

Article

Sentiment Analysis to Measure Quality and Build Sustainability in Tourism Destinations

Fernando Borrajo-Millán ^{1,*} , María-del-Mar Alonso-Almeida ¹, María Escat-Cortes ¹  and Liu Yi ²

¹ Business Organization Department, Faculty of Economics and Business Administration, Autonomous University of Madrid, 28049 Madrid, Spain; mar.alonso@uam.es (M.-d.-M.A.-A.); maria.escat@uam.es (M.E.-C.)

² School of Tourism Management, Sun Yat-sen University, Guangzhou 510275, China; liuyi89@mail.sysu.edu.cn

* Correspondence: fernando.borrajo@uam.es

Abstract: The models used for analyzing and measuring quality in tourist destinations are changing with the incorporation of new techniques derived from data science and artificial intelligence. Recent studies show how social media and e-word of mouth (e-WoM) are playing key roles in the perception and image diffusion of tourist destinations. Thus, it is necessary to look for new methods for analyzing the tourist management and attractiveness of tourist spots. This includes conducting a sentiment analysis of tourists that modifies former research methods based on previously proposed model, supported by a survey, which obtained predefined and incomplete results. This study analyzed the quality of tourism in Spain, a major tourist destination that is considered to be the country with the greatest tourist competitiveness according to the World Economic Forum, and in China, the country with the greatest level of development and potential. A sentiment analysis was carried out to measure the quality of tourist destinations in Spain, and this involved three challenges: (1) the analysis of the sentiments of Chinese tourists obtained from e-WoM; (2) the use of new models to measure the quality of a destination based on information from Chinese social networks, and (3) the use of the latest artificial intelligence analytical technologies. Our findings demonstrate how sentiment analysis can be a determining factor in measuring WoM and identifying areas of development in tourist destinations in order to build a more sustainable destination. The results includes the following aspects: (1) the use of real images with more empirical evidence, (2) the use of impressive and disappointing sentiments, (3) a “no comment status”, (4) elimination of stereotypes, and (5) the identification of new opportunities and segments.

Keywords: tourist attractions; tourist experiences; quality of tourist destination; sentiment analysis; WoM; e-WoM; social media; tourism management



Citation: Borrajo-Millán, F.; Alonso-Almeida, M.-d.-M.; Escat-Cortes, M.; Yi, L. Sentiment Analysis to Measure Quality and Build Sustainability in Tourism Destinations. *Sustainability* **2021**, *13*, 6015.
<https://doi.org/10.3390/su13116015>

Academic Editors: Susan Ryan and Jacob Bethem

Received: 14 April 2021

Accepted: 15 May 2021

Published: 26 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many studies conducted in the last 40 years have examined the quality of tourist destinations at the level of tourist spots by using different theoretical models [1,2]. In the last 12 years, three aspects have presented new challenges: (1) the transfer to new types of international tourists; (2) new ways of contracting trips due to e-commerce and social networks, and (3) the use of new study methodologies. In terms of new international tourists, the recent and progressive opening of China has resulted in new types and a greater number of clients [3]. The modification of behavior guideline changes previously involved quality attributes.

Chinese tourists have altered previous research results on tangible variables [4] in destination infrastructures, tourist administration, historical-cultural tourist attractions, other local attractions, and social, political, security, and even environmental aspects. Psychosocial models [5] also have different applications for Chinese tourists.

The transformation of contracting and consumption models derived from e-commerce and the use of online travel agencies (OTAs) have modified variables such as quality of

perceptions. There has been a move towards the use of WoM, the extension of brands and reputation in social networks, and the formation of new relationship models established with e-commerce and its platforms. Quality models at tourist destination has been altered due the influence of eReferral sharing behavior.

Lastly, the use of new technologies for the treatment of enormous amounts of data collected on social networks and in e-commerce has led to the development of new analytical models of tourists' quality perceptions. In this sense, analytical techniques and the establishment of results have been modified. The use of artificial intelligence techniques, such as machine learning, data science, and the application of sentiment analysis, allows the use of information from natural [6] and written language [3,7], as well as the inclusion of images [8]. In this research, large amounts of data are analyzed in unstructured or text format or in a language that is difficult to treat in text mode; namely, Chinese. The use of these techniques modifies the research system, which is not based on a previous model but is achieved by real contributions from social media users.

In this study, three challenges are addressed: the analysis of the feelings of Chinese tourists in the country with the highest level of tourist competitiveness in the world according to the World Economic Forum [9], Spain; the use of models based on information offered on Chinese social networks, and the use of the latest analytical technologies, artificial intelligence (AI), and sentiment analysis.

This paper is organized as follows: Section 2 presents an analysis of a literature review. Section 3 introduces the research methods and data resources. Section 4 presents the results and findings, and the paper finishes with the presentation of conclusions, limitations, and directions for future research.

2. Literature Review

The literature review analyzed the service quality models applied to WoM and sentiment analysis application studies applied to the tourist service. The service quality model applied to e-WoM was used to establish a previously proposed conceptual model, while the application of the sentiment analysis was carried out to reduce data and introduce new techniques, a process that usually offers results.

The revision of contents, models, and techniques was crucial for this research project. The variables used in the models were mainly based on society, the environment, and the defense of cultural heritage. These are the basis of sustainability in tourism destinations and the tourism industry [10].

2.1. Quality of Service and WoM Models in Tourism

There is an extensive body of literature on the quality of a destination, and on quality, safety, and the environment (QSE) as a specific application of the quality of service. In a bibliometric study [1], three dimensions were proposed: the economic, sociocultural, and environmental dimensions. Tourists' expectations data was collected prior to the experience, and after the experience, perceptions of the tourist services at the destination and intentions to return and recommend the destination were investigated. In addition, the visitors' personal experiences and what the visitors transmit by word of mouth (WoM) were investigated.

Many different variables interact in the conceptualization and modeling of factors, categories, and elements. These include those that are involved prior to the experience in the tourist destination as well as repetition factors, e-WoM, and loyalty. Research relates objective quality and perceived quality [11,12], classifying them into three categories: psychological or intangible, conative or behavior and attitude, and functional or tangible.

Models based on the perception of real and tangible variables focus on concrete and measurable aspects, such as tourist attractions, tourist facilities and services, and the presence of tourists [13]. In these studies, the perceptions and future behaviors of visitors are analyzed. Other models [14,15] classify the tourist-destination experience in terms of service and infrastructure, destination environments, and the preservation of local

resources that result in improved quality and value, as well as determining the tourists' intention to return.

In these models, variables that are derived from tourist and resident perceptions are measured in addition to the joint study of both variables [15]. In this sense, the authors in [16] indicated the influence that residents have as a factor in tourism quality at a destination through a cost and benefit analysis using event attachment and community concerns. This factor can also be influenced by the overtourism of tourist spots, residents' quality of life, crowding [17,18], and administrative quality, which, in conjunction with historical authenticity and experience quality, determine the perceived value, travel, revisit intention, and preservation attitude associated with a destination [19].

The incorporation of intangible variables—psychological, behavioral, attitude, image, and brand quality—is also decisive in establishing the quality of a destination. In [20], the sequence of image–quality–satisfaction was defined for tourist satisfaction. This is the basis of models of destination loyalty and image [5], which are used to establish the satisfaction level of tourists as a result of after-purchase behavior, recommendations, and return. They are also the basis of models of behavioral intention based on previous measures of the effects of customer loyalty indices (CLI) on behavior and studies of WoM and e-WoM [21].

Destination loyalty models [22] include destination image and destination loyalty based on three aspects: directly influenced attribute satisfaction, attribute satisfaction as a direct antecedent of overall satisfaction, and overall satisfaction and attribute satisfaction. Previous research [23–25] based on the experience economy in tourism shows that positive WoM, revisit intention, and loyalty are the result of a QSE based on hedonics, peace of mind, involvement, recognition, and perceived risk.

Following the incorporation of social networks and the extension of tourism contracting by platforms and other e-commerce systems, aspects of brand quality, WoM, and e-WoM are extended [26–30]. This is demonstrated by the effects of social media, through the usage at the cognitive level and how affective images [31], brand quality, and customer value determine WoM and e-WoM [32], promoting tourist engagement and sharing intentions to visit the destination in the interactive eReferral platform, which also results in eReferral sharing behavior.

A great qualitative and quantitative leap has occurred in recent years as a result of two effects [19,26,33,34]. The first is the use of large amounts of e-WoM data that offer more quantitative and qualitative information than data that are traditionally obtained through surveys and other official data sources. The second effect is derived from the use of new analytical techniques that add other points of view to those previously considered. The utilization of these new models has led to a new conceptual model dimension and to a new way of understanding behavior and quality at a destination due to the complex results obtained by using a large number of observations, variables, and techniques, resulting in many new analyses and types of results.

Thus, our research proposition (RP) is that sentiment analysis can be used as a relevant tool to measure WoM and identify possible areas of improvement in tourist destinations to build more sustainable and resilient destinations. The aforementioned results and techniques are discussed and applied in this research to analyze the RP.

2.2. Sentiment Analysis Applied to Quality and Tourism: Comparative Analysis

Sentiment analysis is a set of consolidated techniques with multiple research papers published in many fields [7], including management and tourism [2]. Even though there has been a significant amount of research on sentiment analysis in tourism, there are restrictions that have limited its development, although in recent years, this area has shown strong development. These limitations are precisely what characterize this type of research and its subsequent publications.

The first limitation is the ability to obtain data [35]. Data are often obtained from a survey, which limits the quality and quantity of the data and the technology used, as well as being outside the models used this research. Limited tourism publication data based on

geotagged spatiotemporal, social, and sentiment analyses have been collected [36]. The use of extraction and tracking to obtain data, the length in terms of characters, and the context can make it difficult to obtain a good tourism database.

Researchers use a very limited number of tourism databases, which are namely, TripAdvisor (tool NLTK), Expedia, Booking, Airbnb, Twitter, Amazon, Facebook or Academic Yelp Dataset [2,35,37]. In the case of tourism databases in China, these are obtained from the main OTAs and the social networks used there [3,33,35,36,38]: Baidu Travel, Ctrip Travel, Tongcheng Travel, Qunar Travel, Tu Niu, Qiong You, and Sina Weibo. As a result of these limitations, studies in this area are usually limited to a certain tourist spot or a specific area, and there only a few cases where an investigation has focused on cross-cutting aspects that affect several countries [2,7].

The second limitation is the methodology. The four most commonly used methodologies are dictionary-based methods, machine learning methods, deep learning methods, and hybrids of previous methods [2,33,35–39]. All methodologies require the use of multiple combined technologies, the use of large unstructured databases, and the difficult reduction and acquisition of results. The use of different languages with different dictionaries, semantic rules, and contexts introduces complexity into the process and analysis.

The third limitation on this type of publication is the way results are interpreted, due to the study base and technologies used.

In any case, the use of sentiment analysis techniques to study the quality of the tourist service has clear advantages over previous studies [7]. The main advantages derive from the lack of biases and errors produced by the information, modeling, and data treatment of the previous models. The application to real data has proven benefits that are achieved in all disciplines of data science, machine learning, and artificial intelligence.

3. Research Methods and Data Resources

Based on a review of techniques and data used for sentiment analysis, the authors in [2,3,19,33,38] set two great challenges. The first challenge was to determine how to use the large amounts of data straight from databases, websites, social networks, and e-WoM data. This type of study surpasses classical ones based on surveys and introduces new elements of complexity into data analysis, such as the use of big data and AI techniques including machine learning and neural networks. The results, likewise, show additional effects to those measured with the traditional systems used in previous models.

The second challenge was to use new analytical tools based on sentiment and language analyses. The new techniques are added to statistics, such as the support vector machine (SVM), naive Bayesian (NB), dynamic language model classifier, and lexicon matching methods, or a combination of models is used (SVM with others). Solving these challenges is one of the contributions of this research.

3.1. Data Sources and Information Analysis

This research used a huge amount of data obtained directly from China's specialized tourism social networks from the websites of the largest OTA of China. These were, in alphabetical order, Baidu Travel, Ctrip, Mafengwo, and Qunar, the largest tour operators offering travel packages to visit Spain. Most of the travel packages (75%) are offered in combination with the possibility of traveling to other nearby countries in Europe and North Africa, specifically Morocco. The resulting data offer information that contributes to the analysis of feelings in combination with the offer of travel packages to destinations, allowing data on the quality of tourist destinations in the main tourist spots in Spain from the perspective of inbound Chinese tourists to be collected. This study did not analyze cross sentiments with other countries or other tourist destinations included in the packages that probably influence the total experience and allow cross comparison.

The data obtained from the users of social networks and websites, known as user-generated content (UGC), represent a vast amount of unstructured information which, in many cases, is not directly related to research. In this study, big data techniques were

used to process the enormous amount of information generated by online reviews. The first problem to solve was that the reviews are written in Chinese, while most of the previous analyses conducted with UGC data were performed in English, which has different semantic rules and lexicons.

For the treatment of the data obtained from the UGC of the main OTAs, the lexicon, specialized semantic rules, and lexical filters of the tourist sector were built to categorize emotionally positive or negative sentiments in each sentence. The position of each word in each sentence was taken into account.

Regarding the lexicon analysis, the How Net dictionary was chosen. This has more than 90,000 words—more than other lexicons used, such as the Word Net, Senti Word Net, or VADER. This lexicon was complemented by the thesaurus of the World Tourism Organization, which contains about 8200 terms and 20 hierarchical semantic fields, which are classified into five levels and represent tourist activities in French, English, and Spanish languages. Finally, it was necessary to add new words and expressions that were not previously included to obtain better results, because those words are not classified in the How Net dictionary as sentiments, despite being used in common language.

3.2. Sentiment Analysis for the Quality of Tourist Destinations: Research Method

In sentiment analysis, two groups of methods are used: lexicon-matching and corpus-based machine learning methods [33]. These methods allow the processing of large amounts of information, from language to unstructured text. This information is grouped by semantic logic and word combinations to obtain classifiable feelings in clusters.

Many of the recent tourism sentiment-based studies have used these techniques to conduct an emotional analysis of the image and perception quality of a tourism destination using data obtained from text written by users. To carry out this investigation, an analysis was conducted using the first method, the lexicon-matching and semantic rules method, to complete a network analysis of tourists' sentiments and preferences.

To do this, in the first step, a measurement was made for each sentence of the data obtained from the UGC at the levels of emotions and sentiments. This was performed through identification and cross analysis with a set of semantic logics. This step provides a positive or negative score for the sentiments contained in each phrase, which is assigned to words or word combinations.

The second step consists of the treatment and initial classification of the scores indicating positive and negative emotions. To do this, the scores were added to determine the sentiment inclination, thus completing the filtering of the information. For the filtering process, the Linguistic Inquiry and Word Account (LIWC) program was used. With this program, three analyses were performed: the set of words was separated, the words were filtered with the tourist lexicon, and the positive and negative emotional scores of the reviews were calculated.

The third step consisted of a second filtering and the treatment of the results. For this, the Gephi program was used, and more advanced results were obtained through a routine in which two processes were conducted. In the first process, a score was calculated using the higher-frequency words. In the second process, an analysis of the co-occurrence of the most mentioned words was performed and a cluster network was built.

The results obtained with these research techniques are better than those obtained with other methods because a network analysis of tourist preferences based on sentiment analysis can be obtained, including positive and negative emotions. This method is also better because the analysis can determine the frequencies of the words and the co-occurrence strengths of interconnected objects under different circumstances.

3.3. Contextual Analysis: Spain's Position in Chinese Outbound Tourism

Despite the discontinuity caused by COVID-19 regarding the entry of Chinese and international tourists into Spain in 2020 (UNWTO [40]), China's tourism has enormous potential. Chinese tourism in Spain is low, with 699,000 Chinese residents visiting Spain

in 2019 (0.8% of the 83.7 million foreign tourists who visited in Spain in 2019), although Chinese tourism has greatly developed since 2015, when Spain received 399,000 Chinese tourists (0.3% of the total) [41].

The number of Chinese tourists is low due to three factors: (1) although Spain was the first country in the world in terms of the competitiveness of the tourism sector in recent years according to the WEF as well as the second greatest recipient of tourists in the world after France (UNWTO [42]), it does not hold this position in the European ranking of inbound Chinese tourists; (2) the slow opening of the Chinese population to international tourism, and (3) the historical restriction of touristic offer in terms of suppliers and destinations, whereby travel packages traditionally offered Spain as an additional destination of the package. All of these restrictions are slowly being removed.

The characteristics of Chinese tourists in 2019, according to COTRI [43], were urban and upper-middle class, although those who came to Europe were of the highest class; of an intermediate age (25–44); carried out 90% of their travel in Asia; mainly engaged in tourism activities involving cultural activities, shopping, and urban activities, and traveled on a different tourist calendar from the European one.

The development of the Chinese tourism market in Europe has mainly concentrated on western Europe and northern Europe. The exploitation of tourism business in eastern Europe is not apparent, and Chinese tourists' destinations are mainly Britain, France, Germany, and Austria.

The motivation for Chinese citizens to travel to Europe appears to vary, from sight-seeing to leisure vacations. As there are differences in the level of economic and cultural development and in the diversity of residents' lifestyles, people's psychological pursuit to seek new and different ideas has developed. In addition, many new tourism products, such as academic travel and study tours, have been developed.

Spain is the oldest country in Europe. Its tourism industry is well developed, and it has abundant tourism resources. The climate is pleasant, and spring and autumn are ideal tourist seasons for Chinese tourists. According to online OTA sales data, Spain's most popular tourist products (or selling points) that are favored by Chinese tourists include sunshine coast cities, sports events (La Liga Football League), bullfighting, ethnic dance, ancient architecture, the hometown of the guitar, and religious culture. The quality of Spanish tourism products is relatively high, and most products involve in-depth tours. Still, Spain ranks lower among the popular tourist destinations in Europe, and the reasons for this are worth studying.

Spain enjoys the unique climate of Mediterranean areas with dry summers and moderate temperature winters. It is one of the hottest areas in Europe, and its hot season spans from July to September, a period in which a large number of tourists visit the country. As this period overlaps with the summer holiday in China, there are special promotions during this period, and flight ticket pricing is decreased to some extent.

We created a heat map based on the number of tour packages offered by OTAs, which indicates that certain cities in central, eastern, and southern Spain play leading roles in all tourism products (see Figure 1). Those areas have famous attractions with a high resource quality and mature urban transportation infrastructures.

This map, built from the packages made by the Online Travel Agencies of China, has historically been supervised by the Chinese authorities. It serves to collect specific information on tourist spots. The results of the search for comments on tourist spots or cities are shown in Figure 2. The following section details how these comments were obtained.

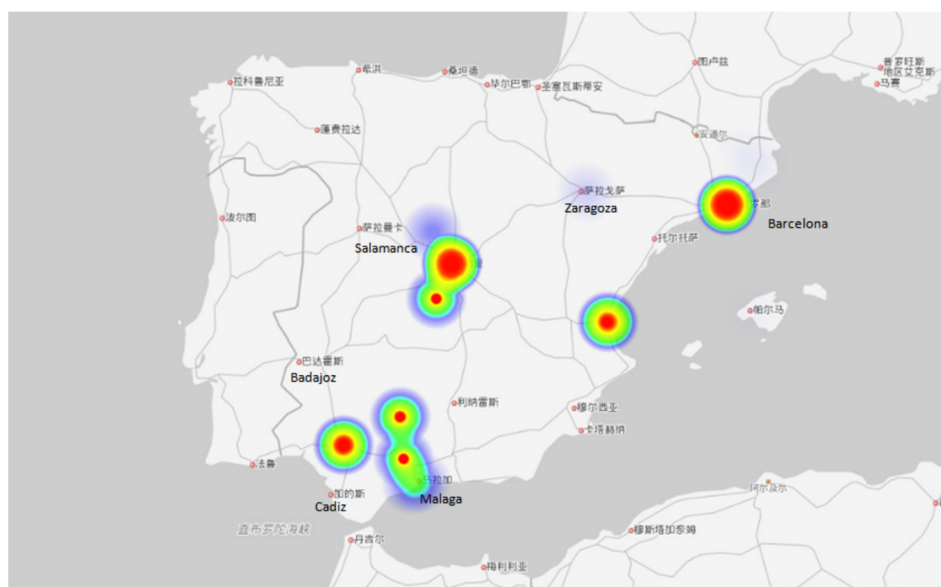


Figure 1. Heat map of popular tourism destinations in Spain.

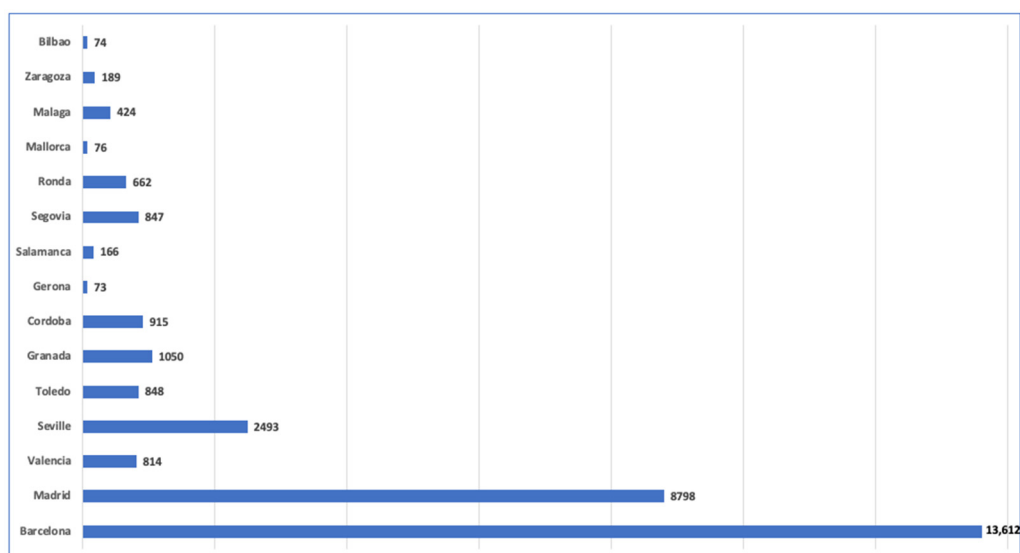


Figure 2. Results of comments obtained on cities in Spain.

3.4. UGC Sentiment Analysis of Spain

This study collected online tourist reviews from some of the main OTAs in China, including Ctrip, Baidu Travel, Qunar, and Mafengwo. The destinations targeted were 18 representative tourism cities in Spain, including Madrid, Barcelona, and Seville. By applying Gephi to visualize comment data, an emotional semantic network was constructed.

The data were obtained from 38,337 reviews of 18 Spanish tourist cities, obtained by tracking the comments of the main Chinese travel websites: Biadu Trip, Qunar, Mafengwo, and C-trip. Through Gephi, the sentiment inclinations of the tourist reviews can be visualized through the emotional scores filtered by the LIWC program with the How Net dictionary.

The most frequent topics contained in the comments used for the sentiment analysis were architecture (44.69%), the city environment (31.65%), museums (14.45%), and football stadiums (5.24%). The most commonly mentioned cities were, in order, Barcelona, Madrid, and Seville. These cities accounted for almost 80% of the comments. A summary of the comments by category can be seen in Table 1.

Table 1. Comments by identified and classified categories.

Comments	Number	Percentage
Architecture	13,871	44.69%
City environment	9824	31.65%
Stadium	1625	5.24%
Monuments	879	2.83%
Museums	4,486	14.45%
Towns	295	0.95%
Restaurants	53	0.17%
Stores	8	0.03%
Total	31,041	103.50%

The most frequently mentioned words and those with co-occurrence were classified as three types: general and neutral results, results related to positive aspects, and negative results. Correctly identified and classified comments can be seen in Table 2.

Table 2. Comments by identified and classified sentiment types.

Comments	Number	Percentage
Total	29,991	100.00%
Positive	25,610	85.39%
Neutral	792	2.64%
Negative	3592	11.98%

3.4.1. General Sentiment Analysis Results

The results identified five clusters, each with a relative strength in a central topic that operated as the center of its cluster, as shown in Figure 3. The five relevant clusters were buildings, Spain in relation to the main cities, local in relation to visits, church in relation to museums, and other specific spots.

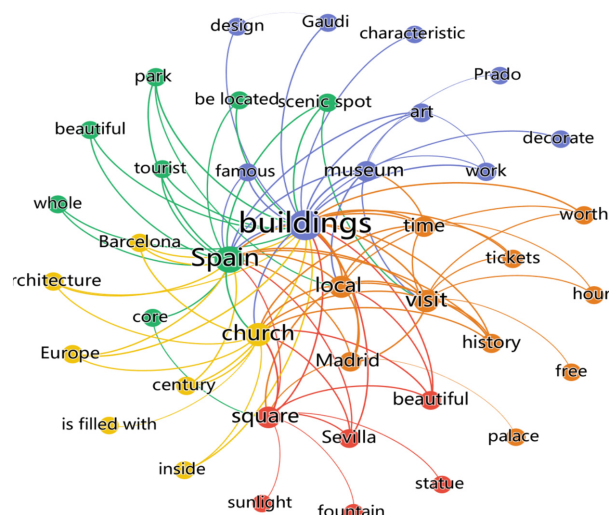


Figure 3. Comments on Spanish tourism cities.

In cluster 1, “buildings” is the center of the image, which means that among all tourist attractions in Spain, Chinese tourists have the strongest feelings towards buildings. Chinese tourists, as previously mentioned, travel to Europe and Spain for cultural, gastronomic, and shopping tourism. Therefore, it is logical that the first thing to appear is buildings and architecture. Tourists attach great importance to this topic not only because these buildings have special features and artistic value, but also because they have historic and religious cultural value.

Clusters 2 and 3, “Spain” and “local”, are related to the most important tourist spots. The results show that the most frequently commented-on spots are in Madrid, Barcelona, and Seville, with small differences among the three sites, as highlighted by the most commented-on places.

Clusters 4 and 5 represent the nodes “church” and “museum”. These clusters explain the tourist attractions and places with the greatest impact, which are also the most visited. They are strongly linked to cluster 1 and clusters 2 and 3, where the visited sites are located.

To allow for a better understanding of the results, in the graphs, (1) the area of the nodes shows us the size of the result at a given point (2) the unions between the nodes show us their relationships with the two previous points, (3) the clusters and their relationships are structured, and (4) the distance from the center of the graph represents the statistical distance from the cluster.

Most tourists want to spend their time experiencing the culture behind these buildings, such as the historical changes in Baroque and Gothic architecture, the religious status of the Christian King’s Castle, and the legends of ancient architects. On the basis of the above analysis, the perceived image of Chinese tourists of Spain is that it is artistic, legendary, honorable, and elegant, a place where tradition collides with modernity.

In summary, (1) among the main tourist cities in Spain, Barcelona, Madrid, and Seville are the top destinations for Chinese tourists. (2) Barcelona is widely recognized as a must-visit place for Chinese tourists. Its buildings, which symbolize typical European cultural elements such as Gaudi, Picasso, and Columbus, are major foci of tourists. (3) Chinese tourists are shocked by the Royal Palace of Madrid, and they tend to show great interest in Madrid’s royal culture. (4) The Giralda Tower near the Seville Cathedral is favored by Chinese tourists, as it provides them with the opportunity to overlook the whole city. Wooden parasols are also attractive to tourists, as their modern style contrasts sharply with the surrounding traditional buildings. (5) Although Spain’s football culture is famous around the world, its related tourist attractions are ignored by Chinese tourists. (6) Chinese tourists’ degree of acceptance of the ticket price of Spain’s cultural attractions is relatively high. Most tourists believe that these attractions are worth the cost. It is also noticed that the “free ticket” strategy used at certain times is highly attractive to tourists. (7) The general education level of Chinese tourists to Spain is high. This group of people includes business tourists and study-oriented tourists, who have a strong desire to gain knowledge from attractions.

3.4.2. Positive and Negative Results

The five main clusters of positive results, shown in Figure 4, coincide with the previous general analysis nodes: “architecture”, “Spain”, “local”, “Madrid”, “church” and “square”, and “history”. In the first cluster, the central node of positive reviews in the Gephi image is also buildings which, in this case, are described as architecture.

The second, third, and fourth clusters—“Spain”, “local”, and “Madrid”—coincide with the vision of cities being centers of tourism. What is new, although it was identified in the general results, is that Madrid is associated with more positive feelings than Barcelona, despite being a tourist destination with a greater number of travelers and accumulating a greater number of visits.

The fourth cluster, which integrates two nodes, “church” and “square”, and the fifth, “history”, reinforce, together with other smaller nodes, the cultural and historical interests of Spain. Chinese tourists are less interested in sun and beach tourism and nature tourism.

In all cases, it was found that the frequencies, weights, and cases offer different results from those of the sentiment analysis, which reinforces the applicability of the analysis methodology used.

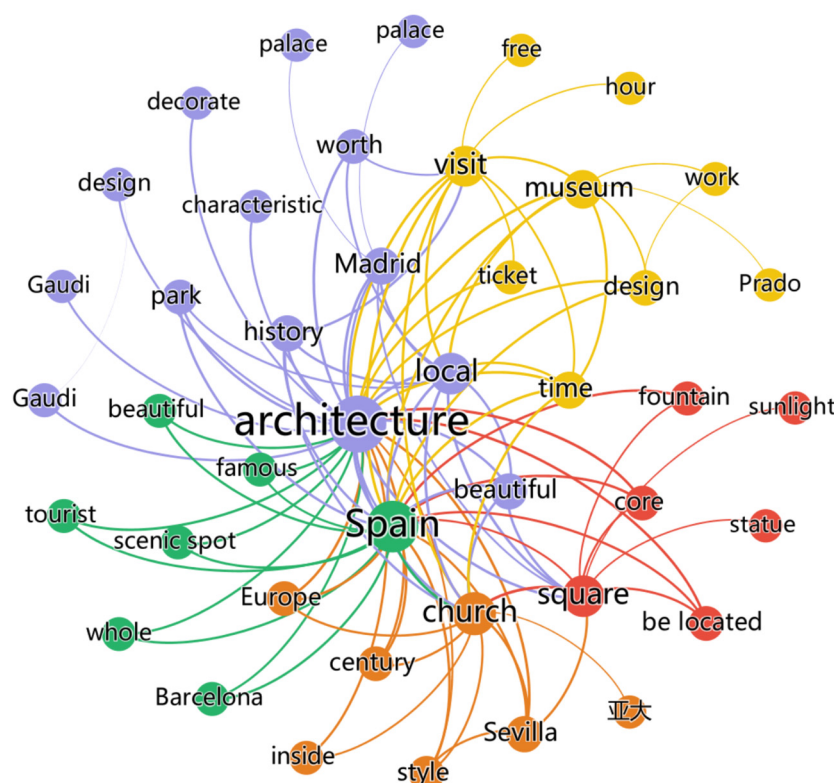


Figure 4. Positive comments on Spanish tourist cities.

Figure 3 shows the five clusters with similar topics to the general ones. Positive comments include the following: (1) Chinese tourists admire the rich history of Madrid's local buildings, including the magnificent Alcatel Gate and the urban change of Madrid. The Almudena Cathedral is known as a "humanistic picture scroll". (2) The famous architect Gaudi is regarded as a national treasure in Spain. Many Chinese tourists visit Spain just to see his work. (3) Intelligent facilities in museums have excellent audiovisual functions. In addition, ticket prices of museums are cheap. Of all museums, the Prado Museum has the most mentions in positive reviews. It is believed to be the best museum in Spain. (4) Chinese tourists comment relatively less on the religious meaning of churches. However, most affirm the aesthetic value of churches and are "shocked at the first sight".

Generally speaking, the nodes presented in Figure 4 are consistent with the structure of the general image. This indicates that most Chinese tourists have a positive attitude regarding Spain's main tourist attractions.

Negative aspects are shown in Figure 5. The nodes could not be extracted in a clear order, and they also occurred with much lower frequencies, which is indicative of the power of positive comments over negative ones. However, the same clusters were repeated, although with a different arrangement. The resulting clusters were “architecture”, “Spain”, “visit”, “church”, and “local”, which represent areas of improvement for the tourist destination.

It was observed that cluster 1, “architecture”, can be combined with clusters 2 and 5, “Spain” and “local”. However, for cluster 3, it can be seen why negative sentiments happen, as these nodes depend on tourist administration. Cluster 4, “church”, is separated from other nodes, such as “square”, and is smaller than the cluster of positive aspects. The reason may be related to religion and the influence that the Church has had in the Spanish history.

For the negative aspects, there is no obvious center node for the negative comments on Spain, and certain key nodes, such as "architecture", "visit", "Spain", and "church", appear frequently. This indicates that although there is no distinct problem in Spain's tourism

development, Chinese tourists still are not satisfied with some aspects, as follows: (1) Some tourists think that Spanish buildings are “old” and “boring” because of their “old age” and the “shabby” internal facilities. (2) Due to the cultural distance between China and Spain, a number of tourists show a lack of interest in the church architecture and historic culture. (3) Some tourists complain that their length of stay in Spain or in some Spanish cities is too short to allow them to visit museums, churches, squares, stadiums, and other main attractions, which causes regret. Most tourists use a package tour. Unreasonable and tight schedules set by travel are the focus of their complaints. (4) Descriptions related to “bullfighting” were found to be mostly negative. Chinese tourists have a strong interest in Spanish bullfighting culture, but because of a local bullfighting ban, the bullring is a site that fails to match the high expectations of visitors. The psychological gap is even larger due to the false advertisements about bullfighting on travel agencies’ websites. A minority of visitors express antipathy, because they consider bullfights barbaric and bloody.

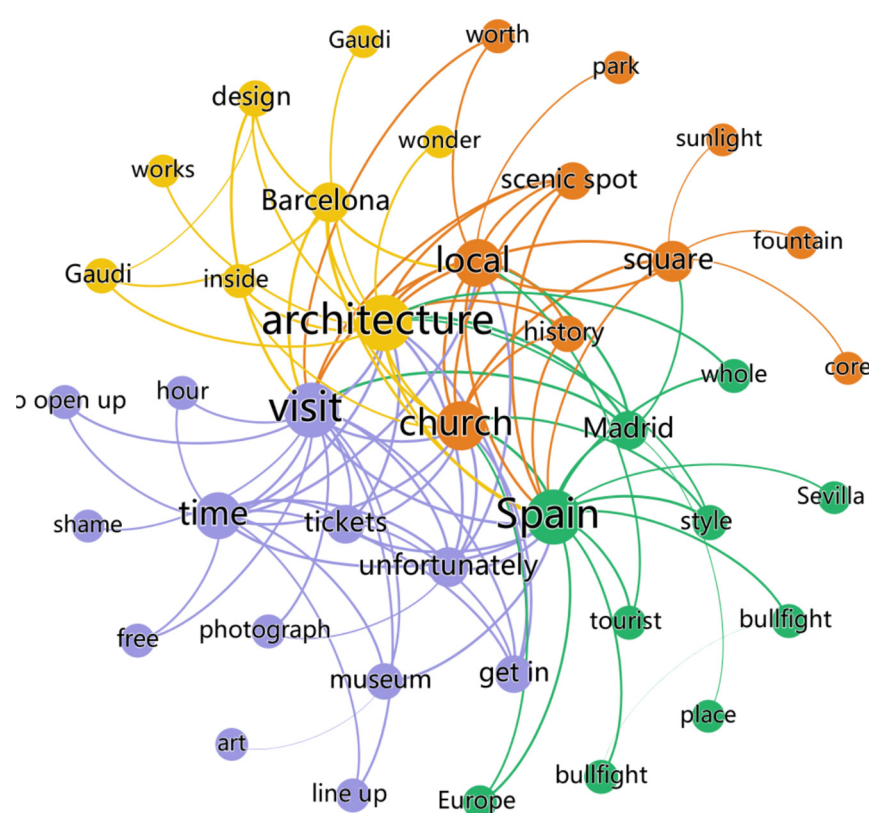


Figure 5. Negative comments in Spanish.

Therefore, the RP was tested. We can confirm that the sentiment analysis provides in-depth results about tourist destinations beyond mere comments. This type of analysis provides useful information that can be used to improve the attractiveness of a tourism destination.

4. Conclusions and Recommendations

It is widely known that tourism is a source of health, and there is competition to attract more visitors. In recent years, social media platforms have played an active role in the construction of the image of tourist destinations [2], creating an opportunity to identify areas of improvement and to reinforce or change perceptions about a certain destination.

As previous research pointed out [9,44], destination sustainability could be improved by means of flow control to avoid overtourism. Thus, the analysis of media data together with other strategic administrative measures could contribute to ensuring sustainability in a destination with a balance among economic, social, and environmental issues. Sentimental

analysis is a tool that can be used to improve the quality and sustainability of destinations, that provides tourism information and data for decision-making. It is very useful to achieve sustainable tourism, that avoids excess tourists, the degradation of natural and cultural resources, and the dissatisfaction of the citizens.

Therefore, in this section, some conclusions derived from the findings are presented. Additionally, some recommendations related to both academia and practitioners are highlighted, and limitations are noted.

First, e-WoM expands opinions about tourist destinations quickly in both positive and negative ways. Positive e-WoM can help to make a destination more visible and attractive. On the contrary, negative e-WoM can dissuade visitors, because social media plays an active role in the formation of the image of the tourist destination. Therefore, tourist destinations should monitor what people are saying about them on social media in order to act quickly and provide answers to reduce the negative effect of bad comments.

Second, sentiment analysis is a powerful tool that can be used to test the quality of a tourist destination. The main reason for this is that travelers speak a lot about things that are more impressive or disappointing to them. Disappointed comments open a window for the improvement of the destination. These comments could be a signal that changes in services, attractions, or infrastructures are needed to maintain the competitiveness of a destination. These comments could advise tourism companies about failures in service and give them the opportunity to correct them. Moreover, they could point out the necessity for changes in tour packages.

Third, a “no comment” status, neither positive nor negative, in a tour package shows little interest in certain sites, trips, or services. Therefore, this also provides evidence that tour operators and travel agencies should evaluate their tour packages and the information provided before the trip. Moreover, it gives tour operators the opportunity to find new places to include in their tour packages to gain competitiveness. It also provides an opportunity to manage the touristic flow better.

Fourth, contrary to the current idea in Spain that certain tourist groups visit Europe and Spain for shopping, sentiment analysis helps to eliminate stereotypes and expand the promotional resources with a real focus on the importance for potential travelers. Thus, the use of sentiment analysis can improve the quality and sustainability of a destination.

Fifth, sentiment analysis identifies new tourism niches, destinations, places within certain destinations, and the necessity for tourists. Therefore, this kind of analysis should be adopted for public tourism and business managers as a management tool. In fact, the current COVID-19 pandemic has almost eliminated traditional paper tourism guides. It has intensified the use of social media and the number of comments about destinations. Thus, sentiment analysis could be a critical way to redefine the new normal in tourism. This type of methodology could be used to measure changes in the tourism industry in the post-COVID-19 age. From an academic point of view, sentiment analysis can be applied to a number of research fields and can open new areas of analysis and findings. Therefore, it could be adopted in a multidisciplinary way.

This work is not free of limitations, mainly because the only source of data was social media. Enrichment of the sentiment analysis with different data sources, sociodemographic data, and data types such as text, photos, videos, or audio would give a more complete picture of the studied phenomena. Sections may be divided by subheadings. This should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Finally, regarding the present and future research on this area, new techniques and new ways of establishing models need to be developed. The future will involve the use of large amounts of unstructured data or data predicted from previously conducted surveys. In addition, data may be presented in natural language and sounds, in written language or texts, in images, or through a combination of all of these types of data. Advances in the treatment of big data through data science and unstructured databases is the first challenge. The next challenge is the analysis of these data using different techniques,

mostly in the disciplines of computer science and artificial intelligence. To do this, the use of machine learning, neural networks, and virtual agents and bots, among other related techniques, is proposed as the present and future direction for research on the quality of tourist destinations.

Author Contributions: F.B.-M., conceptualization, investigation, methodology, resources, visualization and writing. M.-d.-M.A.-A., funding acquisition, project supervision, conceptualization, investigation, methodology, visualization, and writing. M.E.-C., conceptualization, investigation, and writing. L.Y., investigation, methodology, resources, software, validation, and visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the program of the Autonomous University of Madrid–Santander Bank, grant number 2017/Asia01.

Data Availability Statement: Data mining and processing works were funding by the Humanity and Social Science Foundation of Ministry of Education of China (19YJAZH060).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Garrigos-Simon, F.J.; Narangajavana-Kaosiri, Y.; Narangajavana, Y. Quality in Tourism Literature: A Bibliometric Review. *Sustainability* **2019**, *11*, 3859. [CrossRef]
2. Guerreiro, J.; Rita, P. How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *J. Hosp. Tour. Manag.* **2020**, *43*, 269–272. [CrossRef]
3. Alonso-Almeida, M.M.; Borrajo-Millán, F.; Yi, L. Are Social Media Data Pushing Overtourism? The Case of Barcelona and Chinese Tourists. *Sustainability* **2019**, *11*, 3356. [CrossRef]
4. Murphy, P.; Pritchard, M.P.; Smith, B. The destination product and its impact on traveller perceptions. *Tour. Manag.* **2000**, *21*, 43–52. [CrossRef]
5. Bigne, J.E.; Sanchez, M.I.; Sanchez, J. Tourism image, evaluation variables and after purchase behaviour: Inter-relationship. *Tour. Manag.* **2001**, *22*, 607–616. [CrossRef]
6. Lv, X.; Liu, Y.; Luo, J.; Liu, Y.; Li, C. Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure. *Ann. Tour. Res.* **2021**, *87*, 103114. [CrossRef]
7. Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; Kuksa, P. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **2011**, *12*, 2493–2537.
8. Suna, J.; Lv, X. Feeling dark, seeing dark: Mind–body in dark tourism. *Ann. Tour. Res.* **2020**, *86*, 103087. [CrossRef]
9. Available online: <https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2019> (accessed on 30 November 2020).
10. Alonso-Almeida, M.M.; Rocafort, A.; Borrajo, F. Shedding light on eco-innovation in tourism: A critical analysis. *Sustainability* **2016**, *8*, 1262. [CrossRef]
11. Solís-Radilla, M.M.; Hernández-Lobato, L.; Callarisa-Fiol, L.J.; Pastor-Durán, H.T. The Importance of Sustainability in the Loyalty to a Tourist Destination through the Management of Expectations and Experiences. *Sustainability* **2019**, *11*, 4132. [CrossRef]
12. Sanchez-Gonzalez, G.; Gonzalez-Fernandez, A.M. The Influence of Quality on eWOM: A Digital Transformation in Hotel. *Manag. Front. Psychol.* **2021**, *11*. [CrossRef]
13. Terkenli, T.S.; Skowronek, E.; Tucki, A.; Kounellis, N. Toward understanding tourist landscape. A comparative study of locals' and visitors' perception in selected destinations in Poland and Greece. *Quaest. Geogr.* **2019**, *38*, 81–93. [CrossRef]
14. Goffi, G.; Osti, L.; Nava, C.R.; Maurer, O.; Pencarelli, T. Is preservation the key to quality and tourists' satisfaction? *Tour. Recreat. Res.* **2020**, 1–7. [CrossRef]
15. Biagi, B.; Ladu, M.; Meleddu, M.; Royuela, V. Tourism and the city: The impact on residents' quality of life. *Int. J. Tour. Res.* **2019**, *22*, 168–181. [CrossRef]
16. Ouyang, Z.; Gursay, D.; Chen, K. It's all about life: Exploring the role of residents' quality of life perceptions on attitudes toward a recurring hallmark event over time. *Tour. Manag.* **2019**, *75*, 99–111. [CrossRef]
17. Santos-Junior, A.; Almeida-Garcia, F.; Morgado, P.; Mendes, L. Residents' Quality of Life in Smart Tourism Destinations: A Theoretical Approach. *Sustainability* **2020**, *12*, 8445. [CrossRef]
18. Yin, J.; Cheng, Y.; Bi, Y.H.; Ni, Y.S. Tourists perceived crowding and destination attractiveness: The moderating effects of perceived risk and experience quality. *J. Destin. Mark. Manag.* **2020**, *18*, 100489. [CrossRef]
19. Liu, Y.; Huang, K.; Bao, J.; Chen, K. Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destination. *Tour. Manag.* **2019**, *71*, 337–347. [CrossRef]
20. Kotler, P.; Bowen, J.; Makens, J. *Marketing for Hospitality and Tourism*; Prentice-Hall: London, UK, 1996; ISBN 9780132784023.

21. Chen, C.; Tsai, D. How destination image and evaluative factors affect behavioral intentions? *Tour. Manag.* **2007**, *28*, 1115–1122. [CrossRef]
22. Chia, C.G.; Hailin Qu, H. Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tour. Manag.* **2008**, *29*, 624–636. [CrossRef]
23. Cervera-Tauleta, A.; Pérez-Cabañero, C.; Schlesinger, W. Experience destinations. *J. Bus. Res.* **2019**, *101*, 536–547. [CrossRef]
24. Schlesinger, W.; Cervera-Taulet, A.; Perez-Cabanero, C. Exploring the links between destination attributes, quality of service experience and loyalty in emerging Mediterranean destinations. *Tour. Manag. Perspect.* **2020**, *35*, 100699. [CrossRef]
25. Shaykh-Baygloo, R. Foreign tourists & experience: The tri-partite relationships among sense of place toward destination city, tourism attractions and tourists & overall satisfaction- Evidence from Shiraz, Iran. *J. Destin. Mark. Manag.* **2021**, *19*. [CrossRef]
26. Ye, Q.; Law, R.; Gu, B.; Chen, W. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Comput. Hum. Behav.* **2011**, *27*, 634–639. [CrossRef]
27. Tham, A.; Croy, G.; Mair, J. Social media in destination choice: Distinctive electronic word-of-mouth dimensions. *Social Media Tourirm Mark.* **2013**, *30*, 144–155. [CrossRef]
28. Chen, F.W.; Guevara Plaza, A.; Alarcón Urbistondo, P. Automatically extracting tourism-related opinion from Chinese social media. *Curr. Issues Tour.* **2017**, *20*, 1070–1087. [CrossRef]
29. Stojanovic, I.; Andreu, L.; Curras-Perez, R. Effects of the intensity of use of social media on brand equity. An empirical study in a tourist destination. *Empir. Study Tour. Destin.* **2017**, *27*, 83–100.
30. Pandey, A.; Sahu, R. Modeling the relationship between service quality, destination attachment and eWOM intention in heritage tourism. *Int. J. Tour. Cities* **2020**, *4*, 769–784. [CrossRef]
31. Dedeoglu, B.B.; van Niekerk, M.; Kucukergin, K.G.; De Martino, M.; Okumus, F. Effect of social media sharing on destination brand awareness and destination quality. *J. Vacat. Mark.* **2020**, *26*, 33–56. [CrossRef]
32. Al-Htibat, A.; Garanti, Z. Impact of interactive eReferral on tourists behavioral intentions. *Mark. Intell. Plan.* **2018**, *37*, 527–541. [CrossRef]
33. Tao, Y.; Zhang, F.; Shi, C.; Chen, Y. Social Media Data-Based Sentiment Analysis of Tourists' Air Quality Perceptions. *Sustainability* **2019**, *11*, 5070. [CrossRef]
34. Kim, H.; Jae-Joun, H.; Choe, Y.; Schroeder, A. How Can a Destination Better Manage Its Offering to Visitors? Observing Visitor Experiences via Online Reviews. *Sustainability* **2019**, *11*, 4660. [CrossRef]
35. Chen, W.; Xu, Z.Y.; Zheng, X.Y.; Yu, Q.Y.; Luo, Y.L. Research on Sentiment Classification of Online Travel Review Text. *Appl. Sci.* **2020**, *10*, 5275. [CrossRef]
36. Jiang, W.; Xiong, Z.A.; Su, Q.; Long, Y.; Song, X.Q.; Sun, P. Using Geotagged Social Media Data to Explore Sentiment Changes in Tourist Flow: A Spatiotemporal Analytical Framework. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 135. [CrossRef]
37. Ainin, S.; Feizollah, A.; Anuar, N.B.; Abdullah, N.A. Sentiment analyses of multilingual tweets on halal tourism. *Tour. Manag. Perspect.* **2020**, *34*, 2211–9736. [CrossRef]
38. Jiang, Q.M.; Chan, C.S.; Eichelberger, S.; Ma, H.; Pikkemaat, B. Sentiment analysis of online destination image of Hong Kong held by mainland Chinese tourists. *Curr. Issues Tour.* **2021**, 1368–3500. [CrossRef]
39. Luo, Y.Y.; Zhang, X.X.; Qin, Y.; Yang, Z.; Liang, Y. Tourism Attraction Selection with Sentiment Analysis of Online Reviews Based on Probabilistic Linguistic Term Sets and the IDOCRIW-COCOSO Model. *Int. J. Fuzzy Syst.* **2021**, *23*, 295–308. [CrossRef]
40. Available online: <https://www.unwto.org/international-tourism-and-covid-19> (accessed on 30 November 2020).
41. Available online: <https://es.statista.com/estadisticas/1043077/numero-de-turistas-chinos-en-espana/> (accessed on 7 December 2020).
42. UNWTO. *UNWTO Annual Report 2018*; World Tourism Organization: Madrid, Spain, 2018.
43. China Outbound Tourism Research Institute. 2019. Available online: <https://china-outbound.com> (accessed on 7 December 2020).
44. Rodríguez-Antón, J.M.; Alonso-Almeida, M.M. The circular economy strategy in hospitality: A multicase approach. *Sustainability* **2019**, *11*, 5665. [CrossRef]