

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

The use of sentiment analysis to identify visitors' perceptions of Central European capital cities

Ana Luiza Beck Santos

Erasmus Mundus Masters in Tourism Development and Culture

Supervisor:

Dr Ricardo Godinho Bilro, Assistant Professor
ISCTE Business School

May 2023



BUSINESS
SCHOOL

Department of Marketing, Strategy and Operations

The use of sentiment analysis to identify visitors' perceptions of Central European capital cities

Ana Luiza Beck Santos

Erasmus Mundus Masters in Tourism Development and Culture

Supervisor:

Dr Ricardo Godinho Bilro, Assistant Professor
ISCTE Business School

May 2023

Abstract

User-generated content can provide valuable insights into the tourism industry. This dissertation aims to address the tourist perceptions of cultural attractions in Central European capital cities. Despite the tourism relevance of destinations such as Prague, Berlin and Vienna, few studies focus on the tourism attractiveness of the region. The research objectives include (i) identifying the perceptions of the top cultural attractions of the cities in the sample, (ii) underlining similarities and differences in the tourism experience in the different locations, and (iii) providing recommendations for tourism planning organizations.

To achieve that, a netnographic methodology is applied. To do so, TripAdvisor online reviews were extracted from cultural attractions for the six Central European capital cities included in the sample – Berlin, Bratislava, Budapest, Prague, Vienna, and Warsaw. Following that, text and sentiment analyses were conducted with the use of the software Meaning Cloud. The results show the General and the Topic Sentiment analysis for each city, and a cluster Topic Sentiment Analysis, which highlights the main aspects of the tourism experience brought up in the comments.

The results indicate an overall positive perception of all the cities in the sample. The cities with better performance were Budapest, Prague, and Vienna, with a focus on aspects such as city, tourism attractions, and locations. The topic “People” showed the worst performance, aligning with concerns about over-tourism in the main attractions. Berlin presented the lowest average polarity, indicating the need for a closer look at the city’s management as a cultural destination.

Keywords: Central Europe, cultural tourism, text-mining, sentiment analysis, DMOs

JEL Codes: Z33, Z38

Resumo

O conteúdo gerado pelos usuários pode fornecer insights valiosos para a indústria do turismo. Esta dissertação tem como objetivo abordar as percepções dos turistas sobre atrações culturais nas capitais da Europa Central. Apesar da relevância turística de destinos como Praga, Berlim e Viena, poucos estudos se concentram na atratividade turística da região. Os objetivos da pesquisa incluem (i) identificar as percepções sobre as principais atrações culturais das cidades da amostra, (ii) destacar semelhanças e diferenças na experiência turística nos diferentes locais e (iii) fornecer recomendações para organizações de planejamento turístico.

Para alcançar isso, uma metodologia netnográfica é aplicada. Para isso, foram extraídas avaliações online do TripAdvisor de atrações culturais das seis capitais da Europa Central incluídas na amostra – Berlim, Bratislava, Budapeste, Praga, Viena e Varsóvia. Em seguida, foi realizada uma análise de texto e sentimento utilizando o software Meaning Cloud. Os resultados mostram as Análise Gerais e Tópico de Sentimento para cada cidade, além de uma análise de sentimento de cluster de tópicos, que destaca os principais aspectos da experiência turística mencionados nos comentários.

Os resultados indicam uma percepção geralmente positiva de todas as cidades da amostra. As cidades com melhor desempenho foram Budapeste, Praga e Viena, com foco em aspectos como cidade, atrações turísticas e localizações. O tópico "Pessoas" apresentou o pior desempenho, em linha com as preocupações sobre o excesso de turismo nas principais atrações. Berlim apresentou a menor polaridade média, indicando a necessidade de uma análise mais detalhada da gestão da cidade como um destino cultural.

Palavras-chave: Europa Central, turismo cultural, mineração de texto, análise de sentimento, organizações de planejamento turístico.

Códigos JEL: Z33, Z38

General Index

1.	Introduction	1
1.1.	Research question and objectives	2
1.2.	Methodology	3
2.	Literature review	4
2.1.	Cultural and historical heritage as tourism products.....	4
2.2.	Tourism in Central Europe.....	6
2.3.	The role of visitors' satisfaction in tourism loyalty	9
2.4.	The use of user-generated content in decision-making processes	11
3.	Current DMO initiatives for Central European Capital Cities.....	14
4.	Text mining and sentiment analysis	17
5.	Data analysis and results	19
5.1.	Data collection	19
5.2.	Data description and limitations	19
5.3.	Results.....	21
5.3.1.	Berlin	21
5.3.2.	Bratislava.....	24
5.3.3.	Budapest	27
5.3.4.	Prague	30
5.3.5.	Vienna.....	32
5.3.6.	Warsaw	35
5.3.7.	Results summary.....	38
6.	Discussion	41
7.	Conclusions	44
7.1.	Theoretical contributions	44
7.2.	Managerial contributions	44
7.3.	Research limitations.....	45
7.4.	Implications and future research.....	45
	Reference list.....	47
	Annex A.....	55

Table of contents and figures

Table 5.1 – Data description.....	20
Table 5.2 – Polarity conversion table	20
Table 5.3 – General Sentiment Analysis: Berlin	22
Table 5.4 – General Sentiment Analysis metrics: Berlin.....	22
Table 5.5 – Topic Sentiment Analysis: Berlin	23
Table 5.6 – Topic extraction (TSA analysis): Berlin	23
Table 5.7 – General Sentiment Analysis: Bratislava	25
Table 5.8 – General Sentiment Analysis metrics: Bratislava	25
Table 5.9 – Topic Sentiment Analysis: Bratislava.....	26
Table 5.10 – Topic extraction (TSA analysis): Bratislava.....	26
Table 5.11 – General Sentiment Analysis: Budapest.....	28
Table 5.12 – General Sentiment Analysis metrics: Budapest.....	28
Table 5.13 – Topic Sentiment Analysis: Budapest	28
Table 5.14 – Topic extraction (TSA analysis): Budapest	29
Table 5.15 – General Sentiment Analysis: Prague	30
Table 5.16 – General Sentiment Analysis metrics: Prague	31
Table 5.17 – Topic Sentiment Analysis: Prague	31
Table 5.18 – Topic extraction (TSA analysis): Prague	31
Table 5.19 – General Sentiment Analysis: Vienna	33
Table 5.20 – General Sentiment Analysis metrics: Vienna	33
Table 5.21 – Topic Sentiment Analysis: Vienna.....	33
Table 5.22 – Topic extraction (TSA analysis): Vienna	34
Table 5.23 – General Sentiment Analysis: Warsaw.....	36
Table 5.24 – General Sentiment Analysis metrics: Warsaw	36
Table 5.25 – Topic Sentiment Analysis: Warsaw	36
Table 5.26 – Topic extraction (TSA analysis): Warsaw	37
Table 5.27 – Results summary	39

1. Introduction

Tourism represents a relevant economic segment in European countries. There are several well-known and well-established cultural tourism destinations in the continent, such as London, Paris, Barcelona, Rome, Amsterdam, and several others. In 2022, the number of international tourist arrivals doubled after the slower activities registered in 2020 and 2021 due to the COVID-19 pandemic. Although the numbers remain below the levels of 2019, the tourism segment poses a relevant economic force in the European context (Statista, 2023b). Although Central European capital cities cannot be described as new or unknown destinations, there is a gap in literature and studies that analyse and compare the potentialities of the region regarding their touristic potential. Tourism in Central Europe presented significant growth after 1989 when cities started to modernize themselves and market themselves with an attempt of distancing their brand from the socialist history whilst taking advantage of the tourist attractiveness resulting from Soviet and other heritage monuments (Naumov & Weidenfeld, 2019; Rätz et al., 2008).

Cultural and heritage tourism is one of the fastest-growing segments in the sector (Vena-Oya et al., 2021) – therefore, exploring visitors' perceptions of these kinds of attractions holds an opportunity to further develop destinations and explore the effectiveness of the local Destination Management Organisations' marketing strategies and objectives.

There are several determinants in the destination decision-making choice. Among the elements that influence visitors' likelihood to choose a destination or attraction is word of mouth. Jalilvand and Samiei (2012) define word of mouth (WOM) as “communication about products and services between people who are perceived to be independent of the company providing the product or service, in a medium perceived to be independent of the company”. The intrinsic intangibility of tourism products generates a need for other's information regarding the destination. It is not possible to provide a sample of tourism products beforehand, and, therefore, they attempt to measure the satisfaction they will have with that given place or service through the opinion of people that already had that experience (Jalilvand & Samiei, 2012).

Decades ago, that information was limited to tour agencies, blogs, and agency pamphlets - and word of mouth from the surrounding people. Today, the digital sphere allows access to a vast range of information, made available both by service providers and random individuals that share their experiences through blogs, social media posts, comments, and online reviews. The expansion of that to the digital sphere potentialized the impact of word of mouth – or, in

that case, electronic word of mouth (eWOM) – in tourism products (Águeda et al., 2019; Jalilvand & Samiei, 2012).

Although the massive amount of data available makes it more difficult for businesses and stakeholders to control the information regarding their product or service, it also opens up the opportunity to analyse the user-generated content and learn from the visitors' experience, enhancing or fixing the aspects that generate discontent, as well as emphasizing the good practices in place and predict the likelihood of future purchases (Sharda et al., 2014). In order to process this intensive volume of information, text-mining techniques pose themselves as valuable tools. Sharda et al. (2014, p. 230) define text-mining as a “semiautomated process of extracting patterns (useful information and knowledge) from large amounts of unstructured data sources”. User-generated content can be extracted from several sources – social media, forums, blogs, consumer-generated content sites or consumer review websites and others. The biggest consumer review website currently available is TripAdvisor (Mehraliyev et al., 2022).

With that considered, this work proposes the use of text-mining and sentiment analysis to explore reviews on tourism attractions in Central European capital cities, and, through that, identify the elements that impact visitor satisfaction, as well as different perceptions and behaviours regarding the six different cities – Berlin, Bratislava, Budapest, Prague, Vienna, and Warsaw.

1.1. Research question and objectives

The aim of this dissertation relies on understanding the key aspects of the visitors' perceptions and experiences when visiting Central European (CE) capital cities and their main attractions, as well as identifying the most positive aspects in each city to propose best practices and managerial recommendations to the Destination Management Organizations (DMO) in place. Therefore, the research question could be posed: what are the visitors' perceptions about the main attractions of Central Europe's capital cities and which aspects of the tourist experience have the best perceptions among tourists?

Within that scope, three main research objectives were established:

- i. Identify the visitors' perceptions of each CE capital city through sentiment analysis of 2022 TripAdvisor reviews on the top attractions of each capital city.
- ii. Compare and describe similarities and differences in visitors' experiences in the six locations.
- iii. Provide insights for tourism planning decisions to promote regional tourism development.

1.2. Methodology

A netnography method was considered the most fitted methodology to proceed with the visitors' experience towards the Central European capital cities' attractions. The netnography approach was initially contextualized by Konizets publication in 1988, and other studies from the academic were published further on the subject, as the web developed. It relied on the idea that consumers turned to online communication and forums to acquire information on to base their purchase decisions – therefore, analysing that content consisted of a significant market research opportunity (Kozinets, 2002). The author defines netnography as:

a new qualitative research methodology that adapts ethnographic research techniques to study the cultures and communities that are emerging through computer-mediated communications. As a marketing research technique, netnography uses the information that is publicly available in online forums to identify and understand the needs and decision influences of relevant online consumer groups. (Kozinets, 2002, p. 62)

The conceptualization of netnography as a development of ethnography was later adapted, as the methodology developed its features and was adapted to the needs of the researchers (Whalen, 2018). Nevertheless, it remains a tool to identify behavioural patterns, consumption motivations (Kozinets, 2002), cultural aspects, as well as expectations, desires and experiences from both tourists, suppliers, and online platform developers (Tavakoli & Mura, 2018; Tavakoli & Wijesinghe, 2019).

Finally, netnography was employed by a diverse number of tourism and leisure researchers to analyse consumer experiences and satisfaction (Tavakoli & Wijesinghe, 2019). After selecting netnography as the adequate methodology, text reviews from the selected online user-generated content platform (TripAdvisor) were extracted and a text mining analysis was conducted. Within that scenario, a sentiment analysis was performed to interpret the collected information and attribute the visitors' perceptions and feelings towards the destinations and their attractions.

2. Literature review

2.1. Cultural and historical heritage as tourism products

Cultural and heritage tourism is one of the fastest growing segments in the industry – it characterizes itself by the motivation of travellers to learn, discover, visit and experience tangible and intangible aspects of a destination's culture during the visit (Richards e Vena-Oya et al.). The aspects highlighted within the culture of a destination can include lifestyle, architecture, arts, heritage, history, and other elements (Konstantakis et al., 2020). According to Statista reports (Statista, 2023b), culture and heritage are one of the main types of trips planned by Europeans that go on vacation within the continent, overpassed only by sun and beach destinations and city breaks.

Gursoy et al. (2022) and Richards (2018) argue that learning new facts contributes to a more memorable and positive experience for tourists during all stages of the visit (pre-trip, on-sight, and post-trip). Therefore, the search for activities alternative to the sand and sea presents an opportunity to further develop cultural attractions and destinations. Differences between cultural and heritage tourism can be highlighted. The first one bases itself on people, and has a more interactive nature, including the tourist engagement in a vast range of activities, events, and attractions. The attractions and presentation of cultural tourism also evolve as culture transforms and it has a product and process-based form of consumption. Heritage tourism, on the other hand, relies on its location, as its attractions are related to a specific site that is not renewable or replaceable – the nature of the attraction, therefore, is static, and the consumption in process-based (Gursoy et al., 2022).

Richards (2018) argues that cultural tourism relies on the integration and management of three critical factors. First, consider available local resources, which include tangible and intangible heritage, local customs, habits, and cultural representations. Second, the meanings, refer to how the local community and tourist management tie the resources to the local identity, learning, and cultural experience narratives. Finally, the competencies, which represent good cultural tourism practices the cultural capital, and the interpretation of cultural resources.

Historically, tourism activities in European cities, especially the ones not on the Mediterranean coast, rely on city tourism, which most times overlaps with historical and heritage monuments to attract tourists, as well as other elements of local culture (gastronomy, festivals, museums, and others) (García-Hernández et al., 2017). Over the past decades, the perceptions and access to cultural tourism changed, as stated:

Cultural tourism is moving away from its former association with a narrow cultural elite towards becoming a significant global phenomenon. As the cultural tourism market

grows, the focus of cultural tourism is also rapidly changing from a preoccupation with sites and monuments into a much wider phenomenon, covering all aspects of 'high' and 'popular' culture. Cultural tourism consumption is no longer restricted to 'serious' and purposeful visits to revered cultural sites but has also become part of the 'atmosphere' of places, to be soaked up by tourists and residents alike. (Richards, 1999, p. 16)

Nonetheless, the intense concentration of tourists in historical monuments and heritage attractions, usually represented by the European cities' historic centres, brings some negative effects that can impact both the tourist experience and the life of locals. Issues such as overcrowded areas interconnect to overloads on infrastructure systems (such as transport) and intensification of noise pollution. In addition to that, the high incidence of visitors in specific areas of the city generates an increase in the prices for utilities (restaurants, shops, supermarkets) and rent, driving locals from out of these areas, which end up being frequented almost exclusively by tourists – the process called gentrification (García-Hernández et al., 2017; Matoga & Pawłowska, 2018).

Matoga and Pawłowska (2018) argue that these negative impacts of city and cultural tourism create a new niche of visitors that still aim to experience the cultural aspects of the cities, but alternatively seek attractions out-of-the-beaten-track, in a slow-tourism experience. However, even from that perspective, the authors point out that visitors still desire and visit the monuments located in the historic centres, and from there expand their itinerary. Within that context, cultural and heritage attractions are – for many European cities – the key image associated with the destination. Therefore, they are strongly connected to the branding of the destination and the marketing promotions from Destination Management Organizations. With that said, resource and site limitations must be taken into consideration. Even though historic centres will remain as relevant tourist attractions, a better understanding of the tourists' perceptions, experiences and goals when visiting these locations can result in the development of alternative routes within the destination, as well as pointers for better exploring other areas and adjusting marketing and branding strategies to maximize tourist satisfaction and the socio-economic benefits of tourism for the city.

Niche tourism approaches have been vastly adopted in the context of cultural tourism – including food tourism, dark tours, creative destinations, sustainable routes, and others (Richards, 2018). In the present work, the main cultural attractions of the cities in the sample are considered as starting points to understand the overall perceptions of visitors in the city and provide insights on the aspects of the tourism experience that can be improved.

2.2. Tourism in Central Europe

The definitions and distinctions between East and Western Europe have been vastly discussed in literature – from cultural, political, and economic perspectives as well as geography and demographics. The borders between these areas, however, remain fluid and open to discussion. Therefore, the attempted definition of Central Europe, a region in between those two. The most accepted designation of Central Europe relies on the concept of *Mittleuropa*, which is the area of the Austrian-Hungarian empire and the Lithuanian Commonwealth (Okey, 1992).

The *Mittleuropa* concept was geographically established by German academics. The territories included in the definition vary – some encompass all the former countries of the Austria-Hungarian empire (Austria, Hungary, Poland, Czech Republic, Slovenia, and Slovakia) and Germany. From that list, Slovenia can be sometimes left out due to its proximity to the Balkans. Other scholars include Croatia and the north of Italy, Bulgaria, and others. The list of different definitions of Central Europe, mostly based on political and historical elements, is extensive and no consensus was found between scholars (Okey, 1992; Šabič & Drulák, 2012). More recently, in the context of the European Union, the programme called “Central Europe” was established to promote cooperation projects between Austria, the Czech Republic, Germany, Hungary, Italy, Poland, Slovakia, and Slovenia. The list of definitions goes on indefinitely (Šabič & Drulák, 2012).

For this study, the countries included in the Central European analysis are Austria, Czechia, Germany, Hungary, Poland, and Slovakia, and their respective capital cities. For all the countries mentioned, the capital city is the main touristic destination (Statista, 2023b). The literature surrounding the tourism activities in these countries is very limited – there are a few studies published in English concerning some of the countries/cities, such as Vienna, Budapest, and Berlin – and, from a region point of view, published and updated research on tourism in Central Europe is close to inexistent.

On the few studies concerning tourism in Central Europe, the analysis provides an overall review of both Eastern and Central Europe as one homogeneous location. The development of tourism in Central European countries focused mainly on city and historical-culture tourism. After the fall of the iron curtain in 1989, the marketing of Central and Eastern cities attempted an effort to distance them from the past soviet stereotype (Kowalczyk-Anioł, 2023; Rätz et al., 2008). Simultaneously, some of the main attractions in these destinations related to socialist landmarks and icons – while locals tried to distance themselves from these ideals, tourists were attracted to them, creating a dichotomy within these sights (Naumov & Weidenfeld, 2019).

During the 1990s and the beginning of the 21st century, several central European capital cities experiences a modernization process, creating spaces that mixed the historic centres and modern developments (such as shopping stations and modern districts). Although the combined elements might sound appealing from a touristic point of view, it often presents an issue for national residents that cannot afford the new facilities, rent and shopping prices in the areas (Rátz et al., 2008). In addition to investment in infrastructure and revitalization of historic areas, to become tourist destinations, central and eastern European cities also had to invest in developing their branding and marketing strategy, as well as their place identity (Kowalczyk-Anioł, 2023). In 2004, when a lot of the applying countries (including Hungary, Czechia, Slovakia, and Poland) joined the European Union, the ties to the west became even closer, as well as the facilitation of transit within the Schengen area in 2007 (Tracz & Bajgier-Kowalska, 2019).

As one of the few comparative studies for Central European cities, Tracz & Bajgier-Kowalska (2019) propose a comparative study between visitors of Budapest, Prague, and Warsaw. The authors find that Budapest and Prague are more oriented towards international tourists and often benefit from incoming visitors that are in Vienna and Berlin and include the cities in the itinerary. Warsaw, on the other hand, presents a larger share of domestic tourism and competes with Krakow. Nevertheless, in terms of visitors' motivations and demographics, the three cities present similar findings: the main motivation for the visits relies on the desire to discover new places and engage in cultural tourism and entertainment including elements such as historical sites, architecture and activities that allow them to experience the environment and culture of the cities.

Borodako and Rudnicki (2014) and Šauer et al. (2021) discuss important elements in establishing tourism flows, such as transport infrastructure, tourist attractiveness, accessibility, and the local community. The most important aspect of transport infrastructure lies in how to get to the destination. Air connectivity is shown to be the most relevant in the development of tourist destinations, and access to Central European countries increased with the installation of low-cost airlines and new and more frequent air routes, leading to new segments of tourists and higher demand for tourism in the cities. More recently, concerns with climate impacts press to new investments in hard infrastructure, especially in the enhancement of train connections. Even though Central Europe is well-connected in terms of train lines, most of the routes do not include high-speed transport options, which can be considered a downside for visitors that rather not fly or that would like to experience more than one destination in the area at once within their holidays (Šauer et al., 2021).

Another aspect brought up by Šauer et al. (2021) is the identification of source markets in Central European destinations. The authors found that German tourists represent a large share of the tourists visiting Central Europe, as well as an intense flow of visitors between neighbouring countries or places with a shared history and culture background, such as the Czech Republic and Slovakia.

Some studies have an individual city as the focus. Rátz et al. (2008) discuss the importance of cultural tourism in the development of tourism in the city of Budapest from 1989 on, and the changes resulting in the making the city a touristic destination. Other research topics focused on Budapest relate to over-tourism and night-tourism (Olt et al., 2019; Pérez Garrido et al., 2022; Pinke-Sziva et al., 2019; Remenyik et al., 2021; Smith et al., 2019), the impacts of the COVID-19 pandemic in tourism activities and worker demand (Karacsony et al., 2022; Remenyik et al., 2020), and visitors' profile and motivations (Kay Smith et al., 2022).

Jakob (2013) discusses the eventification of the cities of Berlin and New York. The city of Berlin, while a subject of several studies, is not usually approached for its cultural and historical attractions. Despite its historical relevance and vast availability of historical sites (such as Brandenburg Gate, the Berlin Wall, and Check Point Charlie, among others), the city consolidated itself as a festival and alternative destination due to its strong manifestation of contemporary art and establishment as a hub for creative industries. There are, however, some studies focused on specific tourist attractions in Berlin, such as Luisenstadt (Engelbart & Krech, 2016) and Sehlik Mosque (Becker, 2018). Other studies on tourism sustainability (Grube, 2022; Kalandides & Grésillon, 2021), destination marketing and social media (Bonilla-Quijada et al., 2021), and commemorative events (Viol et al., 2018).

Some of the topics of articles on Prague and the Czech Republic include the authenticity of the tourism experience through the analysis of souvenir shops in Prague (Dumbrovská & Fialová, 2020), impacts of COVID on tourism activities in the Czech Republic (Rončák et al., 2021; Tittelbachová et al., 2022) and, the role of social media on visitors' behaviour (Javed et al., 2020). Kádár (2013) discusses tourist dispersion by comparing the cities of Vienna and Prague.

Warsaw, however, is overlooked by scholars, despite the acknowledge of authors as Poland's most visited and consolidated destination. Studies on Poland cities include Warsaw as one of the main cities in the sample, as the case of Widz et al. (2022) on COVID's impact on accommodation and Gonia and Jeziarska-Thöle's (2022) study on sustainable tourism in Polish cities. In addition to that, Solima and Izzo (2018) discuss the use of innovative interpretation alternatives in heritage sites in Warsaw and Naples (Italy). Krakow (instead of Warsaw) is the

subject of a few studies, such as the ones conducted by Kowalczyk-Anioł (2023) and Matoga and Pawłowska (2018), the first elaborating on the social impacts of urban tourism and the later focusing on off-the-beaten-track tourism and the over tourism in historic centres.

For the city of Vienna, Gunter and Önder (2021) develop an exploratory analysis of site popularity by comparing residents' and tourists' perceptions of the city of Vienna through social media posts. Moreover, studies measure the impact of Airbnb on accommodation (Gunter & Önder, 2018), innovations in Vienna's hospitality industry (Binder et al., 2016), and the marketing of Vienna as Austria's top destination (Popescu & Corbos, 2011).

Matzler et al. (2016) use Slovakia's branding identity to compare the impacts of cultural background from incoming tourists on destination choice and satisfaction. Specific tourism research on Slovakia and Bratislava is very limited.

Overall, there is a shortage of studies comparing the visitors' perceptions and attractions of Central European cities. Although the above-mentioned studies provide insights into the tourism development in the region, as well as shared patterns between Central and Eastern European destinations, there is a gap in recent studies evaluating and comparing the tourist attractiveness and tourist experiences in the region. Therefore, the present study presents a theoretical contribution to the field, aiming to provide a comparative study regarding six Central European capital cities.

2.3. The role of visitors' satisfaction in tourism loyalty

Several factors impact the visitors' level of satisfaction and happiness with their tourist experience. From the perspective of the overall travel experience in a given destination, different aspects of the visit must be combined to result in not only a positive but also a memorable experience. The memorability supports the connection and loyalty to the destination, therefore contributing to revisits or recommendations (Garner et al., 2022; Papadopoulou et al., 2023). The level of satisfaction with the visit results from a combination of the different phases of the travel experience. First, the time preceding the visit, in which research, flight and accommodation booking and visit planning contribute to building up a set of expectations on the destination. Second, during the trip, when the experience itself takes place. Third, the post-trip, when people return to their routine and reflect on their experience - that can be reflected in posting online reviews and feedback about the visit, posting photos on social media, and sharing stories with friends, family, and online channels (social media, forums, blogs) (Garner et al., 2022; Gursoy et al., 2022).

Ramseook-Munhurrin et al. (2015, p. 254) argue that “satisfaction is defined as customers’ judgments about products or service fulfilment”. For a wide range of products and services provided by the tourism industry, a set of objective criteria can be evaluated in order to perceive if the service was successfully fulfilled. For restaurants, for instance, there are aspects regarding the food itself (quality, quantity, price), service, and environment. Although the opinion on those can be subjective, the general goal when attending a restaurant is clear: the meal itself. For hotels, the same situation applies – even if customers have subjective opinions on the facilities, service, decoration and environment, the key product remains the overnight stay.

On the other hand, cultural and historic attractions, especially the ones that are public places, can receive visitors with a wide range of goals and expectations. Therefore, determining the elements that lead to tourist happiness poses a challenge. In general, the service encounter (interaction between the service provider and visitors) in these attractions and the experience relies on the attraction relevance itself - which, again, can be subjective to the visitor’s goal, such as aesthetic value, cultural and historical context, architecture, interpretation (informative signs) - and on the combination of a wider range of external factors, such as city infrastructure (cleanness, accessibility, connection to public transport), other visitors (tourists and locals), weather, and others. The encounter between visitors and hosts is mostly limited to the moment of purchase of goods or services, the interactions within the visits to the sites, and less formal interactions, such as the sharing of information, experience, and tips with each other (Gursoy et al., 2022; Marques et al., 2022; Papadopoulou et al., 2023).

Alternatively, due to the control over the entrance at paid cultural and historic attractions, visitors can be profiled and grouped according to their goals and the management can adapt aspects of the experience accordingly. Museums and castles, for instance, can improve and/or adjust interpretation (information displayed or presented at the site) to make it more appealing for a larger share of the public that visits the attraction. Moreover, social media and user-generated content can serve as a tool to get insights into the tourism experience. Other than providing a source of information for tourism service providers and Destination Management Organizations, the use of social media impacts the trip within its duration as visitors constantly turn to it to acquire information about other attractions and services and decide their next steps (Gursoy et al., 2022; Marques et al., 2022).

Notwithstanding, the visitors’ happiness can also be influenced by aspects beyond the control of tourism service providers and destination management. Social factors, such as with whom the trip is being shared (partner, friends, family, or solo trip), psychological aspects such as nostalgia feeling that could create a connection to the sights visited and even meteorological

conditions such as unsuitable weather (Almeida-Santana & Moreno-Gil, 2018; Gursoy et al., 2022). Studies also find that the cultural aspects and self-image of the visitor can influence the destination choice, as well as their perception of aspects of the tourism experience and services. Furthermore, a destination brand or marketing can determine brand identification and influence visitors' loyalty (Matzler et al., 2016; Bilro & Cunha, 2021).

In marketing studies, consumer loyalty most commonly translates to product repurchase. Although researchers in tourism often consider customer loyalty in the industry as revisit potential (i.e., returning to the same destination and attraction several times), factors such as time travel, high costs and the large variety of holiday destinations might limit this behaviour. Alternatively, the satisfaction and loyalty of visitors could additionally be translated into positive word of mouth recommendations, leading to the visit of new groups of people (Wu, 2016). The impact of loyalty impacts both the willingness to buy (or revisit, in tourism's case) and the willingness to pay for the service even if the price increases. From the DMO's point of view, revisits most likely mean higher-spending tourists. Also, according to Almeida-Santana and Moreno-Gil (2018), the cost of serving a re-visitor is inferior when compared to a first-time tourist.

On another perspective on the topic, Almeida-Santana and Moreno-Gil (2018) raise the possibility of marketing destinations vertically instead of horizontally – that is, while most DMOs focus on promoting one destination individually aiming to achieve tourist loyalty and revisits, people are likely to have a set of destinations they are fond of and willing to visit more than one time. In general, even when attending different cities or countries, the motivations for tourists to choose a given location are set. A tourist in search of a sand and beach experience, for instance, will most likely vary its visits within a range of destinations that include that feature – as well as other aspects of the experience that were positive. Therefore, marketing regions or a set of destinations with similar characteristics poses an opportunity to engage visitors' loyalty while they can experience a similar environment.

2.4. The use of user-generated content in decision-making processes

The evolution of the internet created a wide possibility for networks and user interaction through social media platforms, blogs, review sites, online communities, and others. The vast amount of data contained in these sources represents both a tool for analysis and satisfaction measurement and a tool for users to exchange data and their experiences – both positive and unpleasant (Kim et al., 2017). Due to the intangible nature of tourism and its services, people cannot “have a taste” or a sample of the services they hire before going on the trip itself, and

the feedback from previous travellers turns into an asset and a source of information to support decisions. The so-called word of mouth consists of the sharing of opinions and recommendations expressed by consumers that already experienced a given good or service. It is a valuable source that holds the potential of impacting choices or generating the desire to acquire goods, as well as influence visitors' behaviours (Jalilvand & Samiei, 2012; Li et al., 2021; Xu & Li, 2016).

Decades ago, word of mouth was restricted to relatives, friends, and other acquaintances. For further information, people relied on travel magazines, blogs, travel agencies and other materials sourced by tourism service providers (Jalilvand & Samiei, 2012). With the emergence of web-based interactions and the shift from service provider to consumer information flow to peer-to-peer generated content and interactions, the impact of word of mouth was maximized as the flows of data and opinions became faster, more easily accessible, and no longer restricted to one's social circles (Xu & Li, 2016).

In this context, users shifted from spectators of the online content to active participants, both reading and writing, collaborating with other users, and sharing information in several channels, such as social network platforms, blogs, communities, wikis, and other portals (Kim et al., 2017). Electronic word of mouth (eWOM) can be defined as "any positive or negative statement made by potential, actual or former consumers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Filiari et al., 2015, p. 175). It is considered one of the most reliable online sources of information, especially review-based websites such as TripAdvisor and other forums, as the content is out of the control of service providers and destination organizations. In that sense, they might present stronger contributions to destination image and branding than official sources, such as Destination Management Organizations (DMOs) (Jalilvand & Samiei, 2012; Költringer & Dickinger, 2015).

Writing and posting online reviews represent a way for people to revisit their positive travel experiences and prolong the satisfaction and positive memories they experienced. The motivations to engage in eWOM after consuming a tourism service are providing helpful information to others, venting negative experiences supporting local companies, and the enjoyable feeling of sharing one's perceptions. Moreover, research points out that posting reviews after going on holiday provides visitors with psychological satisfaction (Garner et al., 2022; Loureiro et al., 2019).

On the verge of making purchase decisions, word of mouth poses an important element. Researching online reviews on user-generated content websites, such as TripAdvisor and Yelp

allows visitors to adjust their expectations regarding the services being hired (Filiari et al., 2015). Other than impacting visitors' pre-perception of a tourism product, studies have shown that eWOM can directly influence destination choice and changes in demand (Filiari et al., 2015; Jalilvand & Samiei, 2012). By considering the consumer purchase decision in its five steps (recognition of need, search for information, option evaluation, purchase decision and post-purchase behaviour), eWOM plays a role in providing data and experiences from other consumers in addition to contributing to the brand, service, or destination image (Binh et al., 2017).

As mentioned in the previous session, visitors are impacted and make use of social media and user-generated content during the whole journey of their trip (pre-trip, on-site, and post-trip). While it is an important factor in the pre-purchase decision-making process, once the tourists arrive at the destination, they continue resorting to online content to decide their next plans, where to eat and which places to visit, as well as share their experiences as they take place. Therefore, user-generated content on all kinds of attractions (free, paid, where a service was provided or where the visitor had an autonomous time) is created and can be used and analysed by tourism service providers (Bilro & Loureiro, 2023; Gursoy et al., 2022).

Nevertheless, from a tourism stakeholder perspective, several businesses can benefit from the analysis of the compiled user-generated data, including DMOs, hotels, restaurants, attractions and tour guide agencies and even other logistic-related planning services, such as public transport and traffic management (Költringer & Dickinger, 2015). The content in customer feedback and reviews is a step up in comparison to customer ratings, as they allow the identification of the sources of consumer satisfaction and dissatisfaction (Xu & Li, 2016). In this context, analyses concerning eWOM represent a potential tool to infer issues, strengths and good practices and points of concern. They pose an asset to justify decision-making processes for tourism and logistics policies as well as a tool to evaluate how organizations are performing and the appropriate reaction (Kim et al., 2017; Költringer & Dickinger, 2015).

Concerning DMOs, Költringer and Dickinger (2015) argue that the analysis of user-generated content (UGC) allows these organizations to identify topics on which travel blogs, tourist guides and visitors have their attention focused. In addition to that, interacting with potential consumers on these platforms represents a channel to engage with them - solving doubts, and providing advice or suggestions. As for interactions with local tourism industry stakeholders, DMOs can make use of these platforms to share information obtained from reviews and reports on the current tourism situation.

3. Current DMO initiatives for Central European Capital Cities

In this chapter, the main initiatives promoted by the cities' Destination Management Organizations are highlighted. It provides an additional comprehension of the cities' tourism goals and state in addition to the studies and information described in the literature review. To understand the tourist perceptions of the cities, as well as provide suggestions for tourism development and DMO policy making – listed as one of this study's objectives – and compare results within the cities in the sample, the analysis requires the combination of the existent literature review, the netnographic analysis, and the current initiatives foreseen by the Central European capital cities' tourism plans.

About 42.6% of the tourists that visited the Czech Republic in 2022 visited cultural monuments as part of their activities. In Germany, the main activity for domestic tourists relies on nature-based activities, followed by visiting cultural and historical attractions (60%). City breaks, sightseeing and museums and national parks are the tourist and leisure activities preferred by tourists when visiting Poland (Statista, 2023a, 2023d, 2023c). Cultural and city tourism is not a new subject in Central European capital cities. The segment is responsible for a significant share of the tourism demand in the countries and, therefore, should be addressed in its development plans. This chapter highlights the initiatives present in the DMO's tourism plans for cultural and heritage tourism and analysis of user-generated-content.

The *Sustainable and City-Compatible Berlin Tourism Plan 2018+* emphasises the aim to enhance tourism flows qualitatively – not quantitatively – aiming to make tourism more sustainable and beneficial for the local population. In the cultural sphere, Berlin's plan proposes the creation of cultural activities and services in decentralized areas of the city, as well as the intensification of local participation in the elaboration of the activities and policies implemented (Berlin Tourismus & Kongress, 2018).

Regarding the analysis of user-generated-content, the report states that the standard key performance indicators such as the number of arrivals are no longer sufficient to promote insights into the development of tourism policies. It states the need to evaluate “the degree of acceptance of tourism among Berlin's residents, their own evaluation of life quality, guest satisfaction, or the distribution of tourist activities across the city's districts, and the activity patterns of visitors” (Berlin Tourismus & Kongress, 2018), which could be accomplished through the extraction of user-generated-content from several sources.

According to the *Marketing and communications strategy for the destination Bratislava*, the city does not face the issues of over-tourism at this point. The report stresses the importance of user-generated content for destination branding and marketing, mentioning the important

role of personal recommendations, reviews, and stories in the decision-making process of deciding on a holiday destination. Within the targets for 2022 is the creation of a system to manage and share content created by visitors. Another point relates to the recognized potential of extracting and analysing data on visitors and their behaviour during the trip (Visit Bratislava, 2017). A more recent document, Bratislava 2030, was developed to outline the tourism goals and strategies for the years 2022-30 – however, there is no translation to English.

The most recent document published in English for Budapest’s planning, *Integrated Urban Development Strategy – Budapest 2020*, mentions the enhancement of the city’s capacity to receive tourists and of the cultural activities promoted beyond the heritage sites (Municipality of Budapest, 2015).

Prague’s Tourism Plan, entitled *Prague Destination Management: Putting Prague First - Strategy for Sustainable Tourism in Prague*, demonstrates concerns with the concentration of visitors in the historic centre, as well as the need to disperse tourism around the city. It also reiterates the importance of data collection on tourism activities and the need for data to evaluate the overall state of tourism in the city as well as support decision-making processes. One of the issues brought up regards “Insufficient data collection in the area of tourism and the related overall picture of tourist activities in the city” (Prague City Tourism, 2020). There is a lack of information on the insights coming from the analysis of online data, or the specific aimed applications.

Regarding cultural and heritage sites, the plan reiterates in its strategy session the importance of promoting access to culture and the protection of the Heritage Sites and Prague’s Historic Centre, and, with that, safeguarding its position on the UNESCO World Culture and Nature Heritage List. Finally, in the Vision and Goals session, there is a topic dedicated to research and data monitoring, which includes the need of developing surveys for the visitors and residents, the creation of a database for tourism data and “use all the available data and new technologies to direct tourist flows and assess tourism impacts” (Prague City Tourism, 2020).

Warsaw’s Tourism Plan is part of the *#Warsaw2030* initiative, created with the intent to elaborate cohesive urban policies that collaborate to a sustainable tourism approach in the city. Even though Warsaw has not yet reached the limit on the number of visitors and, therefore, is not facing the impact of mass tourism currently, initiatives to decentralize tourism activities are being elaborated. The key concept relies on directing and marketing specific routes and sights to different tourist segments – a task not so easy to accomplish in terms of global marketing. As stated, "Tourism can have a considerable impact on economic growth. However, this progress cannot happen at the expense of deterioration of quality of life for residents, especially,

if the trends of sharing economy and locality are influencing the preferences and patterns of tourism consumption.” (City of Warsaw, 2020).

Vienna’s Tourism Plan, *Reshaping Vienna: Vienna Visitor Economy Strategy 2025*, focuses on making the city a premium destination – several data sources and portals are mentioned to monitor the tourism flows in the city, benchmarking with other European capitals and residents’ perceptions of tourism. There are also initiatives to promote cultural events year-round to disperse the tourists throughout the year, therefore avoiding overcrowding tourist attractions in the peak months (Vienna Tourism Board, 2021).

Overall, the capital cities show concern with providing alternative cultural attractions, both to decentralize the concentration of tourists in historic centres and main attraction sites and to promote tourism throughout the year. The negative impacts of tourism on residents and the social conflicts between tourists and the local population were also frequent topics, and cities with more intense tourism flows, such as Berlin, Prague, and Vienna, include that as a main concern in their tourism plans, as well as rising the issue of sustainability in tourism initiatives. Specific actions or analyses including user-generated content are limited, although some of the plans mention the potential of online sources of information.

4. Text mining and sentiment analysis

The tourism industry relies heavily on the visitors' experiences – their perceptions influence the likelihood of revisiting or recommending an attraction or destination. Hence, analysing user-generated content, such as reviews and posts on social media, presents a significant opportunity to explore patterns regarding customers' perceptions and behaviour and identify aspects that require enhancements (Águeda et al., 2019). Tourism and leisure-related studies represent one of the main topics approached in netnography studies. The online content generated by users not only encompasses large amounts of data that can be translated into insights and used to improve services and policies but also represents a sphere where tourists express their opinions and experiences openly, without the influence of a third party or an inquisitor, as would be the case when conducting interviews or focused groups (Mkono & Markwell, 2014). Therefore, the more naturalistic nature of the methodology is pointed out as one of its advantages in comparison to other methods (Kozinets, 2002; Tavakoli & Wijesinghe, 2019).

Within the context of netnography studies, text mining analysis refers to a semiautomatic process to extract valuable patterns of large unstructured data sets. The unstructured aspect of the data is one of the aspects that differentiate text mining from data mining. The goal is to provide structured data that can be used for the development of analysis that generates insights to fix real-world challenges (Águeda et al., 2019). Due to the large amounts of user-generated content published online regarding tourism and travel services, the tourism industry can benefit highly from taking inputs from these large sets of data. Both DMOs and other stakeholders, such as hotel managers, restaurants, tourism operators, and transport operators (such as airlines, bus, and train companies) can use the analysis provided by these online reviewers to identify areas in which their service is lacking and increase quality and consumer satisfaction.

Text-mining analysis has been used largely in tourism studies. Hospitality and accommodation stand out in the searches – there are several studies focused on the visitor's perception of accommodation and the comparison of similar brands – such as luxury hotels – a few examples are Zhang et al. (2022), Handani et al. (2022), Xiang et al. (2015), Xu and Li (2016). The research on specific tourism products can also be mentioned, as in the study conducted by Barbierato et al. (2022) on wine tours. The analysis can also be an instrument to evaluate destinations, as in the case of Skotis and Livas (2022), focused on urban historic districts, Garner et al. (2022), Kim et al. (2017), and Bilro et al. (2019) on visitors' satisfaction, Költringer and Dickinger (2015) on destination branding and even demand forecast studies (Li et al., 2021).

Within the analysis of non-structured data, the element of Natural Language Processing (NLP) poses a fundamental step. It allows the comprehension and processing of the text beyond syntax, considering context and semantic and grammatical limitations, resembling closely to how humans communicate, as well as translating it into more formal representations (in the form of numeric and symbolic data) that are more easily manipulated by computer programmes (Águeda et al., 2019; Sharda et al., 2014).

For the last decade, the use of NLP tools to integrate sentiment analysis has been crescent. Sentiment analysis aims to attribute a polarity to a given emotional content and it is one of the main analyses conducted within the tourism-related studies above-mentioned. Sharda et al. (2014, p. 235) define sentiment analysis as “a technique used to detect favourable and unfavourable opinions toward specific products and services using a large number of textual data sources (customer feedback in the form of Web postings)”. The following session describes the netnography analysis conducted through the extraction of online reviews from different attractions of Central European capital cities. Furthermore, text-mining and sentiment analyses were conducted to understand the perceptions of tourists on the cities included in the study.

5. Data analysis and results

5.1. Data collection

For this work, reviews of the main attractions of the capital cities of Central European countries were selected: Berlin, Bratislava, Budapest, Prague, Vienna, and Warsaw. The data source is the website TripAdvisor, one of the main portals for tourism services reviews. As of January 2023, TripAdvisor ranked as the second most visited travel and tourism website with over 161 million visits, only behind booking.com. By the end of 2022, the number of ratings and reviews on TripAdvisor surpassed 1 billion (Statista, 2023e). The initial idea consisted of collecting 2022 reviews from the three attractions with the highest number of reviews in each city. Due to data limitations (lower number of observations), for two of the cities analysed – Bratislava and Warsaw – the data collection was expanded to the top attractions.

The following attractions were included in the sample:

- Berlin: Brandenburg Gate, the Reichstag Building (German Parliament), and the Holocaust Memorial
- Bratislava: Bratislava Old Town, Bratislava Castle, The Blue Church, Cumil's statue, Devin Castle, UFO Observation Deck, St. Michael's Tower, Michael's Gate, Bratislava's Main Square, and Slavin Memorial
- Budapest: Hungarian Parliament, St. Stephen's Basilica, and the Fisherman's Bastion
- Prague: Charles Bridge, Prague Castle, and Prague Old Town Square
- Vienna: Schönbrunn Palace, St. Stephen's Cathedral, and Belvedere Museum
- Warsaw: Old Town, The Royal Łazienki Park, Warsaw Rising Museum, Palace of Culture and Science, Museum of the History of Polish Jews, Warsaw's Old Town Market Place, Copernicus Science Centre, The Royal Castle in Warsaw, Castle Square, and Nowy Swiat

Regardless, the two cities (Bratislava and Warsaw) present fewer observations than the others, as presented further ahead. Once the source of the information was defined, as well as the scope of the data to be collected, the information was extracted using a web scraper tool, webscraper.io. Webscraper.io is a Chrome extension that allows the extraction of large amounts of information from websites – a sample of the sitemap used for the data extraction for each city can be found in Annex A.

5.2. Data description and limitations

In total, 2,260 reviews were considered in the analysis – more reviews were extracted, but only the ones referring to visits in 2022 were considered. The data extracted includes the start URL

(attraction URL), reviewer name, place of origin (if available), date of stay/visit, main quote (review title), long quote (full-text review) and bubble rate (from 1-5). Table 5.1 shows the key figures for the data extracted from each city.

Table 5.1 – Data description

City	Berlin	Bratislava	Budapest	Prague	Vienna	Warsaw
Number of attractions Selected	3	10	3	3	3	10
Number of reviews	401	195	518	458	506	182
Average ratings (1-5 bubbles)	4.55	4.31	4.59	4.51	4.47	4.40

Source: Author’s elaboration (2023)

Following the data extraction to Excel, the Excel addon “Meaning Cloud” application was used to proceed with the text analysis. In order to achieve that, a sentiment analysis was conducted. Within the sentiment analysis, Meaning Cloud attributes different sentiments (highly positive, positive, neutral, negative, highly negative or none) to the reviews within two different analyses. The first, General Sentiment Analysis, considers the review text in its entirety, attributing the overall sentiment to the comment. In the second one, Topic Sentiment Analysis, the Application breaks off parts of the text, investigating and attributing the sentiment polarity to parts of the review, as well as grouping similar topics within a categorization. As a method to analyse the results from a numeric and statistical perspective, the sentiment polarity was translated to numbers, as shown in Table 5.2.

Table 5.2 – Polarity conversion table

Polarity	Sentiment	Conversion
P+	Highly positive	2
P	Positive	1
NEU	Neutral	0
N	Negative	-1
N+	Highly negative	-2
None	None	None

Source: Author’s elaboration (2023)

As well as the sentiment polarity, in the Global Sentiment Analysis, the comments are also categorized according to three different metrics: irony (ironic or non-ironic), subjectivity (subjective or objective) and agreement (agreement or disagreement). The last one evaluated customers' feedback and management responses based on the similarity (or not) of the many issues' polarities inside a review, as each aspect had a corresponding sensation. If they were aligned, the comment was considered an agreement.

Regarding data limitations, as mentioned in the previous section, despite the higher number of selected attractions for the cities of Bratislava and Warsaw, the total number of comments

for the cities remains low compared to the other 4 (182 and 195 versus an average of 471). Moreover, factors that could lead to a biased analysis (such as demographic factors such as country of origin, age, and gender) are not accounted for since that data is not publicly available on the TripAdvisor platform. Therefore, there is no alternative to guarantee the sample reliability concerning the demographics of the total number of visitors to the attractions. Nonetheless, despite the TripAdvisor platform being one of the main user-generated review platforms for the tourism industry, its use among countries and nationalities. In January 2023, for instance, over 65% of the visits originated in the United States (Statista, 2023e).

5.3. Results

For the results, the following subsections refer to each city contained in the analysis. Although the reviews are focused on tourist attractions and not the overall perceptions of the cities, it is noticeable that other aspects of the tourism experience come up – such as hospitality and gastronomy, which include restaurants, local foods and the reviewer’s accommodation, city and country reflecting the overall destination and/or comparing it to alternative locations and the general perception of people (both locals and other tourists) in the tourism experience. The results of the General Sentiment Analysis and the Topic Sentiment analysis are presented in the following subsections.

5.3.1. Berlin

The Berlin results consist of the analysis of reviews from three tourist attractions: Brandenburg Gate, the Reichstag Building (German Parliament), and the Holocaust Memorial. A total of 401 online reviews were included in the analysis with an average rating (1 to 5 bubbles on TripAdvisor) of 4.55. From the reviews, 147 refer to Brandenburg Gate (average rating 4.59), 125 to the Reichstag Building (average rating 4.64) and 129 to the Holocaust Memorial (average rating 4.40). It is worth mentioning that due to the tragic nature of the Holocaust Memorial, we cannot rule out the hypothesis that it may lead to more negative perceptions from visitors. Although there is no method to verify this behaviour in comparison to the other cities, the opinion is based on the frequency and polarity of terms such as “history”, “holocaust”, “Second World War”, “jew”, and “victim” included in the Topic Sentiment Analysis.

In Berlin’s General Sentiment Analysis (Table 5.3), most of the reviews were considered positive (72.1%). It is worth mentioning that most of these positive reviews were evaluated with a positive sentiment (P, or 1), while a smaller share scored as highly positive (P+, or 2). 11.7% of the comments were considered neutral, and 12.5% registered a negative sentiment to their

polarity – 11.2% negative and 1.2% highly negative. Meaning Cloud was unable to attribute a polarity sentiment to only 3.7% of Berlin’s online reviews. The overall polarity for Berlin, disregarding the online reviews to which no sentiment was assigned (“none”), is 0.82, indicating a tourist perception in between a neutral and positive position.

Table 5.3 – General Sentiment Analysis: Berlin

Polarity	Scale	Value	%	P-N
P+	2	84	20.9%	72.1%
P	1	205	51.1%	
NEU	0	47	11.7%	11.7%
N	-1	45	11.2%	12.5%
N+	-2	5	1.2%	
None	None	15	3.7%	3.7%
Total	Total	401	100.0%	100.0%

Source: Author’s elaboration (2023)

As for the metrics of agreement, irony, and objectiveness, Table 5.4 shows the results. 64.6% of the comments demonstrated agreement (versus 35.4% of disagreement). Most of the comments were considered non-ironic (98.3%) and subjective (75.8%).

Table 5.4 – General Sentiment Analysis metrics: Berlin

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	259	64.6%	Ironic	7	1.7%	Objective	97	24.2%
Disagreement	142	35.4%	Non-ironic	394	98.3%	Subjective	304	75.8%

Source: Author’s elaboration (2023)

For the Topic Sentiment Analysis, the reviews were broken into 1,595 clusters, from which 647 had their polarity classified, while the remaining 948 (59.4%) had no polarity identified, which indicates a limitation to the analysis. From the clusters to which polarity was attributed, the most significant share was positive (29.5%), 1.5% was neutral and 9.5% reflected negative feelings. The Topic Sentiment Analysis follows the same pattern as the General Sentiment Analysis, in which most of the positive clusters are positive (P, or 1) while a smaller share is highly positive (P+, or 2) as well as the negative sentiments are concentrated in the less extreme polarity (more N than N+), as described in Table 5.5. The overall polarity of the Topic Sentiment Analysis (disregarding “none”) is 0.70, slightly lower than indicated by the General Sentiment Analysis and reinforcing the positioning of Berlin’s visitors of mildly positive perception of the city.

Table 5.5 – Topic Sentiment Analysis: Berlin

Polarity	Scale	Value	%	P-N
P+	2	155	9.7%	29.5%
P	1	316	19.8%	
NEU	0	24	1.5%	1.5%
N	-1	133	8.3%	9.5%
N+	-2	19	1.2%	
None	None	948	59.4%	59.4%
Total	Total	1595	100.0%	100.0%

Source: Author’s elaboration (2023)

Following the analysis of the polarity attributed to the Topic Sentiment Analysis, Table 5.6 describes the main topics contained by the clusters. Note that the topic classified by Meaning Cloud as “Top”, a general grouping that does not indicate the nature of the review, was not included. In addition to that, solely topics with 10 observations (clusters) or more were included, corresponding to a total of 74% of the observations to which a polarity sentiment was attributed. In total, 476 observations were included, with an average polarity number of 0.71.

Table 5.6 – Topic extraction (TSA analysis): Berlin

Topic	Observations	Polarity number	Standard deviation	Variance
Hospitality and Gastronomy	16	1.06	0.68	0.46
Tourist Attractions	153	0.90	1.08	1.16
City	37	0.86	1.03	1.06
Tourist Experience	48	0.67	1.10	1.21
Architecture and History	44	0.66	1.24	1.53
Locations	92	0.63	1.20	1.44
Organization and logistics	19	0.47	1.47	2.15
People	67	0.37	1.10	1.21
Total	476	0.71	1.13	1.29

Source: Author’s elaboration (2023)

As presented above, the Topic Sentiment Analysis resulted in eight main topics. For Berlin, the topic with the highest positive sentiment polarity attributed to it, with an average polarity of 1.06, was “Hospitality and Gastronomy”, which includes comments regarding hospitality services such as hotels, stores, restaurants as well as street markets (and Christmas markets, for instance). The topic with the second highest positivity, “Tourist Attractions”, refers to the attractions themselves. Subsequently, the topic “City”, refers to Berlin in general.

“Tourist Experience”, “Architecture and History” and “Locations” presented a similar polarity average (0.67, 0.66 and 0.63, respectively). The first concerns aspects of the visit experience, service encounters and perceptions of tourists on a service or attraction, as well as mentions to photograph places and social media, as in the comment “*Iconic monument in Berlin.*”

It is a must see. If you visit in the evening, there are less crowds and you have a much better chance of taking an unobstructed photo.”. The second, as it is named, includes review parts on the local architecture and aspects of the history related to Berlin, Germany and the sites included in the attractions selected. Finally, location refers to places that do not correspond to attractions themselves, such as squares, parks, streets, and bridges. They are broader and less specific than the topic “Attractions”.

Next, the topic of “Logistics and Organization”, raking closer to a neutral sentiment than to a positive one, with an average of 0.47. This topic was included exclusively in the Berlin analysis, and it reflects the organization and structure of the attractions. Finally, with the lower polarity (0.37), “People” includes part of the reviews focused on other tourists, locals and, in the case of Berlin, historical figures, such as soldiers, (holocaust and Second World War) victims and other personalities. The average polarity of all the topics combined corresponds to 0.71 – very similar to the combined results of the Topic Sentiment analysis. The result is close to a neutral sentiment of the tourists towards Berlin, although it is positively inclined.

5.3.2. Bratislava

Bratislava’s analysis consists of the compilation of 2022 reviews of ten distinct tourist attractions in the city: Bratislava Old Town, Bratislava Castle, The Blue Church, Cumil’s statue, Devin Castle, UFO Observation Deck, St. Michael’s Tower, Michael’s Gate, Bratislava’s Main Square, and Slavin Memorial. As previously mentioned, even with the expansion of the number of attractions, the total number of reviews extracted is still limited when compared to others. In total, 195 online reviews were included in the analysis, with an average TripAdvisor rating of 4.31. From the 195 reviews, 51 are from the Bratislava Castle (average rating 3.92), 37 from Bratislava Old Town (average rating 4.73), 26 from Devin Castle (average rating 4.54), 20 from Cumil’s statue (average rating 4.20), 17 from the UFO Observation Deck and The Blue Church each (average ratings of 4.47 and 4.27, respectively), 11 from the Slavin Memorial (average rating 4.36), 6 from the St. Michaels’s Tower and Bratislava’s Main Square each (average ratings of 4.17 and 4.33, respectively), and lastly, 4 reviews from Michael’s Gate (average rating of 4.00).

Table 5.7 – General Sentiment Analysis: Bratislava Table 5.7 shows the polarity results of the General Sentiment analysis conducted. Most of the reviews had a positive sentiment attributed to them (82.1%), from which most were positive (68.2%) while a smaller share was considered highly positive. The negative sentiment was attributed to 6.7% of the online reviews, although none was classified as highly negative. Nevertheless, 8.2% of the comments had a

neutral sentiment attributed to them, while the software was unable to classify the remaining 3.1%. The average sentiment attributed to the online reviews (excluding the unclassified ones) rates 0.92, indicating an overall positive sentiment of the visitors regarding Bratislava.

Table 5.7 – General Sentiment Analysis: Bratislava

Polarity	Scale	Value	%	P-N
P+	2	27	13.8%	82.1%
P	1	133	68.2%	
NEU	0	16	8.2%	8.2%
N	-1	13	6.7%	6.7%
N+	-2	0	0.0%	
None	None	6	3.1%	3.1%
Total	Total	195	100.0%	100.0%

Source: Author’s elaboration (2023)

Regarding the metrics contained in the General Sentiment analysis, 122 (62.6%) of the reviews displayed an agreement tone throughout the comment, and the largest part (98.5%) presented non-ironic wording. In addition to that, 84.1% of the opinions expressed were evaluated as subjective (versus 15.9% objective), as displayed in Table 5.8.

Table 5.8 – General Sentiment Analysis metrics: Bratislava

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	122	62.6%	Ironic	3	1.5%	Objective	31	15.9%
Disagreement	73	37.4%	Non-ironic	192	98.5%	Subjective	164	84.1%

Source: Author’s elaboration (2023)

Consecutive to the General Sentiment Analysis, Table 5.9 exhibits the results of the polarities resulting from the Topic Sentiment Analysis for Bratislava. In a similar behaviour to Berlin, more than half of the observations were not classified (54.5%). From the remaining 484, the largest part reflected a positive sentiment (36.6%), most of those positive (score 1) instead of highly positive (score 2). Following a similar pattern to the general analysis, the polarity N (score -1) concentrated the negative spectrum of the results, while highly negative and neutral held about 1% of the comments each. The average polarity for the classified items on Bratislava’s Topic Sentiment Analysis equals 0.83, inferior to the registered in the general one. The score indicates a neutral-to-positive feeling of the visitors toward the city and its attractions.

Table 5.9 – Topic Sentiment Analysis: Bratislava

Polarity	Scale	Items	%	P-N
P+	2	104	9.8%	36.6%
P	1	285	26.8%	
NEU	0	15	1.4%	1.4%
N	-1	68	6.4%	7.5%
N+	-2	12	1.1%	
None	None	580	54.5%	54.5%
Total	Total	1064	100.0%	100.0%

Source: Author’s elaboration (2023)

Lastly, the list of topics extracted from the Topic Sentiment analysis and their respective polarity scores are presented in Table 5.10. Seven topics were highlighted, all of which are common to the ones observed in the Berlin analysis. Under the same criteria, the topic classified as “Top” was not included, as well as topics with less than 10 registered observations. A total of 401 observations were included, 82.85% of all the items with a polarity attributed to them.

Table 5.10 – Topic extraction (TSA analysis): Bratislava

Topic	Observations	Polarity number	Standard deviation	Variance
City	78	1.10	0.89	0.79
Location	64	1.02	0.81	0.65
Tourist Attractions	156	0.83	0.98	0.97
Hospitality and Gastronomy	22	0.73	1.08	1.16
Tourist Experience	46	0.67	1.19	1.42
Architecture and History	15	0.67	0.90	0.81
People	20	0.35	1.23	1.50
Total	401	0.86	0.99	0.99

Source: Author’s elaboration (2023)

As portrayed in the table above, the topic with the most positive sentiments attached to it is “City”, referring to overall Bratislava and comparisons to other cities, with an average of 1.10. As an illustration of the views on the destination, the sample included comments as: *“Not too many tourists come to Bratislava. Maybe this a blessing for the city. The infrastructure might not be able to support too many people and the quality of life might deteriorate. It is a very beautiful city, serene, quiet, and a simple lifestyle. I bet the inhabitants are happier than neighbouring touristic cities.”*.

Note that, for the topic analysis, several topics can be highlighted in the same comment. The review *“Lots of cool old buildings and history, neat shops and trendy cafes and restaurants, but with a more relaxed and slow-paced atmosphere compared with the hectic environment of Vienna, Prague, Budapest or other big European cities.”*, for instance, is classified for its sentiment for each aspect or topic in question: buildings for “History and Architecture”, shops,

cafes, and restaurants for “Hospitality and Gastronomy”, atmosphere for “Tourist Experience”, Vienna, Prague, and Budapest for “Cities”. The subsequent “Location”, areas of the city that are not the attractions themselves, were perceived generally positively, with an average score of 1.02. The visitors’ perceptions of “Tourist Attractions”, the topic with the third highest polarity, scored 0.83, followed by “Hospitality and Gastronomy”.

Both “Tourist Experience” and “Architecture and History” registered an average polarity of 0.67, which indicates a more neutral view of those elements. “People”, the topic that includes opinions on other visitors and locals, presented the lowest polarity score, 0.35, indicating a neutral perception of the visitors. It is worth mentioning that, even though this does not mean an overall dissatisfaction with the experience, polarity scores closer to and above 1 indicate tourism happiness and satisfaction. Overall, the average polarity for the topics analysed was 0.86.

5.3.3. Budapest

The analysis of the city of Budapest results from the processing of online reviews from three tourist attractions: Hungarian Parliament, St. Stephen’s Basilica, and the Fisherman’s Bastion. A total of 518 online reviews from the year 2022 were included in the analysis, with an average TripAdvisor rating of 5.59. 271 reviews refer to the Hungarian Parliament (average rating 4.55), 147 to St. Stephen’s Basilica (average rating 4.53), and 100 to the Fisherman’s Bastion (average rating 4.78).

The General Sentiment Analysis results for Budapest exhibited in Table 5.11 demonstrate the generally positive experience of the reviewers during their visit. Most of the polarities attributed were positive (score 1, 56.9%) and highly positive (score 2, 31.3%) indicating the tourist satisfaction with their visit. 5.8% of the comments were read as neutral, while the smallest share received a negative or highly negative polarity (a total of 4.4% with scores of -1 and -2). Finally, the analyses were unable to attribute a polarity to 8 of the online reviews (1.5%). The average polarity contained in the General Sentiment Analysis of Budapest equals 1.16, indicating an overall satisfaction of the tourists that visit the attractions contained in the sample.

Table 5.11 – General Sentiment Analysis: Budapest

Polarity	Scale	Value	%	P-N
P+	2	162	31.3%	88.2%
P	1	295	56.9%	
NEU	0	30	5.8%	5.8%
N	-1	21	4.1%	4.4%
N+	-2	2	0.4%	
None	None	8	1.5%	1.5%
Total	Total	518	100.0%	100.0%

Source: Author's elaboration (2023)

Regarding the metrics of the General Sentiment Analysis (Table 5.12), over 70% of the reviews demonstrated agreement with their content. The largest part of the sample was perceived as written in a non-ironic form, and most of the comments indicated subjective (82%) opinions.

Table 5.12 – General Sentiment Analysis metrics: Budapest

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	367	70.8%	Ironic	5	1.0%	Objective	93	18.0%
Disagreement	151	29.2%	Non-ironic	513	99.0%	Subjective	425	82.0%

Source: Author's elaboration (2023)

Following, the Topic Sentiment Analysis was conducted. Following the behaviour observed by the other analyses, there was no polarity attributed to the largest share of the observations (53.6%), as displayed below in Table 5.13. From the 958 observations to which polarity was attributed, the largest share concentrated in the positive spectrum (38.8%, 23% being positive and 15.9% highly positive), while only 1.1% classified as neutral, and 3.8% reflected negative opinions, most of which were considered N (polarity -1). As noticed in the previous analysis, the Topic and the Global Sentiment Analysis polarities share a similar pattern. For Budapest in particular, the average polarity for the Topic and General analyses both equal 1.16, an indication of visitors' happiness with the destination.

Table 5.13 – Topic Sentiment Analysis: Budapest

Polarity	Scale	Value	%	P-N
P+	2	348	15.9%	38.8%
P	1	503	23.0%	
NEU	0	24	1.1%	1.1%
N	-1	76	3.5%	3.8%
N+	-2	7	0.3%	
None	None	1233	56.3%	56.3%
Total	Total	2191	100.0%	100.0%

Source: Author's elaboration (2023)

After the polarity scale measurement, the main topics covered in the analysis were clustered, as displayed in Table 5.14. The topics are common to the ones presented in the previous result sessions and, maintaining the criteria, only topics with over 10 observations were included, as well as observations with the topic labelled as “Top” were disregarded. Overall, 84% of the items with a designated polarity were included in the table below.

Table 5.14 – Topic extraction (TSA analysis): Budapest

Topic	Observations	Polarity number	Standard deviation	Variance
Location	136	1.28	0.77	0.59
Tourist Attractions	341	1.26	0.84	0.71
Architecture and History	42	1.19	0.97	0.94
City	60	1.17	0.67	0.45
People	41	1.02	0.88	0.77
Tourist Experience	165	0.97	0.93	0.86
Hospitality and Gastronomy	22	0.86	1.08	1.17
Total	807	1.17	0.86	0.74

Source: Author’s elaboration (2023)

Six out of the eight topics contained in Budapest’s results have an average polarity rating superior to 1. The first one in the ranking, “Location”, refers to areas of the city of Budapest that could not be classified as tourist attractions, as in the comment: *“Our hotel was literally steps away so we passed here several times a day. It is a massive building which you can climb to get views across the city. It was the location of one of the Christmas Markets and the ice rink when we visited which was lovely!”*.

Reviews regarding the “Tourist Attractions” topic also presented a very positive average polarity, 1.26, as well as “Architecture and History” (polarity 1.19). It is worth mentioning that these attributes, “Location”, “Tourist Attractions” and “Architecture and History” are very interconnected, especially when referring to cultural destinations. Therefore, a good perception of one of these elements can lead to the same opinion on the others – nevertheless, the combination of these aspects in a positive setting leads to a positive experience.

Perceptions on the city of Budapest, topic “City”, also presented a high polarity score, 1.17. Next, “People”, which includes both other visitors and locals, held a positive perception in its comments, a different behaviour than the one observed in Berlin and Bratislava, where the topic was classified with the lowest score. The visitors’ perceptions of the “Tourist Experience” came next – the topic includes tourist perceptions, opinions on good spots to take photos and share on social media, and the quality of services offered during the visit. Last, there is “Hospitality and Gastronomy”, with a polarity of 0.87. Although the topic ranks last on the list, the polarity

attributed to it still indicates a generally positive overview of the hospitality services it embraces (hotels, markets, stores, restaurants). As previously mentioned, the overall polarity for Budapest indicates a positive tourist experience – for the topics, the average polarity rate was 1.17, slightly higher than the averages in the General and Topic analysis.

5.3.4. *Prague*

Online reviews from three tourist attractions were included in the Prague analysis: Charles Bridge, Prague Castle, and Prague Old Town Square. In total, 458 reviews from the year 2022 were included, with an average TripAdvisor rating of 5.51. From the 458 comments, 205 refer to Charles Bridge (average rating of 4.61), 185 to Prague Castle (average rating of 4.31) and 68 to Prague Old Town Square (average rating of 4.73).

The results for the polarity scale of the General Sentiment Analysis for Prague are presented in Table 5.15. The outcome from the sentiment perceived in the reviews is mostly positive, 83.2%, from which 22.9% were highly positive. Only a minority of the comments rated as highly negative (0.7%), totalizing 7.2% of results that indicate a negative polarity, and, therefore, tourist dissatisfaction. Moreover, 5.5% of the comments were considered neutral, and the Meaning Cloud software was unable to determine the polarity sentiment of 19 reviews (4.1%). The average polarity for the comments on Prague’s attractions equals 1.03, indicating overall satisfaction with the destination.

Table 5.15 – General Sentiment Analysis: Prague

Polarity	Scale	Value	%	P-N
P+	2	105	22.9%	83.2%
P	1	276	60.3%	
NEU	0	25	5.5%	5.5%
N	-1	30	6.6%	7.2%
N+	-2	3	0.7%	
None	None	19	4.1%	4.1%
Total	Total	458	100.0%	100.0%

Source: Author’s elaboration (2023)

The metrics for Prague’s General Sentiment Analysis are displayed in Table 5.16. About two-thirds of the comments presented agreement within its content. Most of the speech was classified as non-ironic (97.6%). In addition to that, 75.8% of the comments contained subjective opinions, while 24.2% were objective.

Table 5.16 – General Sentiment Analysis metrics: Prague

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	304	66.4%	Ironic	11	2.4%	Objective	111	24.2%
Disagreement	154	33.6%	Non-ironic	447	97.6%	Subjective	347	75.8%

Source: Author’s elaboration (2023)

Henceforth, the Topic Sentiment Analysis was conducted, as presented in Table 5.17. Similar to the others, half of the sample did not get any polarity attributed to it (51.8%), which could indicate a limitation to the results. For the remaining share (1160 observations), positive sentiments were the most prominent, with 12.7% considered highly positive and 27.9% positive. Neutral opinions accounted for 1.5% of the results, and the remaining 6.1% concentrated in the negative part of the spectrum (most of them negative, score -1). The overall polarity for the observations that had their polarity classified corresponds to 0.97, suggesting an overall satisfaction of tourists with the destination.

Table 5.17 – Topic Sentiment Analysis: Prague

Polarity	Scale	Value	%	P-N
P+	2	306	12.7%	40.6%
P	1	670	27.9%	
NEU	0	37	1.5%	1.5%
N	-1	133	5.5%	6.1%
N+	-2	14	0.6%	
None	None	1245	51.8%	51.8%
Total	Total	2405	100.0%	100.0%

Source: Author’s elaboration (2023)

As the final step of the Prague analysis, the main topics highlighted by the Topic Sentiment Analysis were clustered and presented in Table 5.18. As stressed in the previous subsections, only topics with 10 observations were more were included and items classified as “Top” were removed, encompassing 86.9% of the items to which a sentiment polarity was attributed.

Table 5.18 – Topic extraction (TSA analysis): Prague

Topic	Observations	Polarity number	Standard deviation	Variance
Architecture and History	37	1.19	0.91	0.82
Location	248	1.12	0.80	0.64
City	77	1.06	0.78	0.61
Tourist Attractions	449	0.98	0.93	0.87
Hospitality and Gastronomy	30	0.93	1.11	1.24
Tourist Experience	72	0.79	0.95	0.90
People	95	0.66	1.02	1.03
Total	1008	0.99	0.91	0.84

Source: Author’s elaboration (2023)

The elements “Architecture and History”, “Location” and “City” stand out in Prague’s topic analysis, with average polarities of 1.19, 1.12 and 1.06, respectively. The results indicate a high tourist satisfaction with the destination, especially in terms of the city aesthetics and environment, as these topics combined account for physical aspects of Prague in a broader context than solely its tourist attractions.

Still, “Tourist Attractions” is the subsequent topic, presenting a positive average polarity of 0.98. In addition to that, “Hospitality and Gastronomy” also performs well, with a score of 0.93, in line with the image of Prague for its beer heritage, as well as the famous Christmas markets, as expressed in one of the comments “*We were on a vacation to visits some of the best Christmas markets in Europe and Prague was our last stop. The best market is in the Old Town Square. Having visited many historic old towns, I would have to say that the Stare Mesto has to be the best we've seen. In addition to the Christmas market, there were plenty of other things to see and do. Old churches, the astronomical clock tower, horse and buggy rides. It seemed like this fantastic old town was right out of a fairy tale!*”.

The topics with the lowest average polarities, “Tourist Experience” and “People”, signal a not-so-positive overall experience in these aspects – although the ratings of 0.79 and 0.66 point towards a positive view, the elements do not reach the same levels as other aspects of the tourist experience previously highlighted. Nonetheless, the average polarity for Prague’s Topic cluster analysis, 0.99, demonstrates a positive outcome of the visitors’ perceptions of the destination and its attractions.

5.3.5. Vienna

The Vienna analysis results from online reviews extracted from three tourist attractions: Schönbrunn Palace, St. Stephen’s Cathedral, and Belvedere Museum. In total, 506 reviews were retrieved, presenting an average TripAdvisor rating of 4.47. Schönbrunn Palace is the source of 257 comments, with an average rating of 4.42, followed by St. Stephen’s Cathedral with 165 reviews and a 4.58 score, and Belvedere Museum with the remaining 84 online posts and a TripAdvisor rate of 4.39.

Once the data was extracted, a General Sentiment Analysis was conducted. Results for Vienna, presented in Table 5.1, demonstrate that most of the reviews expressed a positive sentiment (86.2%), 20.2% highly positive and 66.0% positive. The neutral comments accounted for 6.9% of the results, while the smallest share of the comments to which polarity was attributed was found negative (5.1%), most of those negative (4.5%) instead of highly negative (0.6%). At last, a small number of observations in the sample (9 comments, 1.8%) did not have

any polarity assigned to them. Overall, the average polarity for Vienna’s attractions equals 1.02, reflecting a favourable opinion from its visitors.

Table 5.19 – General Sentiment Analysis: Vienna

Polarity	Scale	Value	%	P-N
P+	2	102	20.2%	86.2%
P	1	334	66.0%	
NEU	0	35	6.9%	6.9%
N	-1	23	4.5%	5.1%
N+	-2	3	0.6%	
None	None	9	1.8%	1.8%
Total	Total	506	100.0%	100.0%

Source: Author’s elaboration (2023)

In continuance, the metrics for the General Sentiment Analysis are displayed in Table 5.20. Most of the comments were considered non-ironic (98.0%) and subjective (84.2%). As for the agreement aspect, 41.1% of the reviews were considered to have different points of view or polarity in their content, while 58.9% presented a more homogeneous perception.

Table 5.20 – General Sentiment Analysis metrics: Vienna

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	298	58.9%	Ironic	10	2.0%	Objective	80	15.8%
Disagreement	208	41.1%	Non-ironic	496	98.0%	Subjective	426	84.2%

Source: Author’s elaboration (2023)

Hereupon, the Topic Sentiment Analysis was performed. More than half of the observations broken down by Meaning Cloud could not be classified with any polarity (53.2%). From the share that received a polarity sentiment, most observations concentrated in the positive and highly positive scale (39.5%), while 1.0% was considered neutral and the remaining 6.3% negative, in a similar pattern to the General Sentiment Analysis. The average polarity for the Topic Sentiment Analysis sums up to 0.94, lower than the General context, although still indicating a positive outcome in the tourists’ visits.

Table 5.21– Topic Sentiment Analysis: Vienna

Polarity	Scale	Value	%	P-N
P+	2	289	11.6%	39.5%
P	1	695	27.9%	
NEU	0	25	1.0%	1.0%
N	-1	140	5.6%	6.3%
N+	-2	18	0.7%	
None	None	1326	53.2%	53.2%
Total	Total	2493	100.0%	100.0%

Source: Author’s elaboration (2023)

The main topics extracted, and their respective polarities are displayed in Table 5.22. As previously mentioned, observations labelled as “Top” and topics with less than 10 observations with a polarity assigned to them were not included in the analysis. Vienna presents one topic that distinguishes itself from the other Central European capitals, “Weather”. From all the observations with a polarity designated, 82.5% were included in the clusters.

Table 5.22 – Topic extraction (TSA analysis): Vienna

Topic	Observations	Polarity number	Standard deviation	Variance
Architecture and History	41	1.27	0.71	0.50
Location	144	1.15	0.78	0.61
Tourist Attractions	461	0.98	0.91	0.83
Tourist Experience	90	0.87	1.05	1.11
Hospitality and Gastronomy	24	0.75	1.11	1.24
People	130	0.75	1.01	1.03
City	59	0.75	0.96	0.92
Weather	14	0.00	1.30	1.69
Total	963	0.94	0.94	0.89

Source: Author’s elaboration (2023)

The topic with the highest average polarity from the eight was “Architecture and History”, with an average polarity of 1.27 – reflecting the tourist experience on the numerous historical buildings and architecture that characterize the city of Vienna. The second topic to come up, “Location”, refers to broader locations.

Following, the sentiment on “Tourist Attractions” was also found positive by visitors with an average polarity of 0.98. Mentions to specific works of art exposed in the museums (such as The Kiss, by Gustav Klimt, which is a permanent part of the collection of Schönbrunn Palace) were considered within the topic “Tourist Attractions”, as they are part of the attractions themselves, as in the comment “*Glorious buildings and gardens. An exceptional way to spend an afternoon. Of course the Klimt’s are a highlight but definitely not the only reason to visit.*”.

Aspects of the “Tourist Experience”, such as services, infrastructure and visitors’ perceptions presented an average polarity sentiment of 0.87. “Hospitality”, “People” and “City” appear next on the list, with a score of 0.75 each. Some of the negative comments on hospitality include: “*The collection is beautiful. The museum is not too big. It was a little bit crowded, but it was Sunday... On the other hand, the in house restaurant/coffee was disappointing. Mediocre food.*”. In Vienna’s analysis, “People” includes not only visitors and locals but also personalities, and artists' names, similar to the case of Berlin.

Lastly, the topic “Weather” registered an average polarity of 0.00, a performance below the levels of other aspects of the tourism experience in Vienna. It is worth mentioning that the low

number of observations (14) can be a limiting aspect to make inferences about the bad performance of this element. Nonetheless, the weather conditions are an external factor that cannot be controlled by tourism managers, although the facilities can be adapted to provide shelter and an adequate tourist experience. Some aspects of the experience, however, might depend on the seasonality, as expressed in the reviews: “*Saw and enjoyed the upper palace and its museum - the work of Klimt and other artists, most of them Austrian. In winter there is no use walking through the garden.*” versus “*The beautiful place to explore with a family and friends. I fell in love in lovely gardens. Summer is the best time to visit.*”. The overall average for the clusters was 0.94, indicating that tourists were mostly satisfied with their visits to Vienna’s attractions.

5.3.6. *Warsaw*

Warsaw’s results are a combination of the reviews of ten distinct sights: Old Town, The Royal Łazienki Park, Warsaw Rising Museum, Palace of Culture and Science, Museum of the History of Polish Jews, Warsaw’s Old Town Market Place, Copernicus Science Centre, The Royal Castle in Warsaw, Castle Square, and Nowy Swiat. As in the case of Bratislava, the decision to expand the number of attractions relied on the goal to reach a similar number of online reviews to the other capital cities. Despite that, the attractions combined summed up to 182 reviews, below the expected, which could pose a limitation to the analysis, as previously mentioned.

The average TripAdvisor score for the comments extracted equals 4.40. Of the 182 items, 57 refer to Warsaw’s Old Town (average rate 4.74), 32 to the Rising Museum (average rate 4.47), 19 to The Royal Łazienki Park (average rate 4.68), 17 each to the Palace of Culture and Science and Museum of the History of Polish Jews (average rate 3.94 and 4.06, respectively), 15 to Copernicus Science Centre (3.27), 8 to Warsaw’s Old Town Market Place (average rate 4.75), 7 to The Royal Castle in Warsaw (average rate 4.14), 6 to Castle Square (average rate 4.67), and 4 to Nowy Swiat (average rate 4.5).

The polarity summary for Warsaw’s General Sentiment Analysis, displayed in Table 5.23, shows the prominence of positive sentiments on the reviews (82.4%), most of them with a score of 1 (62.6%). The second largest portion refers to neutral comments, and 6.6% had a negative sentiment attributed to it. The remaining 1.6% includes comments to which the software was unable to attribute a sentiment. The average polarity for the General Sentiment Analysis equals 0.96, pointing out an overall satisfied perception expressed in the comments.

Table 5.23 – General Sentiment Analysis: Warsaw

Polarity	Scale	Value	%	P-N
P+	2	36	19.8%	82.4%
P	1	114	62.6%	
NEU	0	17	9.3%	9.3%
N	-1	9	4.9%	6.6%
N+	-2	3	1.6%	
None	None	3	1.6%	1.6%
Total	Total	182	100.0%	100.0%

Source: Author’s elaboration (2023)

As for the General Sentiment Analysis metrics, 57.7% of the comments registered agreement within its content. Most of the text was labelled as non-ironic (98.9%) and 84.6% expressed a subjective idea.

Table 5.24 – General Sentiment Analysis metrics: Warsaw

Metric	Value	%	Metric	Value	%	Metric	Value	%
Agreement	105	57.7%	Ironic	2	1.1%	Objective	28	15.4%
Disagreement	77	42.3%	Non-ironic	180	98.9%	Subjective	154	84.6%

Source: Author’s elaboration (2023)

In segment to that, the Topic Sentiment Analysis was conducted, and the summary of the polarity results is presented in Table 5.25. Almost half of the observations did not have any polarity attributed to them (49.0%), which could be a limiting factor to the analysis. From the remaining 51%, 40.3% were classified as positive (28.0%) and highly positive (12.3%) and 9.6% expressed negative sentiment. The remaining 1.2% stated neutral opinions. The average polarity for the items classified equals 0.83. Although that polarity indicates a neutral-to-positive outcome, it signs out that elements of the attractions in the city could be performing better.

Table 5.25 – Topic Sentiment Analysis: Warsaw

Polarity	Scale	Value	%	P-N
P+	2	127	12.3%	40.3%
P	1	289	28.0%	
NEU	0	12	1.2%	1.2%
N	-1	90	8.7%	9.6%
N+	-2	9	0.9%	
None	None	506	49.0%	49.0%
Total	Total	1033	100.0%	100.0%

Source: Author’s elaboration (2023)

Following that, the main topics were clustered, and the main elements of the tourist opinions are presented in Table 5.26. From the 527 observations with a polarity classified in the previous table, 81.0% were included in the clusters. As mentioned in the previous sessions,

topics labelled “Top” and the ones with less than 10 observations were disregarded. “Country” is the one topic that is exclusive to Warsaw’s analysis, referring to Poland.

Table 5.26 – Topic extraction (TSA analysis): Warsaw

Topic	Observations	Polarity number	Standard deviation	Variance
City	24	1.25	0.68	0.46
Location	66	1.18	0.70	0.49
Tourist Attractions	179	0.99	0.93	0.86
Hospitality and Gastronomy	30	0.97	0.76	0.59
Country	13	0.77	0.93	0.86
Tourist Experience	57	0.56	1.15	1.32
Architecture and History	33	0.09	1.23	1.52
People	25	0.08	1.32	1.74
Total	427	0.85	1.02	1.04

Source: Author’s elaboration (2023)

For the topic with the highest polarity averages, there is “City”, referring to the city of Warsaw, as in the comment: *“This is my fourth time in Poland and my second time in Warsaw, beautiful city, European city with a great history. Food is good.”* As mentioned before, in the topic analysis different aspects of the review are broken down – therefore, the comment could also refer to “Country” in its mention of Poland, “Architecture and History” for the great history clipping, and “Hospitality and Gastronomy” for its food part.

Following “City”, “Location” and “Tourist Attractions” appear, both with scores that indicate a positive overview (1.18 and 0.99, respectively). Henceforth, regarding restaurants, hotels, food, and stores, “Hospitality and Gastronomy” presented an average polarity of 0.97. As for the comments on “Country”, which refer to Poland, the average is 0.77, a more neutral-towards-positive score.

The topic “Tourist Experience” follows with an average polarity of 0.56, in which the services and tourist perceptions are the highlights. On a more controversial comment, there is the example: *“If you never been here but you're planning to I would recommend thinking twice. NOTHING special, boring gloomy streets, guides just intending to make profit of you exaggerating cost of viewing actually free historical spots. Restaurants seeing that you're foreigner intending to make the price higher. IF YOU'RE BORED OF Prague, better return there, it would be more worth it.”*

Lastly, “Architecture and History” and “People” present average polarities very close to neutral (0.09 and 0.08, respectively), indicating that there is potential for enhancing the tourist experience regarding these aspects. It is worth stressing that, for the topic “Architecture and History”, comments reminiscing the horrors of World War II were given a negative polarity –

this, however, does not translate to a bad performance of Warsaw's attractions. Overall, the average polarity for the clusters was 0.85 – as mentioned, although it points out to a neutral-positive tourist experience, the score under one indicates space for enhancement in the tourist experience, as demonstrated by the results of other cities included in the sample.

5.3.7. Results summary

The key results and metrics of all the analyses were compiled and presented in Table 5.27. The results on polarity sentiment and TripAdvisor ratings above the average of the sample were highlighted. The cities of Berlin, Budapest and Prague presented the highest average rating on TripAdvisor for the reviews extracted and included in the analyses – it is worth noting that all the cities presented an overall high rating. The ranking between the cities' TripAdvisor scores, however, was not replicated for the analysis. Both in the General and Topic Sentiment Analysis, Budapest, Prague, and Vienna presented higher polarities, superior to the sample average, indicating a more positive global tourist experience in the three destinations.

As for the clusters in the Topic Sentiment Analysis, Prague and Budapest stand out. Seven topics were common to all the cities: "Architecture and History", "Hospitality and Gastronomy", "Location", "Tourist Attractions", "City", "People", and "Tourist Experience". The perceptions on Budapest's attractions showed a sentiment above average for six of the seven elements of the tourist experience, while Prague's registered five. Following that, Vienna's average polarity was very similar to the one presented in the sample (0.94 vs 0.95) and the overall scores were positive, except for "Weather". Warsaw differentiates itself for its low polarity on "People" and "History and Architecture". Finally, the cities of Berlin and Bratislava present averages below the sample's, with only one element of the clustered topics above the medium each ("Hospitality and Gastronomy" and "City", respectively).

Focusing on the average polarities for the topics, "Location", "City" and "Tourist Attractions" displayed an average above 1, indicating tourist satisfaction in those regards. On the lower part of the scale, there are "People" and "Tourist Experience" (0.62 and 0.82, respectively), as well as the topics presented by only one city, "Organization and Logistics", "Country" and "Weather".

Table 5.27 – Results Summary (Part I)

Analysis	Indicator	Berlin	Bratislava	Budapest	Prague	Vienna	Warsaw	Average
General Analysis	Number of selected attractions	3	10	3	3	3	10	5
	Number of reviews analysed	401	195	518	458	506	182	377
	Average rating on TripAdvisor	4.55	4.31	4.59	4.51	4.47	4.40	4.50
Global Sentiment Analysis	Agreement	64.6%	62.6%	70.8%	66.4%	58.9%	57.7%	64.4%
	Disagreement	35.4%	37.4%	29.2%	33.6%	41.1%	42.3%	35.6%
	Ironic	1.7%	1.5%	1.0%	2.4%	2.0%	1.1%	1.7%
	Non-ironic	98.3%	98.5%	99.0%	97.6%	98.0%	98.9%	98.3%
	Objective	24.2%	15.9%	18.0%	24.2%	15.8%	15.4%	19.5%
	Subjective	75.8%	84.1%	82.0%	75.8%	84.2%	84.6%	80.5%
	Polarity Average	0.82	0.92	1.16	1.03	1.02	0.96	1.01
	Polarity 2 (P+)	20.9%	13.8%	31.3%	22.9%	20.2%	19.8%	22.8%
	Polarity 1 (P)	51.1%	68.2%	56.9%	60.3%	66.0%	62.6%	60.0%
	Polarity 0 (NEU)	11.7%	8.2%	5.8%	5.5%	6.9%	9.3%	7.5%
	Polarity -1 (N)	11.2%	6.7%	4.1%	6.6%	4.5%	4.9%	6.2%
	Polarity -2 (N+)	1.2%	0.0%	0.4%	0.7%	0.6%	1.6%	0.7%
No polarity (None)	3.7%	3.1%	1.5%	4.1%	1.8%	1.6%	2.7%	

Source: Author's elaboration (2023)

Table 5.27 – Results Summary (Part II)

Analysis	Indicator	Berlin	Bratislava	Budapest	Prague	Vienna	Warsaw	Average	
Topic Sentiment Analysis: Overview	Number of topics	1595	1064	2191	2405	2493	1033	1797	
	Polarity Average	0.70	0.83	1.16	0.97	0.94	0.83	0.93	
	Polarity 2 (P+)	9.7%	9.8%	15.9%	12.7%	11.6%	12.3%	12.3%	
	Polarity 1 (P)	19.8%	26.8%	23.0%	27.9%	27.9%	28.0%	25.6%	
	Polarity 0 (NEU)	1.5%	1.4%	1.1%	1.5%	1.0%	1.2%	1.3%	
	Polarity -1 (N)	8.3%	6.4%	3.5%	5.5%	5.6%	8.7%	5.9%	
	Polarity -2 (N+)	1.2%	1.1%	0.3%	0.6%	0.7%	0.9%	0.7%	
	No polarity (None)	59.4%	54.5%	56.3%	51.8%	53.2%	49.0%	54.2%	
Topic Sentiment Analysis: Clusters (average polarity)	Number of topics in the clusters	476	401	807	1008	963	427	680	
	Polarity average among clusters	0.71	0.86	1.17	0.99	0.94	0.85	0.95	
	Architecture and History	0.66	0.67	1.19	1.19	1.27	0.09	0.89	
	Hospitality and Gastronomy	1.06	0.73	0.86	0.93	0.75	0.97	0.88	
	Location	0.63	1.02	1.28	1.12	1.15	1.18	1.09	
	Tourist Attractions	0.90	0.83	1.26	0.98	0.98	0.99	1.02	
	City	0.86	1.10	1.17	1.06	0.75	1.25	1.03	
	People	0.37	0.35	1.02	0.66	0.75	0.08	0.62	
	Tourist Experience	0.67	0.67	0.97	0.79	0.87	0.56	0.82	
	Organization and logistics	0.47	-	-	-	-	-	-	0.47
	Weather	-	-	-	-	0.00	-	-	0.00
	Country	-	-	-	-	-	0.77	-	0.77

Source: Author's elaboration (2023)

6. Discussion

European capitals have their tourist activities closely linked to cultural tourism, especially in the context of Central Europe. Several issues have been rising over the last few decades regarding intense tourism flows in historic centres and cultural attractions, such as overcrowding, insufficient infrastructure and loss of authenticity. In order to keep track of that and evaluate visitors' and residents' perceptions of tourist activities and sites, user-generated content poses a valuable tool. Within the several available sources (blogs, social media, review websites), TripAdvisor has established itself as the most relevant channel for online reviews and research on tourism information in most countries. Although planning and policy-making organizations are aware of the large availability of user-generated content and its potential as a source of information, the analyses conducted on the data are still incipient. In that regard, this work poses an example of the nature of studies that can be conducted using a netnography and text mining approach.

Regarding the research question presented for this thesis, "What are the visitors' perceptions about the main attractions of Central Europe's capital cities and which aspects of the tourist experience have the best perceptions among tourists?", the literature review shows a lack of insight on the tourists' perceptions of the region. The tourism development efforts have been focused on modernization and strengthening ties with Western Europe whilst exploring the potential attractiveness of former-soviet history and monuments. Beyond published studies, DMO's tourism plans demonstrate concerns with overflows of tourism as well as a lack of use of the available data provided by user-generated content.

In order to properly address the research question, three main research objectives were set for this study. Regarding the first, which refers to identifying the visitors' perceptions of Central European capital cities through online comments, the results show an overall positive perception of visitors when visiting the Central European cities included in the sample. The literature regarding tourism in Central Europe is mostly focused on the branding of the destination and modernization process in the post-soviet context. Addressing perceptions from the point of view of eWOM and user-generated content is a novelty in the region, both for research and for managing organizations (DMOs). In consonance with the proposed by Naumov and Weidenfeld (2019) regarding the potential of Soviet architecture and historical buildings and monuments as tourist potentials, the analysis shows the positive perceptions of the elements "City", "Locations", and "Tourist Attractions". Those serve as an indicator of the tourist potential of the cities, as well as the attractiveness of its sites, buildings, and monuments. The inference is highlighted by the best-performing cities (Budapest, Prague, and Vienna)

favourable perception of “History and Architecture”, which adds up to the existent physical and intrinsic attributes of historical and cultural sites of the destinations. Gursøy et al. (2022) and Wu (2016) reinforce the relevance of the tourism experience, as in satisfaction with the provided services and amenities to tourism loyalty – which could translate to revisits or recommendations of the destination and its attractions.

The overall selection of attractions, given the number of TripAdvisor reviews as criteria, confirms the nature of the activities with higher visitor engagement, namely cultural activities, adding up to the proposed by Matoga and Pawłowska (2018) and García-Hernández et al. (2017) regarding the nature of tourism in historical European cities. In addition to that, since the attractions included in the samples refer to the most popular on TripAdvisor (and, therefore, among the most visited), the impacts of over-tourism, could demonstrate negative sentiment results. Nonetheless, the topic “People” performed badly in several cities, indicating either a not-so-positive perception towards other visitors or the local population, which could be a result of the excess in tourism flows.

The second research objective relates to the similarities and differences in the visitor experience between the cities. From the six Central European capitals included in the sample Budapest, Prague, and Vienna presented the best results for both the General and the Topic Sentiment analysis, and the higher average polarities for most of the topics, demonstrating the superior level of infrastructure and tourism planning of the cities, with a highlight to the topics “Architecture and History”, “Location”, and “Tourist Attractions”. Berlin, despite its relevance as a tourist destination, showed the weakest performance in the cluster Topic Sentiment Analysis – as mentioned in chapters 2 and 3, the city established itself as a hub for creative visitors and festival tourism. Therefore, analysing it from the spectrum of cultural destinations, especially when including monuments connected to the Holocaust, may not highlight the city’s overall touristic potential. Still, the results indicate that investments in cultural tourism and attention to the visitor’s experience in that segment must be taken into account and considered more carefully in Berlin’s tourism planning, as the city holds a significant amount of cultural and heritage attractions to explore.

As for providing insights for tourism planning decisions to promote regional tourism development, the third and last objective, this research stresses the potential of analysing user-generated content. As stated by Költringer and Dickinger (2015), user-generated content provides valuable information, allowing service providers and DMOs to identify the aspects in which the visitor experience is lacking and address those elements. As stated in the methodology session, text mining has been used for several tourism industry segments to

determine consumer preferences and satisfaction. The literature including text-mining and destination analysis and comparisons is limited but poses a tool for DMOs to focus on their development plans. Through this study, key elements of the tourist experience in Central European cities can be highlighted.

One of the topics raised both in the literature regarding cultural tourism and in the DMOs' tourism planning documents is the issue of over-tourism. As previously mentioned, that is confirmed by the low polarity of the topic "People". The results for the "Tourism Experience" topic, although not negative, present some of the lowest average polarities throughout the sample. Therefore, the potential regarding the visit memorability and its impacts might be underexploited. The satisfaction and memorability of a tourist destination also impact the likelihood of the visitors sharing their positive experiences online. Hence, DMOs should attempt to collaborate with tourism service providers and develop tourism policies that address the aspects of the tourism experience that need enhancement. As mentioned in Bratislava's marketing and communications strategy, user-generated content has a strong potential to influence purchase decisions and can be utilized as a tool for marketing and promotion strategies. The topic "Hospitality and Gastronomy" does not follow the same pattern as the others, standing out in the cities of Berlin, Warsaw, and Prague. The positive polarity for the topic on these cities indicates a favourable sentiment on the services provided by the hospitality industry, such as hotels, markets, and restaurants – a satisfaction indicator for DMOs that could be reported to local tourism stakeholders to reinforce best practices and strengthen the ties between policymakers and enterprises.

Thus, the results of the present study allow an overview of the visitors' perspectives of Central European capital cities, as well as highlight the positive aspects of their experience alongside identifying elements that should be addressed and enhanced by tourism stakeholders and destinations' DMOs.

7. Conclusions

7.1. Theoretical contributions

As mentioned in the methodology, several published studies make use of text mining and sentiment analysis to gather a better understanding of the aspects of tourism services and their performances. Most research, however, focuses on specific services or service providers, such as restaurants, hotels, and tours. The analysis of attractions as proxies for the destinations and the comparison within a region is less common. Therefore, this study provides three main contributions to the theory.

First, there is a lack of tourism studies focused on Central Europe and its tourism potential. Despite its tourism notoriety in the European context, most studies on Central Europe also include the Eastern part of the continent and provide generalizations that do not account for the particularities of the region or address specific cities. The present work fills a gap in the literature regarding tourism in Central Europe, as well as highlights strengths and aspects of the tourism experience that need to be attended to or enhanced by Destination Management Organizations.

Second, a comparative study among the destinations in CE provides insight into the tourism perceptions of the region and a comparison of several aspects of the tourist experience in different countries. Therefore, the study provides not only an understanding of the region but also the peculiarities and strong aspects of the tourism industry in each capital city.

Third, the use of netnography as a methodology, specifically text mining and sentiment analysis demonstrates the potential of online and user-generated content for tourism studies. The present study successfully demonstrates how text mining can be applied to tourism studies and to gather a more comprehensive understanding of tourists' perceptions of the attractions and the city.

7.2. Managerial contributions

From the perspective of Destination Management Organizations, this study presents several outcomes that demonstrate the potential of utilizing user-generated content to promote a better understanding of tourism services and experiences. In light of the goals proposed in most Central European capital cities' tourism plans, analysing user-generated content from attractions allows the organizations to have a better view of the tourists' perceptions of different areas of the city – therefore, contributing to the development of similar attractions in different neighbourhoods and the adaptation of elements in the surroundings – such as interpretation signs or tourism services. In addition to that, text analysis can assist in identifying patterns of

tourist behaviour, allowing DMOs to promote actions to shift tourism concentration and enhance the visitor experience.

In addition to that, DMOs can leverage the information on their destination and its competitors to market specific features or elements of the city. Alternatively, once the weak spots in the tourism experience are identified, DMOs can provide a combined market strategy with other destinations, either international or within the country, to encourage a more satisfactory and memorable experience.

7.3. Research limitations

This study presents some limitations regarding the sample, data sources and tourist focus. First, concerning the sample, not all Central European capital cities were included in the sample. Furthermore, the use of the main attractions as a proxy for the tourist perceptions of the cities might not illustrate all the aspects of the destination. In addition to that, the number of comments varies from attraction to attraction (and destination) and, therefore, the lower number of reviews for Bratislava and Warsaw can provide biased or less accurate results.

As for data sources, this research relies solely on TripAdvisor reviews. Despite the relevance of the platform, no verification or guarantee in place allows researchers to filter fake reviews or verify if the reviewers visited the site or engage in the services being rated. Moreover, several other sources of user-generated content could be included for a more reliant study, such as other review platforms (Yelp, Google) and social network platforms (Facebook, Instagram, Twitter).

Finally, this study is focused on the tourist perspectives and the implications of their experiences in tourism destination planning. Therefore, the experience of residents is not accounted for in the analysis, even though that is an important factor to be considered in local policies. Another aspect of the focus on the visitors is the lack of data on the demographics of the sample, hence, the impossibility to check for the sample representativeness and reliance. Although that is a common issue in online content analysis, it is an important aspect as it can impact the results.

7.4. Implications and future research

Given the limitations mentioned in the previous subsection, suggestions for future research include the expansion of the number of attractions and data sources, to provide a more comprehensive analysis of the tourist activities in the cities. Furthermore, as the aim of the study is to provide insights for DMOs, including the perceptions of residents in addition to the tourist

experience would provide a more complete understanding of the tourism industry needs and the more adequate policies to be implemented, benefiting both tourists and tourism stakeholders and the local population. Finally, given the fluidity of the definition of Central Europe, other capitals could be included in the analysis, as well as other cities in the countries.

Reference list

- Águeda, M., Rita, P., & Guerreiro, P. (2019). Sentiment analysis in online reviews classification using text mining techniques. *Iberian Conference on Information Systems and Technologies, CISTI, 2019-June*. <https://doi.org/10.23919/CISTI.2019.8760671>
- Almeida-Santana, A., & Moreno-Gil, S. (2018). Understanding tourism loyalty: Horizontal vs. destination loyalty. *Tourism Management, 65*, 245–255. <https://doi.org/10.1016/j.tourman.2017.10.011>
- Barbierato, E., Bernetti, I., & Capecchi, I. (2022). Analyzing TripAdvisor reviews of wine tours: an approach based on text mining and sentiment analysis. *International Journal of Wine Business Research, 34*(2). <https://doi.org/10.1108/IJWBR-04-2021-0025>
- Becker, E. (2018). Tour-guiding as a pious place-making practice: The case of the Sehitlik Mosque, Berlin. *Annals of Tourism Research, 73*. <https://doi.org/10.1016/j.annals.2018.09.009>
- Berlin Tourismus & Kongress. (2018). *Sustainable and City-Compatible Berlin Tourism Plan 2018+*.
- Bilro, R. G., & Cunha, J. F. da. (2021). An exploratory study of Western firms' failure in the Chinese market: a network theory perspective. *Journal of Chinese Economic and Foreign Trade Studies, 14*(2). <https://doi.org/10.1108/JCEFTS-07-2020-0033>
- Bilro, R. G., & Loureiro, S. M. C. (2023). I am feeling so good! Motivations for interacting in online brand communities. *Journal of Research in Interactive Marketing, 17*(1). <https://doi.org/10.1108/JRIM-07-2021-0182>
- Bilro, R. G., Loureiro, S. M. C., & Guerreiro, J. (2019). Exploring online customer engagement with hospitality products and its relationship with involvement, emotional states, experience and brand advocacy. *Journal of Hospitality Marketing and Management, 28*(2), 147–171. <https://doi.org/10.1080/19368623.2018.1506375>
- Binder, P., Mair, M., Stummer, K., & Kessler, A. (2016). Organizational Innovativeness and its Results. *Journal of Hospitality & Tourism Research, 40*(3). <https://doi.org/10.1177/1096348013496277>
- Binh, L. D., Vo, T. H. G., & Le, K. H. (2017). The impact of electronic word of mouth on brand image and buying decision: An empirical study in Vietnam tourism. *International Journal of Research Studies in Management, 6*(1). <https://doi.org/10.5861/ijrsm.2017.1738>

- Bonilla-Quijada, M., Tugores-Ques, J., & Olmo Arriaga, J. L. (2021). Promotion of urban tourism: insights into user engagement on social media. *Information Technology & Tourism*, 23(4). <https://doi.org/10.1007/s40558-021-00213-6>
- Borodako, K., & Rudnicki, M. (2014). Transport accessibility in business travel - a case study of central and East European Cities. *International Journal of Tourism Research*, 16(2), 137–145. <https://doi.org/10.1002/jtr.1908>
- City of Warsaw. (2020). *Tourism Policy of the City of Warsaw*.
- Dumbrovská, V., & Fialová, D. (2020). The city of one hundred souvenir shops: authenticity of souvenirs in Prague. *Journal of Tourism and Cultural Change*, 18(2). <https://doi.org/10.1080/14766825.2019.1606228>
- Engelbart, R., & Krech, V. (2016). The Berlin Site of Luisenstadt and the Figure of St Michael. A Case Study on Iconic Space. *Material Religion*, 12(2). <https://doi.org/10.1080/17432200.2016.1172765>
- Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174–185. <https://doi.org/10.1016/j.tourman.2015.05.007>
- García-Hernández, M., de la Calle-Vaquero, M., & Yubero, C. (2017). Cultural heritage and urban tourism: Historic city centres under pressure. *Sustainability (Switzerland)*, 9(8). <https://doi.org/10.3390/su9081346>
- Garner, B., Thornton, C., Luo Pawluk, A., Mora Cortez, R., Johnston, W., & Ayala, C. (2022). Utilizing text-mining to explore consumer happiness within tourism destinations. *Journal of Business Research*, 139, 1366–1377. <https://doi.org/10.1016/j.jbusres.2021.08.025>
- Gonia, A., & Jezierska-Thöle, A. (2022). Sustainable Tourism in Cities—Nature Reserves as a ‘New’ City Space for Nature-Based Tourism. *Sustainability*, 14(3). <https://doi.org/10.3390/su14031581>
- Grube, N. (2022). Seeking pathways for sustainable urban tourism: a critical deconstruction of Berlin’s approach for *city-compatible tourism*. *European Planning Studies*. <https://doi.org/10.1080/09654313.2022.2094199>
- Gunter, U., & Önder, I. (2018). Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics*, 24(3). <https://doi.org/10.1177/1354816617731196>

- Gunter, U., & Önder, I. (2021). An Exploratory Analysis of Geotagged Photos From Instagram for Residents of and Visitors to Vienna. *Journal of Hospitality & Tourism Research*, 45(2). <https://doi.org/10.1177/1096348020963689>
- Gursoy, D., Akova, O., & Atsız, O. (2022). Understanding the heritage experience: a content analysis of online reviews of World Heritage Sites in Istanbul. *Journal of Tourism and Cultural Change*, 20(3), 311–334. <https://doi.org/10.1080/14766825.2021.1937193>
- Handani, N. D., Williady, A., & Kim, H.-S. (2022). An Analysis of Customer Textual Reviews and Satisfaction at Luxury Hotels in Singapore’s Marina Bay Area (SG-Clean-Certified Hotels). *Sustainability*, 14(15). <https://doi.org/10.3390/su14159382>
- Jakob, D. (2013). The eventification of place: Urban development and experience consumption in Berlin and New York City. *European Urban and Regional Studies*, 20(4), 447–459. <https://doi.org/10.1177/0969776412459860>
- Jalilvand, M. R., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet Research*, 22(5), 591–612. <https://doi.org/10.1108/10662241211271563>
- Javed, M., Tučková, Z., & Jibril, A. B. (2020). The Role of Social Media on Tourists’ Behavior: An Empirical Analysis of Millennials from the Czech Republic. *Sustainability*, 12(18). <https://doi.org/10.3390/su12187735>
- Kádár, B. (2013). Differences in the spatial patterns of urban tourism in Vienna and Prague. *Urbani Izziv*, 24(2). <https://doi.org/10.5379/urbani-izziv-en-2013-24-02-002>
- Kalandides, A., & Grésillon, B. (2021). The Ambiguities of “Sustainable” Berlin. *Sustainability*, 13(4). <https://doi.org/10.3390/su13041666>
- Karacsony, P., Krupánszki, K., & Antalík, I. (2022). Analysis of the Impact of the COVID-19 Crisis on the Hungarian Employees. *Sustainability*, 14(4). <https://doi.org/10.3390/su14041990>
- Kay Smith, M., Pinke-Sziva, I., Berezvai, Z., & Buczkowska-Gołabek, K. (2022). The changing nature of the cultural tourist: motivations, profiles and experiences of cultural tourists in Budapest. *Journal of Tourism and Cultural Change*, 20(1–2). <https://doi.org/10.1080/14766825.2021.1898626>
- Kim, K., Park, O. jung, Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, 123, 362–369. <https://doi.org/10.1016/j.techfore.2017.01.001>

- Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online sources: A web content mining approach. *Journal of Business Research*, 68(9), 1836–1843. <https://doi.org/10.1016/j.jbusres.2015.01.011>
- Konstantakis, M., Alexandridis, G., & Caridakis, G. (2020). A Personalized Heritage-Oriented Recommender System Based on Extended Cultural Tourist Typologies. *Big Data and Cognitive Computing*, 4(2). <https://doi.org/10.3390/bdcc4020012>
- Kowalczyk-Anioł, J. (2023). Rethinking tourism-driven urban transformation and social tourism impact: A scenario from a CEE city. *Cities*, 134. <https://doi.org/10.1016/j.cities.2022.104178>
- Kozinets, R. V. (2002). The Field behind the Screen: Using Netnography for Marketing Research in Online Communities. *Journal of Marketing Research*, 39(1). <https://doi.org/10.1509/jmkr.39.1.61.18935>
- Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. In *Tourism Management* (Vol. 83). Elsevier Ltd. <https://doi.org/10.1016/j.tourman.2020.104245>
- Loureiro, S. M. C., Bilro, R. G., & Japutra, A. (2019). The effect of consumer-generated media stimuli on emotions and consumer brand engagement. *Journal of Product & Brand Management*, 29(3). <https://doi.org/10.1108/JPBM-11-2018-2120>
- Marques, S. H., Bilro, R. G., Gonçalves, M., & Rather, R. A. (2022). Partners we can trust: the role of employee engagement influencing partnerships' relationship quality in the voluntary sector. *Management Research: Journal of the Iberoamerican Academy of Management*, 20(2). <https://doi.org/10.1108/MRJIAM-01-2022-1258>
- Matoga, Ł., & Pawłowska, A. (2018). Off-the-beaten-track tourism: a new trend in the tourism development in historical European cities. A case study of the city of Krakow, Poland. *Current Issues in Tourism*, 21(14), 1644–1669. <https://doi.org/10.1080/13683500.2016.1212822>
- Matzler, K., Strobl, A., Stokburger-Sauer, N., Bobovnický, A., & Bauer, F. (2016). Brand personality and culture: The role of cultural differences on the impact of brand personality perceptions on tourists' visit intentions. *Tourism Management*, 52. <https://doi.org/10.1016/j.tourman.2015.07.017>
- Mehraliyev, F., Chan, I. C. C., & Kirilenko, A. P. (2022). Sentiment analysis in hospitality and tourism: a thematic and methodological review. In *International Journal of Contemporary Hospitality Management* (Vol. 34, Issue 1, pp. 46–77). Emerald Group Holdings Ltd. <https://doi.org/10.1108/IJCHM-02-2021-0132>

- Mkono, M., & Markwell, K. (2014). The application of netnography in tourism studies. *Annals of Tourism Research*, 48. <https://doi.org/10.1016/j.annals.2014.07.005>
- Municipality of Budapest. (2015). *Integrated Urban Development Strategy – Budapest 2020*.
- Naumov, N., & Weidenfeld, A. (2019). From socialist icons to post-socialist attractions: Iconicity of socialist heritage in central and Eastern Europe. *Geographia Polonica*, 92(4). <https://doi.org/10.7163/GPol.0147>
- Okey, R. (1992). Central Europe / Eastern Europe: Behind the Definitions. *Past & Present*, 137, 102–133.
- Olt, G., Smith, M. K., Csizmady, A., & Sziva, I. (2019). Gentrification, tourism and the night-time economy in Budapest's district VII – the role of regulation in a post-socialist context. *Journal of Policy Research in Tourism, Leisure and Events*, 11(3). <https://doi.org/10.1080/19407963.2019.1604531>
- Papadopoulou, N. M., Ribeiro, M. A., & Prayag, G. (2023). Psychological Determinants of Tourist Satisfaction and Destination Loyalty: The Influence of Perceived Overcrowding and Overtourism. *Journal of Travel Research*, 62(3), 644–662. <https://doi.org/10.1177/00472875221089049>
- Pérez Garrido, B., Sebrek, S. S., Semenova, V., Bal, D., & Michalkó, G. (2022). Addressing the Phenomenon of Overtourism in Budapest from Multiple Angles Using Unconventional Methodologies and Data. *Sustainability*, 14(4). <https://doi.org/10.3390/su14042268>
- Pinke-Sziva, I., Smith, M., Olt, G., & Berezvai, Z. (2019). Overtourism and the night-time economy: a case study of Budapest. *International Journal of Tourism Cities*, 5(1). <https://doi.org/10.1108/IJTC-04-2018-0028>
- Popescu, R. I., & Corbos, R. A. (2011). Vienna's branding campaign - strategic option for developing austria's capital in a top tourism destination. *Theoretical and Empirical Researches in Urban Management*, 6(3), 43–60.
- Prague City Tourism. (2020). *Prague Destination Management: Putting Prague First - Strategy for Sustainable Tourism in Prague*.
- Ramseook-Munhurrun, P., Seebaluck, V. N., & Naidoo, P. (2015). Examining the Structural Relationships of Destination Image, Perceived Value, Tourist Satisfaction and Loyalty: Case of Mauritius. *Procedia - Social and Behavioral Sciences*, 175. <https://doi.org/10.1016/j.sbspro.2015.01.1198>

- Rátz, T., Smith, M., & Michalkó, G. (2008). New places in old spaces: Mapping tourism and regeneration in Budapest. *Tourism Geographies*, *10*(4), 429–451.
<https://doi.org/10.1080/14616680802434064>
- Remenyik, B., Barcza, A., Csapó, J., Szabó, B., Fodor, G., & Dávid, L. D. (2021). Overtourism in Budapest: Analysis of spatial process and suggested solutions. *Regional Statistics*, *11*(3). <https://doi.org/10.15196/RS110303>
- Richards, G. (2018). Cultural tourism: A review of recent research and trends. *Journal of Hospitality and Tourism Management*, *36*. <https://doi.org/10.1016/j.jhtm.2018.03.005>
- Richards, G. W. (1999). European Cultural Tourism: patterns and prospects. In D. Dodd & A. van Hemel (Eds.), *Planning cultural Tourism in Europe* (pp. 16–32). Boekmanstichting.
- Rončák, M., Scholz, P., & Linderová, I. (2021). Safety Concerns and Travel Behavior of Generation Z: Case Study from the Czech Republic. *Sustainability*, *13*(23).
<https://doi.org/10.3390/su132313439>
- Šabič, Z., & Drulák, P. (2012). Introduction to ‘Central Europe’ . In Z. Šabič & P. Drulák (Eds.), *Regional and International Relations of Central Europe* (pp. 1–18). Palgrave Macmillan.
- Šauer, M., Vystoupil, J., Novotná, M., & Widawski, K. (2021). Central European Tourist Flows: Intraregional Patterns and Their Implications. *Moravian Geographical Reports*, *29*(4), 278–291. <https://doi.org/10.2478/mgr-2021-0020>
- Sharda, R., Delen, D., & Turban, E. (2014). *Business Intelligence: a managerial perspective on analytics* (S. Wall, Ed.; 3rd ed.). Pearson.
- Skotis, A., & Livas, C. (2022). A data-driven analysis of experience in urban historic districts. *Annals of Tourism Research Empirical Insights*, *3*(2).
<https://doi.org/10.1016/j.annale.2022.100052>
- Smith, M. K., Sziva, I. P., & Olt, G. (2019). Overtourism and Resident Resistance in Budapest. *Tourism Planning & Development*, *16*(4).
<https://doi.org/10.1080/21568316.2019.1595705>
- Solima, L., & Izzo, F. (2018). QR Codes in cultural heritage tourism: new communications technologies and future prospects in Naples and Warsaw. *Journal of Heritage Tourism*, *13*(2). <https://doi.org/10.1080/1743873X.2017.1337776>
- Statista. (2023a). *Travel and tourism in Czechia*.
- Statista. (2023b). *Travel and tourism in Europe*.
- Statista. (2023c). *Travel and tourism in Hungary*.
- Statista. (2023d). *Travel and tourism in Poland*.

- Statista. (2023e). *Tripadvisor: Statista Dossier on Tripadvisor*.
- Tavakoli, R., & Mura, P. (2018). Netnography in tourism – Beyond Web 2.0. *Annals of Tourism Research*, 73. <https://doi.org/10.1016/j.annals.2018.06.002>
- Tavakoli, R., & Wijesinghe, S. N. R. (2019). The evolution of the web and netnography in tourism: A systematic review. *Tourism Management Perspectives*, 29. <https://doi.org/10.1016/j.tmp.2018.10.008>
- Tittelbachová, Š., Čajková, A., Lukáč, M., & Lžičar, P. (2022). Impact of the COVID-19 pandemic on tourism in the Czech Republic. *Terra Economicus*, 20(3). <https://doi.org/10.18522/2073-6606-2022-20-3-133-145>
- Tracz, M., & Bajgier-Kowalska, M. (2019). Profile of tourists visiting cultural heritage cities: Comparative study of selected cities in Central Europe. *Geographia Polonica*, 92(4). <https://doi.org/10.7163/GPol.0159>
- Vena-Oya, J., Castañeda-García, J. A., Rodríguez-Molina, M. Á., & Frías-Jamilena, D. M. (2021). How do monetary and time spend explain cultural tourist satisfaction? *Tourism Management Perspectives*, 37. <https://doi.org/10.1016/j.tmp.2021.100788>
- Vienna Tourism Board. (2021). *Reshaping Vienna: Vienna Visitor Economy Strategy 2025*.
- Viol, M., Todd, L., Theodoraki, E., & Anastasiadou, C. (2018). The role of iconic-historic commemorative events in event tourism: Insights from the 20th and 25th anniversaries of the fall of the Berlin Wall. *Tourism Management*, 69. <https://doi.org/10.1016/j.tourman.2018.06.018>
- Visit Bratislava. (2017). *Marketing and communications strategy for the destination Bratislava – Executive summary*.
- Whalen, E. A. (2018). Understanding a shifting methodology. *International Journal of Contemporary Hospitality Management*, 30(11). <https://doi.org/10.1108/IJCHM-08-2017-0536>
- Widz, M., Krukowska, R., Walas, B., & Kruczek, Z. (2022). Course of Values of Key Performance Indicators in City Hotels during the COVID-19 Pandemic: Poland Case Study. *Sustainability*, 14(19). <https://doi.org/10.3390/su141912454>
- Wu, C.-W. (2016). Destination loyalty modeling of the global tourism. *Journal of Business Research*, 69(6). <https://doi.org/10.1016/j.jbusres.2015.12.032>
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120–130. <https://doi.org/10.1016/j.ijhm.2014.10.013>

- Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, 55, 57–69. <https://doi.org/10.1016/j.ijhm.2016.03.003>
- Zhang, G., Cheng, M., & Zhang, J. (2022). A cross-cultural comparison of peer-to-peer accommodation experience: A mixed text mining approach. *International Journal of Hospitality Management*, 106. <https://doi.org/10.1016/j.ijhm.2022.103296>

Annex A

```
{ "_id": "trip-a-example", "startUrl": ["https://www.tripadvisor.co.uk/Attraction_Review-g274707-d275451-Reviews-or200-Staromestske_namesti-Prague_Bohemia.html"], "selectors": [{"clickElementSelector": ".Hdolsa", "clickElementUniquenessType": "uniqueCSSSelector", "clickType": "clickMore", "delay": 2000, "discardInitialElements": "do-not-discard", "id": "Element Select - Page Changer", "multiple": true, "parentSelectors": ["_root", "Element Select - Page Changer"], "selector": ".LbPSX div._c", "type": "SelectorElementClick"}, {"id": "Name", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "regex": "", "selector": "span a", "type": "SelectorText"}, {"id": "place", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "regex": "", "selector": ".osNWb span:nth-of-type(1)", "type": "SelectorText"}, {"id": "Date of stay", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "regex": "", "selector": "div.RpeCd", "type": "SelectorText"}, {"id": "Main Quote", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "regex": "", "selector": ".BMQDV span", "type": "SelectorText"}, {"id": "Long-Quote", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "regex": "", "selector": ".pZUbB span.yCeTE", "type": "SelectorText"}, {"extractAttribute": "aria-label", "id": "Bubble Rating", "multiple": false, "parentSelectors": ["Element Select - Page Changer"], "selector": "svg.UctUV", "type": "SelectorElementAttribute"}]}
```