

Analyzing tourist data on Twitter: a case study in the province of Granada at Spain

Marlon Santiago Viñán-Ludeña and Luis M. de Campos

Departamento de Ciencias de la Computación e Inteligencia Artificial, ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada, 18071 Granada, Spain

Purpose: The main aim of this paper is to build an approach to analyze the tourist content posted on social media. The approach incorporates information extraction, cleaning, data processing, descriptive and content analysis, and can be used on different social media platforms such as Instagram, Facebook, etc. Our work proposes an approach to social media analytics in traveler-generated content, and we use Twitter to apply our study and examine data about the city and the province of Granada.

Methodology: In order to identify what people are talking and posting on social media about places, events, restaurants, hotels, etc., we propose the following approach for data collection, cleaning and data analysis. We first identify the main keywords for the place of study. A descriptive analysis is subsequently performed, and this includes post metrics with geo-tagged analysis and user metrics, retweets and likes, comments, videos, photos and followers. The text is then cleaned. Finally, content analysis is conducted, and this includes word frequency calculation, sentiment and emotion detection and word clouds. Topic modelling was also performed with latent Dirichlet association (LDA).

Findings: We used our framework to collect 262,859 tweets about Granada. The most important hashtags are #Alhambra and #SierraNevada, and the most prolific user is @AlhambraCultura. The approach uses a seasonal context, and the posted tweets are divided into two periods (spring-summer and autumn-winter). Word frequency was calculated, and again Granada, Alhambra are the most frequent words in both periods in English and Spanish. The topic models show the subjects that are mentioned in both languages and although there are certain small differences in terms of language and season, the Alhambra, Sierra Nevada and gastronomy stand out as the most important topics.

Research limitations/implications: Extremely difficult to identify sarcasm, posts may be ambiguous, users may use both Spanish and English words in their tweets, and tweets may contain spelling mistakes, colloquialisms or even abbreviations. Multilingualism represents also an important limitation since it is not clear how tweets written in different languages should be processed. The size of the dataset is also an important factor since the greater the amount of data, the better the results. One of the largest limitations is the small number of geo-tagged tweets as geo-tagging would provide information about the place where the tweet was posted and opinions of it.

Originality: This study proposes an interesting way to analyze social media data, bridging tourism and social media literature in the data analysis context and contributes to discover patterns and features of the tourism destination through social media. The approach used provides the prospective traveler with an overview of the most popular places and the major posters for a particular tourist destination. From a business perspective, it informs managers of the most influential users, and the information obtained can be extremely useful for managing their tourism products in that region.

Keywords: traveler-generated content; descriptive analysis; content analysis; sentiment analysis; Latent Dirichlet Allocation; social media analytics

Introduction

The enormous growth of social networks has led to their use in a wide variety of research fields, and the use of social media data is an emerging and consolidated field in many different scientific domains. In recent work, for example, Twitter has been used for text mining and exploratory analysis of geospatial data in order to provide a useful guide for developing routes that evoke pleasant emotions (Park et al., 2020); the Twitter situational awareness (TwSA) framework, has been used in disaster management (Karami et al., 2020); a person's social media data (i.e. their profile data, posts, likes, etc.) can be used for identity resolution (D. K. Srivastava & Roychoudhury, 2020); and social media can also be used to evaluate public services provided in a city (Myneni & Dandamudi, 2020), and can be extended to marketing (Cripps et al., 2020; Gloor et al., 2020), public health (Aramburu et al., 2020), and the fashion industry (Casadei & Lee, 2020), etc.

Tourism is one of the most promising fields, representing as it does a major part of the gross domestic product of many countries. A great deal of research has been conducted into the use of online travel reviews or user-generated content (UGC), and such information has been widely used in the tourist and hospitality industry. Whether travelling for business, work or pleasure, tourists and travelers are extremely likely to share their experiences before, during and after their trips. This information is available online and on social media, and has been analyzed extensively by a number of authors: Airbnb (Leoni, 2020; Martí et al., 2020), Tripadvisor (Chang et al., 2020; M. T. Liu et al., 2020; Sudhakar & Gunasekar, 2020), Yelp (Dai et al., 2018; DeAndrea et al., 2018; V. Srivastava & Kalro, 2019), Expedia (Xu et al., 2019), and Ctrip (Wang et al., 2019). User-generated content enables travelers to provide their opinions about hotels, restaurants or places they have visited, and users planning to visit such places can then use UGC to read recommendations and feedback beforehand. Since users are influenced by other people's opinions (Ampountolas et al., 2019; Viñán-Ludeña et al., 2020) it is important to be able to analyze all of this information in order to understand how travelers use and interact with the different social networks (Viñán-Ludeña, 2019). This work shows the perceptions of Twitter users of a tourist destination taking into account the season of the year and the language of each user, this information is useful in the pre-trip phase.

Many studies mention the significant impact that information published on these platforms has on consumer preferences (T. Liu et al., 2020; Tsai & Bui, 2020). It is, therefore, no longer necessary to conduct interviews or surveys to obtain such information since social media data can be used to analyse the image of a tourist destination (Marine-Roig, 2019). However, when users search for information about tourist destinations on social media, they may become confused by the amount of information available and consequently make poor decisions because this type of search usually contains a lot of noise and might be unreliable. It is, therefore, essential for this information to be processed so that useful and reliable results are shown. Text analysis technology can handle large amounts of information for identifying topics and opinion analysis. Some models have been used to perform this type of task, and these include latent Dirichlet allocation, support-vector machines, fuzzy domain ontology, semantic association analysis (Bohr, 2020; Claster et al., 2010; Ovádek, 2020; Tian et al., 2020).

The main aim of this paper is to analyze the tourist content posted on social media, this provides to practitioners how to draw business insights and to travelers, relevant information of the tourist destination. This analysis incorporates information extraction, cleaning, data processing, descriptive and content analysis, and can be used on different social media platforms such as Instagram, Facebook, etc. Our work analyzes traveler-generated content, and we use Twitter to apply this study and examine data about the city and the province of Granada.

There are three main contributions of our study. Firstly, our work performs a descriptive analysis of user data and content analysis of posts. This includes sentiment and emotion detection and topic modelling and is the first time that this type of analysis has been applied in the sphere of tourism; this allows tourists before their trip to review the perceptions of users on Twitter about the destination and practitioners can improve their services by identifying negative comments they have received about a place or tourist service. Secondly, the study makes a methodological contribution to tourism research by proposing an approach to tourism-related data by considering the seasonal context that is, dividing it into two seasons (spring-summer and autumn-winter) or the four seasons (spring, summer, autumn and winter). In the present study, two seasons are considered because, when performing the content analysis using the data from the spring season, the results are very similar to the results using the data from the summer season; the same happened with the autumn and winter data. Thirdly, in order to test the social media analysis, our work uses Twitter posts about Granada, an important tourist destination in Spain. The approach used provides the prospective traveler with an overview of the most popular places and the major posters for a particular tourist destination. From a business perspective, it informs managers of the most influential users, and the information obtained can be extremely useful for managing their tourism products in that region. With this case study, In summary, the study provides insights about tourism in Granada prior to the trip.

Some of the studies reviewed focus on topic models as Latent Dirichlet Allocation (LDA) or are based on sentiment analysis. However, there is no work in the tourism domain that integrates a descriptive analysis of the information collected (posts, retweets, etc.) and a content analysis (word frequency, sentiment analysis, and topic models). In addition, this approach can be used with Twitter data as well as with Instagram information due to the similarity that exists between the two social networks. Finally, the province of Granada has many tourist attractions and can be visited both in autumn-winter and spring-summer. Due to the variety of places and things that can be done in this Spanish province, it has been chosen as a case study.

The remainder of this paper is organized as follows: the next section provides a background for research into traveler-generated content, text mining on social media and methodological approaches; after, this work examines the methodology used and describes how the data was collected, cleaned, and processed; then, this paper shows the application of our approach using tweets about Granada; and finally, we present the conclusions and the limitations of our approach, theoretical and practical implications and outline future lines of research.

Related Work

Traveler-Generated Content

The content of social media generated by travelers on different platforms such as Twitter, Facebook, Instagram, Yelp, etc. enables researchers to identify patterns for tourism management and planning. While some researchers use the term online-travel reviews (OTR), or user-generated content (UGC), all of the tourist content generated on social media can broadly be described as traveler-generated content (TGC), and this encompasses all social media platforms and networks (social networking sites, blogs/microblogs, content communities, forums-discussion) (Viñan-Ludeña, 2019).

Online travel reviews are the feedback posted by social media users based on their experiences, and their comments and opinions about a specific product or service (Litvin et al., 2008). It is important to identify the reasons why someone might want to share their experiences on different social networks since this affects the quality of the content published about tourism in a certain place. Various authors (Assaker, 2020; Bakshi et al., 2019) found that altruism, reciprocity, effort expectancy, habit and hedonic motivations have a positive impact whereas economic rewards have a negative impact on the intentions of travelers when it comes to posting online reviews.

Due to the large amount of OTR information, an ontology has been developed that uses a recommendation system to examine OTR content for popular attractions so that accurate and informed travel decisions may be made (Pai et al., 2019). In the field of gastronomic tourism, there is a framework for analyzing OTR, extracting useful information from textual and para-textual elements, building a matrix of word frequencies, and performing a quantitative and thematic content analysis (Marine-Roig et al., 2019). In another work, semantic association is used to extract thematic words and build a network (Hou et al., 2019). Other authors, meanwhile, perform a hierarchical classification of the services provided by restaurants, and identify four categories: service, restaurant, hospitality and food (Nascimento Filho et al., 2019).

Finally, in their articles (Assaker, 2020; Shin et al., 2019), the researchers examine the effects of trustworthiness, expertise, perceived usefulness, and perceived ease of use regarding the intended use of social media posts by young and older travelers. They show that the perceived ease of use was stronger among females and older travelers; perceived usefulness was stronger among males; and expertise had a significant impact on younger travelers. These results allow entrepreneurs to better manage marketing campaigns by focusing on specific groups of users according to age, perceived usefulness, experience and expertise.

This work shows the importance of content generated by tourists (TGC) because, when analyzing this data through descriptive analysis and content analysis using data mining techniques in social media platforms such as Twitter or Instagram, it allows both to summarize perceptions of social media users and to present the results in a simple way for consumers (travelers) and practitioners.

Text Mining, Twitter and Tourism

One of the most popular microblog platforms is Twitter. Thanks to its application programming interface (API), it is possible to access posted data and this enables a wide variety of research to be conducted in many different domains. One such sector is tourism, and it has become an invaluable tool for tourist sector entrepreneurs for developing marketing campaigns, and for planning and decision-making strategies. Over the past decade, although research has been conducted using all of the social networks, Twitter has been specifically used in tourism mainly by researchers in the United States and Japan (Curlin et al., 2019). Official accounts are also analyzed to explore market segments of companies to detect communities of clients on the network through text mining and cluster analysis. This is not only beneficial for marketing purposes and client strategy, but also allows companies to detect categories and better focus their marketing strategies (Punel & Ermagun, 2018). Additionally, a system has been built to obtain the most up-to-date tourist information from social networks, and extract the most relevant topics by distinguishing between languages and using a method to calculate the best number of topics using topic similarity (Hoshino et al., 2018).

This work uses twitter data due to (i) users publish the perceptions of the place they visit on social networks such as Twitter, Instagram or Facebook, (ii) the ease of accessibility to Twitter data through its API or through different data collection tools, finally, (iii) the studies reviewed in this section reveal that they can complement the findings of other performed research, not only using specific platforms in the tourism domain such as TripAdvisor or Yelp. The analysis performed in this work can also be carried out with other social networks, however Twitter has been chosen for its accessibility and popularity.

Contrasting previous studies which made data analysis in specific tourism social networking sites like TripAdvisor or Yelp to analyze the tourism destination in general context, this article extends the existing knowledge proposing a data analysis framework with a seasonal context to general social media platforms such as Twitter or Instagram, that allows us to extract all the necessary information that travelers need to plan their trip and tourist managers to improve their services.

Methodological Approaches

The present study proposes an approach that allows to generalize the analysis of the data published on social networks such as Twitter and Instagram in the context of tourism. In one of the analyzed works, authors perform a longitudinal study using data from 10 Spanish Destination Management Organizations (DMOs) to forecast hotel occupancy (Bigné et al., 2019); its approach is based on 4 stages: information extraction (Twitter API), content analysis (definition of tweet categories and classification of tweets in categories), text mining (extraction of significant information from keywords in tweets) and artificial neural network to forecast the hotel occupancy. Other researchers explore how Italian DMOs employ Facebook to promote and market their destinations (Mariani et al., 2016); they use an architecture tool that is composed of four modules (i) an extractor module, that extracts data from DMOs Facebook pages through Graph API, (ii) a parser module which processes the result of each query, (iii) an analyzer module which computes aggregated

engagement metrics defined by the user and (iv) a data visualization module. Another study, meanwhile, explores how European DMOs use social media to promote their destinations using many languages, data was collected from Twitter, Facebook, Instagram and YouTube (Uşaklı et al., 2017). On the other hand, taking into account online reviews, some researchers made a comparison between TripAdvisor, Expedia, and Yelp, in terms of information quality about hotel reviews in New York City (Xiang et al., 2017); they use LDA and sentiment analysis in their analysis. In the same direction, other authors propose a framework composed by LDA and sentiment analysis to extract meaning from the comments or reviews provided by hotel visitors (Guo et al., 2017). Using twitter data, other researchers propose an approach that explores theme park visitors' emotions, involving social media analytics and geospatial analytics (Park et al., 2020). Finally, a spatial analytical framework is proposed to understand tourist experiences from geotagged social network data in Beijing (Zhang et al., 2020); the authors researched the effects of air pollution on tourists' experiences.

By using machine learning algorithms to forecast sales and tourist trends, it is possible to predict social and commercial trends in tourism. Nevertheless, there are still gaps in research in terms of identifying potential consumer needs based on word granularity, semantic association and text mining combined with multilingualism, which is an important issue since tourists come from all over the world.

This work therefore performs a text analysis on Twitter posts to explore the main topics discussed integrating a descriptive analysis of user data (i.e. users, likes, retweets, comments) and text mining analysis (word frequency, sentiment analysis and topic model).

Methodology

Research Design

In order to identify what people are talking and posting on social media about places, events, restaurants, hotels, etc., we propose the following approach for data collection, cleaning and data analysis, as shown in Figure 1. We first identify the main keywords for the place of study. The stored data are then used to build a dataset. A descriptive analysis is subsequently performed, and this includes post metrics with geo-tagged analysis and user metrics, retweets and likes, comments, videos, photos and followers. The text is then cleaned to remove user mentions, URLs, redundant characters, emojis, special characters, numbers and punctuation marks before the process of tokenization, stop word elimination and stemming. Finally, content analysis is conducted and this includes word frequency calculation, sentiment and emotion detection and word clouds. Topic modelling was also performed with latent Dirichlet association (LDA) (Blei et al., 2003).

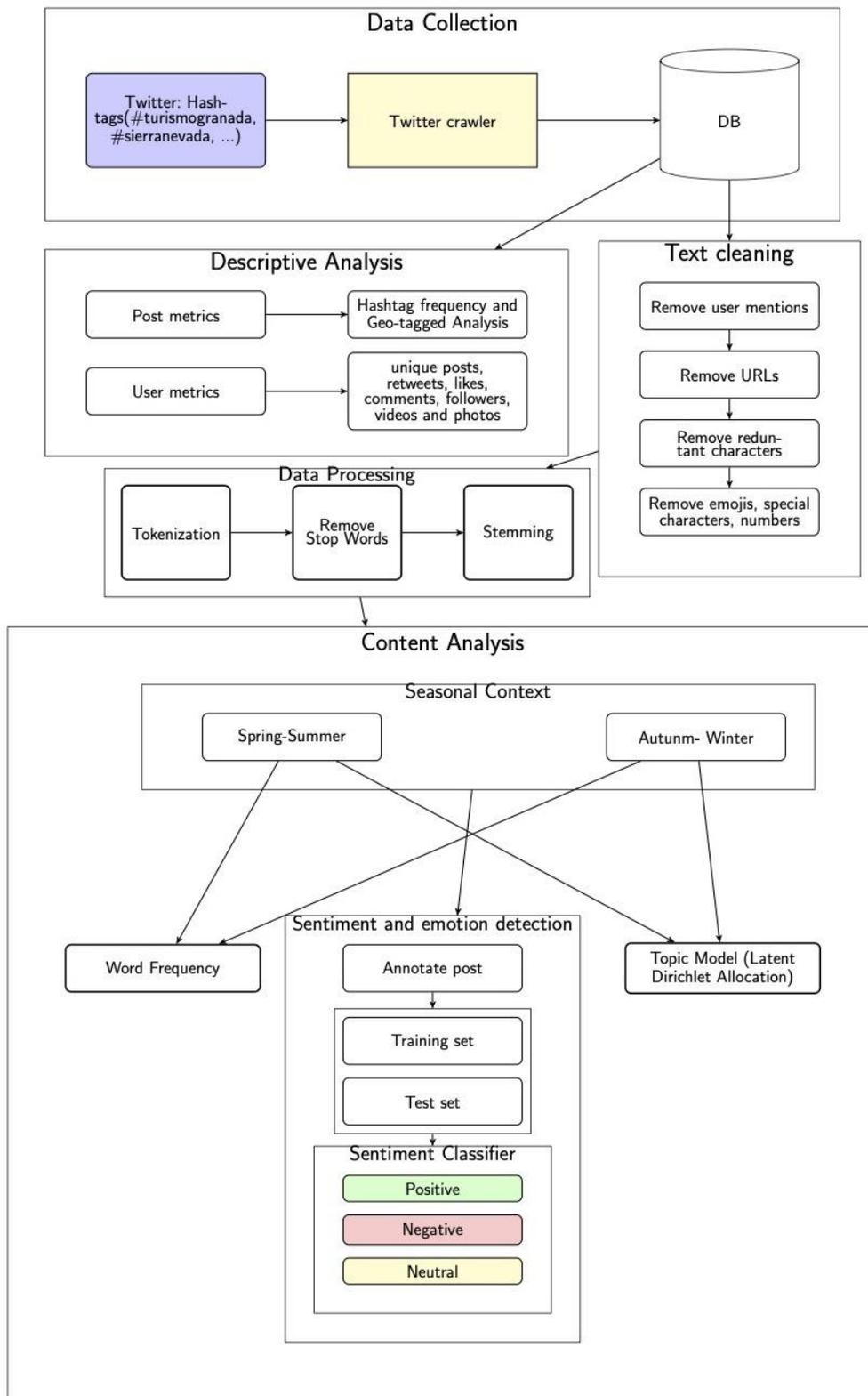


Figure 1 Framework for analysis of social media tourism data

The descriptive analysis allows obtaining the most important users who post about a tourist destination, while the content analysis shows: (i) the perceptions of the users of a certain place that belongs to a tourist destination, either positively or negatively; if they are positive, you can find the popular places and if they are negative, these tweets can show the shortcomings that practitioners can use to improve tourist destinations; (ii) the topics of a destination can be found in the text using word frequency and LDA, using the seasonal context and the difference in perceptions of users who post tweets in Spanish and in English.

Data Collection

Data can be collected by using the crawler API or scraping the social media web page. We have used a crawler to collect Twitter posts about Granada but it can be used with any other type of platform. Twitter is one of the most important social networks in the world and according to Kemp (Kemp, 2020), it has 340 million active users worldwide and is frequently used by people between the ages of 25 and 49. In Spain, 53% of social media users use Twitter. Researchers are able to use collected Twitter data to conduct studies in many different scientific domains, such as public health, law, computer science, etc. (Arthur & Williams, 2019; Colditz et al., 2018; Kim et al., 2020; Monachesi, 2020; Ren et al., 2019; Samoggia et al., 2020). Since the Twitter API prevents published tweets being accessed after seven days, an open source tool called Twint¹ was used since this enables unlimited access to any tweet ever posted and gathers tweets according to username or keyword.

Tweets were collected from Spanish and English posts, and two main accounts were identified for the Spanish tweets: Granadaturismo, which is the official Twitter account for the Granada Tourism Department, with the hashtag #teenseñomigranada in its bio; and Alhambracultura, which is the official Twitter account for the Board of the Alhambra and Generalife, with the hashtag #alhambra. In order to obtain all the Spanish tweets about tourism in Granada, we considered the following search criteria: *granada turismo*, *gastronomia Granada*, *hoteles Granada*, and *restaurantes Granada*. Another criterion that was also taken into account was *What to do in Granada?* identified by the hashtag #planesgranada. In order to obtain additional search criteria, we used the Twint tool to search for tweets about *granadaturismo*. We used these tweets to calculate the frequencies of the hashtags and then manually analyzed the top forty. It was then possible to find #albaicín², #sierranevada³ and #welovegranada (which will be used for tweets written in English). Once all the search criteria have been identified, each is used to extract the tweets with the Twint tool, and a total of eleven search criteria are used to obtain tweets in Spanish. For tweets in English, the search criteria that refer to tourism and the analyzed

¹ Twint is a Twitter crawling Python tool that allows to obtain tweets from Twitter, more information can be found on the Github Page <https://github.com/twintproject/twint>

² an area of Granada that is a World Heritage Site

³ a mountain range near Granada with a ski resort and the highest mountains in the Iberian Peninsula

site were taken into account, and, therefore, #granadatrip and #granadatravel were identified.

By way of summary, the following search criteria were used:

Spanish

1. granadaturismo
2. #teenseñomigranada
3. alhambracultura
4. #alhambra
5. *granada turismo*
6. *gastronomia granada*
7. *hoteles granada*
8. *restaurantes granada*
9. #planesgranada
10. #albaicín
11. #sierranevada

English

1. #welovegranada
2. #granadatrip
3. #granadatravel

Every search criterion is executed separately and the captured tweets are stored in a .csv file. Each file is then read and the tweet data is stored in a global MySQL database. Our work focuses on the city of Granada in Spain, and since there are other cities called Granada, we use exclusions in Twitter search queries that work with the Twitter API. We, therefore, put the character - before the word that we want to exclude: for example, we can use the search criteria *granadaturismo -nicaragua -colombia -california*. The words *Nicaragua*, *Colombia* and *California* are excluded because there are other cities called Granada in Colombia and Nicaragua, and *California* is excluded because there is a mountain range located in the Western United States called *Sierra Nevada*, which has the same name as the mountain range near Granada. These exclusion criteria were used with each of the fourteen search criteria. Although search criteria were taken into account to find tweets written in Spanish and English using search criteria such as granadaturismo,

alhambracultura or #alhambra, tweets in many other languages were also captured, as Table 1 shows.⁴

Table 1 Dataset by Language

Language	Unique posts	Comments	Retweets
Spanish(es)	180215	22301	131
English (en)	24095	1316	4
Portuguese (pt)	10244	652	-
No text (und)	5336	1424	-
Catalan (ca)	4176	486	2
Dutch (nl)	3373	59	-
German (de)	1742	38	-
Italian (it)	1315	281	-
French (fr)	779	56	-
Indonesia (in)	655	27	-
Romanian (ro)	514	66	-
Albanian (sq)	475	29	-
Norwegian Nynorsk (nn)	464	43	-
Norwegian Bokmål (nb)	281	21	-
Pali (pl)	267	7	-
Japanese (ja)	221	3	-
Turkish (tr)	219	16	-
Arabic (ar)	180	16	-
Slovak (sk)	137	8	-
Shona (sn)	129	16	-
Finnish (fi)	117	14	-
Tagalog (tl)	113	8	-

⁴ The data shown in the table refers to the tweets collected using the Twint tool. The number of retweets shown in the table does not refer to retweet rate but to new tweets that have been captured by the tool but with a new owner, obviously specifying that it is a retweet. Therefore, the number of times that a tweet has been retweeted is stored in its metadata that the tool captures, and this work presents it in the table that shows retweets, likes and comments.

Somali (so)	83	3	-
Slovenian (sl)	70	6	-
Haitian (ht)	68	7	-
Czech (cs)	63	1	-
Basque (eu)	57	10	-
Swedish (sv)	48	3	-
Russian (ru)	38	3	-
Korean (ko)	38	2	-
Danish (da)	34	2	-
Chinese (zh)	33	1	-
Lithuanian (lt)	25	4	-
Norwegian (no)	22	1	-
Thai (th)	21	-	-
Estonian (et)	19	-	-
Hungarian (hu)	19	3	-
Indonesian (id)	15	-	-
Bosnian (bs)	9	-	-
Latvian (lv)	9	-	-
Icelandic (is)	9	1	-
Greek, Modern (el)	8	-	-
Tsonga (ts)	8	-	-
Bulgarian (bg)	7	1	-
Hindi (hi)	6	-	-
Persian (fa)	6	-	-
Vietnamese (vi)	5	-	-
Malay (ms)	4	-	-
Urdu (ur)	4	-	-
Welsh (cy)	3	-	-
Hebrew (iw)	2	-	-
Galician (gl)	2	-	-
Swahili (sw)	1	-	-

Pashto (ps)	1	-	-
Belarusian (be)	1	-	-
Xhosa (xh)	1	-	-
Kazakh (kk)	1	-	-
Total	235787	26935	137

It is important to mention that criteria that have been used for the Spanish dataset such as: *granadaturismo*, *alhambracultura*, *#alhambra* and *#sierranevada* contain many English tweets. Therefore, only three more search criteria have been added to have the full English tweet dataset. Furthermore, out of total of tweets collected, 5,336 do not contain text, this means that they only refer to some hashtag (s) (e.g. *#Granada*, *#Alhambra*, etc). The most representative keywords of the province of Granada have been identified, it has also been tested with other hashtags / keywords that could have tweets related to the Spanish province, but no results were obtained, or the tweets were the same since each of them has several hashtags or keywords in its text. We have probably not managed to get close to 100% of tweets about Granada, taking into account the bias in the keywords. However, we have obtained a fairly considerable number of tweets that makes us think that our results are reliable and representative.

The Twint tool allows to obtain most of the tweets that have been published since 2008. Thus, we collected tweets from 2008 to July 2020. The advantage of this tool is that it does not have date restrictions to obtain the data as has the Twitter API.

Descriptive Analysis, Text Cleaning and Data Processing

Once the posts had been collected, they were processed with the following operations: text cleaning, descriptive analysis, tokenization, stop-word removal and stemming.

1. Text-cleaning: Due to the heterogeneity of information, this process is a real challenge for the following reasons: certain tweets share photos or videos, and only contain hashtags or emojis in the text; many comments only mention other users, share photos or videos, and do not contain any text; some posts contain prize draws, product promotion or news about the city; most users do not share their location; some tweets use emojis to express opinions in the comments and posts; and some tweets contain questions for other users. In order to deal with tweets that only contain hashtags or emojis and no text or only mention users, we remove user mentions and emoticons. Any tweet that is not related to tourism in Granada is also discarded by searching for the word *tourism* in the text of the English tweets and *turismo* in the Spanish tweets. Additionally, the geo-location was analyzed for those tweets where this was available.

Finally, punctuation marks, users, and links were removed as was the term *RT* for retweets since when a person retweets, they reiterate the original tweet⁵.

2. Descriptive analysis: This includes post metrics (hashtag frequency) and finds all the tweets with a location to identify the most visited places. In terms of user metrics, we obtain the frequency of unique posts, retweets, comments, followers, videos and photos for each user, and this gives us an initial idea of the user's social influence.
3. Tokenization: We first identify the languages present in our dataset and then divide each tweet into keywords according to the language. In this study, we take tweets written in Spanish and English because these languages have the largest amount of information. A language *und* was detected because the tweet text only includes hashtags, user mentions or links, and does not contain any text.
4. Stop-word removal and stemming: We used the Natural Language Toolkit (NLTK)⁶ to perform this process since it contains both Spanish and English stop-words. The subsequent stemming process also uses NLTK with one stemmer for Spanish and another for English, and reduces inflected words to their word stem, base or root.

Content Analysis

Content analysis was performed in a seasonal context, and the year was divided into the two periods of spring-summer and autumn-winter. The analysis mainly focuses on structuring the data from texts and includes the following processes:

1. Word frequency: this includes a separate word frequency analysis for Spanish and English in addition to the word cloud for all the tweets in an attempt to identify the most commented places. It also enables us to gauge how users perceive these places through the use of words such as *beautiful, love, etc.*
2. Sentiment analysis: this entails identifying and classifying each tweet according to feelings. Since this process does not consider the seasonal context because this division did not reveal any relevant results, it was therefore performed globally. Due to the large number of tweets in both Spanish and English, two tools are used. The first tool for classifying Spanish tweets is called Senti-py⁷. The package uses data to train the classifier from various websites such as Tripadvisor and Twitter. The polarity score is a float within the range [0,1]: values that are less than or equal to 0.3 are negative, values greater than 0.3 and less than 0.6 are neutral, and values greater than

⁵ Retweets have not been removed because they show acceptance of the original tweet and therefore reinforce what the original tweeter posted

⁶ NLTK is a platform to build Python programs to process human language data, and more information can be found on the webpage <http://www.nltk.org>

⁷ Senti-py is a pre-trained classifier in Spanish which is built on scikit-learn and NLTK. More information can be found on the Github page <https://github.com/ayllote/senti-py>

or equal to 0.6 are positive. The second tool for classifying English tweets is called TextBlob⁸. The polarity score is a float within the range [-1.0, 1.0], where values of less than 0 imply negative sentiment, values greater than zero represent positive sentiment, and a value equal to 0 means neutral. This information about positive or negative sentiments is extremely important for tourists when it comes to deciding whether they should visit somewhere or not. These tweets can also be used by tour operators, local authorities and residents to improve the image of the tourist destination.

3. Latent Dirichlet allocation: We use LDA for topic modeling. In order to implement these algorithms, Python 3.7 was used with the Gensim library⁹. LDA attempts to identify the most important underlying topics and places for the tourist destination in order to discover the options available to the tourist according to the seasonal period defined by identifying places that can be visited both in spring-summer and autumn-winter, or all year round.

Findings

In this section, we apply our approach in order to gain greater insights into tourism in Granada, Spain. The raw data is stored in an MySQL server. Tables Table 1 and Table 2 summarize the number of unique posts, retweets and comments which were collected for each language and search criterion (hashtag or keyword), respectively. We then apply the cleaning process, and the descriptive and content analysis are described in the following subsections.

Table 2 Dataset according to Hashtag/keyword

Hashtag / Keyword	Total
alhambra	144036
alhambracultura	58372
granadaturismo	49014
sierranevada	26593
planesgranada	27116
"granada turismo"	42538
"gastronomia granada"	10052

⁸ TextBlob is a Python library for processing textual and provides an API for common natural language processing (NLP) tasks. More information can be found on the webpage <https://textblob.readthedocs.io/en/dev/>

⁹ Gensim is a free library written in Python for topic modelling. More information can be found on the webpage <https://radimrehurek.com/gensim/index.html>

"hoteles granada"	5283
welovegranada	3788
"restaurantes granada"	4459
Albaicín	733
granadatrip	167
Teenseñomigranada	142
granadatravel	153

Table 1 reveals that the most frequent Twitter users tweet in Spanish and English, followed by tweets in Portuguese, Catalan or with no text (i.e., tweets that only contain hashtags, emojis or links). The table also shows, however, that tweets are posted by users anywhere in Europe and Asia, although to a lesser extent. The language *und* means that a tweet has no text. This information provides greater context about the nationalities of the tourists who visit Granada¹⁰ and this enables travel managers to redefine marketing strategies. In a similar way, Table 2 shows the total number of unique posts, retweets and comments for the different hashtags or keywords.

The results presented in Table 2 show the number of tweets for each search criterion and, consequently, the number of occurrences of each criterion in the complete data set. This does not match the number of existing tweets in our dataset because each tweet can refer to many hashtags/keywords/criteria rather than to a single search criterion. One example of this is the tweet *"We invite everybody to visit #baza, its #history and the cultural and natural #heritage #welovegranada #greenwalkes #altiplanogranada @AytoBaza @granadaturismo @JAGGER0779 @turgranada @TurismoSpain ..."*, and since this tweet has the hashtag *welovegranada* and the search criterion *granadaturismo*, it therefore refers to two search criteria.

Descriptive Analysis

Descriptive analysis is a summary that quantitatively describes features from a dataset and our work considers users who publish information about Granada. Certain attributes are considered to be important such as unique posts, number of retweets, likes, followers, photographs, videos and comments. Geographical analysis is possible if users have tweeted with an active geo-location.

Post Metrics

For the purposes of our study, we collected a total of 262,859 tweets which contained 46,555 unique hashtags. The most common hashtags in our dataset are #Granada,

¹⁰ There are, however, certain limitations such as when different countries share the same language.

#Alhambra, #SierraNevada, #Spain, #turismo, #España, #Andalucia, #travel, #planesgranada and #gastronomía. It is clear that the most promoted places and also the most referenced are Granada and Alhambra.

In the context of tourism, it is important to know the location of tweets and these are shown in Figure 2. Out of the 262,859 tweets, only 432 tweets are associated with a location, i.e. tweets that have been geo-tagged.

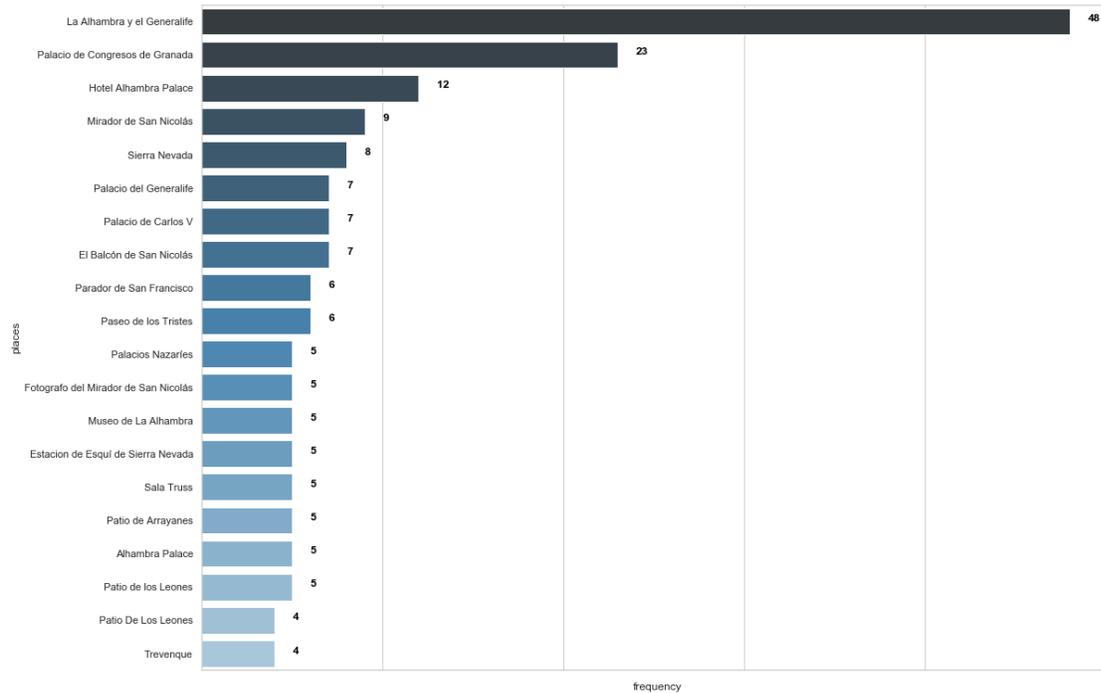


Figure 2 Geo-tagged tweets (top twenty)

Most of the geo-tagged tweets are in Granada. According to these tweets, the most visited places are *La Alhambra y el Generalife*, *Palacio de Congresos de Granada* (a conference centre), *Sierra Nevada* and *Mirador de San Nicolás* (a famous viewpoint in the Albaicín area of Granada).

User Metrics

The 262,859 tweets were posted by 66,409 unique users. Once we had identified the users, we found those users with unique posts i.e. they were neither retweets nor responses to other posts. The most prolific users are @alhambracultura, an official account of the Board of the Alhambra and Generalife with 12,481 tweets; @granadaturismo, an official Granada tourism account about Granada with 12,135 tweets; @planesgranada, an account that provides information about leisure, culture and gastronomy in Granada with 5,270 tweets; @granadaxmundo, an account that promotes the web page granadaporelmundo.com with blog posts about Granada, videos and photos with 2,744 tweets; and @makaralu, a local Granada blogger who promotes her city on Twitter, with 1,652 unique posts. Figure 3 displays the top twenty users who have posted the largest number of unique tweets. User analysis reveals that the account @alhambracultura is the most popular for tourism in the

region and the most active on Twitter with the highest number of tweets compared to other accounts analyzed.

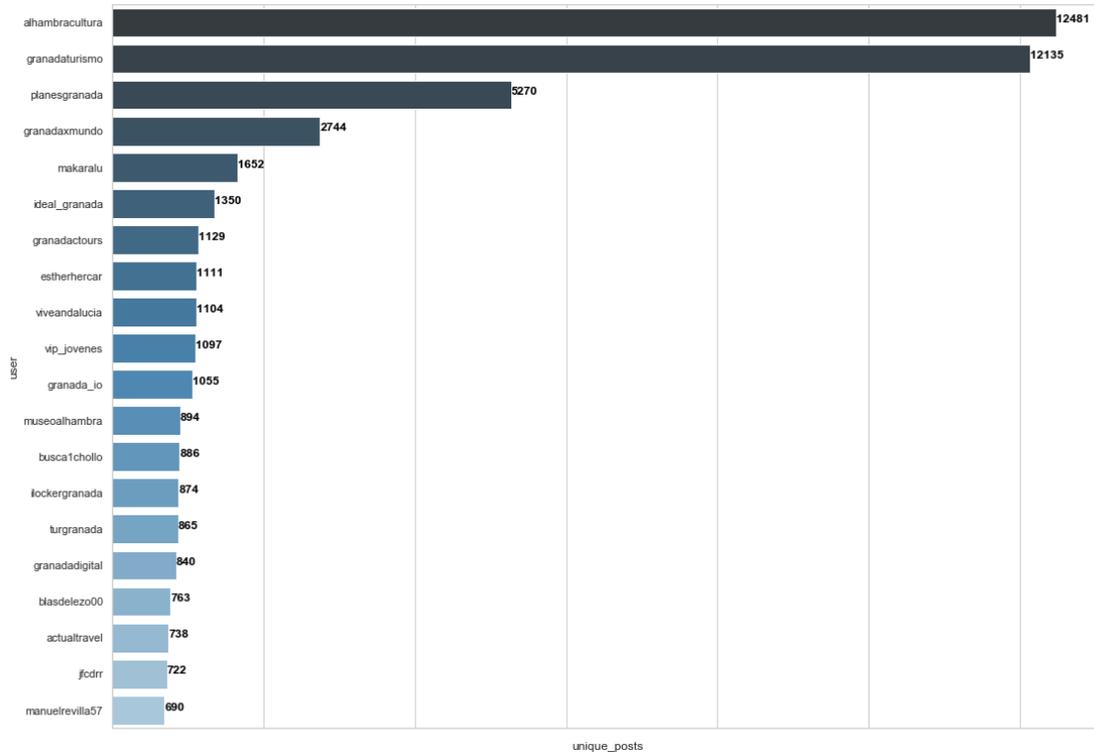


Figure 3 Total unique posts according to user (top twenty)

Identifying the users who have published the greatest number of tweets, firstly allows to know who are the users who have the most presence in the tourism of the studied destination and secondly, it allows obtain other metrics such as likes, retweets and comments. Additionally, these users can contribute to practitioners to promote a certain place or tourist destination.

From the unique post data, we obtain the number of retweets, likes and comments for each post. It is, therefore, possible to find those users who have received the most retweets and these are @ahambracultura with 107,362 retweets; @planesgranada with 97,985 retweets; @granadaturismo with 55,745 retweets; @spain, an account that promotes tourism in Spain, with 16,596 retweets; and @joselop44, a user who lives in Granada, with 13,863 retweets. Table 3 shows the top twenty users with the most retweets, likes and comments in our dataset.

Table 3 Top twenty retweets, likes and comments

user	retweet_count	likes_count	comments_count
ahambracultura	107362	194257	5445
planesgranada	97985	167283	4329

granadaturismo	55745	86821	2992
spain	16596	42839	918
joselop44	13863	23626	913
estherhercar	8080	27635	472
itsmalbert	8074	18856	123
yosoyesssa	12961	13886	793
manuelrevilla57	7050	16850	493
elhuffpost	5813	16369	591
ismaelquesada	9945	10917	430
ferminius	5490	14197	99
ideal_granada	6968	9434	542
aesthevic_	6592	8842	44
davidschez	4259	10998	78
tarandas61	2351	11393	286
viveandalucia	5344	7331	354
blasdelezo00	2389	9344	436
museoalhambra	3938	7403	228
rocio_diazj	3980	7125	327

Popularity on any social media platform is measured by the number of followers per user, i.e. it is generally believed that the more followers a user has, the more popular they are. As shown in Figure 4, @ideal_granada, the local newspaper with news about the city and province of Granada, has the largest number of followers, followed by @viveandalucia, which is an official tourist promotion channel, and @alhambracultura.

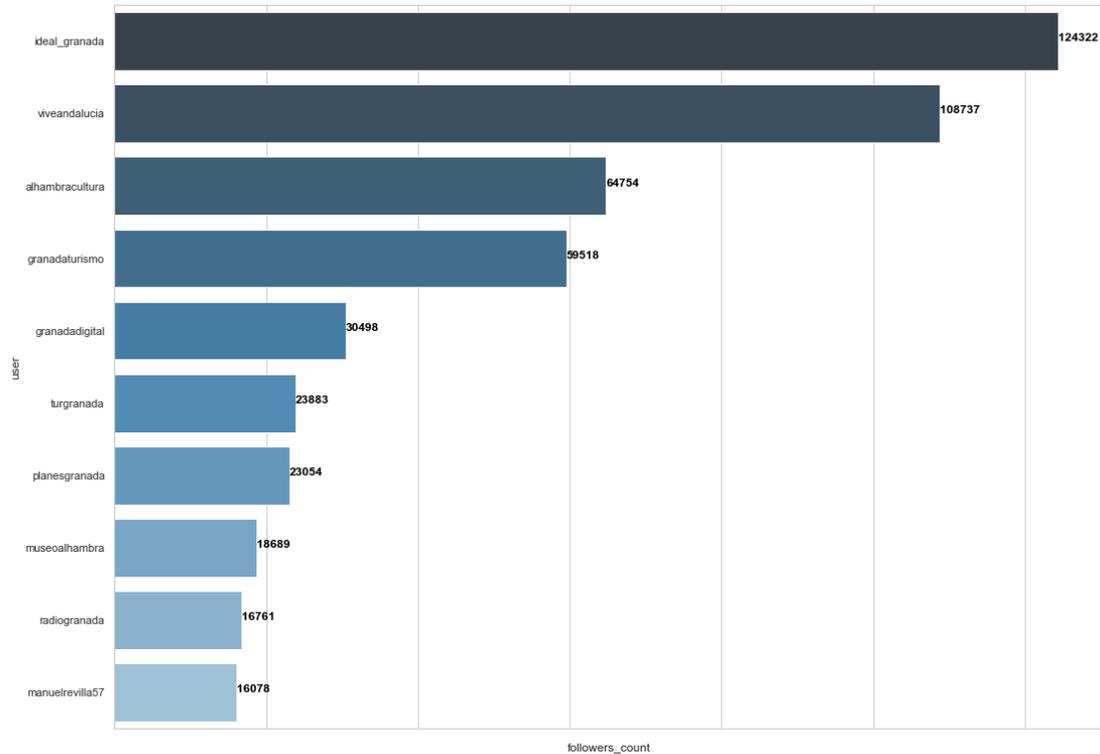


Figure 4 Followers per user (top ten)

It is very easy to obtain the number of followers with the Twitter API. There are, however, a number of users in our dataset who have only commented, posted or retweeted once about Granada tourism and yet have millions of followers, and these users are of little relevance for the purposes of this work. In order to obtain consistent results with our analysis, we consider those users who have published the most unique posts about Granada (i.e. users with more than 500 unique posts). This result was used as a filter to ensure that we obtained the most representative users for our study.

Multimedia metrics were also calculated by examining tweets for the presence of photographs or videos. In all of the 262,859 tweets, there were 123,809 photos or videos, and of the 66,409 users, only 21,753 had published multimedia content. Many users only publish images of places they have visited, and, generally speaking, users tend to take photos or videos of the most important places they have been to and post these on social media. There is no doubt that excellent images on social media will attract, engage, and encourage their friends and followers to visit these destinations in the future. Great visual marketing is also extremely helpful for tourism managers when establishing a personal connection between brand and customer.

Content Analysis

Content analysis or textual data analysis consists in determining the variables of interest and the occurrence frequency of the processed variables and performing a subsequent data dimension reduction. In this paper, we calculate word frequency, and perform sentiment

analysis and topic modelling (Joseph et al., 2017). The unit for text analysis used in this work is the “word”.

Word Frequency

For the purposes of the analysis, word frequency has been divided into Spanish (202,647 tweets) and English (25,415 tweets). For each group, the stemming process is performed separately with the stemmer for the relevant language. Table 4 shows the English words in the spring-summer period, and Table 5 shows the English words in the autumn-winter period. In both cases, it can be seen that the places *La Alhambra* and *Sierra Nevada* are visited all year round (although there are more references to *Sierra Nevada* in autumn-winter). In the case of spring-summer, however, users mention the views and gardens, whereas in the case of autumn-winter, the Andalusia region, snow and skiing stand out. As the words *beautiful* and *love* also appear in both periods, we can assume that tourists really like the viewpoints at different points of the city and different tourist places in the province.

Table 4 Word frequency analysis (Spring-Summer) (en)

Words	Frequency
granada (Granada)	11974
alhambra (Alhambra)	9817
spain (Spain)	4816
palac (palace)	2163
travel (travel)	1434
visit (visit)	1384
beauti (beautiful)	1226
sierranevada (SierraNevada)	1043
citi (city)	916
one (one)	788
day (day)	774
place (place)	713
view (view)	708
love (love)	688
andalucia (Andalucia)	624
garden (garden)	577
see (see)	561
andalusia (Andalusia)	543
go (go)	515
tour (tour)	501

Table 5 Word frequency analysis (Autumn-Winter) (en)

Words	Frequency
granada (Granada)	11132
alhambra (Alhambra)	8343
spain (Spain)	4507
palac (palace)	1709
sierranevada (SierraNevada)	1682
travel (travel)	1540
visit (visit)	1363
beauti (beautiful)	989
citi (city)	825
andalucia (Andalucia)	785
andalusia (Andalusia)	747
day (day)	738
one (one)	678
view (view)	655
love (love)	596
snow (snow)	595
place (place)	570
ski (ski)	524
tour (tour)	520
see (see)	511

Tables Table 6 and Table 7 show the Spanish words for spring-summer and autumn-winter, respectively. The words that stand out in the Spanish tweets are also *Granada*, *Alhambra* and *Sierra Nevada* in both the spring-summer and autumn-winter periods, since these are the most promoted and most visited places in the region, and the *Albaicín* area is also mentioned. The use of positive words such as *bueno*, *gracias*, *disfruta*, *mejor* further confirms tourists' positive opinions. In the tables showing the word frequencies of Spanish and English tweets, the whole word has been placed in brackets since stemming is performed before the counting process.

Table 6 Word frequency analysis (Spring-Summer) (es)

Words	Frequency
gran (Granada)	69752
alhambr (Alhambra)	38859
turism (turismo)	8212
sierranev (SierraNevada)	7250
visit (visita)	7179

espa (España)	6034
graci (gracias)	4612
ciud (ciudad)	4592
buen (bueno)	4317
hoy (hoy)	4206
hotel (hotel)	4071
albaicin (albaicin)	3992
fot (foto)	3905
viaj (viaje)	3683
gastronom (gastronomía)	3650
disfrut (disfruta)	3629
dia (dia)	3496
noch (noche)	3339
mejor (mejor)	3299
hac (hace)	3292

Table 7 Word frequency analysis (Autumn-Winter) (es)

Words	Frequency
gran (Granada)	70431
alhambr (Alhambra)	35178
sierranev (SierraNevada)	11096
turism (turismo)	8424
visit (visita)	8386
espa (España)	6126
ciud (ciudad)	5626
buen (bueno)	4886
graci (gracias)	4536
gastronom (gastronomía)	4457
hoy (hoy)	4145
fot (foto)	4021
dia (dia)	3720
disfrut (disfruta)	3593
viaj (viaje)	3496
mejor (mejor)	3387
conoc (conoce)	3195

Both supervised (Rao et al., 2017; Soong et al., 2019) and unsupervised (Aoudi & Malik, 2019) machine-learning methods have been analyzed to classify emotions. The machine-learning approach or technique that has obtained the best results is the Multinomial Naive Bayes classifier (Farisi et al., 2019). Among the various tools available, and which are freely available, the ones that are best valued by the community have been chosen. Thus, the TextBlob tool is used to classify tweets written in English and this uses the Multinomial Naive Bayes model as the supervised method; for Spanish tweets, we use the Senti-Py tool which is written in Python and which also uses a Multinomial Naive Bayes model as the method. The model is fed with data crawled from various websites which include Tripadvisor and Twitter. Two different tools are used since no single tool allows documents to be processed in both languages.

The main aim of sentiment analysis in our work is to classify the text according to the three categories of positive, negative and neutral. For tweets written in English, our analysis shows (see Figure 6) that of the total 25,415 tweets (unique posts, comments and retweets), 49.55 percent are neutral, 44.01 percent are positive, and only 6.43% are negative. This analysis reveals that although most tweets are concerned with promoting or providing information about Granada, other tweets speak favourably of specific sights, places in the city, and also the city itself. The following examples of tweets express positive and neutral sentiments:

1. Granada is an amazing city with jewels like this monastery <http://bit.ly/jeron> #ttot #travel @granadaturismo (Positive)
2. I love snow :D #snow #cool #great #mountain #cold #jmgt #granada #sierranevada <http://instagram.com/p/jpFILsKRaf/> (Positive)
3. Skiing in the #SierraNevada in #Granada !!! All you need to know before planning your trip <http://ow.ly/8PEx7> (Neutral)
4. We remind you that the shops will be open tomorrow in #Granada! do your last-minute shopping in the city! (Neutral)

A small percentage of the tweets, however, are negative about the region and these include negative tweets about current affairs and Alhambra visits:

1. First bad news: the Rosario and the Nazareno processions won't take place #SSantaGr
2. Alhambra visits are NOT well organized, long wait even with tickets bought in advance. @alhambracultura #Andalucia pic.twitter.com/iXv4ScY2GR

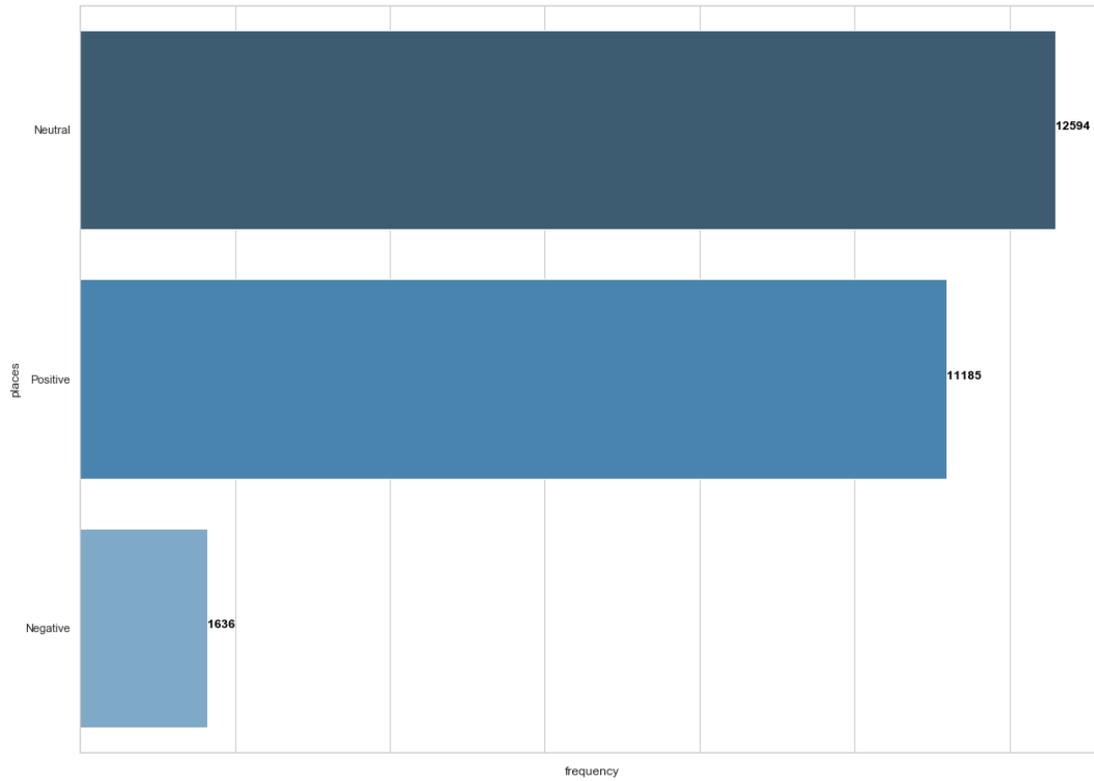


Figure 6 Tweet sentiment distribution (English)

An analysis of the Spanish tweets reveals that 58.70% of the total 202,647 tweets (unique posts, comments and retweets) are positive, 21.34% are neutral, and 19.95% are negative, and this is shown in Figure 7.

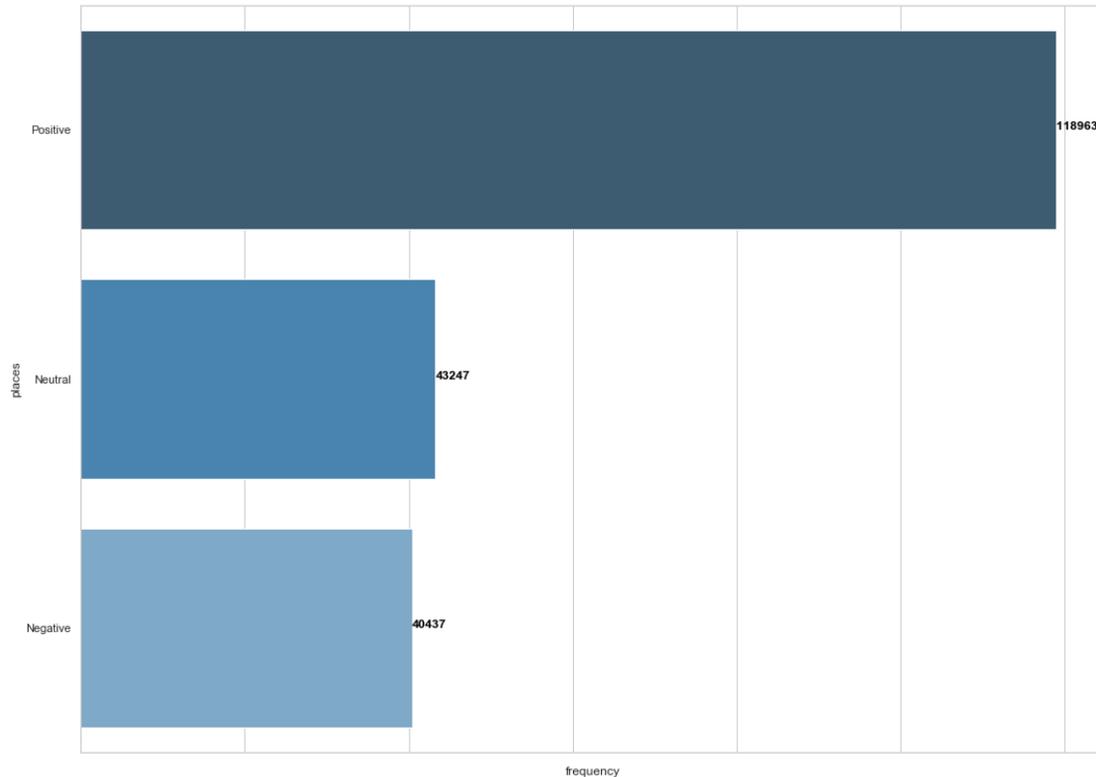


Figure 7 Tweet sentiment distribution (Spanish)

The following tweets are positive and neutral:

1. Ahora a Granada lugar maravilloso por su historia, su gastronomía, su ambiente y por supuesto por el Casting de Somos Sur. [*Now to Granada, a wonderful place for its history, its gastronomy, its atmosphere and of course for the casting of Somos Sur.*] (Positive)
2. Las sombras hacen que sierra nevada parezca estar sin nieve #Granada #SierraNevada [*The shadows make it look as though there is no snow on the sierra nevada #Granada #SierraNevada*] (Neutral)

The following tweets are negative:

1. En serio.....por que #granada esta tan oscura?? Señor Alcalde que no se ve ni la fachada de la catedral!!!!!!(Dicho por turistas) @granadaturismo [*Seriously why is #granada so dark?? Mr. Mayor, you can't even see the facade of the cathedral!!!!!! (Said by tourists) @granadaturismo*]
2. ÚLTIMA HORA: 2 fallecidos y 2 heridos, uno grave, al precipitarse un vehículo por un terraplén en la Carretera de #SierraNevada en #Granada [*STOP PRESS: 2 dead and 2 injured, one seriously, when a vehicle crashed down an embankment on the #SierraNevada Road in #Granada*]

When enough data is available, the most popular places (positive tweets) and the places where negative experiences have been had (negative tweets) are much more evident. In this work, we do not consider the frequencies of geo-tagged tweets that have been identified as negative or positive since the number of these is small.

Topic Modeling (LDA)

Many researchers have used latent Dirichlet allocation (LDA) to extract topics from tweets in different domains (Abd-Alrazaq et al., 2020; Singh et al., 2019; Yoosefi Nejad et al., 2020). As the authors explain in their article (Blei et al., 2003), LDA is a three-level hierarchical Bayesian network model, where each item in a collection is modelled as a finite mixture of an underlying set of themes/topics, and the number of topics is an input parameter that must be fixed. Each topic defines a probability distribution over the keywords, indicating the probability that each keyword is associated to that topic. In our case, having conducted a number of previous experiments, we chose four topics for Spanish tweets for the spring-summer period, and three topics for the autumn-winter period. Four topics were chosen for English tweets for both periods.

Table 8 shows the most representative and most probable keywords and their probabilities¹¹ associated with the different topics for the Spanish tweets.

Table 8 Topics and their keywords (Spanish)

Topics	Keywords
<i>Spanish (Spring-Summer), k = 4</i>	
1	0.056*"alhambr(Alhambra)" + 0.047*"gran(Granada)" + 0.033*"visit(visita)" + 0.013*"monument(monumento)" + 0.010*"conoc(conoce)" + 0.007*"entrad(entrada)" + 0.007*"generalif(generalife)" + 0.007*"palaci(palacio)" + 0.007*"patrimoni(patrimonio)" + 0.006*"ciud(ciudad)"
2	0.070*"gran(granada)" + 0.025*"alhambr(Alhambra)" + 0.021*"viaj(viaje)" + 0.017*"sierranev(sierranevada)" + 0.009*"cuent(cuentame)" + 0.009*"pas(paseo)" + 0.009*"noch(noche)" + 0.008*"buen(bueno)" + 0.008*"sierr(sierra)" + 0.007*"nev(nevada)"
3	0.104*"gran(granada)" + 0.095*"alhambr(Alhambra)" + 0.027*"españ(España)" + 0.018*"ciud(ciudad)" + 0.014*"fot(foto)" + 0.013*"turism(turismo)" + 0.011*"albaicin(albaicin)" + 0.010*"maravill(maravilla)" + 0.008*"volv(volver)" + 0.008*"bonit(bonito)"
4	0.083*"gran(granada)" + 0.041*"turism(turismo)" + 0.014*"restaur(restaurante)" + 0.011*"andaluci(Andalucía)" +

¹¹ For reasons of clarity, both stemmed and non-stemmed words have been included. Also, the numbers next to each word represent the weights or probabilities.

0.011*"sevilla(sevilla)" + 0.010*"spain(spain)" +
 0.010*"gastronom(gastronomia)" + 0.009*"andaluc(Andalucia)" +
 0.009*"hotel(hotel)" + 0.009*"malag(malaga)"

Spanish (Autumn-Winter), k = 3

- 1 0.117*"gran(granada)" + 0.049*"alhambr(Alhambra)" + 0.034*"españ(España)" + 0.024*"sierranev(sierranevada)" + 0.012*"spain(spain)" + 0.011*"fot(foto)" + 0.011*"andaluci(andalucia)" + 0.011*"sierr(sierra)" + 0.011*"nev(nevada)" + 0.009*"oblig(obligado)"
- 2 0.076*"gran(granada)" + 0.066*"alhambr(Alhambra)" + 0.030*"visit(visita)" + 0.020*"ciud(ciudad)" + 0.013*"viaj(viaje)" + 0.013*"turism(turismo)" + 0.010*"monument(monumento)" + 0.010*"disfrut(disfruta)" + 0.009*"conoc(conoce)" + 0.008*"gui(guía)"
- 3 0.060*"gran(granada)" + 0.021*"alhambr(Alhambra)" + 0.017*"turism(turismo)" + 0.013*"gastronom(gastronomía)" + 0.010*"palaci(palacio)" + 0.009*"restaur(restaurante)" + 0.008*"sevilla(Sevilla)" + 0.007*"tap(tapa)" + 0.007*"nuev(nuevo)" + 0.006*"granadin(granadino)"

1. **Spanish (Spring-Summer):** In this case, the fact that the Alhambra is mentioned in almost every topic demonstrates that the Alhambra is synonymous with the city of Granada. The first topic also considers the other adjoining palaces such as the Generalife, which was the summer royal residence in Arab times; the second topic highlights the Sierra Nevada mountains; the third topic is about the Albaicín, the Moorish quarter of Granada; and, finally, the fourth topic is clearly about gastronomy and visiting other famous cities near Granada, such as Sevilla and Málaga.
2. **Spanish (Autumn-Winter):** Once again, the Alhambra appears in all of the three topics (a visit to Granada without visiting the Alhambra is almost inconceivable), with the first topic focusing on Sierra Nevada, the second topic on guided tours of the city, and the third on gastronomy (*tapas* are a well-established tradition in Granada).

The English topics and keywords are shown in Table 9.

Table 9 Topics and their keywords (English)

Topics	Keywords
<i>English (Spring-Summer), k = 4</i>	
1	0.036*"sierranevada(SierraNevada)" + 0.016*"sierra(sierra)" + 0.015*"nevada(nevada)" + 0.012*"mountain(mountain)" + 0.009*"snow(snow)" + 0.006*"ye(yes)" + 0.005*"ski(ski)" + 0.005*"hike(hike)" + 0.004*"alpujarra(Alpujarra)" + 0.004*"away(away)"
2	0.119*"granada(Granada)" + 0.104*"alhambra(Alhambra)" + 0.048*"spain(Spain)" + 0.019*"palac(palace)" + 0.015*"visit(visit)" + 0.015*"travel(travel)" + 0.012*"beauti(beautiful)" + 0.009*"citi(city)" + 0.008*"day(day)" + 0.008*"place(place)"

- 3 0.021*"palac(palace)" + 0.019*"spain(Spain)" + 0.016*"live(live)" +
0.009*"spectacular(spectacular)" + 0.009*"moorish(moorish)" +
0.008*"best(best)" + 0.008*"lion(lion)" + 0.008*"past(past)" +
0.008*"guid(guide)" + 0.007*"alhambra(Alhambra)"
- 4 0.017*"architectur(architecture)" + 0.015*"year(year)" +
0.014*"fortress(fortress)" + 0.009*"restaur(restaurant)" +
0.008*"locat(location)" + 0.008*"complex(complex)" + 0.008*"arab(arab)" +
0.008*"wonder(wonder)" + 0.007*"restaurant(restaurantes)" +
0.007*"granada(Granada)"

English (Autumn-Winter), k = 4

- 1 0.115*"granada(Granada)" + 0.100*"alhambra(Alhambra)" +
0.053*"spain(Spain)" + 0.022*"palac(palace)" + 0.016*"travel(travel)" +
0.015*"visit(visit)" + 0.012*"beauti(beautiful)" + 0.011*"citi(city)" +
0.010*"andalucia(Andalucia)" + 0.008*"andalusia(Andalusia)"
- 2 0.013*"wall(wall)" + 0.011*"arab(arab)" + 0.011*"great(great)" +
0.011*"find(find)" + 0.010*"spectacular(spectacular)" + 0.009*"guid(guide)" +
0.008*"moorish(moorish)" + 0.008*"live(live)" + 0.008*"citi(city)" +
0.007*"past(past)"
- 3 0.010*"restaur(restaurant)" + 0.009*"meet(meet)" +
0.007*"restaurant(restaurantes)" + 0.007*"design(design)" +
0.005*"food(food)" + 0.005*"entranc(entrance)" + 0.005*"learn(learn)" +
0.004*"calligraphi(calligraphic)" + 0.004*"cultur(culture)" + 0.004*"ok(ok)"
- 4 0.046*"sierranevada(SierraNevada)" + 0.021*"granada(Granada)" +
0.016*"sierra(sierra)" + 0.016*"nevada(nevada)" + 0.015*"ski(ski)" +
0.015*"snow(snow)" + 0.014*"mountain(mountain)" + 0.008*"winter(winter)"
+ 0.008*"snowboard(snowboard)" + 0.006*"locat(local)"

1. **English (Spring-Summer):** In this case, the first topic clearly focuses on the Sierra Nevada, whereas the following two topics focus on the Alhambra, with the third topic focusing on Moorish heritage and the famous Courtyard of the Lions, and the fourth focusing both on Moorish architecture and gastronomy.
2. **English (Autumn-Winter):** In this case, the first topic focuses on Alhambra visits; the second topic focuses on Moorish history and guided tours of the city; the third topic mainly focuses on gastronomy; and the fourth focuses on the Sierra Nevada and its ski resort.

From this study of relevant topics, it can be concluded that the most valued aspects of Granada both in terms of domestic and international tourism and throughout the year are the Alhambra, Sierra Nevada (although for different activities at different times of the year) and gastronomy.

Due to the topics and sentiments in English and Spanish tweets could reflect the different opinions of different tourist groups, this study classifies by language the tweets and shows the results for both languages. It is important to highlight that the tweets in Spanish are probably from national tourists (although not necessarily, they may be Spanish speakers but not from Spain) and the tweets in English from international tourists.

Discussion and Conclusions

Conclusions

Our approach proposes a framework to analyze traveler-generated content by identifying keywords, extracting information through crawlers or scrapers, data cleaning, descriptive analysis and content analysis. In order to test our framework, we have collected Twitter posts about the city of Granada in Spain using the platform API, and the inherent limitations of the API have been partially overcome by collecting information on a daily basis so as not to exceed the established limits. The use of Twint has enabled us to obtain a complete dataset of almost every tweet posted about Granada tourism since 2008. The raw data has been stored in an MySQL relational database. The data is subsequently cleaned to remove any special characters, URLs, repeated characters, etc. from the text, and this is the most extensive and important data analysis process. Various aspects which are necessary for the descriptive analysis are then identified (e.g. unique posts, comments, likes, retweets, multimedia content, etc.), and these provide us with a clear idea of the most influential users for the relevant topics. The data is subsequently processed in order to perform tokenization and stemming, and to remove stop words. During the next stage, word frequency is calculated, sentiment analysis is performed and emotions are classified. Finally, a topic model analysis is carried out to identify the most important topics in the dataset.

We used our framework to collect 262,859 tweets about Granada, which is an important tourist destination in Spain. The most important hashtags are #Alhambra and #SierraNevada, and the most prolific user is @AlhambraCultura. The descriptive analysis provides a broad picture of Granada tourism by considering the most prolific users or influencers. The proposed approach uses a seasonal context and the posted tweets are divided into two periods (spring-summer and autumn-winter) in order to calculate word frequency and to perform topic modelling (sentiment analysis is excluded from this seasonal context because the results in this case are not relevant). Word frequency was calculated, and again *Granada*, *Alhambra* and *Spain* are the most frequent words in both periods (spring-summer and autumn-winter) in English and in Spanish, the words *Granada*, *Alhambra* and *Turismo* are the most frequent words in the spring-summer period, and the words *Granada*, *Alhambra* and *Sierra Nevada* in the autumn-winter period. Concerning the LDA results for tweets in Spanish, according to the spring-summer season the topics found were: Alhambra and Generalife, Sierra Nevada mountains, the Albaicín and gastronomy; the autumn-winter season shows topics like: Alhambra, guided tours to the city and gastronomy. Concerning the results for tweets in English, the spring-summer season shows topics such as Sierra Nevada, Alhambra and Moorish history and Moorish architecture and gastronomy; in the autumn-winter season the topics that stand out are:

Alhambra, Moorish history, gastronomy and Sierra Nevada and its ski resort. The topics that LDA shows with the dataset about Granada show the most important places that tourists could visit, such as the Alhambra as well as the Sierra Nevada mountains, which are places that can be visited all year round. One of the advantages of the province of Granada is the diversity of places it has and it can be enjoyed both in spring-summer as well as autumn-winter season, highlighting the importance of including the seasonal context in the proposed framework.

Theoretical Implications

Some theoretical implications are derived from this job. First, the study set up a link between tourism and social media literature in the data analysis context and contributes to a discover patterns and features of the tourism destination through social media. Second, this study proposes a framework that make up the social media analysis to discover the most important places and things to do in a tourism destination.

In social media platforms like Twitter and Instagram it is possible to share anything in any domain; it has been demonstrated that travelers engage and post reviews or experiences of travelers through social media and tourism-related sites(Kanje et al., 2020); this has been utilized by hospitality practitioners sharing hotel prices through social media and developing a promotional prices that is able to drive reservations, which helps to reduce marketing costs and generate better income (Ampountolas et al., 2019). As has been seen in our findings, there is a lot of relevant information that is shared on social networks that reflects different forms of interaction between users before visiting a tourist destination. This information can be used by both practitioners and travelers.

Finally, our study contributes to the integration of multidisciplinary environments that include hotels and tourism, information technology, social networks, marketing and administration (Nusair, 2020) and we focus on integrating big data analytics techniques and digital tourism to discover new insights for both practitioners and travelers.

Practical Implications

This study is considered the first step in the construction of a twitter tourist data analysis tool that shows the results of the descriptive analysis (post metrics, and user metrics) and the content analysis (sentiment analysis, word frequency and topic model) in this case those corresponding to Granada at Spain. It would be possible to carry out the analysis of any tourist destination in a simple way, so that practitioners and consumers can analyze the information that exists about the tourist destination before the trip. This allows to visualize which are the most visited and popular places depending on the season of the year, to classify the positive and negative experiences that users or travelers have had when visiting the city and to verify the most representative users who constantly post about the destination. This work shows in detail the way in which the information analysis is carried out, so it can be difficult for consumers to understand the technical procedures used. However, this approach allows summarizing a large amount of information to be reviewed by the tourist before their trip through a practical and simple way using a web tool that will implement all the techniques used in this study.

It is important that practitioners can perform the analysis for a certain tourist destination by adding the words that were used as search criteria to the list of stop words. In this work it has been tested in both ways and it has been decided to publish the results in which the words that are search criteria are not excluded due to, in the case of the analysis of word frequencies, it is enough to eliminate that words from the list; for LDA, if the words that are the most characteristic and representative sites of a tourist destination are eliminated, the results will be unclear and a bit confusing.

To perform sentiment analysis there are many tools, among them those belonging to companies such as IBM or Microsoft with greater efficiency, however, the ones used in this work are open source; thus, the entire scientific community can use them easily. The two tools used have different training data, consequently the results of the sentiment analysis cannot be compared.

Limitations and Future Research

There are various problems associated with text processing due to the fact that it is extremely difficult to identify sarcasm, posts may be ambiguous, users may use both Spanish and English words in their tweets, and tweets may contain spelling mistakes, colloquialisms or even abbreviations. This work uses a "word" as the unit of the text analysis, furthermore, by each language a Twitter-specific language could be considered (e.g. in the English language words as "b4" instead of "before"; "BFN" instead of "bye for now", etc). Multilingualism also represents an important limitation since it is not clear how tweets written in different languages should be processed. The size of the dataset is also an important factor since the greater the amount of data, the better the results. One of the largest limitations is the small number of geo-tagged tweets as geo-tagging would provide information about the place where the tweet was posted and opinions of it. The data captured with the twint tool focuses on capturing unique tweets and very few retweets and comments. This deficiency does not allow a more detailed analysis of the information published on Twitter, especially to know what users think about a tweet through their comments and which are the users who amplify the original content through a retweet.

This work proposes an approach for a descriptive analysis of the users who post tweets about a tourist destination and a content analysis of these posts. This is, however, only the first step in the development of social media analytics and social media influence in the sphere of tourism. In the future, content analysis can be complemented by analysing images and videos through deep learning techniques which would enable us to identify the most popular places through multimedia content. One additional drawback is that the dataset used in this work does not have sufficient geo-tagged tweets. However, if the users posted their tweets with an active geolocation, sentiment analysis would provide information about the most popular places and events, and negative emotions would provide tour operators with information about areas for improvement in terms of infrastructure and services.

The data collected from these platforms is quite extensive and requires extensive work to clean it, however, they provide information on a tourist destination taking into account relevant users, discovering the most commented sites through content analysis or geo

location of the posts and find the deficiencies that the places have when there are negative comments or complaints about the place they have visited. This information is very useful for industry practitioners so that they can improve their service and focus their strategies according to the season.

Acknowledgement(s)

This work has been funded by the Spanish Ministerio de Economía y Competitividad under project TIN2016-77902-C3-2-P, and the European Regional Development Fund (ERDF-FEDER).

References

- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hai, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. *Journal of Medical Internet Research*, 22(4). <https://doi.org/10.2196/19016>
- Ampountolas, A., Shaw, G., & James, S. (2019). The role of social media as a distribution channel for promoting pricing strategies. *Journal of Hospitality and Tourism Insights*, 2(1), 75–91. <https://doi.org/10.1108/jhti-07-2018-0040>
- Aoudi, S., & Malik, A. (2019). Lexicon Based Sentiment Comparison of iPhone and Android Tweets during the Iran-Iraq Earthquake. *ITT 2018 - Information Technology Trends: Emerging Technologies for Artificial Intelligence*, 233–238. <https://doi.org/10.1109/CTIT.2018.8649509>
- Aramburu, M. J., Berlanga, R., & Lanza, I. (2020). Social media multidimensional analysis for intelligent health surveillance. *International Journal of Environmental Research and Public Health*, 17(7). <https://doi.org/10.3390/ijerph17072289>
- Arthur, R., & Williams, H. T. P. (2019). Scaling laws in geo-located Twitter data. *PLoS ONE*, 14(7). <https://doi.org/10.1371/journal.pone.0218454>
- Assaker, G. (2020). Age and gender differences in online travel reviews and user-generated-content (UGC) adoption: extending the technology acceptance model (TAM) with credibility theory. *Journal of Hospitality Marketing and Management*, 29(4), 428–449. <https://doi.org/10.1080/19368623.2019.1653807>
- Bakshi, S., Dogra, N., & Gupta, A. (2019). What motivates posting online travel reviews? integrating gratifications with technological acceptance factors. *Tourism and Hospitality Management*, 25(2), 335–354. <https://doi.org/10.20867/thm.25.2.5>
- Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. *Tourism Management*, 71, 490–503. <https://doi.org/10.1016/j.tourman.2018.10.013>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.

- Bohr, J. (2020). "Reporting on climate change: A computational analysis of U.S. newspapers and sources of bias, 1997–2017." *Global Environmental Change*, 61. <https://doi.org/10.1016/j.gloenvcha.2020.102038>
- Casadei, P., & Lee, N. (2020). Global cities, creative industries and their representation on social media: A micro-data analysis of Twitter data on the fashion industry. *Environment and Planning A*. <https://doi.org/10.1177/0308518X20901585>
- Chang, Y. C., Ku, C. H., & Chen, C. H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management*, 80. <https://doi.org/10.1016/j.tourman.2020.104129>
- Claster, W. B., Cooper, M., & Sallis, P. (2010). Thailand - Tourism and conflict. Modeling sentiment from twitter tweets using naïve bayes and unsupervised artificial neural nets. *Proceedings - 2nd International Conference on Computational Intelligence, Modelling and Simulation, CIMSIm 2010*, 89–94. <https://doi.org/10.1109/CIMSIm.2010.98>
- Colditz, J. B., Chu, K. H., Emery, S. L., Larkin, C. R., James, A. E., Welling, J., & Primack, B. A. (2018). Toward real-time infoveillance of twitter health messages. In *American Journal of Public Health* (Vol. 108, Issue 8, pp. 1009–1014). American Public Health Association Inc. <https://doi.org/10.2105/AJPH.2018.304497>
- Cripps, H., Singh, A., Mejtoft, T., & Salo, J. (2020). The use of Twitter for innovation in business markets. *Marketing Intelligence and Planning*. <https://doi.org/10.1108/MIP-06-2019-0349>
- Curlin, T., Jaković, B., & Miloloža, I. (2019). Twitter usage in Tourism: Literature Review. In *Business Systems Research* (Vol. 10, Issue 1, pp. 102–119). Sciendo. <https://doi.org/10.2478/bsrj-2019-0008>
- Dai, W. (Daisy), Jin, G., Lee, J., & Luca, M. (2018). Aggregation of consumer ratings: an application to Yelp.com. *Quantitative Marketing and Economics*, 16(3), 289–339. <https://doi.org/10.1007/s11129-017-9194-9>
- DeAndrea, D. C., Van Der Heide, B., Vendemia, M. A., & Vang, M. H. (2018). How People Evaluate Online Reviews. *Communication Research*, 45(5), 719–736. <https://doi.org/10.1177/0093650215573862>
- Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1), 012024. <https://doi.org/10.1088/1742-6596/1192/1/012024>
- Gloor, P., Fronzetti Colladon, A., de Oliveira, J. M., & Rovelli, P. (2020). Put your money where your mouth is: Using deep learning to identify consumer tribes from word usage. *International Journal of Information Management*, 51. <https://doi.org/10.1016/j.ijinfomgt.2019.03.011>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews:

- Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Hoshino, Y., Ishii, E., & Yamada, M. (2018). *A Study of Recommended Tourist Spot Information Extraction Using SNS*. <https://search.proquest.com/docview/2288609482?pq-origsite=gscholar&fromopenview=true>
- Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management*, 74, 276–289. <https://doi.org/10.1016/j.tourman.2019.03.009>
- Joseph, N., Kar, A. K., Ilavarasan, P. V., & Ganesh, S. (2017). Review of Discussions on Internet of Things (IoT): Insights from Twitter Analytics. *Journal of Global Information Management*, 25(2), 38–51. <https://doi.org/10.4018/JGIM.2017040103>
- Kanje, P., Charles, G., Tumsifu, E., Mossberg, L., & Andersson, T. (2020). Customer engagement and eWOM in tourism. *Journal of Hospitality and Tourism Insights*, 3(3), 273–289. <https://doi.org/10.1108/jhti-04-2019-0074>
- Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2020). Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science*, 46(3), 313–324. <https://doi.org/10.1177/0165551519828620>
- Kemp, S. (2020). *Digital 2020: 3.8 billion people use social media - We Are Social*. <https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media>
- Kim, Y., Nordgren, R., & Emery, S. (2020). The story of goldilocks and three twitter's APIs: A pilot study on twitter data sources and disclosure. *International Journal of Environmental Research and Public Health*, 17(3). <https://doi.org/10.3390/ijerph17030864>
- Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (Robert). (2018). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*, 57(8), 1012–1025. <https://doi.org/10.1177/0047287517729757>
- Leoni, V. (2020). Stars vs lemons. Survival analysis of peer-to peer marketplaces: the case of Airbnb. *Tourism Management*, 79. <https://doi.org/10.1016/j.tourman.2020.104091>
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. <https://doi.org/10.1016/j.tourman.2007.05.011>
- Liu, M. T., Liu, Y., Mo, Z., & Ng, K. L. (2020). Using text mining to track changes in travel destination image: the case of Macau. *Asia Pacific Journal of Marketing and Logistics*. <https://doi.org/10.1108/APJML-08-2019-0477>

- Liu, T., Xia, J., & Crowe-Delaney, L. (2020). Tourism Information Diffusion through SNSs: A Theoretical Investigation. *Sustainability*, *12*(5), 1731. <https://doi.org/10.3390/su12051731>
- Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, *54*, 321–343. <https://doi.org/10.1016/j.tourman.2015.12.008>
- Marine-Roig. (2019). Destination Image Analytics through Traveller-Generated Content. *Sustainability*, *11*(12), 3392. <https://doi.org/10.3390/su11123392>
- Marine-Roig, E., Ferrer-Rosell, B., Daries, N., & Cristobal-Fransi, E. (2019). Measuring gastronomic image online. *International Journal of Environmental Research and Public Health*, *16*(23). <https://doi.org/10.3390/ijerph16234631>
- Martí, P., García-Mayor, C., & Serrano-Estrada, L. (2020). Taking the urban tourist activity pulse through digital footprints. *Current Issues in Tourism*, 1–20. <https://doi.org/10.1080/13683500.2019.1706458>
- Monachesi, P. (2020). Shaping an alternative smart city discourse through Twitter: Amsterdam and the role of creative migrants. *Cities*, *100*. <https://doi.org/10.1016/j.cities.2020.102664>
- Myneni, M. B., & Dandamudi, R. (2020). Harvesting railway passenger opinions on multi themes by using social graph clustering. *Journal of Rail Transport Planning and Management*, *13*. <https://doi.org/10.1016/j.jrtpm.2019.100151>
- Nascimento Filho, F. B. do, Flores, L. C. D. S., & Limberger, P. F. (2019). Análise do posicionamento dos restaurantes de São Paulo estrelados pelo guia Michelin com base nas On-line Travel Reviews (OTRS). *Revista Brasileira de Pesquisa Em Turismo*, *13*(2), 1–15. <https://doi.org/10.7784/rbtur.v13i2.1453>
- Nusair, K. (2020). Developing a comprehensive life cycle framework for social media research in hospitality and tourism: A bibliometric method 2002-2018. *International Journal of Contemporary Hospitality Management*, *32*(3), 1041–1066. <https://doi.org/10.1108/IJCHM-09-2019-0777>
- Ovádek, M. (2020). “Popular tribunes” and their agendas: topic modelling Slovak presidents’ speeches 1993–2020. *East European Politics*. <https://doi.org/10.1080/21599165.2020.1756785>
- Pai, M. Y., Wang, D. C., Hsu, T. H., Lin, G. Y., & Chen, C. C. (2019). On ontology-based tourist knowledge representation and recommendation. *Applied Sciences (Switzerland)*, *9*(23). <https://doi.org/10.3390/app9235097>
- Park, S. B., Kim, J., Lee, Y. K., & Ok, C. M. (2020). Visualizing theme park visitors’ emotions using social media analytics and geospatial analytics. *Tourism Management*, *80*, 104127. <https://doi.org/10.1016/j.tourman.2020.104127>

- Punel, A., & Ermagun, A. (2018). Using Twitter network to detect market segments in the airline industry. *Journal of Air Transport Management*, 73, 67–76. <https://doi.org/10.1016/j.jairtraman.2018.08.004>
- Ramanathan, V., & Meyyappan, T. (2019, February 19). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*. <https://doi.org/10.1109/ICBDSC.2019.8645596>
- Rao, Y., Pang, J., Xie, H., Liu, A., Wong, T. L., Li, Q., & Wang, F. L. (2017). Supervised intensive topic models for emotion detection over short text. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10177 LNCS, 408–422. https://doi.org/10.1007/978-3-319-55753-3_26
- Ren, Z., Jiang, B., & Seipel, S. (2019). Capturing and characterizing human activities using building locations in America. *ISPRS International Journal of Geo-Information*, 8(5). <https://doi.org/10.3390/ijgi8050200>
- Samoggia, A., Riedel, B., & Ruggeri, A. (2020). Social media exploration for understanding food product attributes perception: the case of coffee and health with Twitter data. *British Food Journal*, ahead-of-p(ahead-of-print). <https://doi.org/10.1108/bfj-03-2019-0172>
- Shin, S., Chung, N., Xiang, Z., & Koo, C. (2019). Assessing the Impact of Textual Content Concreteness on Helpfulness in Online Travel Reviews. *Journal of Travel Research*, 58(4), 579–593. <https://doi.org/10.1177/0047287518768456>
- Singh, S., Chauhan, A., & Dhir, S. (2019). Analyzing the startup ecosystem of India: a Twitter analytics perspective. *Journal of Advances in Management Research*, 17(2), 262–281. <https://doi.org/10.1108/JAMR-08-2019-0164>
- Soong, H. C., Jalil, N. B. A., Kumar Ayyasamy, R., & Akbar, R. (2019). The essential of sentiment analysis and opinion mining in social media : Introduction and survey of the recent approaches and techniques. *ISCAIE 2019 - 2019 IEEE Symposium on Computer Applications and Industrial Electronics*, 272–277. <https://doi.org/10.1109/ISCAIE.2019.8743799>
- Srivastava, D. K., & Roychoudhury, B. (2020). Words are important: A textual content based identity resolution scheme across multiple online social networks. *Knowledge-Based Systems*, 195. <https://doi.org/10.1016/j.knosys.2020.105624>
- Srivastava, V., & Kalro, A. D. (2019). Enhancing the Helpfulness of Online Consumer Reviews: The Role of Latent (Content) Factors. *Journal of Interactive Marketing*, 48, 33–50. <https://doi.org/10.1016/j.intmar.2018.12.003>
- Sudhakar, S., & Gunasekar, S. (2020). Examining online ratings and customer satisfaction in airlines. *Anatolia*, 31(2), 260–273. <https://doi.org/10.1080/13032917.2020.1747238>

- Tian, Y., Chen, C., Chen, X., Zhang, Q., & Sun, R. (2020). Research on real-time analysis technology of urban land use based on support vector machine. *Pattern Recognition Letters*, 133, 320–326. <https://doi.org/10.1016/j.patrec.2020.03.022>
- Tsai, F. M., & Bui, T. D. (2020). Impact of word of mouth via social media on consumer intention to purchase cruise travel products. *Maritime Policy and Management*. <https://doi.org/10.1080/03088839.2020.1735655>
- Uşaklı, A., Koç, B., & Sönmez, S. (2017). How ‘social’ are destinations? Examining European DMO social media usage. *Journal of Destination Marketing and Management*, 6(2), 136–149. <https://doi.org/10.1016/j.jdmm.2017.02.001>
- Viñan-Ludeña, M.-S. (2019). A Systematic Literature Review on Social Media Analytics and Smart Tourism. In *Smart Tourism as a Driver for Culture and Sustainability* (pp. 357–374). Springer, Cham. https://doi.org/10.1007/978-3-030-03910-3_25
- Viñán-Ludeña, M. S., De Campos, L. M., Jacome-Galarza, L. R., & Sinche-Freire, J. (2020). Social media influence: a comprehensive review in general and in tourism domain. *Smart Innovation, Systems and Technologies*, 171, 25–35. https://doi.org/10.1007/978-981-15-2024-2_3
- Wang, W., Ying, S., Lyu, J., & Qi, X. (2019). Perceived image study with online data from social media: the case of boutique hotels in China. *Industrial Management and Data Systems*, 119(5), 950–967. <https://doi.org/10.1108/IMDS-11-2018-0483>
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Xu, Y., Zhang, Z., Law, R., & Zhang, Z. (2019). Effects of online reviews and managerial responses from a review manipulation perspective. *Current Issues in Tourism*. <https://doi.org/10.1080/13683500.2019.1626814>
- Yoosefi Nejad, M., Delghandi, M. S., Bali, A. O., & Hosseinzadeh, M. (2020). Using Twitter to raise the profile of childhood cancer awareness month. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1). <https://doi.org/10.1007/s13721-019-0206-4>
- Zhang, X., Yang, Y., Zhang, Y., & Zhang, Z. (2020). Designing tourist experiences amidst air pollution: A spatial analytical approach using social media. *Annals of Tourism Research*, 84, 102999. <https://doi.org/10.1016/j.annals.2020.102999>

Analyzing tourist data on Twitter: a case study in the province of Granada at Spain

Journal of Hospitality and Tourism Insights (accepted for publication 26-Feb-2021)

DOI 10.1108/JHTI-11-2020-0209 This Author Accepted Manuscript is deposited under the Creative Commons Attribution Non-commercial International Licence 4.0 (CC BY-NC 4.0). Any reuse is allowed in accordance with the terms outlined by the licence.