Article

Measuring Destination Image Using AI and Big Data: Kastoria's Image on TripAdvisor

Anastasia Yannacopoulou *⬤ and Konstantinos Kallinikos

Department of Communication and Digital Media, University of Western Macedonia, 52100 Kastoria, Greece; kallinikoskon@gmail.com
* Correspondence: ayannacopoulou@uowm.gr

Abstract: In recent years, the growing number of Online Travel Review (OTR) platforms and advances in social media and search engine technologies have led to a new way of accessing information for tourists, placing projected Tourist Destination Image (TDI) and electronic Word of Mouth (eWoM) at the heart of travel decision-making. This research introduces a big data-driven approach to analyzing and measuring the perceived and conveyed TDI in OTRs concerning the reflected perceptive, spatial, and affective dimensions of search results. To test this approach, a massive metadata analysis of search engine was conducted on approximately 2700 reviews from TripAdvisor users for the category "Attractions" of the city of Kastoria, Greece. Using artificial intelligence, an analysis of the photos accompanying user comments on TripAdvisor was performed. Based on the results, we created five themes for the image narratives, depending on the focus of interest (monument, activity, self, other person, and unknown) in which the content was categorized. The results obtained allow us to extract information that can be used in business intelligence applications.

Keywords: destination image; data mining; image recognition; user-generated content; projected image; perceived image; e-WOM image; tourism destination image; visual data mining

1. Introduction

In the last decade, with the rise of information and communication technologies (ICT) and social media, the production of user-generated content (UGC) has increased dramatically. This rise in social media usage has been especially notable in the hospitality and tourism sector, particularly for holiday planning [1]. Destination marketing has become a fundamental pillar for fostering the development and sustainability of tourist destinations in an increasingly competitive and globalized tourism market [2]. Moreover, according to the World Travel and Tourism Council [3], the importance of tourism is reflected in the fact that 10.4% of the global GDP (USD 10 trillion) in 2019 was generated by this industry.

1.1. Online Travel Reviews

Online Travel Reviews (OTRs), which started from travel blogs, have proliferated thanks to the emergence and growth of the digitalized form of travel review websites [4]. In OTRs, a common reaction by the users is to write about their experiences of the destination they visited and rate attractions (such as monuments, museums, lakes, rivers, parks, etc.) and services (such as accommodations, restaurants, expeditions, etc.) and/or give their opinion. Credibility and trust in eWOM are affected by the credibility of the medium providing the review, as it is written by someone the user does not know at all [5]. According

to TripAdvisor [6], there are more than 589 million reviews by travelers about more than 8.6 million accommodations, restaurants, experiences, airlines, and cruises recorded in its database.

The latest Eurobarometer survey [7], to which 25,714 respondents from the 27 countries of the European Union responded by telephone, showed that the main source of information used to plan holidays was Word of Mouth (WoM) recommendations from family, friends, and colleagues (56%), followed by personal experience (37%) and eWoM websites that collect and present comments, reviews, and ratings from travelers (34%). EWoM communication primarily relies on user-generated content (UGC) which is accessible online for free. UGC is the subjective opinion of the users who create it, but it is perceived by others as unbiased and authentic first-hand information. As such, it is considered more trustworthy material [8] than material provided by the webmasters and social media administrators (Webmaster Generated Content-WGC) of the destination itself. This was evident in the above survey where only one in five (21%) consider WGC important for planning their travel.

1.2. Tourist Destination Image

Travel diaries and online travel reviews project an image that, in combination with the image projected by Destination Management Organizations (DMOs), shapes the perceived Tourist Destination Image (TDI). According to [9,10], as cited in [4], the study of TDI has remained a consistent focus in the scientific literature on tourism because "images are crucial in conveying the representation of a region in the minds of potential tourists." It is important to remember that an image is perceived by an individual at a given moment. The image may change over time or shift from one season to another. Therefore, the spatio-temporal dimension of the TDI must be taken into account.

TDI information sources are divided into primary and secondary [11]. Primary sources refer to the information visitors gain directly from their own experiences at a destination, whereas secondary sources originate from information provided by other individuals or organizations. Primary research has relied on communication-focused methods, such as surveys and in-depth interviews, to gather data directly from users and consumers [12]. Nowadays, as a result of the aforementioned, the tourism and hospitality industry appears to be increasingly relying on social media analytics and big data to analyze tourism behavior, public emotions, and attitudes [13]. Therefore, our main objective is to measure and analyze the TDI promoted by TripAdvisor for the attractions of the city of Kastoria, Greece.

1.3. Tourism Research Using Photographs

Photography has been intrinsically linked to tourism [14], especially after the introduction of smart phones, as every traveler engages in some form of photography during their trip. The photos tourists post online are closely tied to their impressions, as taking photographs is found to be a process of preserving personal and family memories, which creates styles and cultural perspectives that allow tourists to classify the scenes they see as "interesting, good, or beautiful" [15]. Research has shown that more than 50% of photos contain information about tourists' faces. Posing in front of a landmark typically follows specific conventions, such as frontal, eye-level photography, smiling, and striking a pose that allows the landmark to remain visible [16]. This seems to be a "been there/done that" statement. According to [17], tourists take photographs to document their experiences, aware that the passage of time can distort their memories and evaluations of those experiences. Thus, although tourist photographs are created for personal use, they can, if published on social media, positively influence a destination's image by effectively evoking

positive emotions (i.e., emotional image) and clearly showcasing the destination's physical and functional characteristics (i.e., cognitive image).

For some time, tourism scholars have been utilizing content analysis to study iconographic material. The literature review in [18] found a 277% overall increase in the number of research papers using social media image data in social science research from 2016 to 2019. With the availability of AI tools and the possibilities they offer to their users, research on color as a fundamental element of visual art and the role it plays in emotional responses has now accelerated [19,20], and the influence of color psychology on branding is also being investigated [21].

2. Materials and Methods

Search engines and OTR sites are powerful tools because of their ability to index and organize vast amounts of information. For our research, we chose to measure reviews from the most popular OTR site, TripAdvisor, the largest online user-generated reviews site in the tourism industry (e.g., Similarweb.com). Our data extraction was performed through the website apify.com.

We collected a sample of 2619 reviews in the category "Attractions", for the destination of Kastoria [22]. This category contains 27 entries. From these entries, after checking, we removed a total of 7 entries (three cafes, one internet cafe, and three fur shops), as they did not belong to the category. In addition, for two entries there was no review. The 2577 remaining reviews are written in 17 different languages by 1631 tourists from at least 45 countries who visited Kastoria between 2012 and 2024. From our sample, we obtained significant results both on the spatio-temporal distribution of tourists and on evaluative and emotional dimensions of TDI.

2.1. Content Analysis

Reviews play a crucial role in sharing opinions about products, places, or experiences. Whether positive or negative, they provide valuable insights that help others make informed decisions. Analyzing content in such a large number of texts written by different people can be a painstaking and lengthy process. However, it is a systematic, reproducible technique that allows keywords or key phrases to be compressed into a few categories of content. This approach enables researchers to systematically navigate large datasets with relative ease [23]. In this study, we used quantitative content analysis in two phases: word splitting and frequency analysis. In addition, we created a Python [24] program to build word clouds for each language (Figure 1). From these word clouds, we automatically removed the stop words of each language using files specifically created for this purpose. We treated the space character and other word separators, such as commas and question marks, as delimiters. We did not differentiate between uppercase and lowercase letters and used two counters: one for the total words in the text, including stop words, and another for the unique keywords.
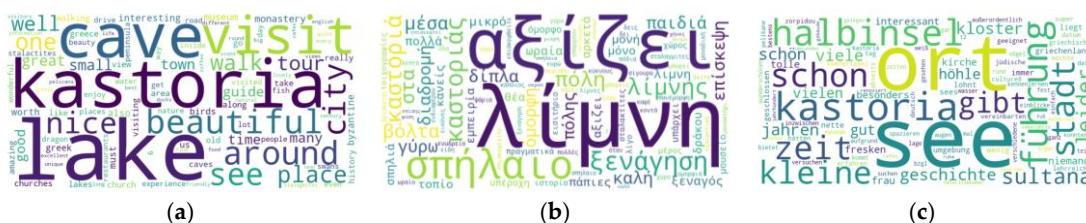

| (**a**) | (**b**) | (**c**) |

Figure 1. Word clouds for (**a**) English, (**b**) Greek, and (**c**) German, based on the reviews' content.

Word frequency analysis has a limitation in that it cannot automatically analyze complex words or word groups with different meanings, whether used together or separately,

using open-source software. To address this, a custom program, such as one developed in Python, is necessary. Despite this limitation, we assume that the most frequently mentioned words highlight the primary concerns of the tourists who wrote the reviews, thereby providing researchers with valuable insights into key points of interest.

In addition, we used TextBlob [25], a Python library developed by Steven Loria, used for natural language processing (NLP) and based on the NLTK (Natural Language Toolkit). For each sentence (input), the analysis provides two outputs: polarity and subjectivity. The polarity score ranges from −1 to 1, where −1 indicates highly negative language (e.g., "disgusting" or "awful") and 1 indicates highly positive language (e.g., "excellent" or "best"). The subjectivity score, ranging from 0 to 1, reflects the degree of personal opinion in the sentence, with a score closer to 1 indicating a predominance of subjective opinions over factual content.

2.2. Image Analysis

Of the 2577 reviews we extracted from TripAdvisor, 537 were accompanied by 2598 photos uploaded by users. For the analysis of the images, we conducted three different checks. First, we sorted the photos into two groups: those containing people and those without. For photos featuring people, we further categorized them into three groups: (a) portraits of the photographers themselves, (b) people close to them, or (c) random passers-by. The second check was done with the help of AI, specifically with the imagerecognize tool which uses Convolutional Neural Network so that the output contains only a single probability score vector, which is organized along the depth dimension. Using an API key we provided the first images from each review for object recognition by AI. This returned a list of up to 10 words describing the image if it had a confidence score above 80%. Finally, we created a Python program, taking advantage of the k-means clustering algorithm and the PIL, tqdm, NumPy and sklearn.cluster libraries, to generate a png file with the 10 most dominant color clusters of each photo. We used these images to associate colors with emotions. Using the above libraries and additional libraries, including pandas, webcolors and collections, and based on the illustration of different emotion correlations and the Mikels' emotion wheel [19], we created a Python program that took as input the palette of 10 colors we had created for each photo and returned the three emotions, out of the 23 emotions we associated with the 10 main colors, that resulted for each palette, as shown in the script below:

```python
def get_emotions_from_colors(colors):
    """It returns the three main emotions associated with a list of colors."""
    color_emotions = {
        "red": ["passion", "energy", "anger"],
        "orange": ["enthusiasm", "creativity", "warning"],
        "yellow": ["happiness", "optimism", "anxiety"],
        "green": ["calm", "nature", "envy"],
        "blue": ["peace", "trust", "sadness"],
        "purple": ["luxury", "mystery", "spirituality"],
        "pink": ["love", "compassion", "immaturity"],
        "brown": ["stability", "nature", "dullness"],
        "black": ["power", "sophistication", "mourning"],
        "white": ["purity", "innocence", "emptiness"]
    }
```

2.3. Research Questions

The proposed method includes the following phases: (a) data collection, (b) metadata mining, and (c) quantitative analysis. As the world's largest source of user-generated content (UGC) in the tourism sector, TripAdvisor offers the advantage of providing a vast open dataset due to the enormous volume of user reviews it hosts. Additionally, TripAdvisor's reputation management system enhances transparency by allowing access to user profiles, other reviews, votes, and ratings, while also encouraging users to submit credible reviews [26].

The research questions that we will try to answer with this paper are the following:

RQ1. Has there been an increase in reviews over time?

RQ2. Is there a change in the number of reviews written about Kastoria depending on the time of year?

RQ3. Which attractions are most visited by tourists from different parts of the world?

RQ4. What appears more often in the photos, people, or landscapes, and which sights do visitors choose most to include in their reviews (e.g., lake, mansions)?

RQ5. Does the existence of photos affect the likelihood of interaction with other TripAdvisor users?

RQ6. What is the general impression tourists get from visiting Kastoria?

3. Results

The analysis of the 2577 reviews revealed a significant focus on two main attractions: the Lake of Kastoria, accounting for 33.1% ($n = 853$) of the reviews, and the Dragon Cave, with 29.2% ($n = 752$), together comprising 62.3% of the total reviews for the region's attractions. The third most-reviewed location was the Panagia Mavriotissa Monastery, representing 14% ($n = 360$) of the reviews, followed by the Kastoria Aquarium with 7.8% ($n = 202$). A detailed breakdown of reviews per attraction is provided in Table 1.

Table 1. Distribution of reviews by place or activity.

| Place/Activity	n (=2577)	%
Kastoria Lake	853	33.1%
Dragon Cave (Spílaio tou Drákou)	752	29.2%
Panagia Mavriotissa Monastery	360	14.0%
Kastoria Aquarium	202	7.8%
Wax Museum Mavrochoriou Kastorias	79	3.1%
Folklore Museum of Kastoria	77	3.0%
Byzantine Museum of Kastoria	49	1.9%
Prophet Elias Church	44	1.7%
Kastorian Byzantine Churches	37	1.4%
Kastoria Outdoors	35	1.4%
Culture 8 Guided Day Tours	30	1.2%
Adventure Kastoria	14	0.5%
Fossilized Forest	13	0.5%
Museum of Costumes (Endymatologiko Mouseio)	11	0.4%
Mountain Lunatics	8	0.3%
Church of the Panagia Koumbelidiki	7	0.3%
Panik Rentals	4	0.2%
Church of St. Taksiarkhov	2	0.1%

The analysis of 2546 reviews, which indicate the year the trip took place, revealed that 19.8% ($n = 503$) of the visits occurred between October 2016 and September 2017. In addressing our first research question (RQ1), we observed that the number of reviews steadily increased from May 2011, when the first review was recorded, until April 2017,

after which it gradually declined until February 2020. In March 2020, with the onset of the COVID-19 pandemic, Kastoria became one of the first regions to enforce strict mobility restrictions due to outbreaks [27]. Notably, from October 2020 to May 2021, no reviews were posted on TripAdvisor for the attractions section in Kastoria. Figure 2 illustrates the temporal distribution of visits to Kastoria.
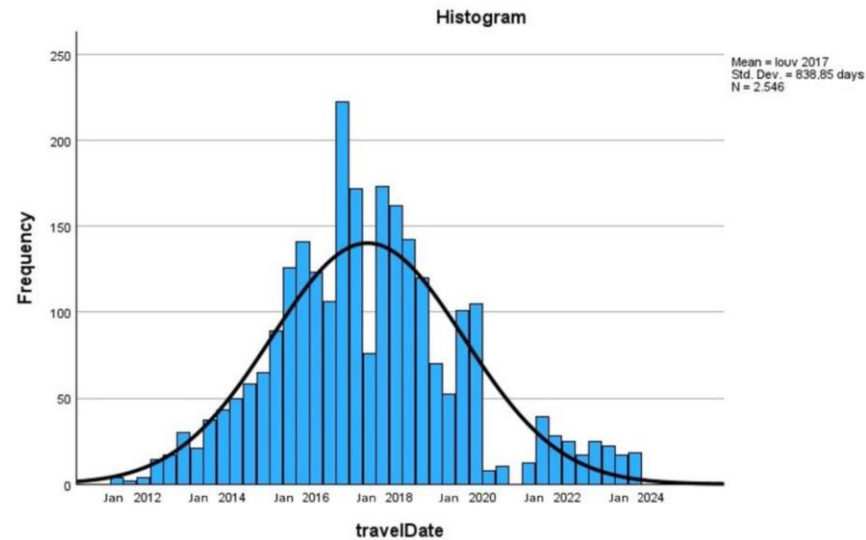
Figure 2. Temporal distribution of visits to Kastoria.

Our second research question (RQ2) examines the period when tourists typically visit Kastoria. Based on the temporal distribution, the months with the highest number of visitors are, in descending order, December (11.5%, *n* = 294), August (10.7%, *n* = 272), October (10.4%, *n* = 264), and January (10.1%, *n* = 257). Conversely, the city experiences the fewest visitors in June (4.5%, *n* = 115) and July (5.9%, *n* = 151). According to our survey data, the peak visitor period is in the winter months of December and January, followed by August, October, and April. This indicates that tourism in Kastoria is highly seasonal. Figure 3 provides a detailed breakdown of visit percentages by month.
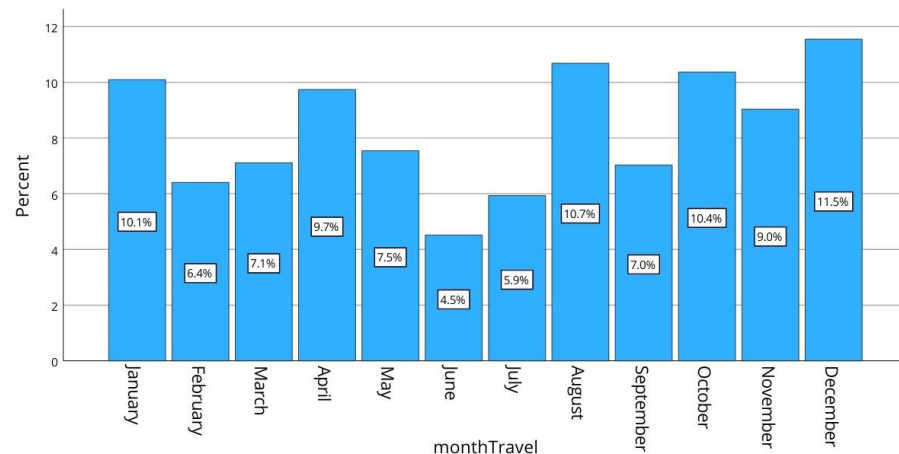
Figure 3. The frequency of trips to Kastoria per month.

To address our third research question (RQ3) regarding the homogeneity of reviews for each attraction and the language in which they were written, we utilized the non-factorial chi-square test after confirming normality. The analysis revealed a statistically significant dependence between the reviews for each attraction and the language used, at a significance level of $\alpha = 0.05$ ($\chi^2 = 20{,}666.420$, df = 16, sig < 0.001). Consequently, we reject the null

hypothesis (which states that there is no correlation between language and the attraction visited) and conclude that tourists from different parts of the world tend to visit different attractions in the city. Detailed data for the top eight languages are provided in Table 2a,b, while Figure 4 illustrates the preferences of Greek and Russian tourists.

Table 2. (**a**) Correlation of review language with the number of visits to each place/activity. (**b**) Correlation of review language with the number of visits to each place/activity (continued).

(a)								
Place/Activity	**Language**							
	Greek		**English**		**Russian**		**Italian**	
	n (=1806)	%	*n* (=496)	%	*n* (=105)	%	*n* (=44)	%
Dragon Cave (Spílaio tou Drákou)	564	31.2%	139	28.0%	16	15.2%	8	18.2%
Kastoria Lake	537	29.7%	195	39.3%	54	51.4%	18	40.9%
Panagia Mavriotissa Monastery	258	14.3%	57	11.5%	18	17.1%	10	22.7%
Kastoria Aquarium	178	9.9%	22	4.4%	1	1.0%	0	0.0%
Wax Museum Mavrochoriou	70	3.9%	5	1.0%	1	1.0%	1	2.3%
Folklore Museum of Kastoria	55	3.0%	19	3.8%	0	0.0%	1	2.3%
Prophet Elias	36	2.0%	5	1.0%	1	1.0%	0	0.0%
Byzantine Museum of Kastoria	24	1.3%	12	2.4%	0	0.0%	1	2.3%
Kastoria Outdoors	22	1.2%	10	2.0%	2	1.9%	0	0.0%
Adventure Kastoria	14	0.8%	0	0.0%	0	0.0%	0	0.0%
Kastorian Byzantine Churches	14	0.8%	7	1.4%	5	4.8%	4	9.1%
Culture 8 Guided Day Tours	10	0.6%	12	2.4%	2	1.9%	0	0.0%
Endymatologiko Mouseio	8	0.4%	3	0.6%	0	0.0%	0	0.0%
Fossilized Forest	7	0.4%	4	0.8%	0	0.0%	0	0.0%
Mountain Lunatics	5	0.3%	3	0.6%	0	0.0%	0	0.0%
Church of the Panagia Koumbelidiki	3	0.2%	0	0.0%	3	2.9%	1	2.3%
Panik Rentals	1	0.1%	3	0.6%	0	0.0%	0	0.0%
Church of St. Taksiarkhov	0	0.0%	0	0.0%	2	1.9%	0	0.0%

(b)								
Place/Activity	**Language**							
	Dutch		**Hebrew**		**French**		**German**	
	n (=31)	%	*n* (=26)	%	*n* (=24)	%	*n* (=23)	%
Dragon Cave (Spílaio tou Drákou)	8	25.8%	10	38.5%	2	8.3%	2	8.7%
Kastoria Lake	11	35.5%	8	30.8%	11	45.8%	8	34.8%
Panagia Mavriotissa Monastery	7	22.6%	1	3.8%	3	12.5%	3	13.0%
Kastoria Aquarium	0	0.0%	0	0.0%	1	4.2%	0	0.0%
Wax Museum Mavrochoriou	0	0.0%	1	3.8%	0	0.0%	1	4.3%
Folklore Museum of Kastoria	0	0.0%	1	3.8%	0	0.0%	0	0.0%
Prophet Elias	0	0.0%	1	3.8%	0	0.0%	1	4.3%
Byzantine Museum of Kastoria	3	9.7%	1	3.8%	5	20.8%	1	4.3%
Kastoria Outdoors	0	0.0%	1	3.8%	0	0.0%	0	0.0%
Adventure Kastoria	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Kastorian Byzantine Churches	2	6.5%	1	3.8%	2	8.3%	1	4.3%
Culture 8 Guided Day Tours	0	0.0%	0	0.0%	0	0.0%	5	21.7%
Endymatologiko Mouseio	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Fossilized Forest	0	0.0%	1	3.8%	0	0.0%	1	4.3%
Mountain Lunatics	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Church of the Panagia Koumbelidiki	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Panik Rentals	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Church of St. Taksiarkhov	0	0.0%	0	0.0%	0	0.0%	0	0.0%

The fourth research question (RQ4) examines the topics most commonly raised by visitors in their reviews, such as people, sights, and landscapes. To investigate, we analyzed 536 of the 2598 photos—the ones that users uploaded first in their reviews. The image analysis involved three checks.
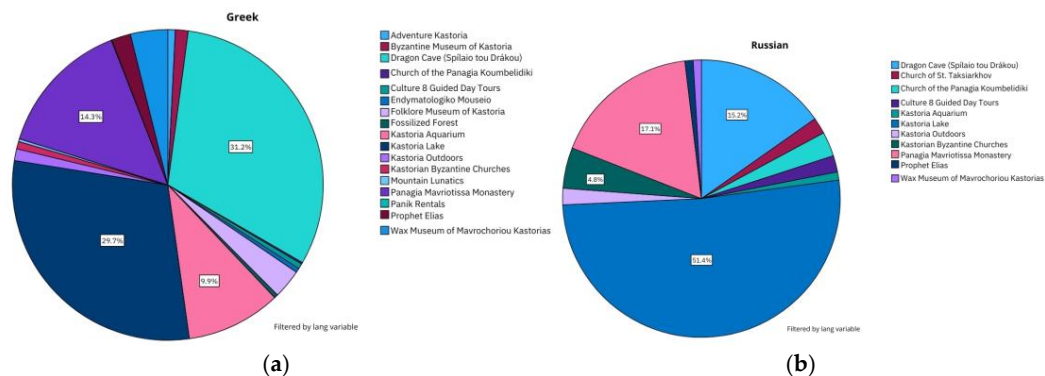
Figure 4. Reviews sorted by language: (**a**) Visit preferences of Greeks. (**b**) Visit preferences of Russians.

First, we manually categorized the photos into those with or without people, finding that 84.9% (n = 455) did not include people. For the 81 photos with people, we divided them into three categories: (a) portraits of the visitors themselves, (b) close companions, or (c) random passers-by.

For the 81 photos with people, we divided them into three categories: (a) portraits of the visitors themselves (7.4%, n = 6), (b) close companions (35.8%, n = 29), and (c) random passers-by (59.3%, n = 48). Among these, only six photos were self-portraits, representing 7.4% of the total, with one being a solo selfie (1.2%), another featuring the visitor with a friend (1.2%), and the remaining four depicting couples (4.9%), either taken as selfies or with assistance. Additionally, 29 photos (35.8%) showed close companions, including five (6.2%) where the visitor also appeared, and three of these (3.7%) featured random passers-by. In 24 photos (29.6%), only close companions were visible in front of an attraction, while in 2 (2.5%), random passers-by were also present. Finally, 48 photos (59.3%) captured random passers-by near the attractions.

The second check used AI to analyze the photos, generating 4248 descriptive words with over 80% confidence. These words included 338 unique terms, with the 10 most frequent ones summarized in Table 3.

Table 3. The objects with the highest frequency of occurrence in photos according to AI visual object recognition.

Word	Views (n = 4248)	Percentage %
Outdoors	367	8.7%
Nature	349	8.2%
Scenery	194	4.6%
Water	189	4.5%
Lake	135	3.2%
Cave	114	2.7%
Person	103	2.4%
Waterfront	102	2.4%
Landscape	99	2.3%
Tree	86	2.0%

Artificial intelligence identified more photos as containing people than we did, as it mistakenly recognized exhibits from the wax museum and hagiographies as people. The words returned by the AI's API to describe people, ranked by frequency, were as follows: Person (n = 103), Female (n = 18), Woman (n = 17), Head (n = 12), Man (n = 11), Male (n = 8), People (n = 4), Child (n = 1), Girl (n = 1), and Lady (n = 1).

Finally, we analyzed the emotions evoked by the colors in the 536 photos (Figure 5), according to different emotion correlations and Mikels' emotion wheel [19]. From a total of 1608 responses identifying the three primary emotions associated with each photo, 72%

(n = 1152) were distributed among the following emotions, ranked by frequency: Nature 17% (n = 272), Dullness and Stability 15% each (n = 249), and Compassion and Love 12% each (n = 191), as the predominant colors are brown and green.


Figure 5. Snapshot of the palette files for each photo.

In response to our fifth research question (RQ5) about whether the presence of photos in a review generates more interaction on TripAdvisor, we found the following results. In our sample, 79.2% (n = 2041) of the reviews were not accompanied by photos, while the remaining 20.8% (n = 536) included a total of 2666 photos (Median = 4.97). Among the reviews with photos, 18.6% (n = 479) contained between one and 10 photos, whereas 2.2% of reviews had between 11 and 63 photos. The total number of helpful votes across all 2577 reviews was 1575. Of these, 47% (n = 737) were attributed to the 20.8% (n = 536) of reviews that included photos. Notably, the 11 reviews with more than 10 helpful votes all had photos, collectively accounting for 27% (n = 426) of the total helpful votes. Conversely, 1826 reviews received no helpful votes, with 19% (n = 341) of these reviews including photos and 81% (n = 1485) having none. These findings highlight the significant role of photos in increasing user interaction and garnering helpful votes for reviews.

In our final research question (RQ6), we aimed to determine the overall impression tourists have of their visit to Kastoria. Analyzing the review scores, we found that the majority of visitors leave with a very good impression of the city (66%, n = 1696). Additionally, 26% (n = 682) leave with a good impression, while smaller percentages leave with neutral (5%, n = 138), bad (1%, n = 33), or very bad impressions (1%, n = 28).

Along with the ratings that visitors assigned to the attractions they visited, we also analyzed the comments they included in their reviews. We developed a Python program that leverages the TextBlob library for natural language processing, along with the langdetect and googletrans libraries. These libraries enable the detection of supported languages (up to 55) through Google and their translation, allowing TextBlob to perform sentiment analysis on the text. The analysis of 2577 guest comments, after mapping polarity values to a 5-point scale for comparison with TripAdvisor scores, revealed the following results: 0.1% (n = 2) of the comments reflected very bad impressions, 1.4% (n = 37) bad impressions, 28.2% (n = 727) neutral impressions, 56.6% (n = 1458) good impressions, and only 13.7% (n = 353) very good impressions. These findings highlight a significant discrepancy between user-assigned scores and AI-extracted sentiment scores from comments, with the AI scores averaging −0.72 points lower than those given by users. Specifically, only 27.4% (n = 706) of the comments showed no difference between the user's TripAdvisor score and the sentiment analysis score derived from their comment. Among the remaining comments, the majority (48.6%, n = 1252) had a one-point lower sentiment analysis score than the TripAdvisor score, 15.6% (n = 402) had a two-point lower score, 0.6% (n = 16) had a three-point lower score, and 0.1% (n = 2) had a four-point lower score. Conversely, 6.1%

(*n* = 157) had a one-point higher sentiment analysis score, and 1.6% (*n* = 42) showed an increase of two to four points. Figure 6 illustrates the distribution of the TextBlob results.
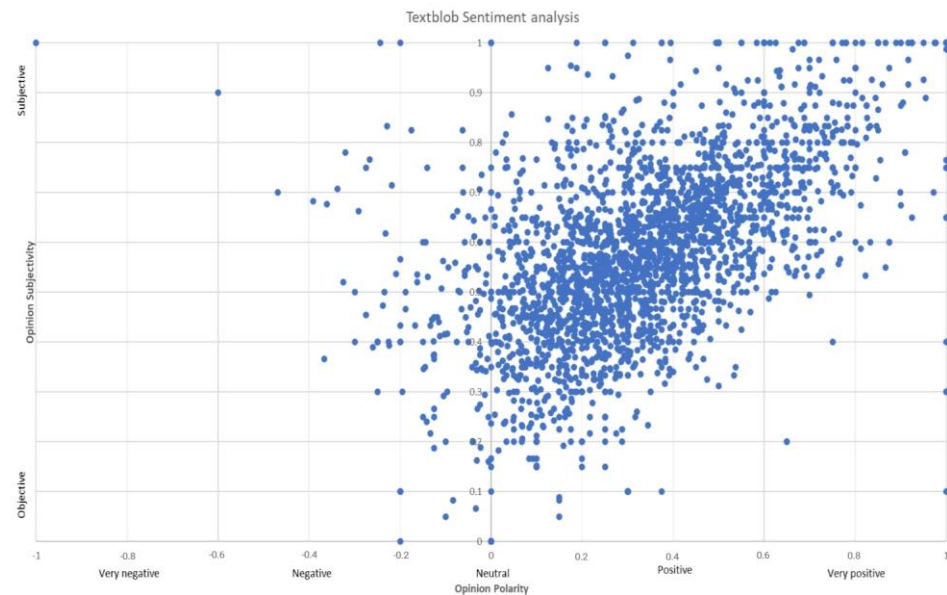
Figure 6. Textblob sentiment analysis results.

4. Discussion

Our research sheds light on visitor trends in the city of Kastoria over time and demonstrates the potential of using big data and AI to enable small municipalities to monitor these trends effectively. This can help them tailor promotional activities to align better with their target audiences.

We also found that Kastoria's tourism traffic has not mirrored the growth observed nationally and globally since the COVID-19 crisis. This finding aligns with the results of [28] but stands in opposition to any optimistic image that may have been conveyed. Our analysis revealed differing attitudes toward the monuments visited and the evaluation of services, influenced by the nationality of the visitors.

Although the reviews are generally positive, focusing on Kastoria's strong points, such as natural beauty and monuments, in many cases, reviewers mentioned that they were unaware of the attractions available in the area. As shown in the word clouds (Figure 1), the prominent words in English are "lake", "Kastoria", and "cave"; in Greek, they are "lake", "worth", and "cave"; and in German, they are "lake", "location", and "Kastoria". Moreover, TripAdvisor is lacking key assets of the region, as none of the Destination Management Organizations (DMOs) have ensured their inclusion. This gap in communicating the region's tourism offerings is evident, highlighting opportunities for professionals working in tourism communication to address and capitalize on these shortcomings.

We also identified distinct patterns in the photos users post on TripAdvisor compared to other social media platforms. Specifically, nearly 90% of the photos in our sample did not feature the users themselves or their close companions. This can be explained by the nature of TripAdvisor, mainly utilized to present attractions and help others with valuable information for other travelers, whereas Instagram would be more relevant in that regard for selfies and self-promotion.

Lastly, our findings on the use of AI reflect both its potential and its limitations at present. While there is a lot of tools and free libraries that allow researchers to take advantage of AI in order to ease their work, the quality of LLMs might be relatively low and full of biases—especially for less common languages like Greek. For instance, TextBlob

assigned a very negative polarity score of −1 to the comment, "Panoramic view of the city and the whole lake!! magic!!!! Visit it for sure for photography! Perfect for climbing on foot or else by road! There is also a perfect cafe to rest in a warm atmosphere and drink a hot coffee in all this cold!!!!!" despite its overtly positive sentiment. On the other hand, the following comment was mistakenly assigned a strongly positive polarity score of 1: "The cave beautiful though small! The guide told us almost nothing, he ran until we got to the end, let us take 2 pictures and gave us directions on how to get to the exit! The shortest tour ever!" As for the discrepancies between the user review ratings and the polarity scores from the review texts, it seems that because of the language nuances it is very challenging for AI tools to decide if a review is good or very good, bad or very bad and most importantly neutral or good/bad. However, 92% reviews are rated by users as good or very good, and around 70% of reviews are categorized by the AI tool as good or very good which can be considered as acceptable. Despite these discrepancies, in the fields of natural language processing, sentiment analysis and computer vision, the state-of-the-art capabilities of AI truly make it a very promising and valuable tool for any researcher.

5. Conclusions

The proposed method allows for the analysis and measurement of the image perceived by travelers as transmitted by other travelers (eWoM). The number of OTRs analyzed (11,328 in total from 137 listings in the Kastoria region, with 2577 specifically in the attractions category) constitutes a robust sample size; therefore, reliable insights can be generated, and actionable business intelligence can be derived.

The method is reliable for several reasons: quantitative content analysis of stored big data has a low probability of error. The source of information, user-generated content (UGC), is widely regarded as reliable. Furthermore, the abundance of readily available information, freely accessible on travel-related websites hosting travel blogs and online travel reviews (OTRs), makes data collection both extensive and straightforward. A significant advantage of UGC data research is its relative freedom from the biases often associated with questionnaire survey results, as the data originate from individuals who voluntarily express their opinions rather than responding to structured survey questions, ensuring more authentic and uninfluenced input for analysis.

The analysis of the data focused on identifying the spatio-temporal distribution of reviews, as well as determining the most popular and best-rated attractions. Our findings revealed that the most popular attractions are concentrated along the lakeside road. This insight could be leveraged by destination management organizations (DMOs) to design targeted promotional strategies for less-visited attractions. The temporal dimension of the data also provides valuable information on the evolution of the tourist destination, including seasonal variations in visitor numbers. Additionally, content analysis of reviews could uncover recurring issues, such as inadequate opening hours, insufficient guide training, or the lack of information available in certain languages, allowing for timely interventions.

The case of Kastoria tourists' impressions shows how this approach can be applied to other destinations worldwide. Using UGC from platforms like TripAdvisor, researchers and DMOs can identify tourist behavior, preferences, and emotional responses to attractions. This approach, merging sentiment analysis, spatio-temporal trends, and content analysis, introduces a framework for understanding global tourism dynamics. It is thus able to show seasonal patterns or attraction underperformance, and these emotional responses are coupled with features, which is extremely important in comparing destinations and thus sending tailored marketing and communication.

In the future, conducting similar research on other platforms that host user reviews, such as Google Maps, would be beneficial, as many regional attractions are not listed on TripAdvisor. Expanding the scope to include these platforms would provide a more comprehensive understanding of the region's tourism landscape.

References

1. Soltani-Nejad, N.; Rastegar, R.; Shahriari-Mehr, G.; Taheri-Azad, F. Conceptualizing Tourist Journey: Qualitative Analysis of Tourist Experiences on TripAdvisor. *J. Qual. Assur. Hosp. Tour.* **2024**, *25*, 343–364. [CrossRef]
2. Pike, S.; Page, S.J. Destination Marketing Organizations and destination marketing: A narrative analysis of the literature. *Tour. Manag.* **2014**, *41*, 202–227. [CrossRef]
3. World Travel & Economic Impact Research. Available online: https://wttc.org/research/economic-impact (accessed on 21 December 2023).
4. Marine-Roig, E. Measuring destination image through travel reviews in search engines. *Sustainability* **2017**, *9*, 1425. [CrossRef]
5. Mahat, N.Z.D.; Hanafiah, M.H. Help me TripAdvisor! Examining the Relationship between TripAdvisor e-WOM Attributes, Trust towards Online Reviews and Travellers Behavioral Intentions. *J. Inf. Organ. Sci.* **2020**, *44*, 83–112. [CrossRef]
6. TripAdvisor. About Tripadvisor. Available online: https://tripadvisor.mediaroom.com/in-about-us (accessed on 10 January 2024).
7. Eurobarometer. *Flash Eurobarometer 499: Preferences of Europeans Towards Tourism*; European Commission: Brussels, Belgium, 2021. Available online: https://europa.eu/eurobarometer/surveys/detail/2283 (accessed on 20 December 2023).
8. Jeacle, I.; Carter, C. In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Account. Organ. Soc.* **2011**, *36*, 293–309. [CrossRef]
9. Li, Y.; He, Z.; Li, Y.; Huang, T.; Liu, Z. Keep it real: Assessing destination image congruence and its impact on tourist experience evaluations. *Tour. Manag.* **2023**, *97*, 104736. [CrossRef]
10. Chon, K.S. The role of destination image in tourism: A review and discussion. *Tour. Rev.* **1990**, *45*, 2–9. [CrossRef]
11. Phelps, A. Holiday destination image-The problem of assessment. *Tour. Manag.* **1986**, *7*, 168–180. [CrossRef]
12. Xiang, Z.; Du, Q.; Ma, Y.; Fan, W. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tour. Manag.* **2017**, *58*, 51–65. [CrossRef]
13. Arabadzhyan, A.; Figini, P.; Vici, L. Measuring destination image: A novel approach based on visual data mining. A methodological proposal and an application to European islands. *J. Destin. Mark. Manag.* **2021**, *20*, 100611. [CrossRef]
14. Deng, N.; Liu, J. Where did you take those photos? Tourists' preference clustering based on facial and background recognition. *J. Destin. Mark. Manag.* **2021**, *21*, 100632. [CrossRef]
15. Urry, J.; Larsen, J. *The Tourist Gaze 3.0*; Sage: London, UK, 2011.
16. Stylianou-Lambert, T. Tourists with cameras: Reproducing or producing? *Ann. Tour. Res.* **2012**, *39*, 1817–1838. [CrossRef]
17. Lee, A.H. What does colour tell about tourist experiences? *Tour. Geogr.* **2020**, *25*, 136–157. [CrossRef]
18. Chen, Y.; Sherren, K.; Smit, M.; Lee, K.Y. Using social media images as data in social science research. *New Media Soc.* **2023**, *25*, 849–871. [CrossRef]
19. Wang, J.Z.; Zhao, S.; Wu, C.; Adams, R.B.; Newman, M.G.; Shafir, T.; Tsachor, R. Unlocking the Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion. *Proc. IEEE* **2023**, *111*, 1236–1286. [CrossRef] [PubMed]

20. Maghraby, T.M.; Elhag, A.E.; Romeh, R.M.; Elhawary, D.M.; Hassabo, A.G. The psychology of color and its effect on branding. *J. Text. Color. Polym. Sci.* **2024**, *21*, 355–362. [CrossRef]

21. Muratbekova, M.; Shamoi, P. Color-emotion associations in art: Fuzzy approach. *IEEE Access* **2024**, *12*, 37937–37956. [CrossRef]

22. TripAdvisor. Kastoria—Sightseeing. Available online: https://www.tripadvisor.com.gr/Attractions-g315844-Activities-oa0-Kastoria_Kastoria_Region_West_Macedonia.html (accessed on 10 December 2023).

23. Stemler, S. An overview of content analysis. *Pract. Assess. Res. Eval.* **2000**, *7*, 17. [CrossRef]

24. Python. Python 3.12.4. Available online: https://www.python.org/downloads/release/python-3124/ (accessed on 12 June 2024).

25. TextBlob. Available online: https://textblob.readthedocs.io/en/dev/ (accessed on 10 December 2023).

26. Yoo, K.-H.; Sigala, M.; Gretzel, U. Exploring TripAdvisor. In *Open Tourism*; Egger, R., Gula, I., Walcher, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 239–255. [CrossRef]

27. Tsakiroglou, V. Koronoios-Kastoria: The Drama of fur—Why the Virus Is Mowing in the Region. *First Theme*. 2020. Available online: https://www.protothema.gr/koronoios-live/article/993568/koronoios-kastoria-to-drama-tis-gounas-giati-therizei-o-ios-stin-periohi/ (accessed on 8 April 2020).

28. Institute of the Association of Greek Tourism Enterprises. Region of Western Macedonia. Annual Report on Competitiveness and Structural Adjustment in the Tourism Sector for the Year 2022. Available online: https://insete.gr/wp-content/uploads/2020/05/23-12_Western_Macedonia.pdf (accessed on 10 November 2024).