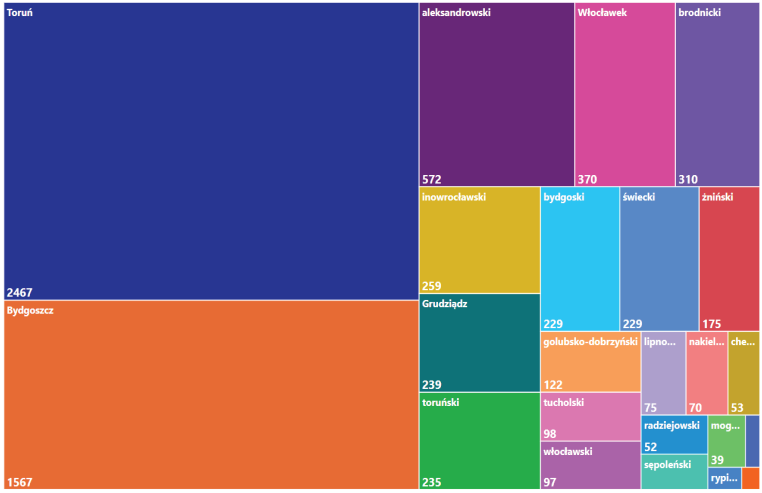
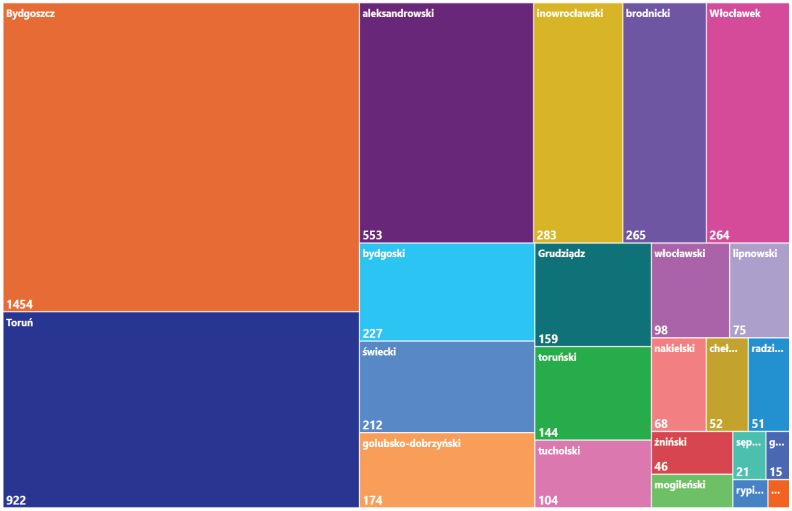



*Figure 42 . Number of Booking.com beds in the counties as of 27-28.11.2024 as of 13.11.2024*



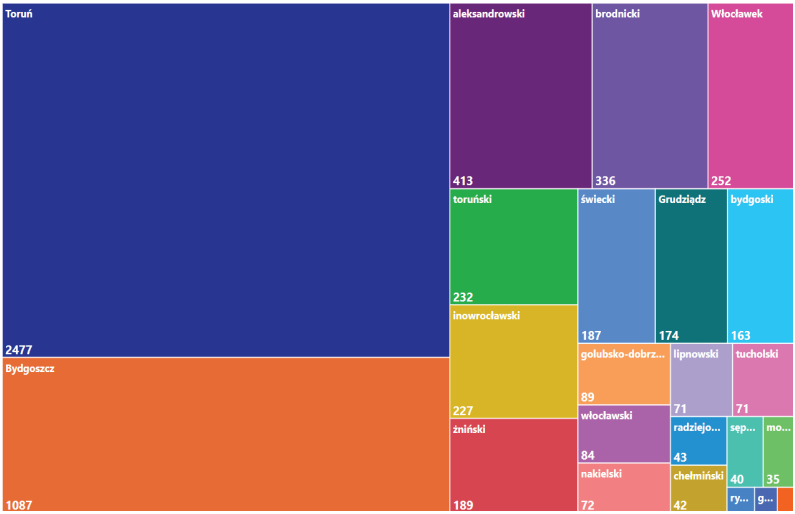

Source: own study.

*Figure 43 . Number of beds on Booking.com in the counties as of 20-21.11.2024 as of 13.11.2024*




Source: own study.

*Figure 44 . Number of beds on Booking.com in the counties as of 13-14.11.2024 as of 13.11.2024*




Source: own study.

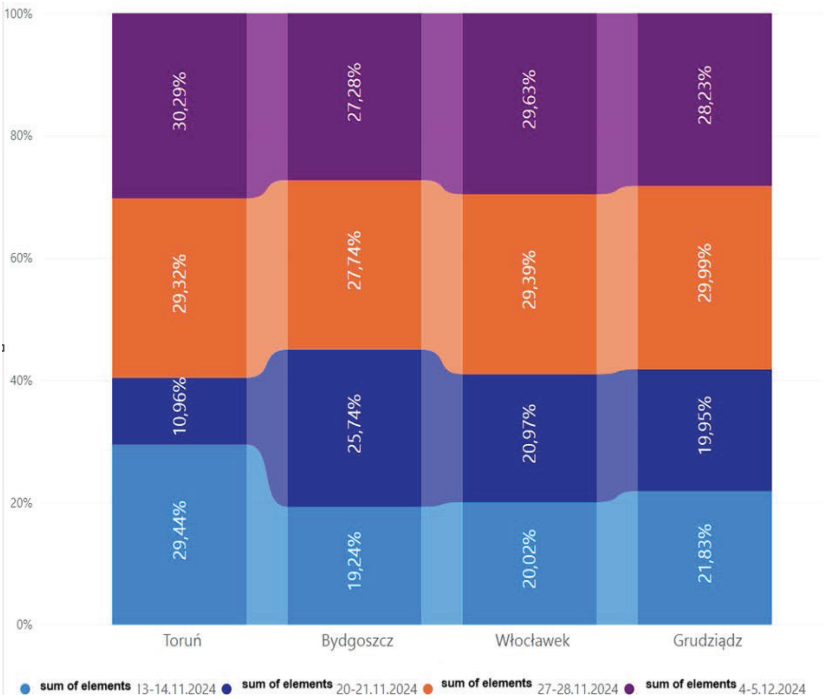
*Table47 . Difference in availability of accommodation on 4-5.12.2024 in facilities as at 13.11.2024 and 4.12.2024.*

| District | Total 4-5/12 (data from 13/11) | Total 4-5/12 (data from 4/12) | Difference (N) | Difference (%) |
|---|---|---|---|---|
| grudziądzki | 15 | 0 | 15 | -100,00% |
| mogileński | 37 | 8 | 29 | -78,38% |
| chełmiński | 61 | 32 | 29 | -47,54% |
| Grudziądz | 225 | 128 | 97 | -43,11% |
| inowrocławski | 304 | 175 | 129 | -42,43% |
| Bydgoszcz | 1541 | 924 | 617 | -40,04% |
| Włocławek | 394 | 272 | 122 | -30,96% |
| golubsko-dobrzyński | 118 | 82 | 36 | -30,51% |
| włocławski | 77 | 58 | 19 | -24,68% |
| sępoleński | 39 | 30 | 9 | -23,08% |
| Toruń | 2549 | 1986 | 563 | -22,09% |
| świecki | 181 | 148 | 33 | -18,23% |
| żniński | 305 | 255 | 50 | -16,39% |
| wąbrzeski | 8 | 7 | 1 | -12,50% |
| toruński | 215 | 192 | 23 | -10,70% |
| brodnicki | 306 | 284 | 22 | -7,19% |
| bydgoski | 199 | 188 | 11 | -5,53% |
| nakielski | 95 | 90 | 5 | -5,26% |
| aleksandrowski | 530 | 523 | 7 | -1,32% |
| tucholski | 72 | 83 | -11 | 15,28% |
| lipnowski | 71 | 96 | -25 | 35,21% |
| radziejowski | 31 | 52 | -21 | 67,74% |
| rypiński | 0 | 6 | -6 | - |

Source: own study.

Table 47 shows the differences in the availability of accommodation in the counties of the Kuyavian-Pomeranian region as at 4-5 December 2024 compared to 13 November 2024. The analysis includes both numerical (N difference) and percentage (% difference) changes. Grudziadz saw a decrease of -100%, meaning that all available accommodation was booked. Significant decreases were also recorded in the districts of Mogileń (-78.38%), Chełmno (-47.54%), Inowrocław (-42.43%) and Bydgoszcz (-40.04%). In these cases, the decreases may indicate a high demand for accommodation in the analysed period, but the values also indicate a low supply. Such a result also indicates the prospects for new offers in these districts. On the other hand, the indicated values are not very high. In Tuchola (+15.28%) and Radziejów (+67.74%) poviats, an increase in the number of available beds was noted. This may be the result of an increase in supply, e.g. the opening of new facilities, seasonal investments or adaptation of the offer to higher demand.

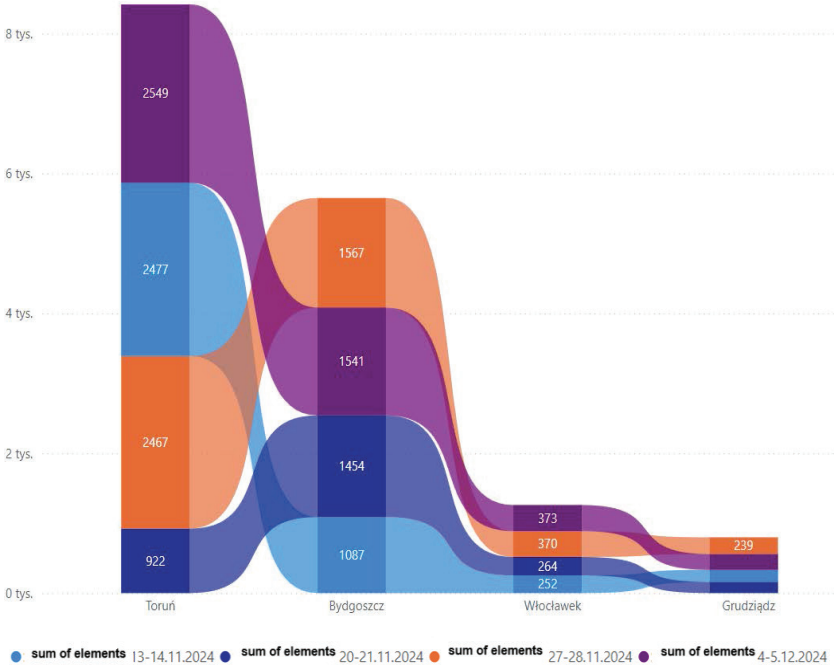
*Figure 45 . Number of beds on Booking.com in cities in the county as of 13-14.11.2024, 20-21.11.2024, 27-28.11.2024 and 4-5.12.2024 as of 13.11.2024*




Source: own study.

Figure 45 and 46 show the changes in the number of beds available on Booking.com in the counties: Toruń, Bydgoszcz, Włocławek and Grudziądz in the four periods analysed: 13-14 November, 20-21 November, 27-28 November and 4-5 December 2024, as of 13 November 2024. Each period is marked with a different colour.

*Figure 46 . Number of beds on Booking.com in cities in the county as of 13-14.11.2024, 20-21.11.2024, 27-28.11.2024 and 4-5.12.2024 as of 13.11.2024*




Source: own study.

*Table48 . Difference (N and %) in the number of reserved places between 13/11 and 4/12 in the district cities*

| District | Total 4/12 (figures from 13/11) | Total 4/12 (data from 4/12) | Difference (N) | Difference (%) |
|---|---|---|---|---|
| Bydgoszcz | 1541 | 924 | 617 | -40,04% |
| Grudziądz | 225 | 128 | 97 | -43,11% |
| Toruń | 2549 | 1986 | 563 | -22,09% |
| Włocławek | 394 | 272 | 122 | -30,96% |

Source: own study.

Table 48 shows the difference in the number of booked beds in the cities with county rights (Bydgoszcz, Grudziądz, Toruń, Włocławek) between data from 13 November and 4 December 2024. The results show changes in number (difference N) and percentage (difference %), indicating the intensity of bookings during the analysed period. The difference represents the number of tourists who visited the city on a specific date. In this way, real-time data can be obtained, as well as in advance. Bydgoszcz recorded the highest number of booked accommodation with 617 places, representing a decrease in availability of -40.04%. This is a positive development, indicating strong interest in the accommodation on offer in this city. Grudziądz showed the largest percentage decrease in availability (-43.11%, 97 places), indicating high tourist interest during this period. Toruń also had a good result, with a decrease of 563 places (-22.09%), indicating high demand for accommodation. Włocławek recorded a decrease of 122 places (-30.96%), which also confirms the high number of bookings, even though the number of available places was initially lower than in Bydgoszcz or Toruń.

*Table49 . Changes in the number of available beds during the period under review*

| District | Total 4/12 (figures from 13/11) | Total 4/12 (figures from 20/11) | Difference (N) | Difference (%) | Total 4/12 (figures from 27/11) | Difference (N) | Difference (%) | Total 4/12 (data from 4/12) | Difference (N) | Difference (%) | Increase in bookings in the last week before the stay date | Average increase in the number of bookings in the last week before the date of stay |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bydgoszcz | 1541 | 1477 | 64 | -4,33% | 1428 | 113 | -7,91% | 924 | 617 | -40,04% | 32,13% | 23,16% |
| Grudziądz | 225 | 223 | 2 | -0,90% | 198 | 27 | -13,64% | 128 | 97 | -43,11% | 29,47% | |
| Toruń | 2549 | 2474 | 75 | -3,03% | 2296 | 253 | -11,02% | 1986 | 563 | -22,09% | 11,07% | |
| Włocławek | 394 | 372 | 22 | -5,91% | 355 | 39 | -10,99% | 272 | 122 | -30,96% | 19,97% | |

Source: own study.

Table 49 shows the changes in the number of available beds in the cities with county rights (Bydgoszcz, Grudziądz, Toruń, Włocławek) during the study period, indicating a gradual increase in the number of bookings as the date of stay approaches. The analysis takes into account the numerical (N) and percentage (%) differences between dates and the increase in bookings in the last week before the stay date. The largest increase in bookings in the last week before the stay was in Bydgoszcz (32.13%), where the number of available places decreased by as many as 504 places. This is the most significant decrease in the cities analysed, highlighting Bydgoszcz's attractiveness as a destination. Grudziądz showed the largest percentage decrease in available places (-43.11%), indicating high demand in relation to the initial number of places (70 reserved places in the last week). In Toruń, the number of seats decreased by 310 in the last week of the survey, indicating significant tourist interest, although the increase in bookings in the last week was 11.07%, lower than in the other cities. Such a result may indicate that in this case arrivals are planned more in advance. Włocławek showed a moderate increase in bookings in the last week (19.97%), which translated into a decrease in availability of 83 places. Although the number of bookings was lower than in Toruń or Bydgoszcz, the percentage increase in interest is significant. The average increase in the number of bookings in the last week for all locations was 23.16%, confirming that decisions to book accommodation are made close to the date of stay, indicating the dynamic nature of tourism in the region.

## 5.5. Challenges and future directions for the application of big data in the tourism sector

The future of big data analytics in the tourism sector, based on actual user behaviour rather than survey results, opens up entirely new possibilities for forecasting and business decision-making. Data obtained through web scraping - a technology that allows the automatic extraction of information from portals such as Booking.com, TripAdvisor or Airbnb - provides detailed and up-to-date information about actual user activities, such as the number of bookings, property ratings or interest in particular destinations.

Unlike traditional surveys, in which respondents declare their intentions or preferences, big data analytics allows real patterns of behaviour to be captured. For example, data on a sharp drop in the number of available beds on a specific date may indicate a sudden increase in interest in a particular location, which enables rapid adjustment of tourism offers, dynamic pricing or identification of seasonal trends. In addition, the analysis of large data sets allows better prediction of future trends,

such as changing tourist preferences in terms of accommodation types (e.g. the rise in popularity of flats) or new destinations

In practice, the data obtained through web scraping can be analysed in real time, enabling a dynamic response to changing market conditions. In the context of tourism, this means better tailoring of offers to customers' needs, more effective marketing campaigns, and precise management of resources, such as the number of available beds or promotion of less busy locations.

Big data and web scraping in the tourism sector not only allow tourism companies and regions to become more competitive, but also support decision-making based on actual data rather than subjective opinions. This makes the tourism sector more precise, flexible and oriented to the needs of the modern consumer, who expects personalised offers and quickly accessible information.

Author of this publication is aware of the fact that the research was conducted on a set of big data on a micro scale, the so-called small big data (only about 42 thousand data), but certainly this experiment on the example of the Booking.com case study makes it possible to confirm the thesis that web scraping of data from booking portals allows for analysing and forecasting demand in tourist reception areas, learning about the dynamics of change, tracking information on the occupancy of facilities, their evaluation or price changes. Therefore, in Figure 47, the author proposes a comprehensive model to look into the future by acquiring data in the tourism sector, combining de facto data on supply with information on demand in a specific tourist reception area. The model can be implemented globally, as the data from the booking portals used to construct the model are also available globally. Other source data, which serve as a reference or framework for the data obtained from the booking portals, such as the Central Inventory of Hotel Facilities (CWOH) or the Central Statistical Office (GUS) data, are the variable points of the model. Although accommodation registers are kept separately in different countries, the example of the planned changes in the regulation of short-term rentals will make the data in the model an even more accurate source of information on the situation and developments in the tourism sector. There are increasing reports (Business Insider, 2024) of the need to implement a short-term rental housing register. Such a solution, similar to the current Central List of Accommodation Facilities operating in Poland, would make it possible to work with a specific operator for web scraping research across all booking portals. In addition, directives such as DAC7 require annual reporting, which in the case of Booking.com may include revenue received per customer (rental property), number of bookings per quarter per property or number of rental days per quarter per property. The directive was introduced by the European Union and is effective from 1 January 2023. It requires digital platforms (such as Booking.com, Airbnb and others) to collect data on users who use these platforms to offer services and to report annually to the tax authorities.

*Figure 47 . Tourism Demand Horizon+ (theoretical assumptions)*

<table>
<tr><td colspan="2" rowspan="2">CWOH</td><td colspan="6">BOOKING PORTALS</td><td rowspan="3">Data from facilities</td><td rowspan="3">STEP 1.<br><br>Definition of available databases</td><td rowspan="6">Research Knowledge DataBase with AI module</td></tr>
<tr><td>Booking.com</td><td>Triadvisor</td><td>Airbnb</td><td>Hotels.com</td><td>Other</td><td>Facility register / EU data (1xQ) (short-term rental)</td></tr>
<tr><td>API (1xM)</td><td>API (1xM)</td><td colspan="5" rowspan="2">Web scraping<br><br>(generation of requests for the following days, weeks, months)</td></tr>
<tr><td colspan="2">Data of government/authorities</td></tr>
<tr><td colspan="8">DATABASE (with AI module)<br>• learn and average the number of sites in a specific area (city, municipality, district, province, country)<br>• learn how many seats (N, %) have been booked on specific dates<br>• matches facilities described on booking portals to those described in the CWOH by supplementing the data with the type and standard of the facility in accordance with the Act<br>• verifies historical data on the number of nights provided with GUS data (KT1)<br>• matches the address of the object to the database of communes and districts from the GUS data</td><td>STEP 2.<br><br>Creation of a database with the AI module</td></tr>
<tr><td colspan="8">EFFECT</td></tr>
<tr><td colspan="8">• Number of tourists using overnight accommodation (day by day)- barometer<br>• Average number of beds in a defined area - barometer<br>• Number of beds (N, %) used - barometer<br>• Distribution of accommodation in a defined area - barometer<br>• Occupancy rate of accommodation in a defined area - barometer<br>• Dynamics of change in the number of beds in a defined area - barometer/horizon<br>• Forecasts the number of tourists in the future based on the number of available beds - horizon</td><td>STEP 3.<br><br>Dasboard<br>www</td></tr>
</table>

Source: own study.

The model presented in Fig. no. 47 is a proposal for how to process, analyse and visualise tourism data in a comprehensive and automated way. The aim of this approach is to understand and forecast the dynamics of tourism in a specific area, such as a city, municipality, district, province or even the whole country. It is based on the integration of different data sources, the use of artificial intelligence (AI) and analytical tools to create forecasts and indicators to aid decision-making. The Tourism Demand Horizon+ concept involves the use of modern technology and analytical methods to get a more complete picture of the tourism situation in a region. It involves three main steps: defining available databases, creating a database with an AI module and creating a visualisation dashboard. An experiment conducted using Booking.com as an example confirmed the feasibility of developing the model in its full formulation.

The model is oriented towards practical application in tourism planning and management. It will provide information not only on the current situation, but also on trends, dynamics and potential changes in the future. Its unique feature is that it can forecast the number of tourists and available accommodation on the basis of historical data and information collected from a number of different sources, which are partly complementary and partly mutually exclusive.

The diagram indicates a multi-dimensional approach to data collection, which comes from a variety of sources. Key elements here are data extracted from booking portals such as Booking.com, Airbnb, Hotels.com and Tripadvisor. These portals provide information on the number of bookings, prices and user reviews. Web scraping, a technique for automatically collecting data from websites, is used to supplement the database with information from these portals. Queries should be generated on a cyclical basis - for specific days, weeks and months on an ongoing basis, but also for selected dates in the future.

The operator for the web-scraped data should be databases of governmental or self-governmental institutions, e.g. in the case of Poland the Central List of Hotel Facilities (CWOH) maintained by the Ministry of Sport and Tourism (MSiT) or the Central Statistical Office, which obtains information on a monthly basis from establishments through KT-1 forms. CWOH is a database containing information on hotel facilities such as hotels, guest houses, flats and other forms of accommodation that are registered in accordance with the law. Its API provides access to information on the location, capacity and categories of establishments. The GUS API, on the other hand, provides access to statistical data, such as the number of tourists, the number of nights provided or the average occupancy rates of accommodation facilities.

Depending on the possibility of obtaining data in a given city, region or country (tourist reception area), it may be possible to consider obtaining data for the operative from individual

facilities, especially in the case of a smaller number of facilities. The register of short-term rental facilities planned for implementation may also prove to be a valuable source.

Data from the CWOH and GUS APIs in model should be retrieved in an automated manner, ensuring that the data is up-to-date and consistent. Data from booking portals, on the other hand, should be collected via web scraping using advanced algorithms that can process dynamic web content such as booking tables and availability data. This data should then be integrated into a single database, where it is validated and processed. A key element of this process should be the use of an AI module that enables:

- learning the average number of sites in specific regions,
- analysis of the percentage of seats booked in a given period,
- matching data from booking portals with data from other sources, including filling in missing information on the standard of facilities,
- verification of historical data on the number of nights provided on the basis of information from public institutions, e.g. the GUS.

The AI module should also act as a 'link' between the various data sources, ensuring their consistency and completeness. This will provide an accurate representation of the state of the accommodation base in a given region.

The construction of the database should start with the collection and organisation of data from the previously mentioned sources, such as CWOH, GUS, booking portals or government institutions. This data should then be processed, e.g. using machine learning algorithms for integration and verification. This process should involve several key steps:

- aggregation of data by geographical area: it is recommended that the AI module learns the number of available accommodation facilities in a specific region, such as city, municipality, county, province or country. This will enable the creation of an accurate map of accommodation infrastructure that takes into account not only formally registered facilities, but also those available on short-term booking platforms such as Airbnb,
- analysis of bookings and occupancy: the AI module should analyse booking data, such as the number of places booked in a given period and the percentage occupancy of the accommodation. This will allow an accurate determination of the seasonality and degree of popularity of a tourist region,
- matching standards and categorisation of facilities: based on data from all sources, AI algorithms will verify the standard of accommodation facilities and their categorisation

according to current regulations. As a result, the database will take into account both high-standard facilities and less formal accommodation such as private flats,

- historical verification: Using historical data, e.g. in the case of Poland from the Central Statistical Office, the AI module will verify the numbers of nights provided in previous years, allowing for better trend forecasting. Historical data will also be used to identify outliers and detect errors in current data.

In this case, the AI module will act as an 'intelligent tool' that not only collects and verifies data, but also provides advanced statistical analysis and predictions. Its self-learning capability means that over time the database becomes more and more precise and useful, as the algorithms constantly optimise their performance based on new data.

The data should be presented in the form of so-called open dashboards that will provide practical information for decision-makers and stakeholders. The Tourism Demand Horizon+ model can be crucial for tourism management as it will provide comprehensive data and forecasts that can be used at different decision-making levels. For tourism reception areas, this means being able to:

- identification of key trends in tourism,
- effective management of accommodation facilities,
- development of promotional strategies based on real data,
- increase competitiveness in the tourism market.

By integrating advanced technologies, the model will enable better decision-making, which will translate into sustainable tourism development and economic benefits for local communities.

The model may also be complemented by the Research Knowledge DataBase, which will be an integrated collection of data and knowledge derived from scientific research, reports, publications, experiments or analyses on tourism research in tourism reception areas, which could be used by researchers, research institutions to create Tourism Demand Horizon+ micro-models. The author of this publication published an experimental model fed with the results of research on tourism in Poland in social media on the assumption that the scientific world is looking for quick access to an overview of research results in the regions.

## Completion

This publication analyses the potential for the use of big data technologies in the context of tourism demand research in tourism reception areas at different scales, city, region or even country. The main findings indicate that the dynamic development of information and communication technologies (ICTs) generates huge amounts of data in the form of so-called big data, which enables the creation of new analytical solutions that are crucial for the tourism sector. Thanks to tools such as web scraping, analysis of accommodation availability or predictive data-based models, regions can more effectively predict tourist behaviour, identify market trends and optimise resource management.

Big data analysis, as shown, allows not only to understand current phenomena, but also to identify potential demand. It is crucial to integrate different data sources, such as booking platforms or social media, to get a more comprehensive picture of tourist behaviour. In this context, the publication proposes a unique 'Tourism Demand Horizon' model, which is assumed to enable more accurate forecasting of market changes.

The Tourism Demand Horizon+ model addresses key questions regarding the application of modern analytical tools in forecasting tourist traffic by integrating data from multiple sources and utilizing artificial intelligence. Its structure is based on three stages: identifying available databases, creating a database with an AI module, and visualizing the results in an interactive panel.

In the context of using Big Data for forecasting tourist traffic in the area of tourist reception, this model enables the collection and analysis of data from various sources, including booking platforms, official databases, and accommodation registers. Through the use of web scraping technology, it is possible to monitor the number of available accommodations and their reservations in real time, allowing for dynamic demand modeling. The transformation of unstructured data into useful economic indicators opens new opportunities for strategic planning in the tourism sector.

The analysis of accommodation availability plays a crucial role in predicting the future volume of tourist traffic in a given region. The model allows for real-time tracking of changes in accommodation availability, enabling the assessment of demand for tourism services. Indicators such as the number of reserved places and the average occupancy rate of accommodation facilities in a specific area form the foundation for modeling tourism trends. Thanks to AI algorithms, it will be possible to predict changes in the volume of tourist traffic while taking into account both seasonal and macroeconomic factors.

Regarding forecasts for future hotel occupancy, the model will provide indicators on the degree of utilization of accommodation infrastructure. The algorithms will allow for the analysis of historical

booking patterns and their projection into the future, enabling the precise determination of the level of infrastructure usage across different periods. This approach is crucial for managing tourism resources, optimizing pricing strategies, and shaping marketing strategies for industry stakeholders.

Accommodation availability forecasts can also indicate changes in tourists' preferences regarding location and type of facilities. The integration of data from multiple booking platforms will make it possible to identify trends in the popularity of specific destinations and types of accommodation. The model analyzes shifts in the distribution of accommodation facilities and their attractiveness among different groups of tourists, allowing for continuous monitoring of consumer preferences and their evolution over time. This information can be used to tailor tourism offerings and guide investment decisions in infrastructure aligned with the changing needs of the market.

The application of the Tourism Demand Horizon+ model in tourism economics makes a significant contribution to the development of modern methods for analyzing tourism demand. The combination of Big Data, artificial intelligence, and predictive analysis will allow for more accurate forecasting of tourist traffic than traditional econometric models. This will lead to increased efficiency in tourism management, better alignment of marketing and investment strategies with real market needs, and the optimization of public policy. The model fills a critical gap between administrative data and dynamic information from booking platforms, providing comprehensive knowledge about the tourism market. In doing so, it lays the foundation for more precise and dynamic strategies for managing tourist flows at both regional and national levels.

The author notes important implications of this for both practice and theory, and also notes opportunities for further research. Among the main implications for practice, he notes a better understanding of tourists' needs, an increase in the competitiveness of regions or the optimisation of business processes. An important aspect of this is the conscious channeling of tourist traffic to regions hitherto less penetrated by tourism, which supports the idea of sustainability. Of course, it will only be possible to support business processes by opening up databases, or rather analyses, to all stakeholders in the sector, because, as already mentioned, obtaining data on one's own is too difficult and costly from the perspective of a single entity. For this reason, it is also important to promote the idea of coopetition in tourism, which creates the win-win principle and encourages cooperation even with competitors.

Among the important theoretical implications, the publication shows that tourism demand can be analysed not only in traditional terms, but also using new research methods such as web scraping. The author is convinced that an approach combining big data, artificial intelligence and traditional research methods can become the standard in tourism analysis, supporting the development of further

models and theories in this area. The model proposed in the publication can be used not only at the micro level but also at the macro level, providing an impetus for the development of demand theory from different angles and perspectives, contributing to new knowledge on, for example, tourism demand segmentation. The model also links the issues of demand and supply, as the information obtained on tourism supply de facto helps to forecast and analyse the demand market situation.

Using the Booking.com case study as an example, the experiment demonstrates how web scraping can be used to forecast tourist traffic in a tourist reception area. The analysis of accommodation availability, on the other hand, provides information on the current status of bookings and trends related to demand for accommodation. If the number of available beds decreases rapidly over a certain period of time, an increase in tourist traffic can be predicted. On the other hand, high availability of places may indicate less interest in a particular location. The data also allows an assessment of the impact of special events, promotions or marketing activities on tourism demand. Based on availability analysis, future hotel occupancy rates can be forecast, taking into account seasonal trends, historical booking data and current changes in tourist behaviour - but this requires data to be collected more frequently and over a longer period of time. Availability forecasts can reveal changing tourist preferences, such as growing interest in luxury hotels, holiday flats or rental homes. Such a holistic approach will make it possible to accurately analyse tourist behaviour and indicate how tourists react to changes in, for example, prices and which dates or locations are most popular. Undoubtedly, such information will allow service providers to better adapt their offerings to changing customer expectations.

Certainly, further research into the use of the model and the collection of data over a longer time horizon are necessary, which will significantly improve the quality of the results obtained. At the same time, the implementation of the model itself in practice (in its full formulation) will require knowledge not only of the tourism economy, but also of technological solutions. Implementation of the solution will allow global comparability of results at regional level, which can support more coordinated tourism strategies. Further experimental work carried out should allow the model to be tested in different regional contexts to confirm its universality and effectiveness.

The author hopes that the publication provides valuable conclusions for tourism practice and theory in economic terms, which can become the basis for more effective tourism management and increasing the competitiveness of tourist reception areas, thanks to the knowledge gained from these analyses.

**Bibliography**


1. Abdirazakov, N. M., Kushebayeva, Z. T., Numanova, F. A., Orynbet, P. Z., & Myltkybayeva, G. E. (2023). Understanding the relationships between hotel variables in Almaty, Kazakhstan: An investigation using booking.com data. *International University of Tourism and Hospitality, University of International Business, National Center of Science and Technology Evaluation, Turan University.* https://doi.org/10.31489/2023Ec4/7-18
2. Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management, 28*(3), 898–910. Retrieved August 30, 2023, from https://www.sciencedirect.com/science/article/abs/pii/S026151770600094X?via%3Dihub
3. Amadeus. (2018). Predictive analytics in the travel industry.
4. American Express Travel. (2022). *Global Travel Trends Report.* Retrieved from https://www.americanexpress.com/en-us/travel/discover/get-inspired/global-travel-trends-en-gb-2022
5. Andereck, K. L., & Caldwell, L. L. (1994). Variable selection in tourism market segmentation models. *Journal of Travel Research, 33*(2), 40–46.
6. Anszperger, A., & Radkiewicz, A. (2009). Monitoring ruchu turystycznego w województwie kujawsko-pomorskim. *Raport z badań.* Kujawsko-Pomorska Organizacja Turystyczna. Retrieved from http://k-pot.pl/wp-content/uploads/2016/10/Raport_2009.pdf
7. Apache Hadoop. (2015, February 2). Retrieved from http://hadoop.apache.org
8. Apache Mahout. (2015, February 2). Retrieved from http://mahout.apache.org
9. Asma. (2021, September 29). The future of web scraping services. Retrieved from Information Transformation Services: https://it-s.com/the-future-of-web-scraping-services
10. Assaf, A. G., & Josiassen, A. (2012). Identifying and ranking the determinants of tourism performance: A global investigation. *Journal of Travel Research, 51*(4), 388–399
11. Baggio, R. (2019). Measuring tourism: Methods, indicators, and needs. In E. Fayos-Solà & C. Cooper (Eds.), *The future of tourism: Innovation and sustainability* (pp. 255–269). Springer.
12. Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46,* 454–464.
13. Barczak, M. (2009). Problematyka wzrostu i rozwoju gospodarczego w gospodarce turystycznej z uwzględnieniem teorii efektów mnożnikowych. W R. Muszkieta i in. (Red.), *Społeczne i ekonomiczne aspekty turystyki i rekreacji*. Poznań-Bydgoszcz: Wydawnictwo Wielkopolskiej Szkoły Turystyki i Zarządzania w Poznaniu, Wyższa Szkoła Gospodarki w Bydgoszczy.
14. Barczak, M. (2014). Koopetycja jako źródło przewagi konkurencyjnej obszaru recepcji turystycznej na przykładzie Bydgoskiego Węzła Wodnego (Doctoral dissertation).
15. Barczak, M. (2023a). Diagnoza – załącznik do Strategii budowy i promocji produktów turystycznych województwa kujawsko-pomorskiego. Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.
16. Barczak, M. (2023b). Ocena narzędzi i oczekiwań pracodawców i pracowników sektora turystyki celem stymulacji na rynku pracy. Instytut Turystyki.
17. Barczak, M., & Brudnicki, R. (2012). Badania ruchu turystycznego w Bydgoszczy. *Raport końcowy.* Wyższa Szkoła Gospodarki w Bydgoszczy. Retrieved from https://visitbydgoszcz.pl/images/stories/pdfy/raport%20z%20brt2012%20press.pdf
18. Barczak, M., & Brudnicki, R. (2015). Monitoring ruchu turystycznego w województwie kujawsko-pomorskim. *Raport z badań dla Kujawsko-Pomorskiej Organizacji Turystycznej.* Retrieved from http://k-pot.pl/wp-content/uploads/2016/10/RAPORT-MONITORING-RT-KPOT-2015.pdf

19. Barczak, M., Brudnicki, R., & Wierzbiańska, S. (2023). Koncepcja metodologii badań ruchu turystycznego dla województwa kujawsko-pomorskiego uwzględniająca nowoczesne technologie pozyskiwania danych źródłowych. Urząd Marszałkowski Województwa Kujawsko-Pomorskiego.
20. Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.
21. Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2019). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*. https://doi.org/10.1080/13683500.2019.1619674
22. Bayram, G. E., Abbasi, A. R., Raza, A., Shaikh, A. H., & Soomro, S. A. (2024). The U-shaped relationship between corruption and international tourism demand: A gravity model approach. *Turyzm/Tourism, 34*(1), 97–108. https://doi.org/10.18778/0867-5856.34.1.09
23. Bean, R. (2017). How companies say they're using big data. *Harvard Business Review.*
24. Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management.* Hoboken, New Jersey: John Wiley & Sons.
25. Bieńkowski, W. (2008). Czynniki i miary międzynarodowej konkurencyjności gospodarek w kontekście globalizacji – wstępne wyniki badań. *Prace i materiały*, (284). Instytut Gospodarki Światowej SGH.
26. Biryukov, V. A., Livson, M. V., Frolova, V. B., Skachko, G. A., & Dukhovnaya, L. L. (2019). Elasticity of tourism demand in the modern world. *Journal of Environmental Management and Tourism, 10*(6(38)), 1317–1321. https://doi.org/10.14505/jemt.v10.6(38)
27. Blum, A., Hopcroft, J., & Kannan, R. (2020). *Foundations of data science.* Cambridge University Press.
28. Bollier, D., & Firestone, C. (2010). *The promise and peril of big data.* Aspen.
29. Borne, K. (2014). *Top 10 big data challenges: A serious look at 10 big data Vs*. [Online]. Retrieved December 31, 2024, from https://www.mapr.com/blog/top-10-big-data-challenges-look-10-big-data-v
30. Borrego-Domínguez, S., Isla-Castillo, F., & Rodríguez-Fernández, M. (2022). Determinants of tourism demand in Spain: A European perspective from 2000–2020. *Economies, 10*(11), 276. https://doi.org/10.3390/economies10110276
31. Brandenburger, A. M., & Nalebuff, B. J. (1995). The right game: Use game theory to shape strategy. *Harvard Business Review, 73*(4).
32. Brenning, A., & Henn, S. (2023). Web scraping: A promising tool for geographic data acquisition. *arXiv.* https://doi.org/10.48550/arxiv.2305.19893
33. Brudnicki, R., & Barczak, M. (2017). Badania ruchu turystycznego – klucz do sukcesu oferty turystycznej. In B. Sowińska & M. Barczak (Eds.), *Zeszyt Naukowy WSG TIR* (No. 14). WSG.
34. Brudnicki, R., Barczak, M., & Wierzbiańska, S. (2023). Koncepcja metodologii badań ruchu turystycznego w województwie kujawsko-pomorskim (unpublished materials).
35. Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management, 21*(1), 97–116.
36. Buhalis, D., Harwood, T., Bogicevic, V., Viglia, G., Beldona, S., & Hofacker, C. (2019). Technological disruptions in services: Lessons from tourism and hospitality. *Journal of Service Management, 30*(4), 484–506.
37. Bulchand-Gidumal, J. (2020). Impact of artificial intelligence in travel, tourism, and hospitality. In Z. Xiang et al. (Eds.), *Handbook of e-Tourism.* Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-05324-6_110-1
38. Bull, A. (1998). *The economics of travel and tourism* (2nd ed.). Addison Wesley Longman Australia Pty Ltd.
39. Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian.*

40. Cao, Z., Li, G., & Song, H. (2017). Modelling the interdependence of tourism demand: The global vector autoregressive approach. Annals of Tourism Research, 67, 1-13.
41. Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. Journal of Urban Technology, 18(2), 65-82.
42. Cavallo, A., & Rigobon, R. (2016). The Billion Prices Project: Using online prices for measurement and research. *Journal of Economic Perspectives, 30*(2), 151–178.
43. Chang, Y. W., & Liao, M. Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan. *Asia Pacific Journal of Tourism Research, 15*(2), 215–231.
44. Chau, S., & Yan, L. (2021). Destination hospitality indicators. *Journal of Destination Marketing & Management, 19,* Article 100537.
45. Choi, H., & Varian, H. R. (2009). Predicting the present with Google Trends. Google Inc.
46. Chomiak-Orsa I., Mrozek B. (2017). Analiza wielkich zbiorów danych w mediach społecznościowych – perspektywa przedsiębiorcy. "Przegląd Organizacji", Nr 8, s. 48-55
47. Cierpiał-Wolan, M., & Stateva, G. (2023). Ocena metod integracji danych dotyczących turystyki z uwzględnieniem big data. *Wiadomości Statystyczne. The Polish Statistician, 68*(12), 25–48.
48. Claveria, O., Monte, E., & Torra, S. (2015). A new forecasting approach for the hospitality industry. *International Journal of Contemporary Hospitality Management, 27*(7), 1520–1538. https://doi.org/10.1108/IJCHM-06-2014-0286
49. Cohen, E. (1972). Towards a sociology of international tourism. *Social Research, 39*, 164–182.
50. Cohen E (2004) Contemporary tourism: diversity and change. Elsevier, Boston
51. Content Marketing Institute (CMI). (n.d.). What is content marketing? Retrieved from https://contentmarketinginstitute.com/what-is-content-marketing
52. Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: Lessons from the ChoicePoint and TJX data breaches. *MIS Quarterly*, 33(4), 673–687
53. Crouch, G. I. (1992). Effect of income and price on international tourism. *Annals of Tourism Research, 19*(4), 643–664.
54. Czakon, W. (2009). Koopetycja – splot tworzenia i zawłaszczania wartości. *Przegląd Organizacji, 12*, 11–14.
55. Dabija, D., Brandusa, B., & Tipi, N. (2018). Generation X versus Millennials communication behaviour on social media when purchasing food versus tourist services. *E&M Economics and Management*, *21*(1), 201.
56. Dagnino, G. B., & Rocco, E. (Eds.). (2009). *Coopetition strategy: Theory, experiments, and cases*. Routledge.
57. DataReportal. (2022). Digital 2022 global digital overview. Retrieved from https://datareportal.com/reports/digital-2022-global-overview-report
58. Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review, 54*(1), 43–46.
59. De Cantis, S., Parroco, A. M., Ferrante, M., & Vaccina, F. (2015). Unobserved tourism. *Annals of Tourism Research, 50,* 1–18.
60. Department of Industry, Tourism and Investment, Government of the Northwest Territories. (n.d.). *Tourism product development: A guide to creating new tourism experiences.* Retrieved from https://www.iti.gov.nt.ca/sites/iti/files/tourism_product_development.pdf
61. Derek, M. (2008). Walory turystyczne jako czynnik turystycznej konkurencyjności regionów. In G. Gołembski (Ed.), *Turystyka jako czynnik wzrostu konkurencyjności regionów w dobie globalizacji* (pp. 89–101). Wydawnictwo AE w Poznaniu, 68.
62. Dibb, S., & Simkin, L. (2008). *Market segmentation success: Making it happen!* Routledge.
63. Divisekera, S. (2013). Empirical estimation of tourism demand models: A review. *Handbook of tourism economics: Analysis, new applications and case studies*, 67-85.

64. Doan, T. D., Quang, T. D., & Nguyen, H. T. (2023). Exploring guests' satisfaction and dissatisfaction with homestay experiences: Evidence from online reviews. *International Journal of Hospitality & Tourism Administration*. Advance online publication. https://doi.org/10.1080/15256480.2024.2350005
65. Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N. (2021). Artificial intelligence: a systematic review of methods and applications in hospitality and tourism. International Journal of Contemporary Hospitality Management, 1-23. Advance online publication. https://doi.org/10.1108/IJCHM-06-2021-0767
66. Domański, R. (1997). *Gospodarka przestrzenna a turystyka*. PWN.
67. Douglas, L. (2011). 3D data management: Controlling data volume, velocity, and variety. *Gartner.* Retrieved May 17, 2024, from http://www.gartner.com/resId=2057415
68. Dutta, S., & Bhattacharya, S. (2015). Segmentation research in tourism: Brief literature review. *TSM Business Review, 3*(1), 69.
69. Dwityas, N. A., & Briandana, R. (2017). Social media in travel decision-making process. *International Journal of Humanities and Social Science, 7*(7), 193–201.
70. Dwivedi, Y. K., Pandey, N., Currie, W., & Micu, A. (2024). Leveraging ChatGPT and other generative artificial intelligence (AI)-based applications in the hospitality and tourism industry: Practices, challenges, and research agenda. *International Journal of Contemporary Hospitality Management, 36*(1), 1–12. https://doi.org/10.1108/IJCHM-05-2023-0686
71. Dwyer, L., & Forsyth, P. (2006). *International Handbook on the Economics of Tourism.* Edward Elgar Publishing.
72. Dziedzic, E., & Skalska, T. (2012). Ekonomiczne uwarunkowania rozwoju usług turystycznych w Polsce. Stowarzyszenie na Rzecz Badania, Rozwoju i Promocji Turystyki.
73. Dzięcielski, M., Radzimski, A., & Woźniak, M. (2020). Bike-sharing system in Poznań – what will web API data tell us? *Transport Geography Papers of Polish Geographical Society, 23*(3). https://doi.org/10.4467/2543859XPKG.20.018.12786
74. EMC². (2011). *Federation Business Data Lake – Enabling Comprehensive Data Services* (White paper 1). Hopkinton, MA.
75. Eugenio-Martin, J. L. (2004). Modelling determinants of tourism demand as a 5-stage process: A discrete choice methodological approach. *Tourism and Hospitality Research, 4*(4), 341–354.
76. European Commission. (n.d.). *Economic forecast for Poland*. Retrieved December 31, 2024, from https://economy-finance.ec.europa.eu/economic-surveillance-eu-economies/poland/economic-forecast-poland_en
77. European Commission. (2020). *Workpackage J Innovative Tourism Statistics Deliverable J5: Final report containing final results and full description of the methodology used*. Retrieved August 31, 2023, from https://ec.europa.eu/
78. European Union. (2024). *Regulation (EU) 2024/1689*. Retrieved July 12, 2024, from https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:32024R1689
79. Eurostat. (2020). *Deliverable K3: Revised version of the quality guidelines for the acquisition and usage of big data*. ESSnet Big Data II. Retrieved from https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en
80. Erevelles, S., Fukawa, N., & Swayne, L. (2015). Big data consumer analytics and the transformation of marketing. *Journal of Business Research, 69*(2), 897–904.
81. Ezeah, C., Fazakerley, J., & Byrne, T. (2015). Tourism waste management in the European Union: Lessons learned from four popular EU tourist destinations. American Journal of Climate Change, 4(5), 431–445.
82. Fan, W., & Bifet, A. (2013). Mining big data: Current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter, 14*(2), 1-5. https://doi.org/10.1145/2481244.2481246
83. Fasel, D., & Meier, A. (2016). *Big data: Grundlagen, Systeme und Nutzungspotenziale.* Springer Vieweg. https://doi.org/10.1007/978-3-658-11589-0

84. Filieri, R., D'Amico, E., Destefanis, A., Paolucci, E., & Raguseo, E. (2021). Artificial intelligence (AI) for tourism: An European-based study on successful AI tourism start-ups. *Audencia Business School* and *Politecnico di Torino*.
85. Fjeldstadt, O. D., Becerra, M., & Narayanan, S. (2004). Strategic action in network industries: An empirical analysis of the European mobile phone industry. *Scandinavian Journal of Management, 20*(1–2), 173–196.
86. Fleischer, A. (2012). A room with a view—A valuation of the Mediterranean Sea view. *Tourism Management, 33*(3), 598–602.
87. Fletcher, J., Fyall, A., Gilbert, D., & Wanhill, S. (2018). *Tourism: Principles and practice* (6th ed.). Pearson Education Limited.
88. Fletcher, R., & Nielsen, R. K. (2024). What does the public in six countries think of generative AI in news? *Reuters Institute for the Study of Journalism, University of Oxford.* DOI: 10.60625/risj-4zb8-cg87
89. Florez-Lopez, R., & Ramon-Jeronimo, J. M. (2009). Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting. *Social Science Computer Review, 27*(1), 96–117.
90. Frederiksen, L. (2012). Big data. *Public Services Quarterly, 8*(4), 345–349.
91. Gancarczyk, J. (2003). Systemy generowania innowacji a małe i średnie przedsiębiorstwa turystyczne. *International Journal of Contemporary Management, 12*(1).
92. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
93. GfK. (2023). Purchasing Power Europe 2023. Retrieved from https://www.gfk.com/press/Purchasing-power-of-Europeans-increases-to-17688-euros-in-2023
94. Glossary of tourism terms. (n.d.). UNWTO. Retrieved August 1, 2024, from https://www.unwto.org/glossary-of-tourism-terms.
95. Glover, P. (2009). Generation Y's future tourism demand: Some opportunities and challengers. In. P. Benckendorff, G. Moscardo, & D. Pendergast (Eds.), Tourism and generation Y (pp. 155-163). CABI Publishing.
96. Główny Urząd Statystyczny. (2023). *Bezrobotni zarejestrowani i stopa bezrobocia – stan w końcu grudnia 2023 r.* Retrieved December 31, 2024, from https://stat.gov.pl/obszary-tematyczne/rynek-pracy/bezrobocie-rejestrowane/bezrobotni-zarejestrowani-i-stopa-bezrobocia-stan-w-koncu-grudnia-2023-r-%2C2%2C137.html
97. Główny Urząd Statystyczny. (n.d.). *Wskaźniki cen*. Retrieved December 31, 2024, from https://stat.gov.pl/obszary-tematyczne/ceny-handel/wskazniki-cen/
98. Gołembski, G. (1998). *Przedsiębiorstwo turystyczne w gospodarce wolnorynkowej*. Wydawnictwo AE w Poznaniu.
99. Gołembski, G. (2000). Culture and development of new tourism product in Poland. In *Tourism and culture* (pp. 123–137). AIEST.
100. Gołembski, G. (2002). *Kompendium wiedzy o turystyce*. Warszawa: PWN
101. Görmüş, Ş., & Göçer, İ. (2010). The socio-economic determinants of tourism demand in Turkey: A panel data approach. *International Research Journal of Finance and Economics, 55*, 145–153.
102. Gorynia, M., & Łaźniewska, E. (Eds.). (2009). *Kompendium wiedzy o konkurencyjności*. Wydawnictwo Naukowe PWN.
103. Gretzel, U., Yoo, K., & Purifoy, M. (2007). *Online travel review study: Role and impact of online travel reviews*. Laboratory for Intelligent Systems in Tourism. Retrieved December 31, 2024, from https://www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf

104. Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google analytics. *Annals of Tourism Research, 61,* 199–212.
105. Habibi, F. (2016). The determinants of inbound tourism to Malaysia: A panel data analysis. *Current Issues in Tourism.* https://doi.org/10.1080/13683500.2016.1145630
106. Han, S., & Anderson, C. K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. *Cornell Hospitality Quarterly, 62*(1), 89–104. https://doi.org/10.1177/1938965520973587
107. Harford, T. (2014, December). Big data: Are we making a big mistake? *The Financial Times.* Retrieved from https://www.ft.com
108. Havitz, M. E., & Dimanche, F. (1990). Propositions for guiding the empirical testing of the involvement construct in recreational and tourist contexts. *Leisure Sciences, 12*(2), 24–31.
109. Hendrik, H., & Perdana, D. H. F. (2014). Trip guidance: A linked data-based mobile tourist guide. *Advanced Science Letters, 20*(1), 75–79.
110. Herschel, R., & Miori, V. M. (2017). Ethics & big data. *Technology in Society*, *49*, 31-36
111. Hillen, J. (2019, November). Web scraping for food price research. *British Food Journal, 121*, 3350–3361.
112. Hintze, A. (2016). Understanding the four types of AI, from reactive robots to self-aware beings. *The Conversation.* Retrieved from http://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616
113. Holloway, J. C. (2009). *The business of tourism*. Pearson Education.
114. Hołderna-Mielcarek, B. (1998). Regionalizacja produktu turystycznego. *Rynek Turystyczny*, (12
115. Huang, H. C. (2014). A study on artificial intelligence forecasting of resort demand. *Journal of Theoretical and Applied Information Technology, 70*(2), 265–272.
116. Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9
117. Hunziker, W. (1951). *Le tourisme social*. Berne: Imprimerie fédérative. Alliance Internationale de Tourisme.
118. Hunziker, W., & Krapf, K. (1961). *Grundriss der allgemeinen Fremdenverkehrlehre.* Zakład Prawnych i Ekonomicznych Zagadnień Turystyki przy SGPiS.
119. Ibragimov, K., Perles-Ribes, J. F., & Ramon-Rodriguez, A. B. (2017). The economic determinants of tourism in Central Asia: A gravity model applied approach. *Tourism Economics, 23*(6), 1242–1253.
120. Inoua, S. M., & Smith, V. L. (2020). The classical theory of supply and demand. ESI Working Paper 20-11. https://digitalcommons.chapman.edu/esi_working_papers/305/
121. Ivanov, S. and Webster, C. (2019), "Economic fundamentals of the use of robots, artificial intelligence, and service automation in travel, tourism, and hospitality", Robots, Artificial Intelligence and Service Automation in Travel, Tourism and Hospitality, Emerald Publishing, Bingley, pp. 39-55
122. Iwasiński, Ł. (2022). Big data a problem reprezentacji poznawczej. In A. Wawrzynowicz (Ed.), *Człowiek i Społeczeństwo, t. LIII: Paradygmaty metafilozofii* (pp. 241–261). Poznań: Uniwersytet im. Adama Mickiewicza. https://doi.org/10.14746/cis.2022.53.14
123. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415
124. Jungherr,A.,&Jürgens,P. (2013). Forecastingthepulse. How deviations from regular pattern sinonline data can identify offline phenomena. Internet Research, 23(5),589–607.

125. Juszczak, A. (2021). Zastosowanie danych scrapowanych w pomiarze dynamiki cen. *Folia Oeconomica. Acta Universitatis Lodziensis, 1*(352). https://doi.org/10.18778/0208-6018.352.02
126. Kachniewska, M. (2014a). Big data analysis jako źródło przewagi konkurencyjnej przedsiębiorstw i regionów turystycznych. *Folia Turistica, 32,* 35–54.
127. Kachniewska M. (2014b), Wpływ digitalizacji kanałów dystrybucji na strukturę rynku usług pośrednictwa turystycznego*, [w:] „E-mentor" nr 1 (53), 2014, s. 86–91.
128.
129. Kachniewska, M. (2019). Big data analysis as a tool for predictive intelligence and experience personalization in tourism. *Przedsiębiorczość i Zarządzanie, 20*(2), 39–52.
130. Kachniewska, M. (2020b). Boundary conditions for the implementation of smart management systems in tourist destinations. *International Entrepreneurship Review, 6*(4). https://doi.org/10.15678/IER.2020.0604.03
131. Kachniewska, M. (2020a). The use of big data in tourism sales forecasting. International Journal of Contemporary Management, 19(2), 7-35. https://doi.org/10.4467/24498939IJCM.20.004.12669
132. Kaczmarek, J., Stasiak, A., & Włodarczyk, B. (2010). *Produkt turystyczny. Pomysł, organizacja, zarządzanie*. PWE.
133. Kamarazaman, N., Ali, N. M., & Arshad, H. (2024). Leveraging web scraping to gather tourism information data. *Journal of Event, Tourism and Hospitality Studies, 4*(4), 16–29. https://doi.org/10.32890/jeth2024.4.2
134. KANTAR. (2019). *Segmentacja turystów krajowych*. s. 8
135. Karl, M. (2016). Risk and uncertainty in travel decision-making: Tourist and destination perspective. *Journal of Travel Research, 55*(3), 1–18.
136. Kaszuba, K. (2023). Lecznictwo uzdrowiskowe: Ruch uzdrowiskowy na terenie województwa kujawsko-pomorskiego w latach 2015-2022. Kujawsko-Pomorskie Biuro Planowania Przestrzennego i Regionalnego we Włocławku.
137. Katal, A., Wazid, M., & Goudar, R. (2013). Big data: Issues, challenges, tools, and good practices. *Proceedings of the International Conference on Contemporary Computing,* 404–409.
138. Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies.* Cambridge, Massachusetts: MIT Press.
139. Keynes, J. M. (2003). *Ogólna teoria zatrudnienia, procentu i pieniądza* (M. Kalecki, Trans.). PWN.
140. Kisiołek, A. (2018). Analiza wpisów na portalu Twitter z wykorzystaniem narzędzi Big Data zawartych w pakiecie R. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 362,* 43–55.
141. Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, 31(3), 471–481. DOI: 10.3233/SJI-150906.
142. Klub Podróżników Soliści. (2021). *Ogólnopolskie Badanie Podróżników 2021*. Retrieved December 31, 2024, from https://badaniepodroznikow.pl/
143. Kornak, A., & Rapacz, A. (2001). *Zarządzanie turystyką i jej podmiotami w miejscowości i regionie*. Wydawnictwo AE we Wrocławiu, 110-111.
144. Kotler, P. (1967). *Marketing management. Analysis, planning, and control*. Prentice Hall.
145. Kotler, P., Bowen, J., & Makens, J. (2010). *Marketing for hospitality and tourism*. Pearson Prentice Hall.
146. Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2017). *Marketing Management* (3rd ed.). Pearson.
147. Krkonošský národní park. (n.d.). *Oficiální web Krkonošského národního parku*. Retrieved December 1, 2024, from https://www.krnap.cz/

148. Kruczek, Z. (2011). *Kompendium wiedzy o turystyce*. Proksenia.
149. Kujawsko-Pomorskie Biuro Planowania Przestrzennego i Regionalnego we Włocławku, Oddział w Bydgoszczy. (2022). Baza noclegowa województwa kujawsko-pomorskiego. Opracowanie zespołu w składzie: M. Antkowiak-Dubińska, P. Biczkowski, K. Kaszuba. Kujawsko-Pomorskie Biuro Planowania Przestrzennego i Regionalnego.
150. Kurek, W. (Ed.). (2008). *Turystyka*. Warszawa: Wydawnictwo Naukowe PWN.
151. Kureková, L. M., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyze the labor market: A methodological inquiry. *IZA Journal of Labor Economics, 4*, Article 18. https://doi.org/10.1186/s40172-015-0034-4
152. Kuźmar, S. (2019). Big data – istota i możliwości zastosowania na przykładzie prognozowania bezrobocia. *Studia Oeconomica Posnaniensia, 7*(4), 43–56. https://doi.org/10.18559/SOEP.2019.4.3
153. Lado, A. A., Boyd, N. G., & Hanlon, S. C. (1997). Competition, cooperation, and the search for economic rents: A syncretic model. *Academy of Management Review, 22*(1), 110–141.
154. Lee, H., & Sohn, I. (2016). *Fundamentals of big data network analysis for research and industry.* Hoboken, NJ: John Wiley & Sons.
155. Lee, K., James, J. P., Zhang, M., Chester, J. i Williams, J. D. (2016). Big data and policy issues. W: Marketing and public policy conference proceedings (s. 83–87).
156. Lerman, J. (2013). Big data and its exclusions. *Stanford Law Review Online, 66,* 55–63. Retrieved from https://review.law.stanford.edu/wp-content/uploads/sites/3/2013/09/66_Stan_L_Rev_Online_55.pdf
157. Life expectancy and healthy life expectancy. (n.d.). World Health Organization. Retrieved from https://www.who.int
158. Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research,* 24(4), 835–849.
159. Lunnemann P., Wintr L. (2006), Are Internet Prices Sticky?, ECB Working Paper, no. 645
160. Łazarek R. (1999), Ekonomia turystyki, Wyższa Szkoła Ekonomiczna, Warszawa.
161. Ma, H. (1999). Creation and preemption for competitive advantage. *Management Decision*, *37*(3), 259–266.
162. Machado, V., Contreiras, J., & Duarte, A. P. (2019). Local tourist accommodation and institutional strengthening in the interior of the Algarve, Portugal: The role of legislation and technology. *Journal of Place Management and Development, 13*(3), 255–272.
163. Macias, P., & Stelmasiak, D. (2018). Food inflation nowcasting with web scraped data. *NBP Working Paper, 302.*
164. Majewska, J. (2008). Samorząd terytorialny w kształtowaniu funkcji turystycznej gminy (Doctoral dissertation). Uniwersytet Ekonomiczny w Poznaniu, 25
165. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The Next Frontier for Innovation.* McKinsey Global Institute.
166. Markham, A. N., Tiidenberg, K., & Herman, A. (2018). Ethics as methods: Doing ethics in the era of big data research—Introduction. *Social Media+ Society, 4*(3), Article 2056305118784502.
167. Martin, R. L. (2003). *A study of the factors of regional competitiveness: A draft final report for the European Commission, Directorate-General Regional Policy*. University of Cambridge.
168. Martins, L. F., Gan, Y., & Ferreira-Lopes, A. (2017). An empirical analysis of the influence of macroeconomic determinants on world tourism demand. *Tourism Management, 61,* 248–260.
169. Maślankowski, J. (2015). Analiza jakości danych pozyskiwanych ze stron internetowych z wykorzystaniem rozwiązań Big Data. *Roczniki Kolegium Analiz Ekonomicznych SGH, (38),* 167–177.

170. Maślankowski, J. (2019). Pozyskiwanie i analiza danych na temat ofert pracy z wykorzystaniem big data. *Wiadomości Statystyczne. The Polish Statistician, 64*(9), 60–74.
171. Mastercard Economics Institute. (2024). Travel trends 2024: Breaking boundaries.
172. Matczak, A. (1992). Model badania ruchu turystycznego. Studium metodologiczne. *Acta Universitatis Lodziensis*. Wydawnictwo UŁ.
173. Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* New York: Houghton-Mifflin Harcourt.
174. McCabe, S., Li, C. (Spring), & Chen, Z. (2015). Time for a radical reappraisal of tourist decision-making? Toward a new conceptual model. *Journal of Travel Research, 55*(1), 3–13.
175. McGrath, R. G. (2013). *The end of competitive advantage: How to keep your strategy moving as fast as your business*. Harvard Business Review Press.
176. McKinsey & Company. (2024). The way we travel now.
177. McKinsey & Company. (n.d.). *What is artificial general intelligence (AGI)?* Retrieved December 31, 2024, from https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-artificial-general-intelligence-agi
178. Meeker, W. Q., & Hong, Y. (2014). Reliability meets big data: Opportunities and challenges. *Quality Engineering, 26*(1), 102–116.
179. Meyer, B. (2006). Obszarowy produkt turystyczny jako efekt relacji pomiędzy grupami interesu. In S. Wodejko (Ed.), *Gospodarka turystyczna a grupy interesu* (pp. 78–91). Szkoła Główna Handlowa, 342.
180. Meyer, B. (2008). *Ekonomika turystyki*. PWN.
181. Middleton, V. T. C. (1994). *Marketing in travel and tourism*. Butterworth-Heinemann.
182. Middleton, V. T. C. (1996). *Marketing w turystyce*. PART.
183. Monitoring sentiment for domestic and intra-European travel. (2023). European Travel Commission.
184. Moore, K., Smallman, C., Wilson, J., & Simmons, D. (2012). Dynamic in-destination decision-making: An adjustment model. *Tourism Management, 33*(3), 635-645.
185. Morrison, A. M., Hsieh, S., & O'Leary, J. T. (1994). Segmenting the Australian domestic travel market by holiday activity participation. *Journal of Tourism Studies, 5*(1), 39–56.
186. Murphy, J., Hofacker, C., & Gretzel, U. (2017). Dawning of the age of robots in hospitality and tourism: Challenges for teaching and research. *European Journal of Tourism Research, 15,* 104–111.
187. Muryani, M., Permatasari, M. F., & Esquvias, M. A. (2020). Determinants of tourism demand in Indonesia: A panel data analysis. *Tourism Analysis, 25*, 157–174. https://doi.org/10.3727/108354220X15758301241666
188. Nelson, B. K. (1998). Time series analysis using autoregressive integrated moving average (ARIMA) models. *Academic Emergency Medicine, 5*(7), 739–744.
189. Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. Journal of Marketing Research, 56(6), 960–980.
190. Niemczyk, J., & Stańczyk-Hugiet, E. (2014). Cooperative and competitive relationships in the higher education sector in Poland. *Journal of Economics and Management, 17*, 5–23.
191. Niezgoda, A. (2012). Popyt turystyczny – uwarunkowania i perspektywy rozwoju. *Zeszyty Naukowe Uniwersytetu Szczecińskiego, 697,* 12–21.
192. Niezgoda A., Zmyślony P. (2003) Popyt turystyczny. Uwarunkowania i perspektywy rozwoju, Akademia Ekonomiczna, Poznań, s. 118-141.
193. Nunan, D., & Di Domenico, M. (2017). Big data: A normal accident waiting to happen? *Journal of Business Ethics*, 145(3), 481–491. https://doi.org/10.1007/s10551-015-2904-x
194. *On economic contribution of tourism: TSA 2008.* (n.d.). UNWTO. Retrieved September 14, 2024, from https://www.unwto.org

195. Osei, B.A. (2022). Travel to Emerging Tourist Destinations: Motivations and Decision-Making Processes. In: Mensah, I., Balasubramanian, K., Jamaluddin, M.R., Alcoriza, G., Gaffar, V., Rasoolimanesh, S.M. (eds) Marketing Tourist Destinations in Emerging Economies. Palgrave Studies of Marketing in Emerging Economies. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-83711-2_5
196. Outeda, C. C. (2024). The EU's AI act: a framework for collaborative governance. Internet of Things, 101291.
197. Pan,B.,&Yang,Y.(2016).Forecasting destination weekly hotel occupancy with big data. Journal of Travel Research. https://doi.org/10.1177/0047287516669050
198. Page, S. (2009). *Tourism management*. Routledge.
199. Panasiuk, A. (2016a). Asymetria informacji na rynku turystycznym. *Marketing i Zarządzanie, 1*(42), 149–158. https://doi.org/10.18276/miz.2016.42-10
200. Panasiuk, A. (2016b). Polityka Unii Europejskiej w oddziaływaniu na wielkość i strukturę popytu turystycznego. *Współczesne Problemy Ekonomiczne, 13,* 116–123.
201. Panasiuk, A. (2018). Relacje rynku turystycznego z rynkami substytucyjnymi. *Ekonomiczne Problemy Usług, 130*, 99–108. https://doi.org/10.18276/epu.2018.130-10
202. Panasik, A. (2019). Produkt turystyczny. Pojęcie, struktura, uwarunkowania kreowania. Przedsiębiorczość i Zarządzanie, 20(2, cz. 2), p. 11-24.
203. Patel, R. (2021). *Principles and practices of tourism*. Dr. Babasaheb Ambedkar Open University.
204. Petrosillo,I.,Zurlini,G.,Grato,E.,&Zaccarelli,N.(2006).Indicatingfragilityofsocio-ecologicaltourism-basedsystems.EcologicalIndicators,6,104–113.
205. Pierścionek, Z. (ca. 2013). Ewolucja koncepcji konkurencyjności przedsiębiorstwa. Retrieved December 31, 2024, from https://www.centrumwiedzy.edu.pl
206. Pitt, C. S., Bal, A. S., & Plangger, K. (2020). New approaches to psychographic consumer segmentation: Exploring fine art collectors using artificial intelligence, automated text analysis, and correspondence analysis. *European Journal of Marketing.*
207. Pociecha, M. (2002). *Metody statystyczne w zarządzaniu turystyką*. Albis.
208. Pokharel, R., Poudel, J., & Grala, R. K. (2018). Impact of economic factors and a political conflict on international tourism demand in Nepal: A vector error correction model. *International Journal of Tourism Policy, 8*(1), 18–41.
209. Polski Instytut Ekonomiczny. (2023). *Wpływ wojny w Ukrainie na działalność polskich firm*. Warszawa. Retrieved December 31, 2024, from https://pie.net.pl/wp-content/uploads/2023/02/Wplyw-wojny-na-PL-firmy.pdf
210. Popeangă, J., & Lungu, I. (2012). Real-time business intelligence for the utilities industry. *Database Systems Journal*, 3(4), 15.
211. Porter, M. E. (2001). *Porter o konkurencji*. Warszawa: PWE.
212. Porter, M. E. (2010). *Strategia konkurencji. Metody analizy sektorów i konkurentów*. MT Biznes.
213. Projekt Planu Zagospodarowania Przestrzennego. (2015). Kujawsko-pomorskie Biuro Planowania Przestrzennego i Regionalnego we Włocławku, Urząd Marszałkowski Województwa Kujawsko-Pomorskiego. Retrieved from https://kujawsko-pomorskie.pl/wp-content/uploads/2015/07/02_plan-cc8.pdf
214. Quaresma, S., Maślankowski, J., Salgado, D., Ascari, G., Brancato, G., Di Consiglio, L., Righi, P., Tuoto, T., Daas, P., Six, M., & Kowarik, A. (2020). Revised version of the quality guidelines for the acquisition and usage of big data (Deliverable K3: Work package K: Methodology and quality). ESSnet Big Data II, Eurostat. https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en
215. Quintais, J. P. (2024). Generative AI, copyright and the AI Act. Available at SSRN.

216. Radzikowski, B., & Śmietanka, A. (2016). Online CASE CPI. First International Conference on Advanced Research Methods and Analytics, Valencia.
217. Rapacz, A. (1994). Przedsiębiorstwo turystyczne. Podstawy i zasady działania. Wydawnictwo Naukowe PWN.
218. Rapacz, A. (2007). *Turystyka i polityka turystyczna*. Wydawnictwo UE.
219. Rapacz, A., Gryszel, P., & Jaremen, D. E. (2010). Klaster turystyczny jako forma koopetycji w obszarze recepcji turystycznej. In J. Sala (Ed.), *Konkurencyjność miast i regionów na globalnym rynku turystycznym* (pp. 55–70). Polskie Wydawnictwo Ekonomiczne.
220. Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, 49, 393–432.
221. Ritchie, J. R. B., & Crouch, G. I. (2000). The competitive destination: A sustainable tourism perspective. CABI Publishing.
222. Rogers EM (2010) Diffusion of innovations. Simon and Schuster. New York
223. Rojewska, K., Dobrowolska, K., & Kreft, P. (2023). *Analiza pokoleń – jak zrozumieć współczesnego turystę?* Qtravel.ai. https://www.qtravel.ai/wp-content/uploads/2023/06/Analiza-pokole-jak-zrozumie-wspczesnego-turyst-1.pdf
224. Rojszczak, M. (2019). Definicja i granice prawnej ochrony prywatności w epoce analityki big data. *Ruch Prawniczy, Ekonomiczny i Socjologiczny, 81*(1), 118–130.
225. Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018). How Trump consultants exploited the Facebook data of millions. *The New York Times.*
226. Russom, P. (2011). *Big data analytics* (TDWI Best Practices Report, Fourth Quarter). The Data Warehousing Institute.
227. Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *Collaboration Technologies and Systems (CTS), International Conference on IEEE,* 42–4
228. Saluveer, E., & Raun, J. et al. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research, 81*, Article 102895. https://www.sciencedirect.com
229. Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: A systematic literature review. *Journal of Hospitality and Tourism Technology, 11*(2), 343–367.
230. Saunders JA (1980) Cluster analysis for market segmentation. Eur J Mark 14(7):422–435
231. Saura, J. R., & Bennett, D. R. (2019). A three-stage method for data text mining: Using UGC in business intelligence analysis. *Symmetry, 11*(4), 519.
232. Schumpeter, J. (1991). *The economic and sociology of capitalism*. New York.
233. Scism, L., & Maremont, M. (2010). Insurers test data profiles to identify risky clients. Retrieved from www.wsj.com/articles
234. Scott, D. (2017). Why Veblen matters: The role of status seeking in contemporary leisure. In K. Spracklen, B. Lashua, E. Sharpe, & S. Swain (Eds.), *The Palgrave Handbook of Leisure Theory* (pp. 435–449). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-56479-5_22
235. Selectivv. (2021). *Badanie ruchu turystycznego w województwie kujawsko-pomorskim*.
236. Seweryn, R. (2017). *Pokolenie Baby Boomers na rynku turystycznym na przykładzie odwiedzających Kraków*. Marketing i Zarządzanie, 2(48), 105–116. https://doi.org/10.18276/miz.2017.48-10.
237. Sharma, C., & Pal, D. (2020). Exchange rate volatility and tourism demand in India: Unraveling the asymmetric relationship. *Journal of Travel Research, 59*(7), 1282–1297. https://doi.org/10.1177/0047287519878516
238. Singh, A. (2021). Methods for data lakes implementation. *Intercathedra, 3*(48), 154. https://doi.org/10.17306/J.INTERCATHEDRA.2021.00130

239. Skalska, T. (Red.). (2020). Segmentacja rynku turystycznego : wybrane aspekty. Akademia Finansów i Biznesu Vistula
240. Smallman, C., & Moore, K. (2010). Process studies of tourists' decision-making. *Annals of Tourism Research, 37*(2), 397-422
241. Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. Retrieved from http://www2.hn.psu.edu/faculty/jmanis/adam-smith/wealth-nations.pdf
242. Smith CS (2019) Dealing with bias in artificial intelligence. Available at https://www.nytimes.com/ 2019/11/19/technology/artificial-intelligence-bias.html.
243. Smith, S. L. J., & Xiao, H. (2008). Product market analysis for tourism. *Tourism Management, 29*(3), 624–636.
244. Smith, K. (2015). Big data discoveries: insurers are just scratching the surface of what's possible with big data. Best's Review, 53–56. Pobrane z www. highbeam.com/doc/1G1–433879430.html
245. Smith, W. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing, 21,* 3–8.
246. Snell, J., & Menaldo, N. (2016). Web Scraping in an Era of Big Data 2.0. Bloomberg BNA, 2–5
247. Sobko, R., & Kolonowska-Matynia, M. (2019). Determinants of demand for tourist products: The example of a local tourist market. *Zeszyty Naukowe Wydziału Nauk Ekonomicznych Politechniki Koszalińskiej, 23,* 89–108.
248. Sokołowski, W., Ficoń, K., & Florczyk-Żółtowska, M. (2024). Data is the fuel of generative artificial intelligence: Dane paliwem generatywnej sztucznej inteligencji. Systemy Logistyczne Wojsk, 60, 199-216.
249. Song,H.,Li,G.,Witt,S.F.,&Fei,B.(2010).Tourism demand modelling and forecasting: How should demand be measured? TourismEconomics,16(1),63–81.
250. Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management, 29*(2), 203–220.
251. Song, H., Qiu, R. T. R., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management, 94,* Article 104655. https://doi.org/10.1016/j.tourman.2022.104655
252. Song, H., P. Romilly and X. Liu (2000), An empirical study of outbound tourism demand in the UK, Applied Economics, 611–24
253. Stabler, M., Papatheodorou, A., & Sinclair, T. M. (2010). *The economics of tourism.* Routledge.
254. Stamberger, L., van Riper, C. J., Keller, R., Brownlee, M., & Rose, J. (2018). A GPS tracking study of recreationists in an Alaskan protected area. *Applied Geography, 93*, 92–102. https://doi.org/10.1016/j.apgeog.2018.02.007
255. Stanton, J. (1975). *Fundamentals of marketing*. McGraw-Hill.
256. Strategor. (2001). *Zarządzanie firmą*. Warszawa: PWE.
257. Stępień, A. (2018). Bezpieczeństwo danych osobowych w cyberprzestrzeni – Big Data. *Przedsiębiorczość i Zarządzanie*, 19(2/3), s. 56.
258. Suganya, E., & S. V. (2019). Sentiment analysis for scraping of product reviews from multiple web pages using machine learning algorithms. *International Conference on Intelligent Systems Design and Applications*.
259. Sun, S., Wei, Y., Tsui, K. L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1–10.
260. Swan, M. (2013). The quantified self: Fundamental disruption in big data science. *Big Data, 1*(2), 85–99.

261. Szafrańska, E., Włodarczyk, B., & Dziedzic, E. (2015). Pomiar wielkości ruchu turystycznego w regionie. In E. Dziedzic (Ed.), *Badania konsumentów usług turystycznych w regionach*. Polska Organizacja Turystyczna.
262. Takahashi, K. (2020). Comparing the determinants of tourism demand in Singapore and French Polynesia: Applying the tourism demand model to panel data analysis. *Tourism Analysis, 25,* 175–181. https://doi.org/10.3727/108354220X15758301241585
263. Tene, O. (2011). Privacy: The new generations. *International Data Privacy Law, 1*(1), 15–27.
264. The End of Covid-19-related travel restrictions. (2023). UNWTO.
265. Tung, V.W.S. and Au, N. (2018), "Exploring customer experiences with robotics in hospitality", International Journal of Contemporary Hospitality Management, Vol. 30 No. 7, pp. 2680-2697
266. Top 10 global consumer trends. (2022). Euromonitor International. Retrieved from https://go.euromonitor.com/white-paper-EC-2022-Top-10-Global-Consumer-Trends
267. Tourism and COVID-19: Leading tourism recovery. (2023). UNWTO.
268. Tourism and Hospitality Market Forecast. (2023). Tourism and hospitality market size (2022–2027). Retrieved from https://www.marketdataforecast.com/market-reports/tourism-and-hospitality-market
269. Tovmasyan, G. (2023). Factors that influence domestic tourism demand: Evidence from Armenia. Economics and Sociology, 16(2), 75-88. DOI: 10.14254/2071789X.2023/16-2/5.
270. *Travel + Leisure Magazine*. (n.d.). Retrieved December 31, 2024, from https://www.travelandleisure.com/
271. Travolution. (2018). Research identifies four key holiday booking trends. Retrieved from https://bit.ly/2zEAhWB
272. Tribe, J. (2016). *The economics of recreation, leisure and tourism*. Routledge.
273. Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big Data*, 2(1), 21. DOI: 10.1186/s40537-015-0030-3, s. 1.
274. Tussyadiah, I. (2020), "A review of research into automation in tourism: launching the annals of tourism research curated collection on artificial intelligence and robotics in tourism", Annals of Tourism Research, December 2018, Vol. 81, p. 102883.
275. Tussyadiah, I., & Miller, G. (2019). Perceived impacts of artificial intelligence and responses to positive behavior change intervention. In *Information and Communication Technologies in Tourism 2019* (pp. 359–370). Springer, Cham.
276. United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. Retrieved from https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981
277. Urban G. (2014), *Marketing wielkich możliwości: jak sprzedawać więcej w szumie informacyjnym*, Wydawnictwo Helion, Gliwice.
278. Urząd Statystyczny w Bydgoszczy. (2022). *Turystyka w województwie kujawsko-pomorskim w 2022 r.* Bydgoszcz: Urząd Statystyczny w Bydgoszczy.
279. Ustawa z dnia 25 czerwca 1999 r. o Polskiej Organizacji Turystycznej (Dz. U. 1999 Nr 62 poz. 689).
280. Uysal, M. (1998). The determinants of tourism demand: A theoretical perspective. In D. Ioannides & K. Debbage (Eds.), *The economic geography of the tourist industry* (pp. 79–98). Routledge.
281. Uysal, M., & Crompton, J. L. (1984). Determinants of demand for international tourist flows to Turkey. *Tourism Management*, 5(4), 288-297.
282. Valls, A., Gibert, K., Orellana, A., & Anton-Clave, S. (2018). Using ontology-based clustering to understand the push and pull factors 49J. of the Acad. Mark. Sci. (2021) 49:30–50 for

British tourists visiting a Mediterranean coastal destination. Information & Management, 55, 145–159.
283. Vankevich, A., & Kalinouskaya, I. (2021). Better understanding of the labor market using big data. *Ekonomia i Prawo. Economics and Law, 20*(3), 677–692. https://doi.org/10.12775/EiP.2021.040
284. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017) *arXiv*. https://doi.org/10.48550/ARXIV.1706.03762
285. Var, T., Mohammad, G., & Icoz, O. (1990). Factors affecting international tourism demand for Turkey. *Annals of Tourism Research*, 17(4), 606-610.
286. Walas B., Segmenty turystów, obszary produktowe oraz rynki emisyjne - weryfikacja, WSTiE, Warszawa 2021
287. Wąsowicz-Zaborek, E. (2019). Treści generowane przez użytkowników Internetu jako źródło wiedzy o wizerunku miejsca docelowego podróży na przykładzie miasta Kołobrzeg**.** *Folia Turistica*, nr 53. DOI: 10.5604/01.3001.0013.7512 za Sweeney J., Soutar G., Mazzarol T. (2008), Factors influencing word of mouth effectiveness: receiver perspectives, "European Journal of Marketing", Vol. 42, s. 344-364 oraz Yoo K-H., Gretzel U. (2016), Web 2.0: ew Rules for Tourism Marketing, Travel and Tourism Research Association: Advancing T ourism Research Globally. 26.
288. Wedel, M. & Kannan, P.K. (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6), 97–121.
289. Włodarczyk, B. (Ed.). (2011). Ruch turystyczny w Łodzi i w województwie łódzkim w 2010 roku. Łódzkie Towarzystwo Naukowe.
290. Wołkowicz, A. (2015). Modele ekonometryczne jako narzędzie sterowania procesami technologicznymi. Ekonometria, (48), 67-77. Uniwersytet Ekonomiczny we Wrocławiu.
291. World Economic Forum. (2024). Tourism is back to pre-pandemic levels, but challenges remain. Retrieved May 21, 2024, from https://www.weforum.org/publications/travel-tourism-development-index-2024/press-release-and-related-contributions-12ce9d1d22
292. World Population Prospects – Population Division. (n.d.). United Nations.
293. World Tourism Organization. (2017). *Tourism and economic growth*. UNWTO.
294. Pierścionek, Z., ca. 2013, Ewolucja koncepcji konkurencyjności przedsiębiorstwa, www.centrumwiedzy.edu.pl
295. Woźniczka, J. (2018). Big data w marketingu: Szanse i zagrożenia. *Studia Oeconomica Posnaniensia, 6*(6). https://doi.org/10.18559/SOEP.2018.6.3
296. WTTC. (2021). *Economic Impact Report*.
297. Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research, 69*(5), 1562–1576.
298. Yazdi, S. K., & Khanalizadeh, B. (2017). Tourism demand: A panel data approach. *Current Issues in Tourism, 20*(8), 787–800. https://doi.org/10.1080/13683500.2016.1170772S.
299. Wodejko, S. (1998). *Ekonomiczne zagadnienia turystyki*. Warszawa: Wyższa Szkoła Handlu i Prawa.
300. Yu, G., & Schwartz, Z. (2006). Forecasting short time-series tourism demand with artificial intelligence models. *Journal of Travel Research, 45*(2), 194–203.
301. Zawadzka, M. (2014). Rola i zadania regionalnych organizacji turystycznych na przykładzie województwa mazowieckiego. *WSTiJO, Zeszyty Naukowe*, *2*(14), 33.
302. Zhang, L., Sun, X., & Wagner, C. (2018). Exploring the group holiday decision-making process with the support of technology. *Information Processing & Management, 54*(3), 397-409.
303. Zhan, Y., Tan, K., Ji, G., Chung, L. & Tseng, M.L. (2017). A big data framework for facilitating product innovation processes. Business Process Management Journal, 23(3), 518–536.

304. Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2012). Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems, 14*(1), 113–123. Retrieved August 28, 2023, from https://ieeexplore.ieee.org/document/6266748
305. Żuchowski, I. (2007). Segmentacja rynku – ujęcie teoretyczne. *Zeszyty Naukowe Wyższej Szkoły Ekonomiczno-Społecznej w Ostrołęce*, (4), 119–132.

Other online sources:


1. Historical monuments: https://www.prezydent.pl/aktualnosci/polityka-historyczna/pomniki-historii/obiekty-wpisane-na-liste-pomnikow-historii
2. Unesco: https:
3. Opera Nova in Bydgoszcz: https:
4. Bydgoszcz Jazz Festival: http://bydgoszczjazzfestival.com.pl/
5. Eurovelo: https://en.eurovelo.com/
6. Polish Tourist Organisation: https:
7. Tourist Guarantee Fund: https://ewidencja.ufg.pl/ewidencja/obywatel/wyszukiwanie
8. Booking: Booking.com
9. Central List of Hotel Facilities: Public Register System for Tourism - CWOH Register
10. Business Insider: https://businessinsider.com.pl/gospodarka/zmiany-w-najmie-krotkoterminowym-powstanie-rejestr-mieszkan/9bcrkgx
11. BlueWhite: https://bluewhite.pl/czy-historia-przegladania-ma-wplyw-na-cene-biletu-lotniczego-lub-hotelu/?utm_source=chatgpt.com

## INDEX OF FIGURES

Figure 1 . Destinations according to UNWTO ..... 10
Figure 2 . Tourism economy diagram ..... 24
Figure 3 . Conceptualising competitiveness ..... 30
Figure 4 . Types of demand ..... 49
Figure 5 . Word cloud showing factors relevant to tourism demand ..... 56
Figure 6 . Decision-making stages ..... 58
Figure 7 . Formula for income elasticity of demand ..... 65
Figure 8 . Market segmentation process in the tourism sector ..... 78
Figure 9 . Overview of research on tourism demand segmentation ..... 78
Figure 10 . Key elements in defining a tourist's behavioural profile ..... 80
Figure 11 . Package minimalist - example of a behavioural profile for Warsaw ..... 81
Figure 12 . STP model ..... 83
Figure 13 . Attributes of big data (traditional view) ..... 85
Figure 14 . Attributes of big data (contemporary view) ..... 87
Figure 15 . Analysis of the benefits of big data analytics. ..... 88
Figure 16 . The process of linking data sources with consumers ..... 92
Figure 17 . The concept of data matching ..... 96
Figure 18 . The potential of big data according to McKinsey ..... 99
Figure 19 . Breakdown of big data ..... 102
Figure 20 . Relationship between AI, ML and DL ..... 103
Figure 21 . Model and application of an artificial neuron ..... 104
Figure22 . Integration of big data and AI ..... 105
Figure 23 . The idea of big data analytics in the cloud ..... 107
Figure 24 . Tourism demand forecasting methods ..... 110
Figure 25 . Data acquisition methods and sources ..... 112
Figure 26 . Structure of tourist accommodation in Bulgaria and Poland ..... 116
Figure 27 . Basic legal acts governing data processing ..... 122
Figure 28 . Structure of AI systems and models ..... 124
Figure 29 . The data trail and its potential use in tourism ..... 131
Figure 30 . AI systems for tourism ..... 136
Figure 31 . Data sources planned for use in the Tourism Barometer ..... 152
Figure 32 . The idea behind the Tourism Barometer ..... 153
Figure33 . Basic parameters of tourist traffic ..... 155
Figure 34 . Use of web scraping in academic literature since 2000 ..... 157
Figure 35 . The process of generating information extracted through web scraping ..... 160
Figure 36 . The Toruism Demand Horizon idea ..... 163
Figure 37 . Simplified diagram of how the Torusim Demand Horizon works (…) ..... 164
Figure 38 . Structure of WKP hotel facilities by standard in the CWOH ..... 172
Figure 39 . Occupancy rate of accommodation in the Kujawsko-Pomorskie voivodship ..... 173
Figure 40 . Average number of beds in the districts during the period under review ..... 175
Figure 41 . Number of beds on Booking.com in the counties (…) ..... 176
Figure 42 . Number of Booking.com beds in the (…) ..... 176
Figure 43 . Number of beds on Booking.com in the counties (…) ..... 177
Figure 44 . Number of beds on Booking.com in the counties (…) ..... 177
Figure 45 . Number of beds on Booking.com in cities in the county (…) ..... 179
Figure 46 . Number of beds on Booking.com in cities in the county (…) ..... 180

Figure 47 . Tourism Demand Horizon+ (theoretical assumptions) ..... 185

**TABLE OF TABLES**

Table 1 . Overview of types and forms of tourism ..... 12
Table 2 . Typology of tourist distribution ..... 12
Table 3 . Categories of tourists according to P.L. Pearce and characteristic (…) ..... 13
Table 4 . Overview of selected tourism product definitions ..... 15
Table 5 . Overview of definitions tourism economy ..... 18
Table 6 . Direct and indirect tourism economy ..... 22
Table 7 . Tourism balance ..... 25
Table 8 . Overview of competition definitions ..... 28
Table 9 . Competitive advantage in the literature ..... 31
Table 10 . Impact of Covid-19 on international tourist arrivals (YTD rate - %) ..... 37
Table 11 . Average number of trips by age group ..... 38
Table 12 . Value of accommodation bookings by type of stay ..... 39
Table13 . Accommodation and travel expenditure 2023-2030 (forecast) ..... 39
Table 14 . 7 major travel trends for 2024 ..... 40
Table 15 . Perspective on changes in travel plans as a result of the war in Ukraine ..... 41
Table 16 . Trends in the tourism market ..... 42
Table 17 . Interpretive differences between VR and AR ..... 46
Table 18 . Overview of definitions of tourism demand ..... 48
Table 19 . Research issues of tourism demand ..... 57
Table 20 . Differences in decision-making by different generations of tourists ..... 60
Table 21 . Economic sources of impact on tourism demand ..... 61
Table 22 . Phrases of the tourist decision-making process ..... 73
Table 23 . Group decision-making process ..... 75
Table 24 . Segmentation criteria ..... 79
Table 25 . Advantages and disadvantages of the STP model (…) ..... 84
Table 26 . Overview of big data definitions ..... 89
Table 27 . Advantages and disadvantages of using big data ..... 93
Table 28 . Differences between ML and DL ..... 103
Table 29 . Main machine learning models ..... 105
Table 30 . Overview of selected tools for big data analysis ..... 106
Table 31 . Advantages and disadvantages of webscraping ..... 114
Table 32 . Overview of big data analysis tools ..... 117
Table 33 . Privacy risks in relation to big data ..... 120
Table 34 . Big data analysis from an ethical perspective ..... 125
Table 35 . Dimensions of Big Data Analysis in Tourism ..... 128
Table 36 . Importance of image attributes of Kolobrzeg ..... 134
Table 37 . Factors transforming AI into AGI and implications for tourism ..... 138
Table 38 . Overview of research methodologies for the tourism sector ..... 139
Table 39 . Process for capturing data for Tourism Demand Horizon (…) ..... 165
Table 40 . Data acquisition process in research ..... 167
Table41 . Overview of the facilities available on the booking portal by locality ..... 168
Table 42 . Summary of the indicated number of facilities and accommodation (…) ..... 169
Table 43 . Number of sites appearing on the CWOH and Booking portal ..... 170
Table 44 . Overview of kujawsko-pomorskie voivodeship accommodation (…) ..... 171
Table 45 . Counties best ranked in raiting ..... 173
Table 46 . Average number of available beds by county (as of 13.11.2024) ..... 174

Table47 . Difference in availability of accommodation (…) ..... 178
Table48 . Difference (N and %) in the number of reserved places (…) ..... 181
Table49 . Changes in the number of available beds during the period under review ..... 182

Annex 1. Example of data capture process dated 13.11.2024

