

## Research Article

# Use Social Media Knowledge for Exploring the Portuguese Wine Industry: Following Talks and Perceptions?

**Gael Pérez-Rodríguez** <sup>1,2</sup> **João Pedro Baptista** <sup>3,4</sup> **Gilberto Igrejas** <sup>5,6,7</sup>  
**Florentino Fdez-Riverola** <sup>1,2</sup> and **Anália Lourenço** <sup>1,2,8</sup>

<sup>1</sup>CINBIO, Universidade de Vigo, Department of Computer Science, ESEI-Escuela Superior de Ingeniería Informática, 32004 Ourense, Spain

<sup>2</sup>SING Research Group, Galicia Sur Health Research Institute (ISS Galicia Sur), SERGAS-UVIGO, Vigo, Spain

<sup>3</sup>Department of Communication, Philosophy and Politics, University of Beira Interior (UBI), Covilhã, Portugal

<sup>4</sup>LabCom-Communication and Arts, UBI, Covilhã, Portugal

<sup>5</sup>Department of Genetics and Biotechnology, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

<sup>6</sup>Functional Genomics and Proteomics Unit, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

<sup>7</sup>LAQV-REQUIMTE, Faculty of Science and Technology, Nova University of Lisbon, Lisbon, Portugal

<sup>8</sup>Centre of Biological Engineering (CEB), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Correspondence should be addressed to Florentino Fdez-Riverola; riverola@uvigo.es

Received 10 January 2022; Accepted 21 May 2022; Published 11 June 2022

Academic Editor: Kashif Zia

Copyright © 2022 Gael Pérez-Rodríguez et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work presents an exploratory study that retrieves, processes, and analyses Twitter data to gain insights about the relevance and perceptions of the wine industry in the Douro Portuguese region (including Porto and Douro wines), as well as other regions in the country. The main techniques and algorithms used in our work belong to the families of natural language processing and machine learning, and the practical relevance of the proposed methodology has been proven in the analysis of 1.2 million unique messages from more than 764,000 distinct users retrieved from the Twitter platform. Derived results from this study are valuable to provide insights that can be further used in the context of Business Informatics to promote better and more efficient marketing campaigns, for example, centering the topic on the most interested people or communicating with the most appropriate words.

## 1. Introduction

The universe of wine (i.e., consumption, marketing, and tourism) is a social-cultural phenomenon accompanying the development of western societies throughout history [1], being currently considered a sector typically conditioned by social influences [2]. With the growing advancement of new technologies and the emergence of a more digital universe, the wine sector, due to the characteristics that define it, is increasingly dependent on the interaction and sharing of experiences among companies, professionals and consumers in online communities. Social media has changed the way companies and consumers interact, creating a greater exchange of information and a closer, more active, and loyal

relationship, essential for adding value to the product or brand [3–5]. In recent years, there has been an exponential increase in the number of tourists using the Internet (especially social media) to book travel and accommodation, buy products and services, and choose holiday destinations.

Social media has become an opportunity not only for users or consumers, but also for brand managers. Social media allow the production and dissemination of experiences among users, while simultaneously serving as a fundamental tool for brand managers to collect information and assess the visibility of their product or service. Furthermore, several studies have already demonstrated the impact that social media can have on the decision to shop online [6], including the purchase and improvement of wines [7, 8].

Indeed, the use of social media is impacting the global discussion of the wine sector, in relation to not only wine brands but also wine tourism [9]. It has always been known that the wine sector is very dependent on sharing experiences and that there is great trust in people's testimony regarding the assessment of product quality. Taking this argument into account, it becomes increasingly urgent to understand what is being said in a social network about a particular brand or product of wine. In addition to the fact that personal testimonies can impact a user's decision-making [10], we also know that users trust the experiences of their peers [11] and take feedback provided by users in its online community very seriously [12]. For example, with regard to the wine sector, several studies (e.g., [13]) recognize the importance of Twitter for beverage bloggers to express their opinion and share tips and advice. In fact, these bloggers may be becoming public opinion leaders as they are considered to be trustworthy and credible [13, 14]. Based on these arguments, Wilson and Quinton [15] considered Twitter a fundamental tool to add soft value to wine-focused businesses. In this sense, what is said about wine on a network such as Twitter, from a personal or business perspective, positive or negative, can influence the public's opinion about a certain brand.

The main objective of this study is to verify what is being said on Twitter about Portuguese wines and wine regions, including Porto and Douro. Despite the various studies that have been published, some questions remain unanswered by either academia or wine professionals. This study does not aim to make a correlation between the content posted and the purchase of wines. Instead, we are interested in collecting information and understanding the public (consumers and tourists), which we consider to be a prerequisite for the success of a product, brand, or tourist destination.

The wine industry plays a fundamental role in the economy of many countries, such as Portugal. In the year 2018, the volume of wine produced in this country reached approximately 6.1 million hectoliters, while approximately 3 million hectoliters were exported to the rest of the world with an invoicing of more than 800 million euros [16, 17]. In Portugal, there are several wine regions and protected designations of origin (DOP), with the Douro region being the most recognized and prolific [18]. Located in the north of Portugal, this region is one of the most important wine-producing regions in the country and is internationally known for being the origin of the famous Port Wine brand [19]. The Douro Valley is also the first demarcated wine-producing region in the world (since 1756), and Alto Douro Wine Region was classified as a World Heritage Site by the United Nations Educational, Scientific and Cultural Organization (UNESCO) in 2001 [20]. This distinction sparked growth in tourism, wine tourism, and an appreciation of Douro wine production [21, 22]. Like other world-famous wine regions, the Douro Valley has become a place of interest to visit (due to its scenic beauty), combining tourism and wine production as an economic and social driver at a regional and local level [23–25].

The importance of our study is twofold. First, the existing literature on the Douro and other Portuguese wine regions

has focused mainly on business management methodologies with little regard for the perception tourists have of the region and their consumer behavior [22]. Studies that focus on the perception of tourists, including business travellers, essentially resort to the analysis of questionnaires and interviews [19, 26]. In contrast, our research offers a vision focused on social media with applied machine learning techniques, which we believe can be a very important contribution to the existing literature.

Second, our study offers an international perspective on what is being said and commented on about the Portuguese wine sector on Twitter. Not only can this network add value to the wine business, [15] but its total number of users continues to increase. Furthermore, it is important to emphasize that our research is the first study that uses Twitter to explore the impact of the Douro and Porto wine regions.

The number of active users on these platforms keeps growing on a daily basis and is estimated to reach 3.1 billion by 2021 [27]. The ease of use, high accessibility, and broad audiences of social media combined with the natural socialization around wines pave the way for the promotion of new discussion forums and a better perception of consumer preferences towards novel marketing strategies designed around wine and tourism [12, 28]. For example, one study demonstrated the association between luxury wine brands and the social media visibility indicators, suggesting that some of these brands should consider these platforms in their marketing strategies [29]. Notably, a study on the impact of social media practices on wine sales in US wineries showed that 87 percent of the wineries reported a perceived increase in sales due to social media [30]. Regarding wine and tourism partnership, a previous study about the relevance of social media on the wine tourism in Greece suggested that customer relationship management could be improved using these channels of communication [31]. Other studies pointed out the fundamental role that social media may have the promotion and valuation of brands with consumers; for example, social media can be important in assessing customer satisfaction [32, 33] by quickly and closely verifying the relationship between brand and consumer, with greater involvement between both [34, 35], and by capturing consumer experiences or guiding potential buyers, or by understanding the quality of service through, as an example, negative or positive comments related to a certain product. Moreover, the presence of a brand on social media can affect the company's competitiveness and its value or reputation.

The present research contributes to this line of work by creating a methodology to perform an exploratory analysis on Twitter to gain insight about the relevance and the perceptions of the wine industry in the Douro Portuguese Region (including Porto and Douro wines) as well as the other regions in the country.

Several studies have already pointed out that although Twitter is considered a generalist social network (i.e., its users talk about any type of conversation), it is becoming increasingly important in carrying out market and business research [36–38]. It is also currently one of the social networks with the largest number of registered and active users

who produce a large amount of content (i.e., tweets) daily [39]. Finally, Twitter provides an application programming interface (API) to easily access the information posted by users. For these reasons, Twitter was the media platform selected to carry out this research, although the presented methodology is not bound to this specific social network and can be applied to others. Extracting data based on posted information (tweets and retweets) allows us to reduce the cost of collecting data by other means. In addition, the act of tweeting is inherently associated with the user's search to engage with other people and to disseminate and share ideas and opinions, which represents an excellent object of study to understand consumer perception. We consider all types of tweets, whether from professionals or consumers, related to the wine sector.

In short, this work seeks to fill the lack of studies focused on this topic, taking into account the problems identified in the literature. As we mentioned earlier, no study to date in Portugal has made such a comprehensive survey of social media posts about the Douro Valley and Douro and Porto wines. Correia et al. [19] carried out a global survey of tourists' experiences of the Douro Valley in Portugal, but essentially using interviews and surveys. In addition, the authors performed a content analysis of tourists' comments on the brands' pages on Facebook and TripAdvisor. Correia et al. [19] also found that many wine companies did not have social media pages or profiles. In a similar study, Vieira et al. [40] applied a sentiment analysis to 814 posts and found that the Port wine brand receives, in general, a very positive appreciation by online stakeholders. However, the authors found that the young generation (Millennials) has lost interest in Port wine.

Allied to these problems—(1) lack of online presence of brands and (2) lack of interest in the product among younger people—it should also be noted that sales of Port wine declined in the last decade, between 2006 and 2016, in terms of volume and value, a trend that has not been observed in terms of tourist attraction in the Douro Valley [40]. However, and to reinforce the importance of the online market, in 2020, the COVID-19 pandemic served to elucidate the relevance of online presence in the commercialization of Douro and Port wines. With the confinement imposed by the pandemic situation, there was an exponential increase in the online purchase of wine (information available online: <https://www.tsf.pt/lifestyle/confinamento-leva-a-aumento-exponencial-dacompra-de-vinho-online-12123858.html> (Accessed on April 8, 2022)). Aware of the importance of the online presence of the Douro brand, the Instituto dos Vinhos do Douro e do Porto, the entity responsible for controlling the quality and quantity of Port wines, increased the budget dedicated to promoting the brand to the international market, reinforcing its commitment to digital market (Information available online: <https://www.dinheirovivo.pt/economia/vinhos-do-douro-e-porto-reforcampromocao-para-os-25-me-em-2021-13346251.html> (Accessed on April 8, 2022)).

In view of the current situation, in which, first of all, the importance of social media in promoting the territory and wines of the Douro is recognized, although there is still a

reduced presence of producers and brands on social media, and, secondly, the online market has been discovered as an opportunity to combat the increasing drops in sales over the previous years, our study emerges as an unprecedented and very relevant approach, which will allow us to identify the most influential users, the perception of buyers about the brand and territory, and the most discussed topics on social media. It is for these reasons that our study is necessary.

The main contribution of this research lays in unveiling the general knowledge extracted from the people who use Twitter to communicate with each other and focusing on the topic of the Portuguese wine industry. With the analysis of this knowledge, it could be possible to develop the capacity to better and more efficiently promote marketing campaigns (e.g., centering the topic on the most interested people or communicating with the most appropriate words), an objective of the field of business informatics.

The methodological design of our study is structured in three complementary phases. First, we collected and filtered data to build three corpora related to wine and wine regions. Only English-language tweets were considered to obtain an international dimension. Then, the corpora were processed according to a set of criteria that allowed filtering the pertinent information. Subsequently, we proceed to the content analysis of the data, highlighting the main topics, the most relevant words, the most active users, and the associated text segments.

## 2. Materials and Methods

The methodology developed to retrieve, process, and study wine-related Twitter data consists of three fundamental steps: (i) data collection and filtering; (ii) corpus processing; and (iii) information analysis (Figure 1).

*2.1. Data Collection and Filtering.* The Twitter4J Java library was used to retrieve the tweets related to the topic of wine [41]. The query used to retrieve the tweets was “wine.” The idea of using such a general query was to obtain as many tweets as possible about the topic in a small window of time. Twitter4J was set to retrieve only English tweets, that is, tweets that the Twitter API considered to be written in the English language, while tweets from suspended user accounts were excluded. The collection of posts dates from 8 January 2020 to 8 May 2020, obtaining a collection of 1,200,428 unique tweets from 764,108 distinct users. From this general collection, three corpora were constructed using more specific keywords to focus on the conversations about Portugal. In particular, the keywords “Douro,” “Porto” (including variants like “port”) and the “Others” PDO and wine regions (e.g., “vinho verde,” “bairrada,” or “minho”) were searched for within the content of the tweets. A tweet is considered to belong to the “Douro” corpus, for example, if the user cited the keyword “douro” in the tweet. Figure 2 shows the number of unique tweets and users for the three generated corpora, as well as the common tweets and users among them (i.e., a tweet may cite both “Porto” and “Douro” and a user may tweet about two or more regions).

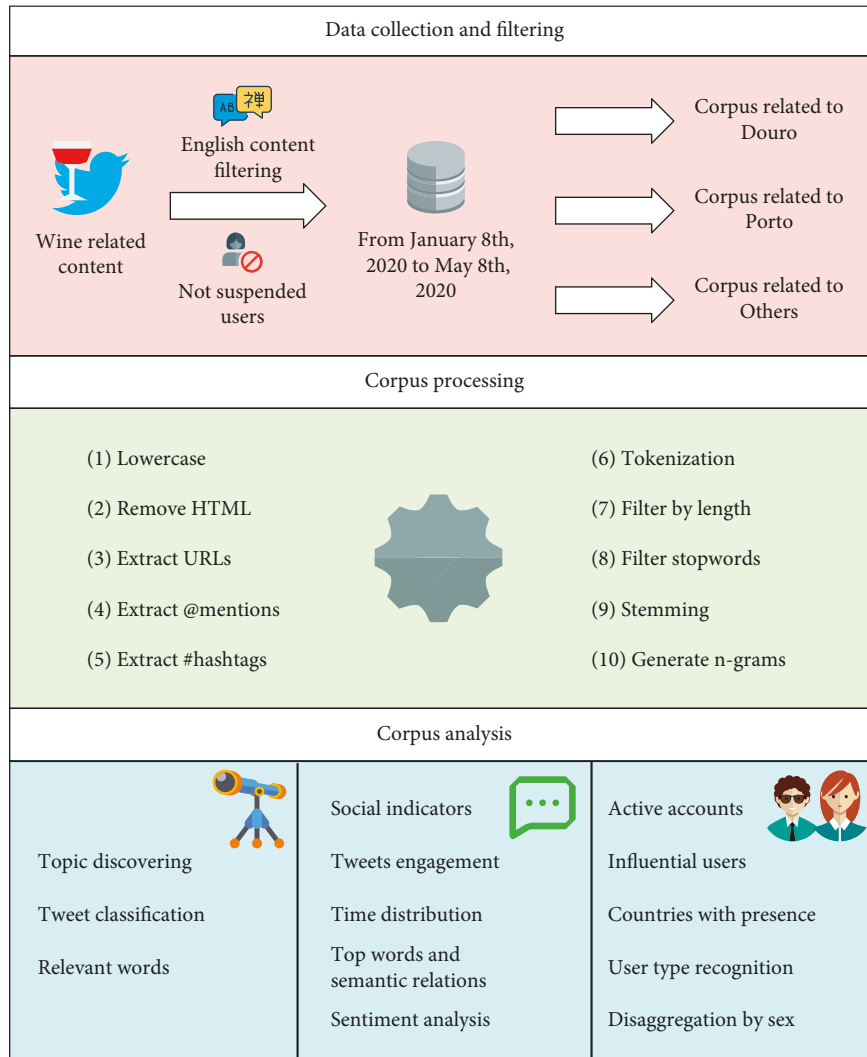


FIGURE 1: The workflow implemented to retrieve, process, and analyze the wine-related Twitter data.

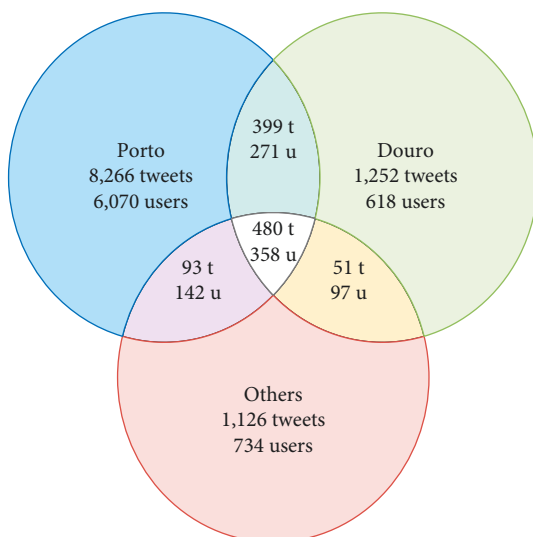


FIGURE 2: Statistics for the Douro, Porto, and other corpora.

**2.2. Corpora Processing.** Every tweet in the corpora was processed according to the following steps: (i) transform to lowercase; (ii) remove HTML tags; (iii) extraction of URL entities; (iv) recognition and extraction of user mentions (e.g., @ivdp\_ip); (v) identification and extraction of hashtags (e.g., #winelover); (vi) tokenization (i.e., split a set of text up into words); (vii) filter tokens by length (i.e., between 3 and 20 characters); (viii) filter English stopwords and generic domain words (e.g., wine); (ix) stemming (i.e., tasting to tast); and (x) generation of n-grams (up to bigrams).

While the extraction of different elements from the tweets (i.e., URLs, mentions and hashtags) denotes the omission of these values from the text, they were kept for other analyses.

**2.3. Topic Discovery.** The latent Dirichlet allocation (LDA) algorithm was applied to discover the topics of conversation most discussed the three corpora [42]. LDA is a probabilistic topic modeling method widely used to identify possible topics in a general set of documents. The key assumption

behind LDA is that each given document is a mix of multiple topics [43]. LDA models have two hyperparameters,  $\alpha$  and  $\beta$ , used to tune the document-topic distribution and the topic-word distribution, respectively. Therefore, a low value of  $\alpha$  places less weight on documents that may be high in number but address secondary topics. Similarly, a low value of  $\beta$  places less weight on topics composed of a high number of only secondary  $n$ -grams (i.e., uni-, bi-, or tri-grams).

To improve the model outcomes, a grid search approximation was applied to find the optimal hyperparameters of the LDA models. The objective was to maximize both the model perplexity and the topic coherence score [44]. To increase the scoring of the topic coherence, the model only considered nouns and proper nouns for the Porto corpus [45].

**2.4. User Characterization.** The following analyses were conducted for every user in the three corpora: (i) identify the most active and cited users inside and outside the topic; (ii) identify the influential users presented in the corpora; (iii) detect countries with the most presence, that is, countries that presented the most activity of users; (iv) discover the type of user account; and (v) analyze gender distribution.

The most active users were identified by taking into account the number of tweets posted about the topic (i.e., tweets containing the term “wine” and one of the Portuguese keywords) and outside the topic (i.e., those tweets not containing the term “wine” nor the Portuguese keywords). The most cited users were discovered by counting the number of mentions made by other users.

The identification of influential people, that is, users whose content spreads easily through a large part of the community, is a topic already discussed in previous works [46, 47]. These works proposed several valid ways to calculate this metric depending on the definition of influence applied. For example, to identify the influential users in the three analyzed corpora, the users were ranked by their follower ( $F/f$ ), retweet ( $Rt/t$ ), and favorite ( $Fav/t$ ) ratios. The equation of the  $F/f$  ratio can be used as an indicator of the potential that the user has to spread their content and to estimate the number of people interested in the user’s status updates:

$$\frac{F}{f} \text{ ratio} = \frac{\#Followers}{\#Followings}, \quad (1)$$

where  $\#Followers$  refers to the number of followings of the user and  $\#Followings$  indicates the number of people who follow the user.

The equation of  $Rt/t$  ratio and  $Fav/ratio$  can be used as an indicator of the potential engagement that the user can reach inside the network and it can be useful when comparing different users in the same subject:

$$\frac{Rt}{t} \text{ ratio} = \frac{\#Retweets}{\#Tweets}, \quad (2)$$

where  $\#Retweets$  refers to the total number of retweets and  $\#Tweets$  indicates the total number of tweets.

$$\frac{Fav}{t} \text{ ratio} = \frac{\#Favorites}{\#Tweets}, \quad (3)$$

where  $\#Favorites$  refers to the total number of favorite tweets, whereas  $\#Tweets$  indicates the number of total tweets related to the topic.

Note that equations (2) and (3) can be adapted to only consider the tweets related to a specific topic, for example, tweets related to the Douro region. With this consideration, it was possible to identify users who might have influence in a small specific community rather than the whole network. So, the adapted equations are the following:

$$\frac{Rt}{t} \text{ ratio} = \frac{\#Retweets\_topic}{\#Tweets\_topic}, \quad (4)$$

where  $\#Retweets\_topic$  refers to the number of retweets reached in the tweets related to the studied topic and  $\#Tweets\_topic$  indicates the number of tweets related to the topic.

$$\frac{Fav}{t} \text{ ratio} = \frac{\#Favorites\_topic}{\#Tweets\_topic}, \quad (5)$$

where  $\#Favorites\_topic$  refers to the number of favorites reached in the tweets related to the topic under study, whereas  $\#Tweets\_topic$  indicates the number of tweets related to the topic.

The detection of active countries and the gender of the users enabled a demographic analysis of the accounts. To perform the detection of the user location, the proposed method accounted for the information related to time-zone, the UTC-offset, and the location text introduced by the user on Twitter. This information was used in combination with the GeoNames database API [48]. A strategy based on a gender-name dictionary and a convolutional neural network model to identify faces were used to identify the gender of the user. Finally, the classification of user accounts considered three types of users, namely, individuals, organizations, and unknown. To perform this classification, the proposed strategy used the information of the user profile (i.e., name and description) provided by Twitter. The algorithms and techniques employed to infer this knowledge were originally proposed by the works of Pérez-Pérez et al. [49, 50].

**2.5. Content Analysis.** Taking into consideration all the tweets stored in the corpora, the following steps were performed: (i) identification and analysis of most common hashtags, resources, and mentions; (ii) study of the tweets with larger engagement (i.e., favorites and retweets); (iii) analysis of posting distribution over time; (iv) discovery of relevant words and semantic associations; and (v) sentiment analysis.

The identification of the most common hashtags, shared resources, and mentions led to the idea of performing a descriptive analysis of these data to better understand what people like to share or like to point out about the theme under consideration. The study of the tweets with high values of engagement was meant to discover which content became popular in the network and why it had reached that popularity. The analysis of the posting time distribution

TABLE 1: Topics discovered in the Douro corpus.

Topic ID	Topic label	Topic terms	#Tweets	Fav. tweet	RT. tweet
0	Quinta do Crasto	Quintadocrasto, crasto, quinta, winelove, reserve	152	Quinta do Crasto view from the other bank of the #Douro river. ??? Photo: ©Pedro Guedes de Almeida (Team #QuintadoCrasto) #Portugal #Vinho #Wine #Wein #Vin #Vino #ワイン #와인 #ВИНО #Viini #WinesofPortugal #PortugueseWines #DouroBoys #Travel #Enoturismo #WineTourism <a href="https://t.co/jWR2jYB0s9">https://t.co/jWR2jYB0s9</a>	The #harvest is about to end ! Last team efforts at Vinha Maria Teresa ! ?? ? #working in the #Douro is hard, requires perseverance and an endless passion.? fx1. photos: © Vasco Maia Lopes/Col. Quinta do Crasto. All rights reserved. . #Vindima #Vinho #Vinhos #Wine #Wines <a href="https://t.co/Zp24SMIUQB">https://t.co/Zp24SMIUQB</a>
1	Types of wine	Red, white, touriga, fruit, bottle	280	Today's lockdown wine. A charming Douro red made in a controversial, "burgundian" way. Also, Douro 15 s are all rather open and drinkable at the moment. <a href="https://t.co/Dtrq4l7xtK">https://t.co/Dtrq4l7xtK</a>	First shout for my new #wine of the week @SymingtonFamily red Altano Organic Douro ? brimful of flavour and character. Just right for the winery weather, though just as brilliant with barbecues <a href="https://t.co/rIEFxxOFzA">https://t.co/rIEFxxOFzA</a> ?? #wine #organic #redwine #winelovers <a href="https://t.co/PmwV5KM6aT">https://t.co/PmwV5KM6aT</a>
2	Consumption	Quinta, region, vineyard, grape, product	142	It is definitely one place that I am never tired of visiting <sup>🇵🇹</sup> ? Quinta Napoles from @NiepoortWines-the place where some of the best #Douro dry wines are crafted, and what could be better than attending a tasting of port wines with Dirk Niepoort and Nick Delaforce ?? #wine <a href="https://t.co/6RiifuJQDZ">https://t.co/6RiifuJQDZ</a>	Port Wine cartoon art funny clipboard for oenophiles and Wine salesmen is FIFTEEN PERCENT OFF today with code ZCARDSGALORE <a href="https://t.co/6cnYT4NX3">https://t.co/6cnYT4NX3</a> #portwine #oenophile #winelover #winelovers #winesalesman #winesales #dourovalley #portlover
3	Tourism	Valley, travel, region, tour, visit	513	??♡A view from the Douro river and the Porto wine vines-#Portugal♡?? fx2A delightful day to you ALL fx2 fx3? <a href="https://t.co/wCrMuzGr2p">https://t.co/wCrMuzGr2p</a>	??♡A view from the Douro river and the Porto wine vines-#Portugal♡?? fx2 A delightful day to you ALL fx2fx3 ? <a href="https://t.co/wCrMuzGr2p">https://t.co/wCrMuzGr2p</a>

revealed the peak posting times, that is, hours when the users usually post more tweets, as well as changes in the habit of posting due to the impact produced by external events (e.g., COVID-19). The discovery of relevant words and semantic associations aimed at discovering keywords used by the users in their conversations (e.g., brands or places) and understating how the users perceived the main wine regions of Portugal (i.e., what other words used when they cite Douro, Porto, and others regions).

Finally, sentiment analysis was performed to identify the social sentiments of the community about each topic. The sentiment score ranged between  $-1.0$  and  $1.0$  for negative and positive tweets, respectively. To adjust this range to the processed corpora, a tweet was considered negative when its score was less than  $-0.7$  and positive when its score was greater than  $0.7$ . The remaining values considered the tweet as neutral.

All the previous analyses were performed using the tools provided by the Stanford CoreNLP pipeline [51] in combination with the Valence Aware Dictionary and sEntiment Reasoner (VADER) [52].

### 3. Results

*3.1. Topic Discovering.* Tables 1–3 describe the topics discovered by the LDA model for the “Douro,” “Porto,” and

“Others” corpora in terms of their five most representative terms. The tables also illustrate the number of optimal topics, the topic label, and a representative tweet sorted by the number of favorites or retweets. The presented tweets retained their original content, that is, hashtags, symbols, URLs, etc. Additionally, all identified terms were stemmed and every set of topics was manually labeled based on its most weighted terms. For example, in the case of the “Douro” corpus, the topic “Tourism” was so labeled because the terms that most contributed to the topic were “valley,” “travel,” “region,” “tour,” and “visit.”

In general, the distribution of the tweets per topic was not balanced; that is, there was always a more dominant topic per corpus, although the particularities of the LDA algorithm should prevent the outcomes from being affected [53]. Regarding the results obtained for the three corpora, a total of 5 different topics were produced: “Quinta do Crasto,” “Types of wine,” “Consumption,” “Tourism,” and “Gastronomy.” The content of the tweets related to the topic “Quinta do Crasto” was focused on advertising this famous Quinta situated in Sabrosa (Douro), while the tweets related to the topic “Types of wine” were focused on presenting different styles of Portuguese wines to the community, with special emphasis on red and white wines. On the other hand, the topic “Consumption” had tweets related to the habit of consuming an alcoholic product, usually accompanied by

TABLE 2: Topics discovered in Porto corpus.

Topic ID	Topic label	Topic terms	#Tweets	Fav. tweet	RT. tweet
0	Tourism	Wine, vintage, travel, food, quinta	1,530	Me in Porto, Portugal?? . . . now on my fav cites list! #Porto #Portugal #port #wine #travel #tourism #duororiver #actress #film #tv <a href="https://t.co/Ahp5txST01">https://t.co/Ahp5txST01</a>	? Porto Visits You! ? #Porto has a charming old town, excellent wine bars, cobbled streets, terracotta rooftops & dreamy architecture Read more ? <a href="https://t.co/JaVOWThhZT">https://t.co/JaVOWThhZT</a> ? #Travel #TravelTuesday #Photo #Photography #Portugal ? ?? <a href="https://t.co/COF03YhFGD">https://t.co/COF03YhFGD</a>
1	Gastronomy	Cheese, chocolate, mushroom, sauce, cream	1,258	Tonight I'm having port wine cheese spread on crackers, corn muffins with warm butter and fresh cherry pie with vanilla ice cream. dairy and carbs with my carbs and dairy	This very easy pork Wellington with Port wine sauce recipe is delicious and so flavorful! Perfect for entertaining, you can assemble it on the previous day, then you'll only need to bake it 25 minutes on d-day! <a href="https://t.co/BSANh65FZk">https://t.co/BSANh65FZk</a> #feedfeed #recipes #Cooking #SundaySupper <a href="https://t.co/hsNuYXIFFW">https://t.co/hsNuYXIFFW</a>
2	Consumption	Barrel, taylor, drink, glass, beer	3,113	I Have slept i have showered and shaved and lotioned and conditioned my hair twice and finally brushed it and then worked through a hair mask of oil and now I AM GOING TO SNUGGLE MY DOG, GET DRUNK ON PORT WINE, AND FINALLY WATCH THE LAST EPISODE OF MANDALORIAN!!!!	Feel good story: Porto's legendary Port wine producers used their stock of distilled alcohol to make 55,000 litres of disinfectant gel that is now being distributed among 15 Portuguese hospitals. <a href="https://t.co/H2JpCkqi7B">https://t.co/H2JpCkqi7B</a>

TABLE 3: Topics discovered in the Others corpus.

Topic ID	Topic label	Topic term	#Tweets	Fav. tweet	RT. tweet
0	Consumption	Madeira, drink, bottle, glass, time	200	Minho: can i get a glass of wine, please? Employee: sir, this is mcdonalds. Minho: oh fuck i am so sorry. . . can i get a mcwine?	Minho: can i get a glass of wine, please? Employee: sir, this is mcdonalds. Minho: oh fuck i am so sorry. . . can i get a mcwine?
1	Tourism	Alentejo, madeira, region, food, lisbon	527	Madeira is renowned for its landscapes, lore—and its legendary wine, which graced the tables of America's Founding Fathers <a href="https://t.co/cP0wiiZCnl">https://t.co/cP0wiiZCnl</a>	Madeira is renowned for its landscapes, lore—and its legendary wine, which graced the tables of America's founding fathers <a href="https://t.co/cP0wiiZCnl">https://t.co/cP0wiiZCnl</a>
2	Types of wine	Red, white, dry, sweet, fruit	164	Port, sherry, Madeira wines from Massandra are all symbols of #Crimean?? Viticulture ? <a href="https://t.co/730UFYL8T0">https://t.co/730UFYL8T0</a> No sanctions, no pandemic will stop Massandra to make a special gift to veterans-#Victory75 edition of Port red Crimean & South Coast Rose Muscat produced since 1944 <a href="https://t.co/3oJ7VDTSI2">https://t.co/3oJ7VDTSI2</a>	Port, sherry, Madeira wines from Massandra are all symbols of #Crimean?? Viticulture ? <a href="https://t.co/730UFYL8T0">https://t.co/730UFYL8T0</a> no sanctions, no pandemic will stop Massandra to make a special gift to veterans-#Victory75 edition of Port red Crimean & South Coast Rose Muscat produced since 1944 <a href="https://t.co/3oJ7VDTSI2">https://t.co/3oJ7VDTSI2</a>

some type of food. The topic “Tourism” was focused on extolling the virtues of the different Portuguese regions, and in many cases, tweets were related to people presenting their experiences in these regions. Finally, the tweets related to the topic “Gastronomy” were focused on showing pairings between different wines and foods, and the presentation of food recipes using several Portuguese wines.

According to the “Gastronomy” topic results, people usually preferred to use Porto wines to prepare food recipes.

The reasons may be that Port wines are usually fortified and sweet so they can add sweetness and balance to the dish, in addition to the fact that these wines also present a wide variety of flavors. On the other hand, the “Types of wine” topic showed that people usually talked about red fruity wines belonging to the Touriga grape variety in the “Douro” corpus, while in the “Others” corpus, people usually talked about dry or sweet wines, which are more common in regions like Madeira. Finally, there were two topics in

common among the overall corpora, that is, “Consumption” and “Tourism.” The “Consumption” topic in the “Douro” corpus was centered on consuming and producing wine products in the quintas, while in “Porto” and “Others” corpora, the content was more general and the places more diverse (e.g., events or parties). Lastly, the “Tourism” topic in “Douro” corpus was focused on visiting and making tours around the Douro Valley and the Douro River, while in the “Porto” corpus was more centered in the city of Oporto and its surroundings and in the “Others” corpus was especially focused on the island of Madeira and the region of Alentejo.

**3.2. User Characterization.** The following section is used to describe the analyses performed on the user accounts presented in the corpora and is structured as follows: (i) identification of both the most active and cited users inside and outside the topic; (ii) identification of the influential users presented in the corpora; and (iii) demographic analysis based on the countries with most presence, the sex of the users, and their type of account.

**3.3. Active Users in the Network.** Table 4 shows the most active users obtained from the general corpus of tweets related to wines. The column “#Tweets in topic” depicts the users sorted by the number of tweets posted in the topic of wines in descending order, whereas the column “#Tweets in Twitter” shows the users sorted by the number of tweets posted in the network.

Regarding the most active users in the topic, the user accounts of “RealWineGuru” and “alawine” were very similar in terms of number and type of publications. Both users looked like wine professionals and their content was centered in the publications of news related to the world of wines (e.g., reviews of different wines, news about the wine industry in several countries, or wine pairings). These accounts would be of interest to someone looking for the latest news related to the topic. In contrast, the user account of “BeerDescriber” was a Twitter bot whose function was to post tweets describing a diverse variety of alcoholic beverages, including wines. Although the number of followers and followings of this account was relatively low compared to the previously commented users (1 follower and 12 followings compared to 15k–35k and 53k–107k, respectively), its activity in the network is very elevated due to the automatic propagation of messages from time to time.

Similar to the previous data, Table 5 shows the most active users sorted by the number of tweets posted in the network (i.e., column “#Tweets in Twitter”), ensuring that they have published at least one tweet related to the topic of wines. The reason for exploring these users was to be able to identify some recognized users in the social network (e.g., famous people) who had also posted content related to the topic at some point. However, the users found were mainly associated with big commercial companies, organizations related to news and media (i.e., “OccurWorld” and “PulpNews”) or social media accounts from different websites (i.e., “urbandictionary”). The content of the tweets posted by these accounts was mainly related to the dissemination of

TABLE 4: Most active users related to the topic of wines and in the network. The users are sorted by the number of tweets posted in the topic.

Username	#Tweets in topic	#Tweets in Twitter
RealWineGuru	833	110,306
Alawine	721	75,393
BeerDescriber	715	25,724

TABLE 5: Most active users related to the topic of wines and in the network. The users are sorted by the number of tweets posted in the network.

Username	#Tweets in topic	#Tweets in Twitter
Urbandictionary	3	5,007,095
OccuWorld	4	3,460,459
PulpNews	12	3,400,490

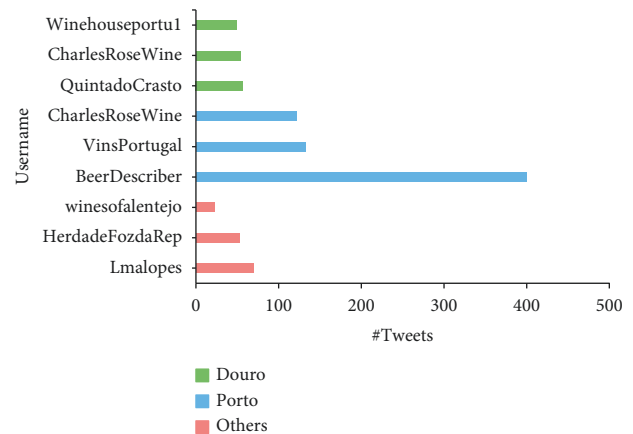


FIGURE 3: Most active users for each corpus.

news related to wine around the world, spamming offers or events, and users asking different types of questions about wine such as the meaning of the term or the available deliveries for online purchases.

From a more specific point of view, Figure 3 shows the most active users for each corpus (i.e., “Douro,” “Porto,” and “Others”) instead of the general corpus of tweets related to wines. The X-axis “#Tweets” depicts the number of tweets posted by each user in the corresponding corpus.

A review of the results indicates that the most active users presented in the three corpora were mainly accounts related to organizations (i.e., hotels/quintas and wine sellers) or individuals associated with some type of business (e.g., “Lmalopes” offers gastronomic experiences). The main content of the tweets posted by these users was focused on the direct (i.e., offering a product with a description and a direct link to buy it) or indirect (i.e., talking about the virtues of a product and including a media related to it) advertisement of some kind of product. This falls in line with previous studies pointing out the advantages of social networks for this type of marketing [7, 37], even more so with such a social topic and as linked to the tourism as wines.



TABLE 6: Most influential users based on  $F/f$  ratio,  $Fav/t$  ratio, and  $Rt/t$  ratio for each corpus.

Corpus	Username	$F/f$ ratio	Username	$Fav/t$ ratio	Username	$Rt/t$ ratio
Douro	JComm_NewsFeeds	1,013	Fernanda_FRocha	483	Fernanda_FRocha	509
	KeepOnTravelin	317	Winewankers	170	DemiCassiani	54
	AMAN_Alliance	188	DemiCassiani	101	Winewankers	52
Porto	mcbourbonnais1	10,693	dmarie116	1,553	Fernanda_FRocha	509
	WineSpectator	5,590	AnneScottlin	801	InsiderFood	188
	TravelMagazine	3,630	InsiderFood	640	Thinkerks	177
Others	NatGeo	297,703	PortugueseIn	9,014	PortugueseIn	7,046
	NatGeoTravel	20,943	MickFerry	1,351	Incorrectstrkid	291
	madeiraholiday7	7,018	Incorrectstrkid	1,132	Global_SHINee	289

TABLE 7: Distribution of users by sex and type for each corpus. The “number” column indicates the number of tweets posted.

Corpus	#Males	Number	#Females	Number	#Individuals	Number	#Organizations	Number
Douro	218	428	152	270	499	904	114	290
Porto	2,070	3,073	1,658	1,951	5,359	6,978	656	1,117
Others	189	407	278	261	624	851	101	229

**3.4. Influential People.** Table 6 shows the most influential users in each corpus based on the calculation of their ratios (i.e.,  $F/f$ ,  $Fav/t$ , and  $Rt/t$ , respectively). According to the results, there were 6 users among the three corpora that were present in both  $Fav/t$  and  $Rt/t$  ratios, that is, “Fernanda\_FRocha,” “winewankers,” “DemiCassiani,” “InsiderFood,” “PortugueseIn,” and “incorrectstrkid.” Although these accounts did not present high activity in their specific topic (the maximum number of tweets posted by these users was 5) and their  $F/f$  ratio is not as elevated as other users, they managed to get high values of engagement compared to the average value of the other accounts. For this reason, these users seem interesting enough to be monitored over time to see whether the behavior is recurring or, on the contrary, an isolated tweet that went viral. Some of these accounts, such as “DemiCassiani,” were users who were highly interested in wines, as stated in their descriptions. Regarding the content of their messages, they were mainly centered on commenting on the beauty of the Portuguese regions (e.g., the vineyards or landscapes), as well as making humorous remarks about wines.

On the other hand, the accounts that were only present in one of the columns related to the engagement ratios (e.g., “dmarie116” or “AnneScottlin”) presented similar characteristics to the users mentioned previously, that is, low activity in their corresponding topic and a  $F/f$  ratio below than the average. However, the content of their tweets was more diverse, being especially focused on banal conversations rather than on showing something to the readers. This may be a solid reason to explain why their tweets did not have as much impact on the community.

Finally, the user accounts with the highest values of  $F/f$  ratios were mainly associated with organizations and companies. These user accounts showed a slight increase in the number of posted tweets compared to previous commented users. Additionally, the content of their tweets was mainly focused on offering various events or products, mainly centered on tourism. This was the expected behavior

for these types of accounts that are mostly focused on organizations, given that people usually follow them on social networks to be in touch with possible offers or events that may be advertised.

**3.5. Demographic Analysis.** Table 7 depicts the distribution of users by sex and type of account for each corpus. The “number” column indicates the number of tweets posted by each sex or type of user. According to these results, there were no significant differences between males and females when talking about wine. It was a gender-parity topic (with a ratio of 1.18:1), even considering the average number of tweets per user ( $\sim 1.39$ ). The usual topics of conversation were also similar and were mainly related to user experiences, product advertising or recommendations, and gastronomy.

Most users were recognized as individuals; that is, there was an average of 1,186 individuals and 196 organizations per corpus with an average ratio of 6:1. Therefore, the organizations were more active in the network with an average of 1.8 tweets per account compared with the 1.34 tweets posted by individuals. Most of these tweets were related to the advertisement of different tours, products, and events offered by the different organizations. This behavior denoted the importance of social networks like Twitter to be a primary channel of communication and marketing [15, 30].

Figure 4 depicts a world map showing the distribution of the user accounts with a recognized location for the three corpora. The color represents the number of users located in each country (i.e., the redder the color, the higher the number of users). As can be seen, the distribution of users by country was uniform, with the top countries being the United States (1,717 users), Portugal (278 users), the United Kingdom (258 users), and Canada (99 users), respectively. These results could reflect a bias due to the limitation of obtaining tweets only in the English language.

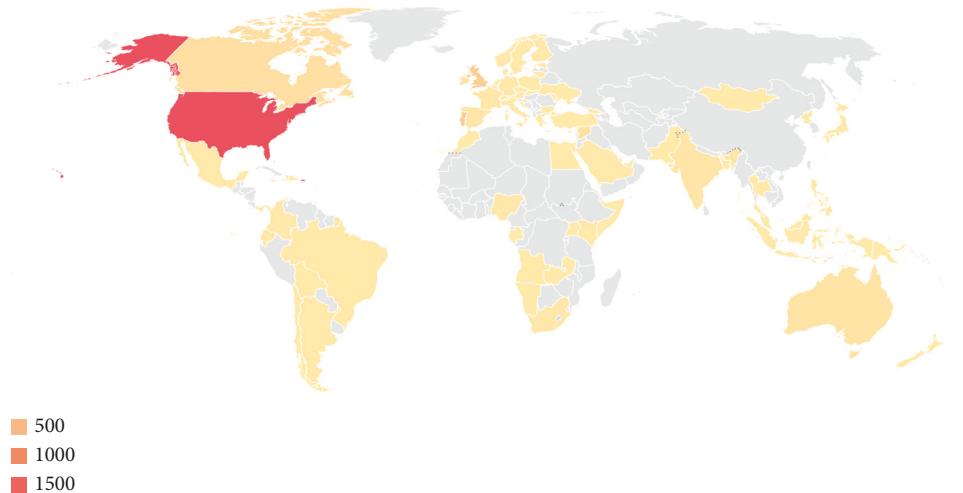


FIGURE 4: World map showing the distribution of users related to the topic of the wines on Twitter. The color represents the number of users located in each country (i.e., the redder the color, the higher the number of users).

**3.6. Content Analysis.** The following section is used to describe the analyses performed on the content of posted tweets and is structured as follows: (i) identification and analysis of most common hashtags, resources, and mentions; (ii) study of the tweets with larger engagement (i.e., favorites and retweets); (iii) analysis of posting distribution over time; (iv) discovery of relevant words and semantic associations; and (v) sentiment analysis.

**3.7. Hashtags, Mentions, and Resources.** Table 8 shows the top 5 common hashtags, mentions, and shared resources (i.e., URLs by domain) for each corpus. Several generic hashtags related to the domain under study (e.g., wine) or the corpus (e.g., douro) were not considered because they were present in most of the messages without providing any extra knowledge. The “number” column indicates the occurrences of each value in a specific corpus.

According to the results, there were no differences among the used hashtags in the corpora. Considering the content of the tweets, the hashtags were mainly used to focus the conversations on the country of Portugal or one of its wine regions, for example, Alentejo or Madeira. There was a significantly high number of appearances of the hashtag “#travel”; that is, it was in the top in the three corpora, pointing out the tourist attraction of the country for foreigners [54]. Similarly, the hashtag “#wineLover” was at the top for both “Porto” and “Others” corpora. According to the content of the tweets in which this hashtag was present, it was used to denote the proven quality of the wines inside the country as well as showing a strong tourist attraction [23]. There were also hashtags with commercial connotations like “#Spain” in the “Porto” corpus and “#quintadocrasto” in the “Douro” corpus. The former may have been used to capture the attention of people from other countries, and it was mainly used by the user account “Vins-Portugal,” a French wine boutique offering Portuguese wines through Twitter. The latter was mostly used by the user account “QuintadoCrado” to advertise its products or events or for people commenting on their experiences in this quinta.

When analyzing the content of the tweets to search for the common user mentions, “@QuintadoCrado” was found to be the most mentioned user account in the “Douro” corpus. The rest of the mentions were related to accounts and tweets talking about trips to the region of Douro, especially to see the valley, the river, or the vineyards. Conversely, the mentions found in the “Porto” corpus were related to more diverse topics, although most of them were centered on alcoholic products and food. For example, the mentions to the user accounts “@papaspilar” and “@ryeandrivet” talked about a rum that was aged in Port wine casks, whereas the mentions to the accounts “@HoarseWisperer” and “@laurenthehough” talked about consuming a Port wine cheese. Lastly, the mentions encountered in “Others” corpus were mainly centered in the gastronomy of the country (e.g., “@TurismodeLisboa”), wines (e.g., “@wines\_Portugal”), and different events (e.g., “@ivdp\_ip”).

Regarding the most shared resources, the social network Instagram was the undisputed leader compared to the other resources, with a total of 572 appearances (i.e., 8% of the total) among all corpora. This was in concordance with the conclusion of multiple studies pointing out the social behavior of wine [1]. Another commonly shared resource was different webpages whose purpose was to sell Portuguese wines (especially from the region of Douro and Porto), such as “vinsportugal” who appeared a total of 133 times (i.e., 3% of the total) in the Porto corpus and “Winehouseportugal” who had a total of 49 occurrences (i.e., 4% of the total) in the Douro corpus. With these results, it is possible to say that there may be an interest in the acquisition of Portuguese products by foreigners. Lastly, people also shared different blogs and news related to wine, food recipes, or tourism (e.g., “cookinglisbon,” “cookandcorks,” or “visitevora”).

**3.8. Content Engagement.** Table 9 shows the most retweeted and favorite tweets in each corpus. The most retweeted and favorite tweet in the “Douro” corpus, and most retweeted in the “Porto” corpus, was originally posted by the account of

TABLE 8: Top 5 hashtags, mentions, and shared resources for every generated corpus.

Corpus	Hashtag	Number	Mentions	Number	URL domain	Number
Douro	Portugal	320	Quintadocrasto	23	Instagram	118
	Dourovalley	98	Visitportugal	12	Charlesrosewines	54
	Port	88	Roquettecazes	11	Winehouseportugal	49
	Quintadocrasto	68	SSDouroValley	11	Winemag	25
	Travel	67	Winewankers	10	Cookinglisbon	21
Porto	Portugal	418	Papaspilar	63	Instagram	392
	Winelover	192	Ryeandrivet	63	Vinsportugal	133
	Travel	187	TalkingPicsTV	44	Charlesrosewines	122
	Visitportugal	156	HoarseWisperer	29	Ryeandrivet	63
	Spain	127	Laurentheough	28	Schnitzdeli	60
Others	Portugal	243	wines_Portugal	9	Instagram	62
	Alentejo	133	ivdp_ip	7	Cookinglisbon	52
	Madeira	52	Portugal	7	Cookandcorks	33
	Winelover	46	WPTUK	7	HFRresort	33
	Travel	41	Turismodelisboa	6	Visitevora	21

TABLE 9: Most retweeted tweets by keyword and corpus.

Corpus	Tweet	#Retweets	Tweet	#Favorites
Douro	?? ♡ A view from the Douro river and the Porto wine vines-#Portugal ♡?? fx2 a delightful day to you ALL fx2 fx3? <a href="https://t.co/wCrMuzGr2p">https://t.co/wCrMuzGr2p</a>	509	?? ♡ A view from the Douro river and the Porto wine vines-#Portugal ♡?? fx2 a delightful day to you ALL fx2 fx3? <a href="https://t.co/wCrMuzGr2p">https://t.co/wCrMuzGr2p</a>	483
Porto	?? ♡ A view from the Douro river and the Porto wine vines-#Portugal ♡?? fx2 a delightful day to you ALL fx2 fx3? <a href="https://t.co/wCrMuzGr2p">https://t.co/wCrMuzGr2p</a>	509	We visited Quinta de Vargellas to see how traditional port wine is made ? <a href="https://t.co/fSVxJQpRem">https://t.co/fSVxJQpRem</a>	2,794
Others	Madeira is renowned for its landscapes, lore—and its legendary wine, which graced the tables of America’s founding fathers <a href="https://t.co/OvISK9dujA">https://t.co/OvISK9dujA</a>	165	Madeira is renowned for its landscapes, lore—and its legendary wine, which graced the tables of America’s Founding Fathers <a href="https://t.co/cP0wiiZCnl">https://t.co/cP0wiiZCnl</a>	826

“Fernanda\_FRocha.” This tweet also contained a picture of the river along with the text. The  $F/f$  ratio of the user had a value of 2.5, whereas her general  $Fav/t$  ratio had a value of 0.72, meaning that it was not common that a tweet posted by this user would reach this engagement. Although the most probable reason for achieving such a strong engagement could be due to the content of the tweet instead of the user, that is, the community likes content that promotes the beauty of the country, it would be useful to keep monitoring this account to determine the exact reason for this high-level response.

The most favorite tweet for the “Porto” corpus was originally posted by the user “InsiderFood,” an account dedicated to posting news about worldwide foods. In this case, the tweet contained a short documentary showing the process of producing traditional Port wine. This user account posted several tweets in the topic under study (i.e., 16 original tweets) and had a  $F/f$  ratio of 3,010, a  $Fav/t$  ratio in the topic of 319 and a  $Rt/t$  ratio in the topic of 100, meaning that the account could be very valuable due to the high power of diffusion and reception of information inside the social network. This was in concordance with the high number of favorites, reaching a value of 2,794, much higher than the other tweets. It is worth noting that both this tweet and the previously commented one presented media (i.e., a picture or a video) that was related to the topic discussed in their messages, which can help achieve a high level of engagement [55].

Finally, the most retweeted and favorited tweet in the “Others” corpus was originally posted by the user “NatGeo,” the official account of National Geographic. The content of the tweet was also accompanied by an external link to an article posted on the National Geographic website. A total of 5 different tweets posted by this user and related to the topic under study were found. Its  $F/f$  ratio had a value of 297,703, whereas the  $Fav/t$  ratio and the  $Rt/t$  ratio in the topic had values of 917 and 238, respectively. As expected, due to the entity of the user account, most of the tweets posted by this user usually reach high levels of engagement.

**3.9. Posting Time Distribution.** The time chart in Figure 5 illustrates how the number of tweets evolved over time by month, hour, and day of the week. In particular, Figure 5(a) depicts tweets related to the “Douro” corpus, Figure 5(b) shows the number of tweets related to the “Porto” corpus, and Figure 5(c) illustrates tweets related to the “Others” corpus. The plots related to the distribution by month have a dark green dashed line showing the estimated number of tweets for all of May (i.e., the current number of tweets from 1 May to 8 May is multiplied by the number of weeks until the end of the month).

The evolution of tweets by months had a similar distribution for “Douro” and “Others” corpus with a peak in January–February (i.e., 235 and 272 tweets, respectively) and a continuous decrease until May. This behavior may be

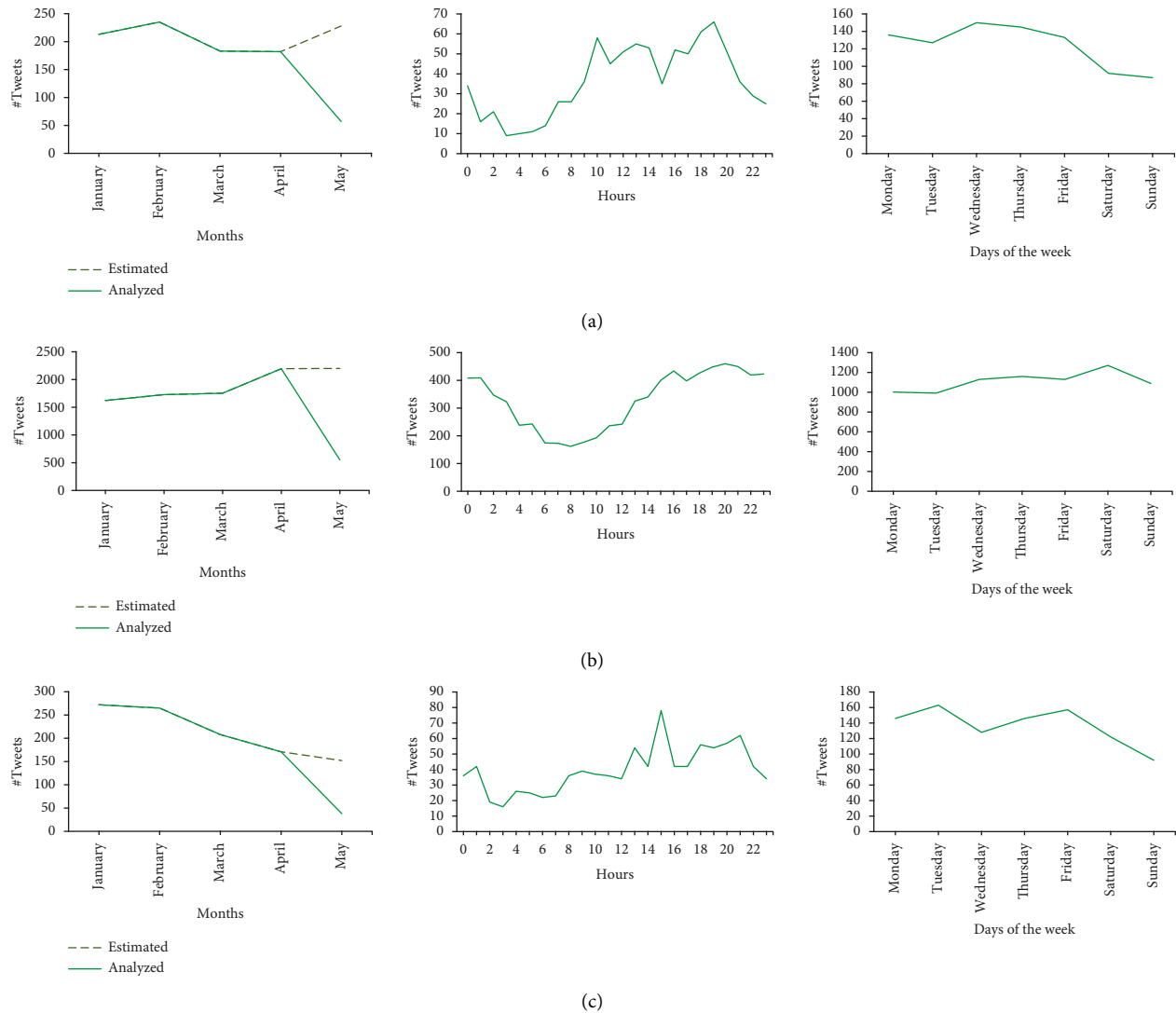


FIGURE 5: Time charts showing the number of tweets posted by the corpus in the year 2020. (a) Tweets related to the Douro corpus. (b) Tweets related to the Porto corpus. (c) Tweets related to the Others corpus.

somewhat expected due to the start of the COVID-19 pandemic in mid-March, after which activities related to tourism and events (e.g., travel, celebration of events, or staying at quintas) plummeted, while tweets related to the imposed quarantine greatly increased [56, 57]. Interestingly, this behavior did not seem to have occurred for the “Porto” corpus, which had its peak of activity in April (i.e., 2,195 tweets), although a possible explanation is the fact that the tweets related to the “Porto” corpus were more varied in nature, with many tweets from individuals talking about their experiences with Porto wines, in addition to those from organizations in the other two corpora. It is worth noting that during the quarantine imposed by the COVID-19, there was no noticeable increase in the tweeting activity related to online purchases of wine, which may imply various conclusions including the possibility that the Portuguese wine producers do not take full advantage of this kind of market as did other countries [58, 59].

The distribution of tweets by hours followed a similar pattern for the three corpora. On the one hand, the number of

tweets had a decrease in activity during the early hours (i.e., from 1 h to 8 h) while experiencing an increase in activity during the mornings and especially the afternoons (e.g., 15 h and 19 h were the peak activity hours). This behavior was expected as it matched the most common working schedules of European organizations. Similarly, the distribution by days of the week had its peak of activity during the weekdays with a decrease in the number of tweets during the weekends for the “Douro” and “Others” corpora, while the “Porto” corpus experienced an increase on Saturdays. Again, this may be explained because of the higher activity of individual accounts in this corpus compared to the other two. People tend to have more spare time on weekends, which can be translated into more time spent in social networks.

**3.10. Relevant Words and Semantic Associations.** Table 10 depicts the most occurring unigrams and bigrams in the tweets of each studied corpus (i.e., words that appeared next

TABLE 10: Most dominant n-grams in the conversations of each corpus as well as the most common semantic relationship with the main corpus keyword (e.g., in Douro corpus, the most common relations with the term Douro).

Corpus	Dominant unigrams	Dominant bigrams	Relations
Douro	Valley, quinta, river, region, vineyard	Douro valley, quinta crasto, red wine, award win, buy online	Douro valley, douro superior, douro river, cruise douro, red douro
Porto	Taylor, drink, red, like, cheese	Taylor port, red wine, wine cheese, tawny port, sherry cask	Taylor port, tawny port, vintage port, drink port, double port, love port
Others	Madeira, Alentejo, green, red, food	Green wine, madeira wine, Alentejo wine, red wine, white wine	Alentejo wine, Algarve beach, Algarve wine, island madeira, madeira dramatic

to each other in the tweet). The column “Relations” is similar to the previous ones although, in this case, it shows the semantic relations for the desired term that is manually selected for each corpus (i.e., it shows the words that appeared next to a selected term). Therefore, for the “Douro” corpus, this column illustrates the relationship with the term “Douro”; for the “Porto” corpus, it depicts the relationship for the term “Port”; and for the “Others” corpus, it shows the relationship for its most common regions (i.e., the terms Madeira, Alentejo, and Algarve).

An analysis of the table and matching the results to those presented in the Topic Discovering section, the most dominant n-grams in the “Douro” corpus were centered on the topic of “quintas” (i.e., primarily rural properties usually used as an appellation for agricultural estates, such as wineries) and tourism. Moreover, considering the semantic relationship, the tourist attraction of the Douro region became even more evident. People usually emphasize the Douro Valley and the Douro River as well as the activities derived from them, such as the catamaran used for cruising the river. The importance of this Portuguese area is so universally acknowledged that the UNESCO recognized it as world heritage site in the year 2001 [20].

Regarding the results of the “Porto” corpus, the dominant n-grams were mainly centered on the consumption of several types of Port wines (e.g., Taylor or Tawny). Similarly, the relationship with the word “Port” also emphasized these kinds of conversations, including even more varieties (e.g., vintage port or double port). Of all the types commented, the one that stood out above the rest was “Taylor port” with 969 appearances among all tweets, indicating that it is an internationally known type of wine.

Finally, regarding the results of “Others” corpus, the most dominant n-grams were focused in the different Portuguese regions (e.g., the island of Madeira or Alentejo) and their exclusive wines (e.g., green wine). Regarding the semantic relations with these regions, people tended to comment about the wines in Alentejo and the Algarve as well as the beaches situated in the region of Algarve, which attracts among the highest number of summer tourists in the country.

**3.11. Sentiment Analysis.** Figure 6 depicts the social sentiment in the corpora. As other studies have pointed out, the predominant sentiment when talking about wines was mainly neutral-positive (with an average score of  $\sim 0.3$  among all corpus). This was something expected due to the social and enjoyable nature of the product [2, 60].

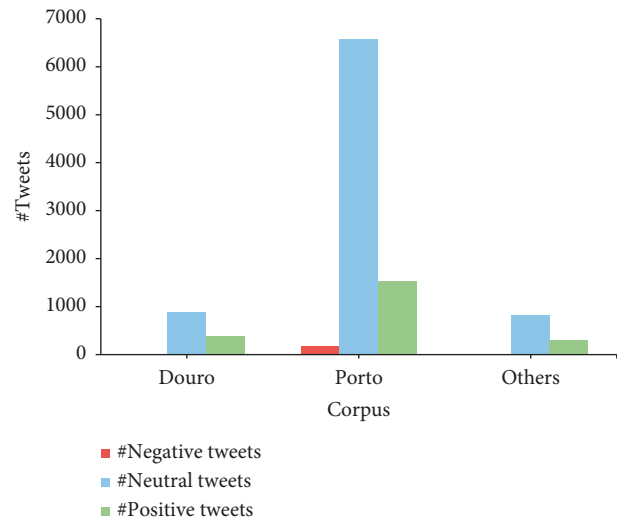


FIGURE 6: Sentiment analysis by corpus.

The negative tweets were primarily related to the sharing of bad news (e.g., “. . .Harsh Heat and Drought Breaks an Over 150 Year Tradition for Port Producers. . .”), to be part of a discussion among users (e.g., “. . .DO NOT drink the port in a jug. . .that stuff is seriously wrong. . .”) or recent emergencies like COVID-19 (e.g., “. . .The Coronavirus emergency is having a devastating economic impact on the Mid Coast’s wine industry. . .”). However, in the “Porto” corpus, there was a significant number of tweets related to specific complaints about Taylor Port wine. It did not appear to be appreciated by foreigners (e.g., “That TAYLOR PORT wine is so damn NASTY !!,” “Taylor port is not wine it’s colored death” or “Taylor port wine is disgusting”).

## 4. Discussion

This work presents a preliminary study of the knowledge of social networks to explore the Portuguese wine industry and its tourism potential. Our study focused on the analysis of data extracted from Twitter. This digital media platform has been little studied in Portugal, namely, with regard to digital marketing and the interaction between brands and tourists interested in the Douro Valley region. In the literature, we found several studies that evaluated and monitored the publication of content related to the Douro Valley on social media [9, 19, 40, 61]. However, none of the studies included a comprehensive reading of what is being said on social media about the Douro Valley and Douro and Port wines.

Existing studies use specific case studies, analyzing the presence of a specific brand, producer, or entity on media platforms. Vieira et al. [40] analyze the impact of a hotel group's campaigns on Instagram, analyzing the engagement generated by the publications and the importance of sharing content by digital influencers. Almeida et al. [61] focus their study on the importance of using hashtags to promote brands in the hotel sector. In a different perspective, Jorge et al. [9] explore the presence of several international destinations, including the Douro valley, through the pages or accounts of tourist entities on Facebook, Instagram, YouTube, and Twitter, measuring the engagement achieved per publication. Additionally, Correia et al. [19] use a quinta in the Douro Valley as a case study, applying content analysis on Facebook and TripAdvisor, as well as interviews and surveys. This study makes a qualitative analysis of the main motivations for visiting this farm, as well as surveying various aspects associated with location, landscape, and wine consumption.

Given this bibliographic scenario, we believe that our study may fill a gap in the literature on the importance of social media, namely, Twitter, in relation to the interaction between consumers or tourists and companies or brands in the Douro Valley region.

The analyzed data are based on a four-month dataset comprising 1,200,428 unique tweets from 764,108 distinct users. The methodology developed to retrieve, process, and analyze this data was mainly focused on the use of algorithms and techniques related to machine learning and the natural language processing family. This allowed for an analysis of the content posted about the topic by the different users, as well as a study of user accounts within the wine community, making it possible to respond to several questions such as what the main topics of conversations are among people, how users express themselves within the topic, or which accounts participating in these communities are the most interesting.

The identified topics of the conversation clearly identified the close relationship that exists between the universe of Portuguese wines (e.g., consumption or types) and tourism and how both mutually feed back into each other. Gastronomy was also a topic that appeared in many conversations within the user community, especially the sharing of recipes in which one of the ingredients was a Portuguese wine. The presence of the "Quinta de crasto" as a topic of conversation was also notable. This indicates how well this organization is using the social networks to apply marketing strategies on many users.

This study also brought to light a set of user accounts that could be useful for future analysis. Although these users had poor activity in terms of tweets related to the topic, some of their tweets managed to go viral. Additionally, while there were fewer organizations than individuals, they were found to be more active and generally achieved greater impact with their messages. This finding demonstrates the importance of planning a communication strategy together with influencers or users with a large following on social media. These can be decisive not only in the process of promoting the tourist destination, but also during consumer decision-

making, conveying credible and realistic experiences and messages.

The topic of wines was also found to be a gender-parity topic with no significant differences between the number of males and females. This implies that, for example, when launching marketing campaigns both genders should be taken into account to reach a larger audience. With a geographical distribution concentrated in English-speaking countries, and the United States in particular, a market analysis in these countries would be useful for future advertising campaigns.

Regarding the content of the messages, the conversations related to tourism were mainly focused on different areas of the Douro (e.g., the Valley or the River) as well as specific tours and events in the area. This indicates a strong tourist attraction of the region throughout the world and effective advertisement outside the country. When talking about the different types of wines, the Taylor port stood out from the rest. However, taking into account the sentiment analysis of the tweets that cited this type of wine, it is possible to observe that people did not speak particularly well of it. An attempt should be made to solve this issue by analyzing the root of the problem and, for example, launching an advertising campaign to address the negative points mentioned by the community.

Finally, the distribution of the tweets over time showed a general decrease in activity from March to May. This period coincides with the strong spread of the COVID-19 disease throughout several European countries, which led to a reduction in tourism and events. However, there was no noticeable increase in tweets related to online shopping. This could be an interesting point for the wine brands to address, especially because of a possible long-term resurgence of the disease.

## 5. Conclusions

In general, our study allowed us to understand the dimension of the universe associated with Port wine and wine tourism in the Douro Valley on Twitter. The main topics of discussion were raised within the three corpora analyzed: "Douro," "Porto," and "Others." An analysis of these topics made it possible to verify the types of wine and the most important characteristics of a region, while at the same time allowing us to follow consumer feedback in relation to their tourist experience or about a certain product or service.

We also concluded that large commercial companies related to the wine and hotel sector are the most active users on this social media platform. This demonstrates the importance that industry organizations place on Twitter to promote the territory and their brands. In addition, we found that there are independent users who can be influential and thus play a very important role in the development of digital marketing strategies.

Our findings also found very positive evidence regarding consumers' perception of the Douro Valley and its wines, as the sentiment analysis showed user comments and experiences with neutral-positive assessments more frequently than publications with content with more negative connotation.

## 6. Practical Implications

The different analyses performed on the tweets and user accounts allowed the creation of a first straightforward methodology, which permits a continuous evaluation of the current impact that exists in social networks for the topic under study. The proposed approach and the application of different analyses (some of them briefly commented in this document) may pave the way to the creation of an information system centered in the real-time monitoring, analysis, and redirection of the information in social media.

Therefore, this hypothetical system would be able to perform more complex analyses in addition to long time and continuous monitoring of a specific topic. With this kind of information, it would be possible to build a more complex methodology used to infer new interesting and hidden knowledge from social networks. This would lead to the publication of novel studies including, for example, (i) providing useful first-hand information to propose interested stakeholders with the possibility of rebranding their products; (ii) gaining a better understanding of the worldwide impact of relevant concepts; (iii) obtaining knowledge about the strengths and weaknesses of a given product; (iv) generating marketing activity in the network with the ability to measure its acceptance among users; or (v) understanding how people feel and react to a specific campaign.

Finally, this system could be used by any interested stakeholder to gain insight into any other business-related topic and, most notably, it can be used to improve the capacity of stakeholders to promote more focused marketing campaigns and to communicate with the target consumers in a more efficient way, two outcomes derived from the field of business informatics.

## Data Availability

All data are available in the article.

## Additional Points

The main limitations of this work are the following: (i) the filtered data based on specific Portuguese keywords (e.g., Douro) may not fully cover the totality of the tweets related to that topic; (ii) the language of the tweets are restricted to English, so it is assumed that information is being lost, especially when analyzing information related to tourism; and (iii) other social networks should be taken into account since Twitter does maybe not accurately or entirely reflect the attitudes of the users about the topic.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this article.

## Acknowledgments

This work was supported by the Associate Laboratory for Green Chemistry—LAQV, financed by the Portuguese

Foundation for Science and Technology (FCT/MCTES) Ref. UID/QUI/50006/2020; the Portuguese Foundation for Science and Technology (FCT/MCTES) under the scope of the strategic funding of UIDB/04469/2020 unit and Bio-TecNorte operation funded by the European Regional Development Fund (ERDF) under the scope of Norte2020—Programa Operacional Regional do Norte. Ref. NORTE-01-0145-FEDER-000004; the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding of ED431C2018/55-GRC Competitive Reference Group, the “Centro singular de investigación de Galicia” (accreditation 2019-2022) funded by the European Regional Development Fund (ERDF)-Ref. ED431G2019/06; and Portuguese Foundation for Science and Technology for a PhD Grant (SFRH/BD/145497/2019).

## References

- [1] M. Trevisan, “Wine and society-welcome editorial,” *Wine Studies*, vol. 1, no. 1, 2011.
- [2] S. Higgs and J. Thomas, “Social influences on eating,” *Current Opinion in Behavioral Sciences*, vol. 9, pp. 1–6, 2016.
- [3] A. Zarei, H. Farjoo, and H. B. Garabollagh, “How social media marketing activities (SMMAs) and brand equity affect the customer’s response: does overall flow moderate it?” *Journal of Internet Commerce*, vol. 21, no. 2, pp. 160–182, 2022.
- [4] G. Tesfom and C. H. M. Lutz, “Consumers’ advertising media use: a cross-cultural study,” *Services Marketing Quarterly*, vol. 43, no. 1, pp. 32–47, 2022.
- [5] I. Khan, “Do brands’ social media marketing activities matter? A moderation analysis,” *Journal of Retailing and Consumer Services*, vol. 64, Article ID 102794, 2022.
- [6] C. McClure and Y.-K. Seock, “The role of involvement: investigating the effect of brand’s social media pages on consumer purchase intention,” *Journal of Retailing and Consumer Services*, vol. 53, Article ID 101975, 2020.
- [7] T. Pucci, E. Casprini, C. Nosi, and L. Zanni, “Does social media usage affect online purchasing intention for wine? The moderating role of subjective and objective knowledge,” *British Food Journal*, vol. 121, no. 2, pp. 275–288, 2019.
- [8] M. Fiore, D. Vrontis, R. Silvestri, and F. Conto, “Social media and societal marketing: a path for a better wine?” *Journal of Promotion Management*, vol. 22, no. 2, pp. 268–279, 2016.
- [9] F. Jorge, M. Sérgio Teixeira, C. Fonseca, R. J. Correia, and R. Gonçalves, “Social media usage among wine tourism DMOs,” in *Marketing and Smart Technologies. Smart Innovation, Systems and Technologies*, Á. Rocha, J. Reis, M. Peter, and Z. Bogdanović, Eds., Springer, New York, NY, USA, pp. 78–87, 2020.
- [10] S. Moon and D. Iacobucci, “Social media analytics and its applications in marketing,” *Foundations and Trends® in Marketing*, vol. 15, no. 4, pp. 213–292, 2022.
- [11] G. Szolnoki, D. Tait, M. Nagel, and A. Fortunato, “Using social media in the wine business: an exploratory study from Germany,” *International Journal of Wine Business Research*, vol. 26, no. 2, pp. 80–96, 2014.
- [12] J. Jin, T. Zhang, and X. Yan, “Why Do Users Continually Seek Knowledge in Online Q&A Communities? an Empirical Investigation,” *Information Discovery and Delivery*, 2022.
- [13] B. Marlowe, E. A. Brown, T. Schrier, and T. Zheng, “Beverage bloggers: a developing relationship between wine blogger

- expertise and twitter followers,” *International Journal of Hospitality Beverage Management*, vol. 1, no. 1, p. 2, 2017.
- [14] B. Urrutikoetxea Arrieta, A. I. Polo Peña, and C. Martínez Medina, “The moderating effect of blogger social influence and the reader’s experience on loyalty toward the blogger,” *Online Information Review*, vol. 43, no. 3, pp. 326–349, 2019.
- [15] D. Wilson and S. Quinton, “Let’s talk about wine: does Twitter have value?” *International Journal of Wine Business Research*, vol. 24, no. 4, pp. 271–286, 2012.
- [16] W. Portugal, “Exports-Wines of Portugal,” 2019, <http://www.winesofportugal.com/en/press-room/statistics/export/>.
- [17] Statista, “Portugal: Wine Production 2006-2018 | Statista,” 2019, <https://www.statista.com/statistics/420509/volume-of-wine-produced-in-portugal/>.
- [18] J. Caldas and J. Rebelo, “Portuguese wine ratings: an old product a new assessment,” *Wine Economics and Policy*, vol. 2, no. 2, pp. 102–110, 2013.
- [19] A. I. Correia, R. Cunha, O. Matos, and C. Fernandes, “Wine tourism experiences and marketing: the case of the Douro valley in Portugal,” in *Wine Tourism Destination Management and Marketing*, M. Sigala and R. Robinson, Eds., Palgrave Macmillan, London, UK, pp. 203–220, 2019.
- [20] A. Guedes and J. Rebelo, “Merging wine and tourism-related services: evidence from the Douro (Portugal) Wine Region,” *Journal of Wine Research*, vol. 30, no. 4, pp. 259–274, 2019.
- [21] J. Martins, R. Gonçalves, F. Branco, L. Barbosa, M. Melo, and M. Bessa, “A multisensory virtual experience model for thematic tourism: a Port wine tourism application proposal,” *Journal of Destination Marketing & Management*, vol. 6, no. 2, pp. 103–109, 2017.
- [22] F. A. D. N. Santos, N. Vavdinis, L. F. Martinez, U. D. L. Portugal, and U. N. D. L. Portugal, “Avances y perspectivas para la investigación del turismo del vino en Portugal,” *Pasos: Revista De Turismo Y Patrimonio Cultural*, vol. 18, no. 1, pp. 159–170, 2020.
- [23] A. Lavrador da Silva, M. João Fernão-Pires, and F. Bianchi-de-Aguiar, “Portuguese vines and wines: heritage, quality symbol, tourism asset,” *Ciencia e tecnica vitivinicola*, vol. 33, no. 1, pp. 31–46, 2018.
- [24] J. Rebelo, A. Guedes, L. Lourenço-Gomes, and M. T. Sequeira, “Balanço de concretização do programa de ação,” *Avaliação do Estado de Conservação do Bem Alto Douro Vinhateiro-Paisagem Cultural Evolutiva Viva*, vol. 2, pp. 3–74, 2013.
- [25] J. Rebelo, J. Caldas, and A. Guedes, “The Douro region: wine and tourism,” *Almatourism-Journal of Tourism, Culture and Territorial Development*, vol. 6, no. 11, pp. 75–90, 2015.
- [26] S. Lins and C. Campos, “Wine, sweet wine”: significados e determinantes da compra por impulso de vinho do Porto,” *Innovar*, vol. 32, no. 83, pp. 35–50, 2021.
- [27] Statista, “Number of Social media Users Worldwide 2010-2021,” 2019b, <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>.
- [28] A. Galati, M. Crescimanno, S. Tinervia, and F. Fagnani, “Social media as a strategic marketing tool in the Sicilian wine industry: evidence from Facebook,” *Wine Economics and Policy*, vol. 6, no. 1, pp. 40–47, 2017.
- [29] D. Creevey, J. Coughlan, and C. O’Connor, “Social media and luxury: a systematic literature review,” *International Journal of Management Reviews*, vol. 24, no. 1, pp. 99–129, 2022.
- [30] L. Thach, T. Lease, and M. Barton, “Exploring the impact of social media practices on wine sales in US wineries,” *Journal of Direct, Data and Digital Marketing Practice*, vol. 17, no. 4, pp. 272–283, 2016.
- [31] D. Belias, E. Velissariou, D. Kyriakou et al., N. Kakkos, “The importance of customer relationship management and social media in the Greek wine tourism industry,” in *Innovative Approaches to Tourism and Leisure*, V. Katsoni and K. Velander, Eds., Springer, Berlin, Germany, pp. 249–259, 2018.
- [32] Z. R. Santos, C. M. K. Cheung, P. S. Coelho, and P. Rita, “Consumer engagement in social media brand communities: a literature review,” *International Journal of Information Management*, vol. 63, Article ID 102457, 2022.
- [33] U. Ramanathan, N. Subramanian, and G. Parrott, “Role of social media in retail network operations and marketing to enhance customer satisfaction,” *International Journal of Operations & Production Management*, vol. 37, no. 1, pp. 105–123, 2017.
- [34] A. Lopez, E. Guerra, B. Gonzalez, and S. Madero, “Consumer sentiments toward brands: the interaction effect between brand personality and sentiments on electronic word of mouth,” *Journal of Marketing Analytics*, vol. 8, no. 4, pp. 203–223, 2020.
- [35] M. Parent, K. Plangger, and A. Bal, “The new WTP: willingness to participate,” *Business Horizons*, vol. 54, no. 3, pp. 219–229, 2011.
- [36] M. J. Culnan, P. J. Mchugh, J. I. Zubillaga, M. Q. Uarterly, and E. Xecutive, “How large U.S. Companies can use twitter and other social media to gain business value,” *MIS Quarterly Executive*, vol. 9, pp. 243–259, 2010.
- [37] A. Pourkhani, K. Abdipour, B. Baher, and M. Moslehpour, “The impact of social media in business growth and performance: a scientometrics analysis,” *International Journal of Data and Network Science*, vol. 3, no. 3, pp. 223–244, 2019.
- [38] M. Zhang, B. J. Jansen, and A. Chowdhury, “Business engagement on Twitter: a path analysis,” *Electronic Markets*, vol. 21, no. 3, pp. 161–175, 2011.
- [39] Statista, “Twitter: Number of Active Users 2010-2019,” 2019c, <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- [40] J. M. C. Vieira, M. Magalhães, and E. Serra, “Why aren’t younger consumers attracted to Port Wine? A Sensory gap analysis,” in *Proceedings of the british academy of management conference bam 2021*, London UK, April 2021.
- [41] Y. Yamamoto, “Twitter4J-A Java Library for the Twitter API,” 2018, <http://twitter4j.org/en/index.html>.
- [42] D. Newman, A. Asuncion, P. Smyth, and M. Welling, “Distributed algorithms for topic models,” *Journal of Machine Learning Research*, vol. 10, pp. 1801–1828, 2009.
- [43] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [44] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, “Evaluation methods for topic models,” in *Proceedings of the 26th annual international conference on machine learning-icml’09*, pp. 1–8, ACM Press, New York, NY, USA, June 2009.
- [45] F. Martin and M. Johnson, “More efficient topic modelling through a noun only approach,” in *Proceedings of the australasian language technology association workshop*, pp. 111–115, Parramatta, Australia, December 2015.
- [46] I. Anger and C. Kittl, “Measuring influence on twitter,” in *Proceedings of the 11th international conference on knowledge management and knowledge technologies*, pp. 1–4, New York, NY, USA, September 2011.
- [47] B. Tzolmon and K. S. Lee, “A graph-based reliable user classification,” in *Lecture Notes in Electrical Engineering*, vol. 285, pp. 61–68, Springer Verlag, 2014.



- [48] GeoNames Database, "GeoNames Database," 2018, <http://www.geonames.org/>.
- [49] M. Pérez-Pérez, G. Pérez-Rodríguez, F. Fdez-Riverola, and A. Lourenço, "Using twitter to understand the human bowel disease community: exploratory analysis of key topics," *Journal of Medical Internet Research*, vol. 21, no. 8, Article ID e12610, 2019.
- [50] G. Pérez-Rodríguez, M. Pérez-Pérez, F. Fdez-Riverola, and A. Lourenço, "Mining the sociome for Health Informatics: analysis of therapeutic lifestyle adherence of diabetic patients in Twitter," *Future Generation Computer Systems*, vol. 110, pp. 214–232, 2020.
- [51] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. Mcclosky, "The Stanford CoreNLP natural language processing toolkit," in *Proceedings of the 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60, Association for Computational Linguistics (ACL), Baltimore, Maryland, January 2014.
- [52] C. J. Hutto and E. Gilbert, "VADER: a parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the eighth international aaai conference on weblogs and social media*, pp. 216–225, Ann Arbor, MI, USA, June 2014.
- [53] J.-H. Xue and D. M. Titterton, "Do unbalanced data have a negative effect on LDA?" *Pattern Recognition*, vol. 41, no. 5, pp. 1558–1571, 2008.
- [54] A. P. Lopes and F. Soares, "Tourism planning and development: the case of Portugal's Norte region," *Tourism & Management Studies*, vol. 13, no. 4, pp. 20–29, 2017.
- [55] Y. Li and Y. Xie, "Is a picture worth a thousand words? An empirical study of image content and social media engagement," *Journal of Marketing Research*, vol. 57, no. 1, pp. 1–19, 2020.
- [56] N. A. Bakar and S. Rosbi, "Effect of Coronavirus disease (COVID-19) to tourism industry," *International Journal of Advanced Engineering Research and Science*, vol. 7, no. 4, pp. 189–193, 2020.
- [57] S. Gössling, D. Scott, and C. M. Hall, "Pandemics, tourism and global change: a rapid assessment of COVID-19," *Journal of Sustainable Tourism*, vol. 29, pp. 1–20, 2020.
- [58] S. Colbert, C. Wilkinson, L. Thornton, and R. Richmond, "COVID -19 and alcohol in Australia: industry changes and public health impacts," *Drug Alcohol Rev*, Blackwell Publishing, vol. 39, no. 5, pp. 435–440, 2020.
- [59] T. Watanabe and Y. Omori, "Online Consumption during the COVID-19 Crisis: Evidence from Japan. CARF F-Series CARF-F-487," *Center for Advanced Research in Finance, Faculty of Economics*, The University of Tokyo, Tokyo, Japan, 2020.
- [60] S. Charters and S. Pettigrew, "Is wine consumption an aesthetic experience?" *Journal of Wine Research*, vol. 16, no. 2, pp. 121–136, 2005.
- [61] S. Almeida, A. Ramires, and J. Marinho, "Hashtags—a keystone of Instagram in the hotel digital strategy. An Iberian case study," *Journal of Quality Assurance in Hospitality & Tourism*, pp. 1–19, 2022.