

TOURIST BEHAVIOUR IN FOUR-AND FIVE-STAR HOTELS IN THE ALGARVE BEFORE AND DURING THE COVID-19 PANDEMIC: AN ANALYSIS OF ONLINE REVIEWS

Ana Ferreira^{*}

ESGHT, Universidade do Algarve
<https://orcid.org/0009-0001-6764-0294>

Marisol B. Correia^{**}

Universidade do Algarve, CiTUR-Algarve
<https://orcid.org/0000-0002-1788-6114>

Ana Isabel Renda^{***}

Universidade do Algarve, CiTUR-Algarve
<http://orcid.org/0000-0002-1279-0703>

Nuno António^{****}

Universidade Nova de Lisboa, CiTUR-Algarve
<https://orcid.org/0000-0002-4801-2487>

ABSTRACT

This study aims to examine the behaviour of tourists staying in four- and five-star hotels in Portugal's Algarve region before and during the pandemic and to identify differences between these periods. To achieve this, 17,601 online reviews published on TripAdvisor

Fecha de recepción: 11 de febrero de 2025.

Fecha de aceptación: 7 de noviembre de 2025.

* School of Management, Hospitality and Tourism, Universidade do Algarve, 8005-139 FARO (Portugal).
E-mail: ana.o.ferreira24@gmail.com

** School of Management, Hospitality and Tourism, Universidade do Algarve, 8005-139 FARO (Portugal); Centre for Tourism Research, Development and Innovation - CiTUR (Portugal); Research Centre for Tourism, Sustainability and Well-Being - CinTurs (Portugal); CEG-IST, Instituto Superior Técnico, Universidade de Lisboa (Portugal). E-mail: mcorreia@ualg.pt

*** School of Management, Hospitality and Tourism, Universidade do Algarve, 8005-139 FARO (Portugal); Centre for Tourism Research, Development and Innovation - CiTUR (Portugal); Research Centre for Tourism, Sustainability and Well-Being - CinTurs (Portugal). E-mail: arenda@ualg.pt

**** NOVA Information Management School, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 LISBOA (Portugal); Centre for Tourism Research, Development and Innovation - CiTUR (Portugal). E-mail: nantonio@novaims.unl.pt

between January 2018 and April 2023 were analysed using NVivo 14 software. The findings reveal that guest behaviour during the pandemic closely mirrored pre-pandemic patterns, except for variations related to tourists' continent of origin.

Keywords: Algarve; COVID-19; Four- and five-star hotels; Online reviews; Text mining.

Comportamiento de los turistas en los hoteles de cuatro y cinco estrellas del Algarve antes y durante la pandemia de COVID-19: Un análisis basado en las reseñas en línea

RESUMEN

Este estudio investiga el comportamiento de los turistas alojados en hoteles de cuatro y cinco estrellas en el Algarve, Portugal, antes y durante la pandemia de COVID-19. Para ello, se analizaron 17,601 reseñas publicadas en TripAdvisor entre enero de 2018 y abril de 2023, utilizando el software NVivo 14. Los resultados revelan que el comportamiento de los huéspedes durante la pandemia siguió un patrón muy similar al del período previo a la pandemia, con la excepción de variaciones relacionadas con el continente de origen de los turistas.

Palabras clave: Algarve; COVID-19; Hoteles de cuatro y cinco estrellas; Reseñas en línea; Minería de textos.

1. INTRODUCTION

The travel and tourism sector is a key economic activity in Portugal. As the country's leading service-exporting activity, tourism receipts in 2024 accounted for 47.9% of service exports, 20.8% of total exports, and contributed 9.7% to the Gross Domestic Product (TravelBI, 2025d). This sector's good performance enabled Portugal to rank 12th in the Travel & Tourism Development Index in 2024, an index that measures indicators and policies related to tourism and travel in 117 economies (World Economic Forum, 2024).

One of Portugal's main tourist regions is the Algarve, which, in the first half of 2025, boasted the largest accommodation capacity among the country's tourist accommodation establishments and was the leading destination in terms of overnight guest stays (TravelBI, 2025a, 2025c). In the first quarter of 2025, 16.7% of all employees in the Algarve had jobs related to the tourism sector, making it the Portuguese location with the highest percentage of employment linked to this activity sector (TravelBI, 2025b).

In 2019, tourism in the Algarve region experienced its best year ever, with 5 million guests staying in tourist accommodation establishments (+7.0% compared to 2018). Of these, 29.1% were residents and 70.9% were non-residents. In the same year, the Algarve recorded 21 million overnight stays (+2.5%), the majority of which were made by foreign guests (76.1%). As a result, the region's tourist accommodation establishments generated a total revenue of 1.2 billion, of which 910 million was attributable to accommodation profits (Instituto Nacional de Estatística, 2020).

However, the tourism industry is highly vulnerable to unexpected occurrences such as pandemics (Hu *et al.*, 2021). On 31 December 2019, the World Health Organization (WHO) was alerted to several unknown pneumonia cases in Wuhan, China, which were later identified as a new type of coronavirus. On 30 January 2020, the WHO declared the new coronavirus outbreak a Public Health Emergency of International Concern (PHEIC). On 11 March 2020, it was characterised as a pandemic caused by the COVID-19 disease (Organização Pan-Americana da Saúde, 2023). Due to generalised lockdowns and travel restrictions imposed by countries to contain the virus's spread, the pandemic has caused a huge drop in international demand, causing an unprecedented interruption in tourism (UNWTO, 2020).

In 2020, the number of guests in Algarve's tourist accommodation establishments fell by 3.1 million compared to 2019 (-60.7%), and overnight stays totalled 7.9 million (-62.3%), of which foreign guests made 51.7%. With this outcome, the total revenue of the region's tourist accommodation establishments and profits from accommodation totalled 462 million euros (-62.3%) and 346 million euros (-61.9%), respectively (Instituto Nacional de Estatística, 2021).

The year 2021 was also impacted by the constraints of the COVID-19 pandemic due to confinement measures imposed, which had negative effects on the tourism sector. Despite having grown compared to 2020, it was still below the levels of 2019 (Instituto Nacional de Estatística, 2022). In 2022, a year significantly marked by the return of foreign tourists to Portugal, the tourism sector came close to record figures in 2019 (Instituto Nacional de Estatística, 2023).

On 5 May 2023, WHO declared the end of PHEIC and determined that COVID-19 is now an established and ongoing health problem (World Health Organization, 2023). In 2023, Algarve's tourism sector reached and surpassed levels of some statistical indicators obtained in 2019, with 5.1 million guests (+1.4% compared to 2019) and 20.4 million overnight stays (-2.5%) in the region's tourist accommodation establishments, which led to 1.6 billion in total revenue (+29.9%) and 1.2 billion in revenue from accommodation (+29.1%) (Instituto Nacional de Estatística, 2024).

In 2024, tourism in the Algarve region continued to show strong growth, further consolidating its position as one of Portugal's leading tourist destinations. The region welcomed around 5.3 million guests staying in tourist accommodation establishments, of which 27.5% were residents and 72.5% were non-residents, highlighting the Algarve's sustained international appeal. During the same year, the Algarve recorded 20.7 million overnight stays, the majority (77.5%) made by foreign visitors. As a result, the region's tourist accommodation establishments generated a total revenue of approximately 1.7 billion, with 1.3 billion attributable to accommodation profits (Instituto Nacional de Estatística, 2025). The positive trend continued into the first half of 2025, with strong occupancy rates, a predominance of international tourists, particularly from the United Kingdom, and continued growth in employment and tourism-related revenue, confirming the Algarve's resilience and sustained attractiveness as a top European destination.

Given that the tourism sector has been severely affected by the COVID-19 pandemic, there is a need to investigate its impact on the hotel sector, particularly regarding changes

in guest behaviour. In this context, this study seeks to answer the following research question: How did the COVID-19 pandemic influence the behaviour of tourists in four- and five-star hotels in the Algarve, as reflected in online reviews? To address this question, the study has the following objectives: (i) to identify the behaviour of tourists in four- and five-star hotels in the Algarve region before and during the COVID-19 pandemic, through the analysis of online reviews published between January 2018 and April 2023; and (ii) to examine the possible changes in the behaviour of these tourists as a result of the pandemic. In addition, it is intended to provide hoteliers with theoretical and practical contributions that may be useful for their business performance, as well as contribute to existing literature by expanding current research on the use of online reviews in the hotel industry.

The study is divided into five sections. After the introductory section, we present a literature review on the subject under study and an explanation of the methodology. The fourth section presents analyses and discusses the results. Finally, the study presents its main conclusions, along with management implications, limitations, and recommendations for future research.

2. LITERATURE REVIEW

2.1. Online reviews: User characteristics and hotel attributes

It has been proven that analysing online reviews enables the assessment of hotel performance, consumer behaviour, improvement needs and market demands (Anderson S. Oliveira *et al.*, 2020). Additionally, it can be confirmed that, alongside textual content, review metadata - such as users' age, trip purpose, comments' language, stay date, hotel location and category - has the potential to help hoteliers assess various factors that influence reviews publication (A. S. Oliveira *et al.*, 2022; Ye *et al.*, 2018). The duplication of online reviews can skew consumer perceptions, so it is important that online review platforms implement strategies to mitigate this (António *et al.*, 2023).

The systematic literature review conducted previously by Ferreira *et al.* (2024), which compiled studies using online review analyses, made it possible to understand that in four- and five-star hotels, guest satisfaction varies according to country of origin (Tsiotsou, 2019, 2022; Wu *et al.*, 2017; Zhou *et al.*, 2014), comments' language (Antonio *et al.*, 2018; Pacheco, 2016; Schuckert *et al.*, 2015) and traveller type, i.e. solo, couple, family, friends or business travellers (Ahani *et al.*, 2019; Ahn *et al.*, 2017; Tsiotsou, 2019; Wu *et al.*, 2017). In addition to these factors, studies have also shown that users' experiences in writing online reviews influence ratings assigned to four- and five-star hotels (Amatulli *et al.*, 2019; Liang *et al.*, 2016).

On the other hand, online reviews provide a wealth of information on how different hotel attributes are evaluated, and it has been shown that the importance given by customers to attributes varies according to the hotel's star rating (Rhee & Yang, 2015). In four- and five-star hotels, the main attribute responsible for customer satisfaction is value for money, according to studies by Li *et al.* (2013), Limberger *et al.* (2014), Rhee & Yang (2015) and Wong *et al.* (2020), and it was indicated by Li *et al.* (2013) that this attribute plays an important role in customers' booking decisions and post-purchase satisfaction.

Nevertheless, in a study by Cibere *et al.* (2020), which investigated the performance of six attributes, value for money ranked last, with cleanliness being the top-rated attribute by customers. In addition, cleanliness appears to be customers' third most valued attribute in investigations by Wong *et al.* (2020) and Rhee & Yang (2015), proving intangible services' impact on customer experience.

Service was a determining factor for customer satisfaction in four- and five-star hotels in Pacheco (2017) and Ríos-Martín *et al.* (2019) studies. Although it was not the most relevant attribute, Cibere *et al.* (2020), Kim *et al.* (2016), Limberger *et al.* (2014), Rhee & Yang (2015), and Wong *et al.* (2020) also highlight it as an attribute that contributes to guest satisfaction. It has been verified that service quality can directly influence hotel recommendations to third parties and the desire to return to the same hotel (Padma & Ahn, 2020).

Similarly, rooms have been identified as an important attribute for guest satisfaction in four- and five-star hotels in many studies (Cibere *et al.*, 2020; Kim *et al.*, 2016; Kitsios *et al.*, 2021; Limberger *et al.*, 2014; Pacheco, 2017; Padma & Ahn, 2020; Rhee & Yang, 2015; Wong *et al.*, 2020). According to Kitsios *et al.* (2021), for guest satisfaction to be achieved, hotel rooms must be pleasant, clean, comfortable and cosy. Additionally, room size is also a factor that customers value (Kim *et al.*, 2016; Pacheco, 2016).

On the other hand, in investigations by Athanasopoulou *et al.* (2023), Kim *et al.* (2016), Kitsios *et al.* (2021) and Li *et al.* (2013), the location was considered the most valued attribute by customers of four- and five-star hotels. Kitsios *et al.* (2021) recommend that, when building future hotels, location and accessibility should be considered. For existing hotel companies, they suggest improving access to tourist attractions and/or transportation hubs. In contrast, in studies by Cibere *et al.* (2020), Wong *et al.* (2020) and Rhee & Yang (2015), location was the least important attribute for customers, possibly because it is only relevant when choosing a hotel (Limberger *et al.*, 2014; Pacheco, 2017; Wong *et al.*, 2020).

In addition to the attributes mentioned above, staff were also a key factor in customer satisfaction at four- and five-star hotels in some studies (Athanasopoulou *et al.*, 2023; Kim *et al.*, 2016; Kitsios *et al.*, 2021; Padma & Ahn, 2020; Ríos-Martín *et al.*, 2019). As the only element that can immediately adapt to each customer's needs, hotel staff gain particular importance by providing an individualised experience tailored to customers' specific needs and expectations (Ríos-Martín *et al.*, 2019). According to Athanasopoulou *et al.* (2023) and Kitsios *et al.* (2021), friendly, well-trained, and efficient staff enhance customer satisfaction.

Lastly, sleep quality is identified as a satisfaction attribute in studies by Cibere *et al.* (2020), Rhee & Yang (2015) and Wong *et al.* (2020). However, Ríos-Martín *et al.* (2019) and Rhee & Yang (2015) note that online reviews rarely mention sleep quality. Customers consider it a basic expectation, similar to cleanliness, since a higher-category hotel is generally expected to provide better sleep quality.

2.2. Online reviews in the COVID-19 pandemic

Since the hotel industry is highly vulnerable to pandemics, it has been suggested that COVID-19 may alter consumer behaviour in the long term (Hu *et al.*, 2021). Given that

text analysis provides valuable insights into trends in the hotel industry (Bi *et al.*, 2024), several studies have used online hotel reviews to compare consumer evaluations before and during the pandemic.

For instance, Xu *et al.* (2023) found that the pandemic had a positive effect on customer ratings, although this effect was less pronounced in higher-star hotels. By contrast, Raedts *et al.* (2023) reported lower ratings for four-star hotels during the same period. Despite these differences, both studies confirmed that the attributes most valued by customers remained relatively stable, particularly services, rooms, breakfast, and location. Similarly, Cheng *et al.* (2021) also identified services, rooms, and location as key drivers of satisfaction and argued that, in five-star hotels, the main determinants of guest experience did not change significantly during the pandemic. At the same time, new priorities emerged. Several studies observed that cleanliness, hygiene, and safety measures gained greater importance (Cheng *et al.*, 2021; Hu *et al.*, 2021; Raedts *et al.*, 2023; Song *et al.*, 2022). Moreover, tourists who explicitly valued health-related measures and positively mentioned how hotels managed the pandemic tended to give higher ratings to hotels (Song *et al.*, 2022; Srivastava & Kumar, 2021; Xu *et al.*, 2023).

Although online reviews have proven to be a valuable and real-time source of information for understanding post-pandemic dynamics (Alamoodi *et al.*, 2021; Bi *et al.*, 2024), there is still a need for research that examines luxury hotels in Southern Europe with a longitudinal perspective and large-scale datasets. This study seeks to address this gap by analysing 17,601 online reviews of four- and five-star hotels in the Algarve, Portugal, comparing pre-pandemic and pandemic periods to provide new insights into tourist behaviour during times of crisis and recovery.

3. METHODOLOGY

3.1. Sampling and data collection

TripAdvisor is the world's largest travel platform, helping 463 million travellers each month make the most of every trip. The platform hosts over 859 million reviews and opinions covering 8.6 million accommodations, restaurants, experiences, airlines, and cruises. (TripAdvisor, 2025). This platform was selected as the source for extracting online reviews of four- and five-star hotels in the Algarve.

Before data extraction, the Registo Nacional de Turismo (2023) was consulted to identify the hotels to be analysed, resulting in a total of 104 four- and five-star establishments in the Algarve (as of 12 June 2023). However, since this research focuses on reviews published from the beginning of 2018, 11 hotels were excluded because they opened in 2018 or later. Therefore, only hotels in operation until 2017 (inclusive) were considered. The final sample comprised 93 hotels (64 four-star and 29 five-star), which were identified on TripAdvisor using their names, addresses, postal codes, and official websites.

An Excel database was then created, assigning each of the 93 hotels an identification number and linking it to the corresponding TripAdvisor page. To collect reviews pub-

lished from 1 January 2018, in five languages (English, Portuguese, Spanish, German, and French), a C# based bot was employed.

To facilitate the collection of user-generated content for research purposes, this bot was developed to extract reviews from the TripAdvisor platform. Given the dynamic and anti-scraping nature of the website, the automation process was implemented using the Selenium library (Selenium, 2025), which enables browser-level interaction and supports the rendering of JavaScript-heavy content. This approach was essential for accurately capturing review data that is often loaded asynchronously.

The bot was configured to operate with the Firefox browser, managed via GeckoDriver (Mozilla, 2025), which serves as a mediator between Selenium scripts and the browser environment. This configuration provided robust control over browser actions such as navigation, page scrolling, and element interaction, which were crucial for accessing paginated review content and simulating human behaviour.

To minimise the risk of being blocked by TripAdvisor's anti-scraping mechanisms, several behavioural mimicry techniques were integrated. Specifically, randomised time delays were introduced between user-like actions. These delays were drawn from uniform distributions designed to reflect typical human browsing intervals, thus simulating realistic interaction patterns.

Furthermore, to obfuscate the automated nature of the requests, the bot incorporated dynamic user-agent switching. This was achieved through the use of the Scrapy Fake UserAgent library (Alecxe, 2025), which enables the emulation of various browser signatures. By rotating user-agent strings across sessions, the bot effectively diversified its digital fingerprint, thereby reducing the likelihood of detection and throttling.

Nonetheless, when the HTML format is changed or TripAdvisor implements A/B testing, errors occur, and manual intervention is required to update the bot code.

The data extraction process began on July 27, 2023, and lasted three days, encompassing reviews published during that period. A total of 39,851 reviews were collected and stored in the Excel database. Along with the review texts, additional variables were extracted, including hotel location (municipality), user location (city, province, or country), number of user contributions, type of traveller, date of stay (different from the review date), average ratings, and review language.

The data collection was conducted solely from publicly available TripAdvisor reviews, for strict academic purposes. No private or confidential data was extracted, ensuring compliance with the platform's Terms of Use and the European General Data Protection Regulation (GDPR), and respecting user privacy.

3.2. Data Refining

Before analysing the online reviews, it was necessary to refine and preprocess the dataset, which involved converting unstructured text data into structured data suitable for analysis (Ban *et al.*, 2019). This process was carried out in several stages. First, all rows in the Excel sheet containing one or more empty cells were removed ($n = 18,988$). Second, a manual cleaning procedure was applied to the user location column, ensuring

that cities and tourist provinces were standardised and presented at the country level. Rows containing data that did not correspond to a valid location were also excluded ($n = 18,943$). Third, the stay date column was reformatted into a date variable, and all entries were sorted in ascending order.

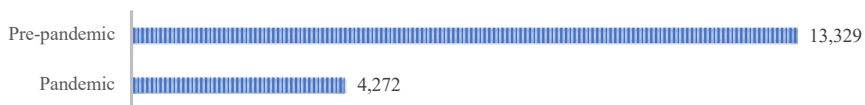
Since the study aimed to analyse reviews from January 2018 onwards, all entries with a stay date before this period were excluded ($n = 18,596$). This refinement was necessary because some reviews written in 2018 actually referred to stays in 2017, which would have introduced inconsistencies into the temporal scope of the research. Similarly, reviews with a stay date in May 2023 or later were also removed ($n = 17,601$), as on May 5, 2023, WHO officially declared the end of the PHEIC related to COVID-19 (World Health Organization, 2023).

It is important to emphasise that TripAdvisor records stay dates only by month and year, without indicating the exact day. As a result, it was not possible to determine whether reviews with a stay date listed as “May 2023” referred to stays that occurred before or after 5 May 2023. Consequently, all entries associated with May 2023 were excluded, which may have led to the omission of a small number of reviews corresponding to the first five days of that month. Nonetheless, this decision was methodologically necessary to ensure a clear and consistent demarcation between the pandemic period and the post-pandemic context.

All reviews were translated into English using an online translator (Online Doc Translator) to ensure that the analysis was conducted in a single language. Automatic translation was deemed appropriate given the large dataset and the need for consistency in language processing. Previous studies have shown that machine translation can be a reliable tool for academic research, particularly in large-scale text mining and sentiment analysis (Kathunia *et al.*, 2024; Lee, 2024). To validate semantic accuracy, a random sample of translated reviews was manually compared with the original texts, confirming that the translations preserved the intended meaning and were suitable for this study. In addition, spelling and grammar checking tools (Grammarly and Microsoft Word’s Spelling & Grammar, integrated with Excel) were applied to the dataset to correct typographical, grammatical, and minor linguistic errors, thereby further enhancing consistency and reliability.

After refining the data, a total of 17,601 reviews were obtained. Two databases were then created in Excel: the first covering the pre-pandemic period, consisting of 13,329 reviews made between January 2018 and January 2020, as WHO declared the coronavirus outbreak a PHEIC on January 30, 2020 (Organização Pan-Americana da Saúde, 2023); and the second covering pandemic period, with 4,272 reviews recorded between February 2020 and April 2023, as shown in figure 1. As mentioned earlier, since TripAdvisor displays stay dates only in terms of month and year, the pre-pandemic database includes reviews from guests who stayed on January 30 and 31, 2020 - dates that technically belong to the pandemic period. On the other hand, the database covering reviews during the pandemic does not include reviews from the first five days of May 2023, due to the official end of the COVID-19 PHEIC declared by the WHO on May 5, 2023 (World Health Organization, 2023).

Figure 1
NUMBER OF ONLINE REVIEWS CONSIDERED BY PERIOD



Source: Own work.

3.3. Data analysis

The data analysis was structured into two main stages: quantitative and qualitative. In the first stage, a quantitative analysis was conducted in Excel, based on two distinct databases: the first referring to the pre-pandemic period and the second to the pandemic period. Through the construction of graphs, it was possible to observe the distribution of the number of reviews across the different metadata considered, namely: hotel location (municipality), review language, user location (country and continent of origin), date of stay, traveller type, number of user contributions, and average ratings assigned. After the individual analysis of each database, the results were integrated, enabling a comparison between the two periods.

In the second stage, a qualitative analysis was conducted, specifically examining the textual content of online reviews. Since handling a large volume of data can be demanding, the use of appropriate software to analyse and categorise online reviews becomes necessary (Anderson S. Oliveira *et al.*, 2020). Following the approach of A. S. Oliveira *et al.* (2022), NVivo 14 was employed, as it is designed to assist researchers in qualitative data analysis. It is important to note that this software can only process one language at a time and does not recognise sarcasm, double negatives, slang, linguistic variations, idiomatic expressions, or ambiguity. Nonetheless, it remains an effective tool for obtaining an overall summary of opinions (A. S. Oliveira *et al.*, 2022). Within this context, the two databases were imported into the software, resulting in two separate projects. Subsequently, an automatic sentence-level sentiment analysis was conducted to detect the polarity expressed in each textual comment within both projects. Sentiment analysis is characterised by its capacity to determine the polarity of a document, comment, opinion, or expressed emotion, which may be positive, neutral, or negative (Alaei *et al.*, 2019). In this study, it was applied to identify the sentiment present in each sentence of the guests' reviews. This procedure aimed to capture the overall sentiment of tourists in both the pre-pandemic and pandemic periods, as expressed in online reviews. At this stage, both the total number of reviews and their full textual content were considered.

Subsequently, a frequency analysis was performed to rank the 30 most frequent nouns associated with hotel attributes. As a result, it was possible to identify which hotel attributes were most frequently discussed in online reviews for each period under study. Furthermore, frequency analysis allowed for the identification of themes, which helped determine which hotel attributes could later be compared. Consequently, aspect-

level sentiment analysis was carried out to identify the polarity of opinions expressed regarding the four hotel attributes most frequently mentioned in the reviews during each study period.

Finally, sentence-level sentiment analysis was applied to the metadata associated with each set of reviews, namely: review language, user location (continent of origin), date of stay, traveller type, and number of user contributions. This procedure not only enabled the identification of the polarity of opinion expressed but also provided insights into how different tourist profiles conveyed distinct sentiments concerning their hotel experiences. Lastly, the results were compared across the two periods under analysis, allowing for the identification of behavioral variations in the pattern of online reviews.

4. RESULTS

4.1. Quantitative analysis

As Figure 2 shows, the Algarve region is composed of 16 municipalities, with most of the four- and five-star hotels located along the coastal areas. No hotels were observed in São Brás de Alportel and Alcoutim.

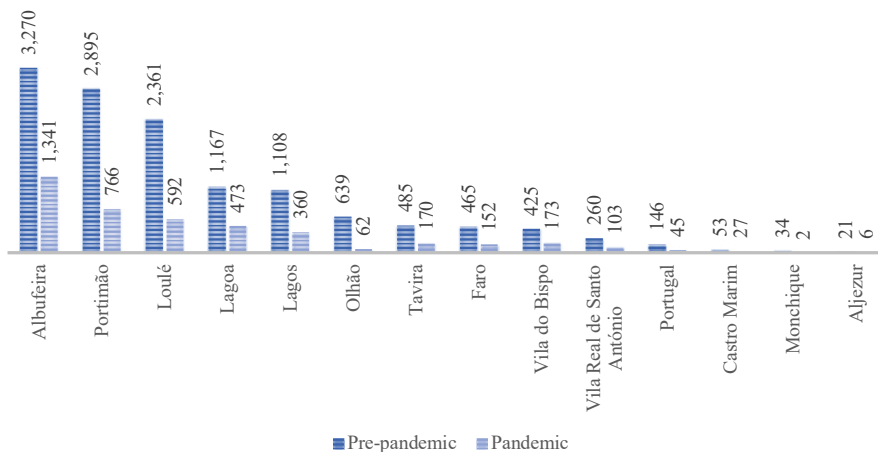
Figure 2
NUMBER OF FOUR- AND FIVE-STAR HOTELS BY MUNICIPALITIES IN THE ALGARVE



Source: Adapted from Associação Nacional de Municípios Portugueses (2020).

Regardless of the study period (Figure 3), most online reviews collected come from tourists staying in hotels located in the municipalities of Albufeira, Portimão, Loulé, Lagoa, and Lagos, which, as previously verified, have a higher number of four- and five-star hotels (Figure 2).

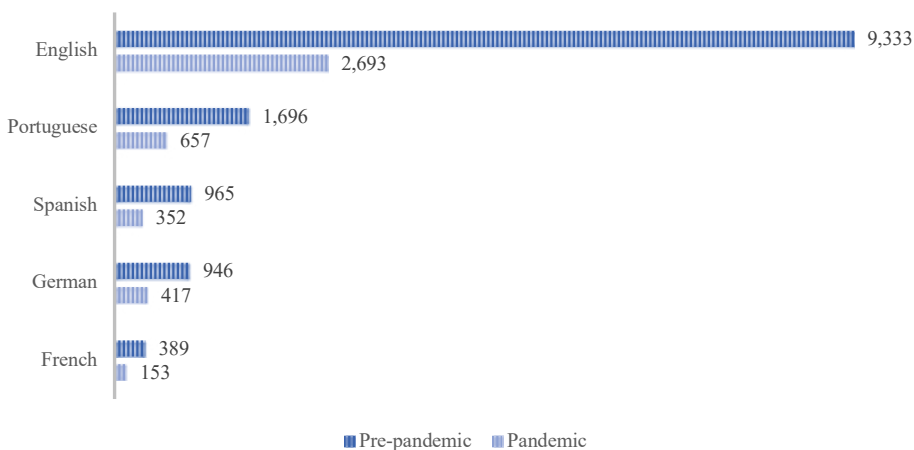
Figure 3
NUMBER OF ONLINE REVIEWS BY ALGARVE MUNICIPALITIES



Source: Own work.

For the two periods in question, more than half the comments were written in English (Figure 4). This was followed by comments written in Portuguese, Spanish, and German in the pre-pandemic period and in Portuguese, German, and Spanish in the pandemic period. French was the least used language by tourists in both periods.

Figure 4
NUMBER OF ONLINE REVIEWS BY LANGUAGE



Source: Own work.

In Table 1, fifteen countries with the highest number of tourists who submitted online reviews under study are listed, with the United Kingdom standing out, accounting for 51.8% and 45.2% of total tourists in pre-pandemic and pandemic periods, respectively, and Portugal, with 12.7% and 16.7% in the first and second periods under study. The same table also shows the origin of tourists by continent, highlighting that in both periods, over 90.0% come from Europe. It is important to mention that reviews from tourists originating from Azerbaijan, Russia, and Turkey were counted twice, as these countries are part of both Europe and Asia, resulting in the total number of reviews in the table being higher than the total number of extracted reviews (an additional 20 reviews). In smaller numbers, tourists from Africa were also present during both periods.

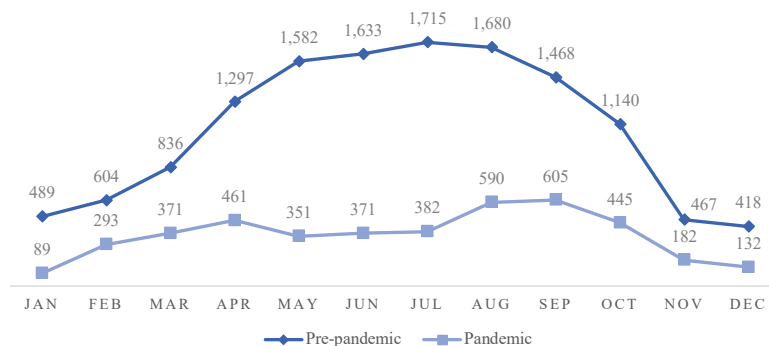
Table 1
NUMBER OF ONLINE REVIEWS BY COUNTRY AND CONTINENT OF TOURISTS' ORIGIN

Country of origin	Pre-pandemic	%	Pandemic	%
United Kingdom	6,898	51.8	1,930	45.2
Portugal	1,687	12.7	714	16.7
Spain	979	7.3	363	8.5
Germany	779	5.8	380	8.9
Ireland	717	5.4	186	4.4
United States	425	3.2	141	3.3
France	403	3.0	154	3.6
Brazil	226	1.7	37	0.9
Switzerland	208	1.6	83	1.9
Canada	162	1.2	49	1.1
Belgium	90	0.7	29	0.7
Netherlands	79	0.6	28	0.7
Australia	77	0.6	8	0.2
Austria	76	0.6	18	0.4
Italy	44	0.3	19	0.4
Others	479	3.6	133	3.1
Continent of origin				
Europe	12,205	91.5	3,977	93.0
North America	613	4.6	199	4.7
South America	284	2.1	44	1.0
Asia	107	0.8	41	1.0
Oceania	89	0.7	11	0.3
Africa	45	0.3	6	0.1

Source: Own work.

Regarding stay dates, tourists primarily chose to stay between April and October in the pre-pandemic period, while they mainly opted for April, August, September, and October in the pandemic period (Figure 5).

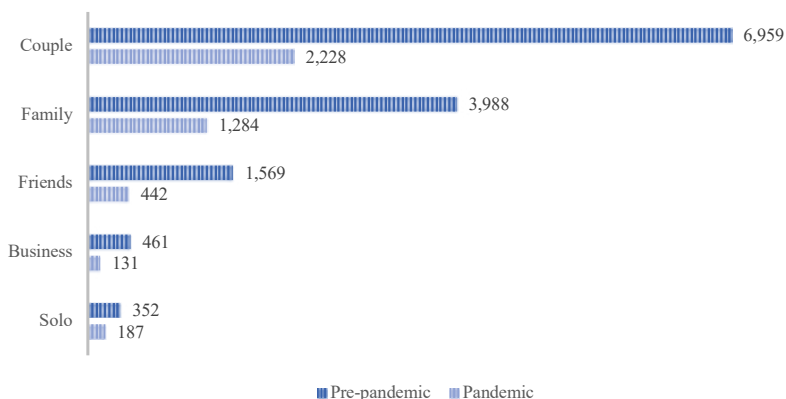
Figure 5
NUMBER OF ONLINE REVIEWS BY STAY DATE



Source: Own work.

Based on Figure 6, more than half the tourists who stayed in four- and five-star hotels in the Algarve travelled as a couple in both periods. This category of hotels also saw a considerable number of tourists travelling with family and friends. Guests travelling for business and those travelling alone were fewer, with more solo travellers than business travellers during the pandemic, which was not the case in the pre-pandemic period.

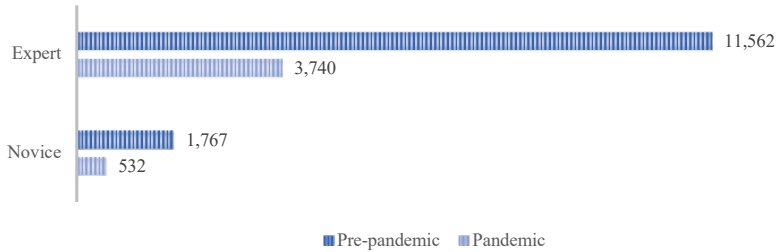
Figure 6
NUMBER OF ONLINE REVIEWS BY TRAVELLER TYPE



Source: Own work.

Regarding tourists’ experience writing online reviews, the majority are experienced, meaning it is not their first time writing an online review for both periods under study (Figure 7).

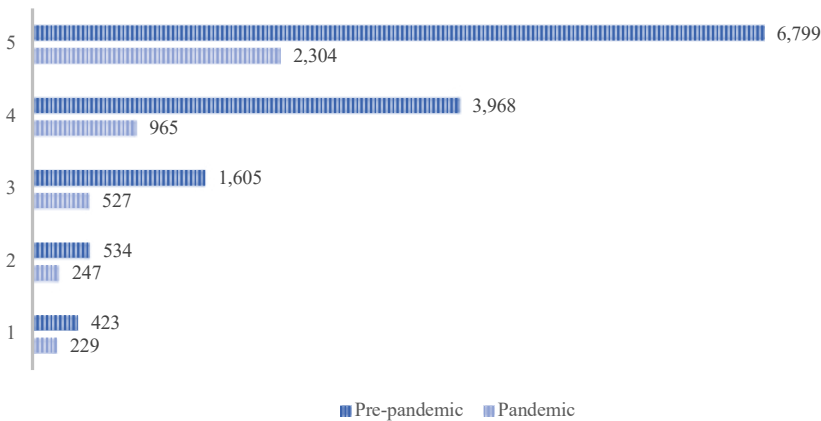
Figure 7
NUMBER OF ONLINE REVIEWS BY USER CONTRIBUTIONS



Source: Own work.

Since TripAdvisor allows hotels to be rated up to five points, four- and five-star hotels in the Algarve region were mostly rated positively by tourists, with ratings of 4 and 5 points in both periods (Figure 8).

Figure 8
NUMBER OF ONLINE REVIEWS BY AVERAGE RATING



Source: Own work.

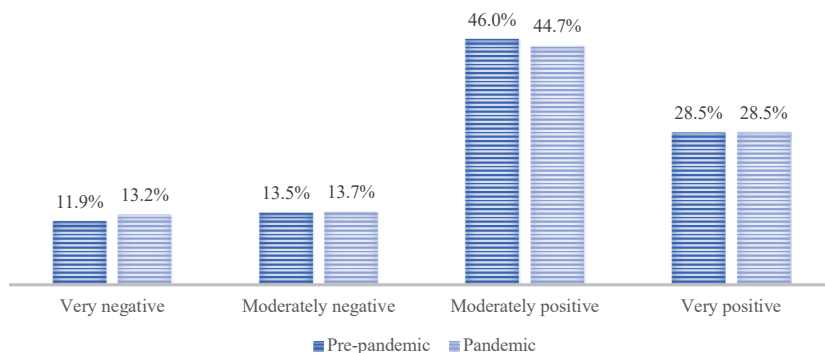
4.2. Qualitative analysis

To simplify qualitative analysis reading, including sentiment and frequency analyses, positive sentiment encompasses moderately positive and very positive sentiments, analysed together. The same applies when discussing negative sentiment, which refers to both moderately negative and very negative sentiments grouped simultaneously.

4.2.1. Global sentiment analysis

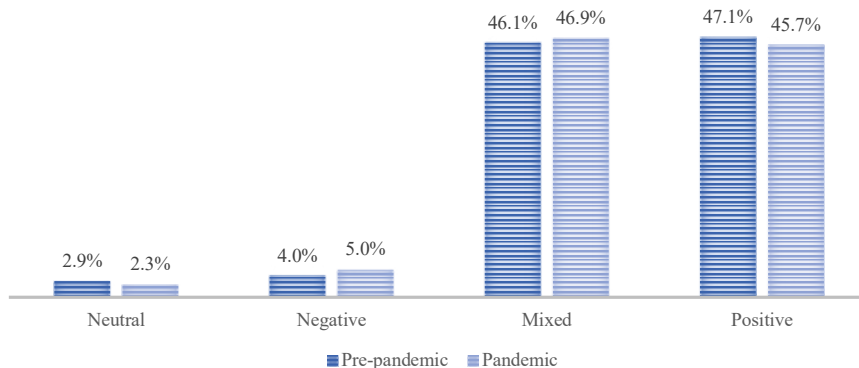
Across both periods analysed, tourists staying in four- and five-star hotels in the Algarve region predominantly wrote comments with positive sentiment, accounting for 74.5% in the pre-pandemic period and 73.2% during the pandemic. Of these, 28.5% were classified as having a very positive sentiment in each period (Figure 9). In contrast, comments with negative sentiment made up 25.4% in the pre-pandemic period and 26.9% during the pandemic, with 11.9% categorised as very negative in the pre-pandemic period and 13.2% during the pandemic.

Figure 9
OVERALL SENTIMENT OF ONLINE REVIEWS



Source: Own work.

Figure 10
OVERALL SENTIMENT OF CODING REFERENCES



Source: Own work.

From another perspective, Figure 10 shows that nearly half of the total comments written by tourists reflect positive or mixed sentiment (containing both positive and negative emotions), with 47.1% and 46.1% in the pre-pandemic period, and 45.7% and 46.9% during the pandemic period. In contrast, comments with negative or neutral sentiment (lacking any emotion) make up a smaller percentage, at 4.0% and 2.9% in the pre-pandemic period, and 5.0% and 2.3% in the pandemic period.

4.2.2. Frequency analysis

Table 2 lists the ten most frequently mentioned words in the comments analysed for both periods. It is evident that, in both the pre-pandemic and pandemic periods, the top four words are equal and appear in the same order (rooms, staff, pool, and breakfast). These four hotel attributes will be given special attention in the following section, where they will be analysed for sentiment. The next three words (food, restaurants, and beaches) appear in both periods but in a different order. In the pre-pandemic period, customers wrote more comments about food, followed by those referring to restaurants and beaches. In the pandemic period, beaches took the spotlight, with food and restaurants being mentioned afterwards. Lastly, “areas” was a commonly mentioned attribute in the pre-pandemic period, while during the pandemic, customers emphasised “views.”

Table 2
FREQUENCY ANALYSIS RESULTS

	Pre-pandemic	%	Pandemic	%
1	rooms	1.94	rooms	1.93
2	staffs	1.29	staffs	1.34
3	pool'	1.21	pools	1.22
4	breakfasts	0.87	breakfasts	0.92
5	foods	0.76	beaches	0.77
6	restaurant'	0.74	foods	0.77
7	beach'	0.73	restaurants	0.74
8	cleans	0.67	cleaning	0.68
9	bar'	0.64	bars	0.66
10	areas	0.63	view'	0.59

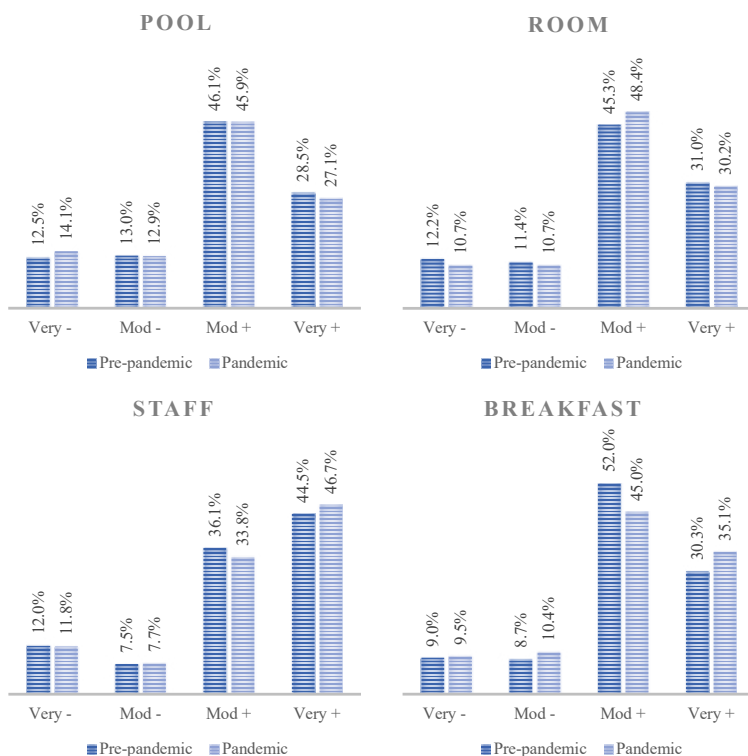
Source: Own work.

It is important to note that only words related to hotel attributes were retained in the frequency table. Words that did not meet this condition, such as “hotel,” which refers to the hotel unit itself, and “day,” which is not a hotel attribute, were removed.

4.2.3. Sentiment analysis of select hotel attributes

Regarding the four hotel attributes most frequently mentioned by customers in the comments (Figure 11), it is evident that, in the pre-pandemic period, breakfast was the attribute that generated the highest number of comments with positive sentiment (82.3%), followed by staff (80.5%), rooms (76.4%) and pools (74.6%). In contrast, during the pandemic period, the largest proportion of positive comments shifted to staff (80.5%), with breakfast (80.1%), rooms (78.5%) and pools (73.0%) following. Despite these findings, in both periods, most comments with a very positive sentiment were related to staff, accounting for 44.5% in the pre-pandemic period and 46.7% in the pandemic period, followed by comments about rooms, breakfast and pools. Conversely, regardless of the period, the pools were the hotel attribute generating the highest percentage of negative sentiment reviews, at 25.5% in the pre-pandemic period and 27.0% in the pandemic period. Additionally, pools received the highest number of very negative comments, with 12.5% in the pre-pandemic period and 14.1% during the pandemic.

Figure 11
SENTIMENT TOWARDS SELECTED HOTEL ATTRIBUTES

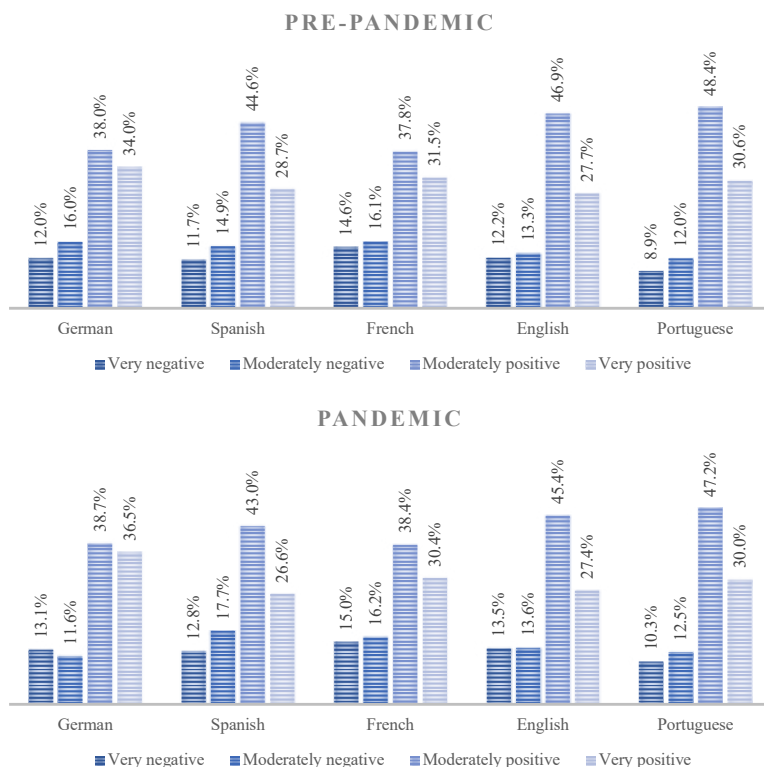


Source: Own work.

4.2.4. Sentiment analysis of metadata

Considering the language of the comments, it is observed that comments written in Portuguese exhibit the highest levels of positive sentiment, with 79.1% before the pandemic and 77.2% during the pandemic (Figure 12). Following these are comments written in English (74.6%), Spanish (73.3%), German (72.1%), and French (69.3%) in the pre-pandemic period. This order shifted with the pandemic, as comments in German (75.3%) ranked second, followed by those in English (72.9%), Spanish (69.5%), and French (68.8%). Although Portuguese comments showed the highest overall positive sentiment, German comments recorded the highest percentages of very positive sentiment in both periods, reaching 34.0% in the pre-pandemic period and 36.5% during the pandemic. On the other hand, French-written comments exhibited the highest percentages of negative sentiment, with 30.7% in the pre-pandemic period and 31.2% during the pandemic. Of these, 14.6% and 15.0% corresponded to comments with very negative sentiment, respectively.

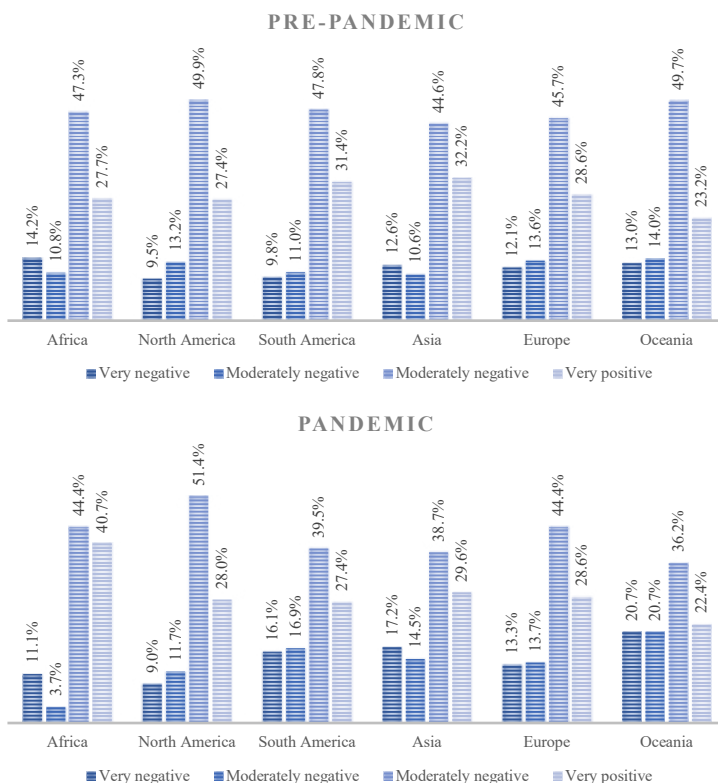
Figure 12
TOURIST SENTIMENT BY LANGUAGE



Source: Own work.

In view of tourists' continent of origin (Figure 13), it can be observed that in the period before the pandemic, the comments with the highest percentage of positive sentiment came from guests in the following order: South America (79.2%); North America (77.3%); Asia (76.8%); Africa (75.0%); Europe (74.3%); and Oceania (73.0%). However, tourists from Asia were the ones who wrote the highest number of comments with very positive sentiment (32.2%). During the pandemic, tourists from Africa wrote the most comments with positive sentiment (85.2%), followed by comments from clients from North America (79.3%), Europe (73.0%), Asia (68.3%), South America (66.9%), and Oceania (58.6%). In addition, tourists from Africa also wrote the highest number of comments with very positive sentiment (40.7%). In contrast, in both periods analysed, comments from clients from Oceania exhibited the highest percentage of negative sentiment, with 27.0% in the first period and 41.4% in the second. Moreover, clients from Oceania were also the ones who wrote the most comments with very negative sentiment during the pandemic (20.7%), whereas, in the pre-pandemic period, clients from Africa stood out in this regard (14.2%).

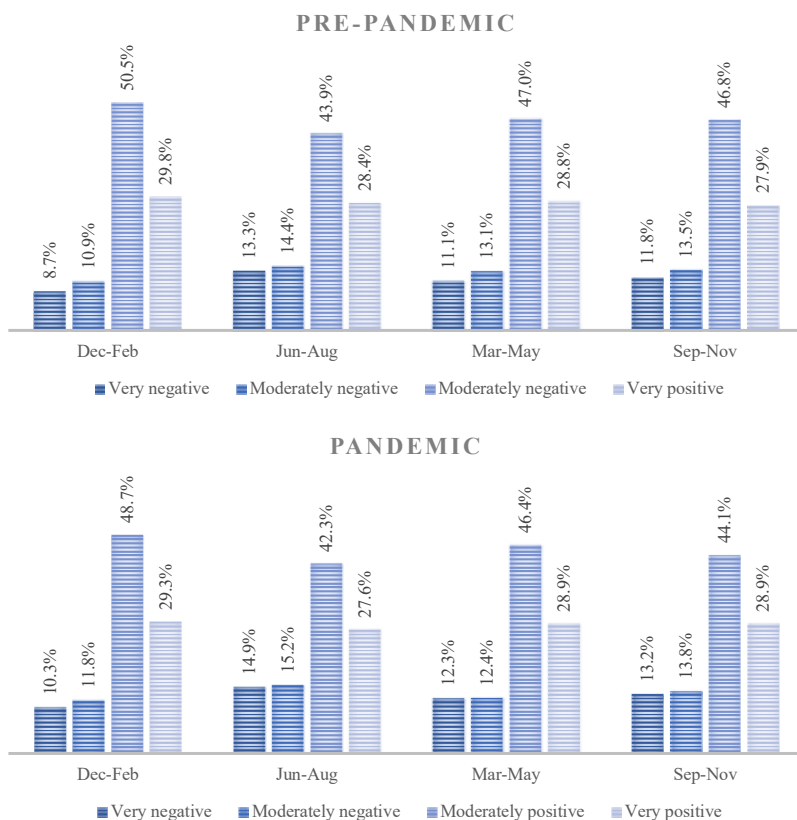
Figure 13
TOURIST SENTIMENT BY CONTINENT OF ORIGIN



Source: Own work.

Considering the dates of the tourists' stay, it is evident that the period from December to February was the one that contributed most to positive sentiment, with 80.3% and 77.9% of positive comments in the pre-pandemic and pandemic periods, respectively. This was also the period when tourists wrote the highest percentage of very positive comments, with 29.8% and 29.3% in the first and second periods, respectively (Figure 14). Following the December to February period, the months from March to May showed the highest percentages of positive sentiment in both periods (75.8% in the pre-pandemic period; 75.3% in the pandemic period), followed by the months from September to November (74.7%; 73.0%) and from June to August (72.3%; 69.9%). In contrast, the majority of negative sentiment comments were written during the months from June to August, with 27.7% and 30.1% in the pre-pandemic and pandemic periods, respectively. This was also the period when the highest number of very negative comments were recorded, with 13.3% and 14.9% in the pre-pandemic and pandemic periods.

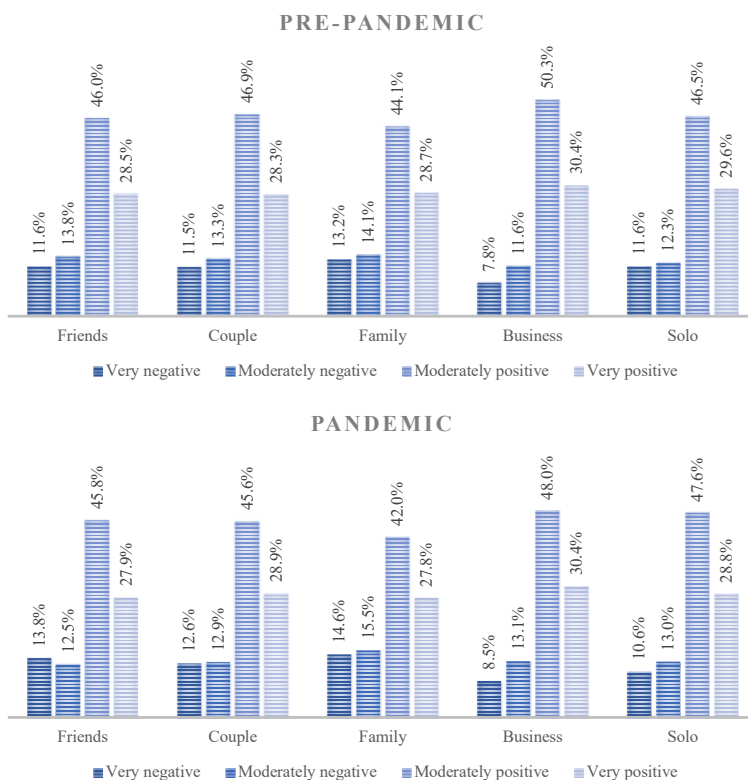
Figure 14
TOURIST SENTIMENT BY STAY DATE



Source: Own work.

Regarding the type of traveller (Figure 15), it is clear that guests travelling for business purposes wrote most positive comments in both periods, with 80.6% in the pre-pandemic period and 78.4% in the pandemic period. This was followed by comments from guests who travelled alone, with a partner, with friends, and with family. Furthermore, business travellers also wrote the highest percentage of comments with very positive sentiment, with 30.4% in each period. On the other hand, guests who travelled with family wrote the highest number of comments with negative sentiment in both periods, with 27.3% and 30.2% in the pre-pandemic and pandemic periods, respectively. This group also expressed the highest levels of very negative sentiment, with 13.2% and 14.6% in the first and second periods.

Figure 15
TOURIST SENTIMENT BY TRAVELLER TYPE

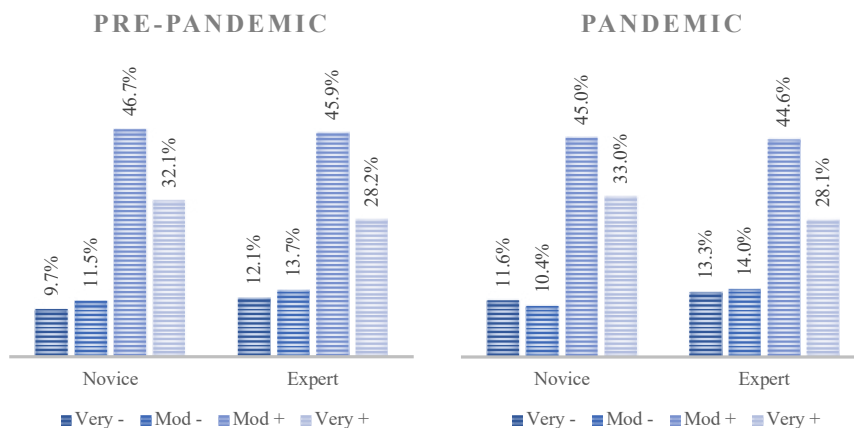


Source: Own work.

Considering the tourists' experience in writing online reviews, the highest number of positive comments was written by those posting an online review for the first time (novices), with 78.8% and 78.0% in the pre-pandemic and pandemic periods, respectively

(Figure 16). Of these, 32.1% and 33.0% wrote comments with very positive sentiment in the first and second periods, respectively. Conversely, tourists with experience in writing online reviews (experts) were the ones who wrote the highest volume of comments with negative sentiment, with 25.8% in the pre-pandemic period and 27.3% during the pandemic, of which 12.1% and 13.3% correspond to comments with very negative sentiment, respectively.

Figure 16
TOURIST SENTIMENT BY CONTRIBUTION NUMBER



Source: Own work.

5. DISCUSSION

The results of the quantitative analysis show that most four- and five-star hotels in the Algarve are in the region's coastal municipalities - particularly in Albufeira, Portimão, Loulé, Lagoa, and Lagos - which also account for the largest number of online reviews analysed in this study. In both the pre-pandemic and pandemic periods, more than half of the reviews were written in English, followed by those written in Portuguese. Spanish and German reviews ranked next during the pre-pandemic period, although their order was reversed during the pandemic. French remained, in both periods, the least frequently used language in the reviews.

Regarding tourists' country of origin, visitors from the United Kingdom stood out the most, followed by Portuguese, Spanish, and German tourists in both periods. Overall, in both timeframes, more than 90% of the reviews were written by European tourists, with a slight increase during the pandemic period.

Stays in the pre-pandemic period occurred mainly between April and October, whereas during the pandemic they were concentrated in April, August, September, and October. In both periods, the majority of tourists travelled as couples, although family and friends'

trips were also significant. Business and solo trips were less frequent; however, during the pandemic, there was a higher number of solo travellers compared to business travellers, the opposite of what was observed in the pre-pandemic period.

With regard to tourists' experience in writing online reviews, it was found that in both periods, most tourists already had prior experience with online reviewing. Furthermore, it is worth noting that four- and five-star hotels in the Algarve were generally rated very positively, predominantly receiving scores of 4 and 5 on TripAdvisor, both before and during the pandemic.

Through qualitative analysis, and in line with the findings of Emir & Bayer (2017), A. S. Oliveira *et al.* (2022), and Tontini *et al.* (2017), who demonstrated that tourists tend to share more positive than negative reviews, the results of this study reveal that tourists staying in four- and five-star hotels in the Algarve region predominantly wrote positive reviews. Specifically, 74.5% of reviews in the pre-pandemic period and 73.2% in the pandemic period were positive. Of these, 28.5% in each period were categorised as highly positive. Conversely, negative reviews accounted for 25.4% during the pre-pandemic period and 26.9% during the pandemic period. This represents a 1.5% increase in negative reviews following the onset of the COVID-19 pandemic, corroborating the findings of Raedts *et al.* (2023), which showed that customers rated hotels less positively during the pandemic. Among the negative sentiment reviews, 11.9% and 13.2% were categorised as highly negative in the pre-pandemic and pandemic periods, respectively.

Based on the word frequency analysis, the hotel attributes most frequently mentioned by tourists in their reviews were identified. Across both periods, the four most commonly cited terms were "rooms," "staff," "pools," and "breakfast," partially corroborating the findings of Kitsios *et al.* (2021), A. S. Oliveira *et al.* (2022), and Padma & Ahn (2020), who also highlighted these attributes in their studies. Furthermore, "rooms" and "breakfast" remained among the most frequently mentioned attributes by customers during the pandemic, as reported by Hu *et al.* (2021) and Raedts *et al.* (2023).

The identified attributes were further analysed for sentiment, enabling the assessment of sentiment associated with each attribute, particularly the percentages of highly positive and highly negative sentiment. The key findings from this analysis are summarised in Table 3.

Table 3
SUMMARY OF SENTIMENT TOWARD SELECTED HOTEL ATTRIBUTES BY STUDY PERIOD

	Very positive sentiment		Very negative sentiment	
	Pre-pandemic (%)	Pandemic (%)	Pre-pandemic (%)	Pandemic (%)
Rooms	31.0	30.2	12.2	10.7
Staff	44.5	46.7	12.0	11.8
Pools	28.5	27.1	12.5	14.1
Breakfast	30.3	35.1	9.0	9.5

Source: Own work.

Staff stood out the most in both periods, as they received the highest number of comments with a very positive sentiment: 44.5% in the pre-pandemic period and 46.7% during the pandemic. Furthermore, staff were more highly valued by customers during the pandemic, reflected in a 2.2% increase in positive feedback. Similarly, breakfast emerged as a highly appreciated attribute during the pandemic, with very positive sentiment rising from 30.3% in the pre-pandemic period to 35.1% during the pandemic, marking a 4.8% improvement. In contrast, swimming pools were the attribute that drew the most very negative comments, increasing from 12.5% in the pre-pandemic period to 14.1% during the pandemic, indicating a slight rise in customer dissatisfaction with this attribute during the pandemic.

A sentiment analysis was performed on the comments, considering specific metadata such as the comments' language, the tourists' continent of origin, their stay date, traveller type, and their experience in writing online reviews. The key findings were consolidated in Table 4, which summarises the percentage of comments with very positive and very negative sentiments for each metadata category and study period.

Table 4
SUMMARY OF SENTIMENT BY METADATA AND STUDY PERIOD

	Very positive sentiment				Very negative sentiment			
	Pre-pandemic (%)		Pandemic (%)		Pre-pandemic (%)		Pandemic (%)	
Language	German	34.0	German	36.5	French	14.6	French	15.0
Date of stay	Dec-Feb	29.8	Dec-Feb	29.3	Jun-Aug	13.3	Jun-Aug	14.9
Type of traveller	Business	30.4	Business	30.4	Family	13.2	Family	14.6
User experience	Novice	32.1	Novice	33.0	Expert	12.1	Expert	13.3
Continent of origin	Asia	32.2	Africa	40.7	Africa	14.2	Oceania	20.7

Source: Own work.

Among the five languages analysed (German, Spanish, French, English, and Portuguese), reviews written in German had the highest percentage of very positive sentiment, with 34.0% in the pre-pandemic period and 36.5% during the pandemic, reflecting a notable increase of 2.5%. In contrast, reviews written in French had the highest percentage of very negative sentiment, rising from 14.6% in the pre-pandemic period to 15.0% during the pandemic, showing a slight decline of 0.4% in sentiment during this time.

Across the four time periods analysed (Dec–Feb, Jun–Aug, Mar–May, and Sep–Nov), tourists who stayed between December and February wrote the highest percentage of reviews with very positive sentiment, reaching 29.8% in the pre-pandemic period and 29.3% during the pandemic. Conversely, tourists who stayed between June and August wrote the most reviews with very negative sentiment, at 13.3% in the pre-pandemic period and 14.9% during the pandemic. During the pandemic, the percentage of reviews with

very positive sentiment decreased by 0.5%, while those with very negative sentiment increased by 1.6%.

When analysing travellers by type (friends, couples, families, business, and solo), business travellers expressed the highest levels of very positive sentiment, with 30.4% in both periods. In contrast, family travellers stood out for having the highest percentage of very negative sentiment, at 13.2% pre-pandemic and 14.6% during the pandemic, reflecting a 1.4% increase in dissatisfaction.

On the other hand, tourists writing an online review for the first time (novices) expressed the highest levels of very positive sentiment, with 32.1% in the pre-pandemic period and 33.0% during the pandemic. These tourists wrote an online review for the first time, with no record of other reviews on their profiles. In contrast, experienced tourists - those who had written at least two online reviews - expressed the highest levels of very negative sentiment, with 12.1% in the pre-pandemic period and 13.3% during the pandemic, highlighting a 1.2% increase in dissatisfaction.

Finally, considering the tourists' continent of origin (Africa, North America, South America, Asia, Europe, and Oceania), it is notable that their behaviour changed with the pandemic. In the pre-pandemic period, tourists from Asia led in very positive comments (32.2%), whereas during the pandemic, African tourists took the lead (40.7%). Interestingly, African tourists also recorded the highest percentage of very negative comments pre-pandemic (14.2%). During the pandemic, this trend shifted, with tourists from Oceania showing the most negative sentiment (20.7%).

In summary, the results of this study revealed that the behaviour of guests leaving online reviews during the pandemic period followed a pattern quite similar to that observed in the pre-pandemic period, except for the results related to the tourists' continent of origin.

6. CONCLUSION AND IMPLICATIONS

This study set out to examine tourist behaviour in four- and five-star hotels in the Algarve region of Portugal, before and during the COVID-19 pandemic, and to assess the changes that occurred between these two periods. Based on an analysis of 17,601 online reviews from 93 hotels, covering the period from January 2018 to April 2023, the findings indicate that, regardless of the timeframe, Algarve's four- and five-star hotels were predominantly evaluated positively by guests. Although the volume of reviews declined sharply during the pandemic, the behaviour of those who continued to write reviews largely mirrored pre-pandemic patterns, with the notable exception of differences associated with tourists' continent of origin.

The results highlight that priority should be given to the quality of staff, breakfast, rooms, and pools, as these were the hotel attributes most frequently mentioned in the reviews across both periods. Hoteliers are therefore encouraged to invest primarily in human capital, since comments about staff exhibited the highest levels of very positive sentiment. Particular attention should also be devoted to hotel pools, which were the attribute most often associated with very negative comments in both periods.

On the other hand, hoteliers should focus on retaining and attracting customers who write online reviews in German, as well as those who travel between December and

February, as they tend to write the most positive reviews about the hotels, regardless of the period analysed. Additionally, guests travelling for business purposes and those who wrote an online review for the first time were the ones who left the most favourable comments about the hotels in both periods. It is important for hotels to offer business-oriented services and facilities, and for staff to encourage customers, particularly those who have never left a review, to write online reviews. Regarding the continent of origin, tourists from Asia wrote the most positive reviews about the hotels in the pre-pandemic period. However, with the onset of the pandemic, it was customers from Africa who took the lead in writing the most positive reviews. It is therefore important for hotel managers to develop strategies to attract clients from these markets.

From a different perspective, hoteliers should aim to provide a tailored service that meets the needs of French-speaking guests and those travelling with families. It is also vital to develop strategies to maintain high service quality during peak months (June to August), as these periods saw the highest number of negative reviews in both analysed timeframes. Additionally, although tourists from the African continent wrote the largest number of highly positive reviews during the pandemic, they were also responsible for the highest number of highly negative reviews in the pre-pandemic period. During the pandemic, however, tourists from Oceania generated the highest volume of highly negative reviews. Therefore, targeted measures should be implemented to cater to the specific needs and expectations of customers from these two markets.

This research offers both practical and theoretical contributions that can assist four- and five-star hospitality professionals in the Algarve region in gaining a deeper understanding of their customers' behaviour. Moreover, the study highlights the importance of leveraging text analysis software to examine online reviews, positioning it as a valuable tool for hotel businesses. It also provides industry professionals with insights on how to effectively capitalise on online reviews. Lastly, in addition to contributing to the existing scientific literature on online reviews, this study paves the way for future complementary research by academics.

As with any study, this research has limitations. Firstly, it focuses solely on one online review platform, namely TripAdvisor. Additionally, the study was restricted to reviews written in Portuguese, English, Spanish, German, and French about four- and five-star hotels in the Algarve region from January 2018 to April 2023, limiting the generalizability of findings. Another limitation concerns the division between the pre-pandemic and pandemic periods. In this study, the division was based on the dates of stay rather than the dates on which the reviews were published, since guests may post their reviews on TripAdvisor at any time, not necessarily close to the actual date of their stay, which could distort the temporal analysis. Nevertheless, TripAdvisor provides only the month and year of the stay, without specifying the day, which introduces a certain degree of ambiguity. For instance, stays that occurred on 30 and 31 January 2020 were classified as pre-pandemic, despite the WHO declaring the outbreak a Public Health Emergency of International Concern on 30 January. Similarly, it was not possible to include in the pandemic period those reviews from guests who stayed during the first five days of May 2023, given that the WHO declared the end of the PHEIC on 5 May 2023. Despite this limitation, the study remains relevant and makes a significant contribution to understand-

ing online hotel reviews. By incorporating a large number of evaluations, the analysis ensures that minor inaccuracies in period classification do not compromise the validity of the findings. Accordingly, the results continue to provide robust insights into tourist behaviour and perceptions during the pre-pandemic and pandemic periods. Future studies could validate and/or compare the results obtained in this research using other review platforms or employing different data mining techniques. Additionally, this study could be expanded to include other languages and regions of Portugal, as well as other countries and different hotel categories.

Further research could also focus on the post-pandemic period, analysing and comparing online reviews from the pre-pandemic, pandemic, and post-pandemic phases. Such an approach would make it possible to capture more recent trends in tourist behaviour and to identify whether the patterns observed during the pandemic have persisted, changed, or given rise to new consumption dynamics. Finally, a more targeted investigation into hotel attributes and the metadata of online reviews could be conducted to understand better the drivers of satisfaction and dissatisfaction within each market.

ACKNOWLEDGMENTS

This work is financed by national funds through FCT - Foundation for Science and Technology, IP, under the Projects CiTUR Doi.org/10.54499/UID/04470/2025, UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS - <https://doi.org/10.54499/UID/04152/2025> (2025-01-01/2028-12-31) and UID/PRR/04152/2025 <https://doi.org/10.54499/UID/PRR/04152/2025> (2025-01-01/ 2026-06-30).

AUTHORSHIP STATEMENT

The authors declare no conflict of interest. All authors contributed equally to all aspects of this research, including the conceptualisation of the study, methodology development, data collection and analysis, and preparation of the manuscript.

BIBLIOGRAPHICAL REFERENCES

- Ahani, A., Nilashi, M., Yadegaridehkordi, E., Sanzogni, L., Tarik, A. R., Knox, K., Samad, S., & Ibrahim, O. (2019). Revealing customers' satisfaction and preferences through online review analysis: The case of Canary Islands hotels. *Journal of Retailing and Consumer Services*, 51, 331–343. <https://doi.org/10.1016/j.jretconser.2019.06.014>
- Ahn, D., Park, H., & Yoo, B. (2017). Which group do you want to travel with? A study of rating differences among groups in online travel reviews. *Electronic Commerce Research and Applications*, 25, 105–114. <https://doi.org/10.1016/j.elerap.2017.09.001>
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., Hameed, H., & Alaa, M.

- (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 167, 114155. <https://doi.org/10.1016/j.eswa.2020.114155>
- Alecxe. (2025). *Scrapy Fake UserAgent*. *GitHub repository*. <https://github.com/alecxe/scrapy-fake-useragent>
- Amatulli, C., De Angelis, M., & Stoppani, A. (2019). Analyzing online reviews in hospitality: data-driven opportunities for predicting the sharing of negative emotional content. *Current Issues in Tourism*, 22(15), 1904–1917. <https://doi.org/10.1080/13683500.2019.1594723>
- António, N., Correia, M. B., & Perdigão Ribeiro, F. (2023). Understanding the impacts and motivations of duplicate reviews on TripAdvisor. *Cuadernos de Turismo*, 52(52), 219–238. <https://doi.org/10.6018/turismo.593611>
- Antonio, N., de Almeida, A., Nunes, L., Batista, F., & Ribeiro, R. (2018). Hotel online reviews: different languages, different opinions. *Information Technology & Tourism*, 18, 157–185. <https://doi.org/10.1007/s40558-018-0107-x>
- Athanasopoulou, P., Giovanis, A. N., & Ioakimidis, M. (2023). How Are Guests Satisfied? Exploring the Asymmetric Effects of Hotel Service Attributes on Customer Satisfaction by Analyzing Online Reviews. *Tourism*, 71, 8–28. <https://doi.org/10.37741/t.71.1.1>
- Ban, H.-J., Choi, H., Choi, E.-K., Lee, S., & Kim, H.-S. (2019). Investigating Key Attributes in Experience and Satisfaction of Hotel Customer Using Online Review Data. *Sustainability*, 11, 6570. <https://doi.org/10.3390/su11236570>
- Bi, J.-W., Zhu, X.-E., & Han, T.-Y. (2024). Text Analysis in Tourism and Hospitality: A Comprehensive Review. *Journal of Travel Research*, 1–23. <https://doi.org/10.1177/00472875241247318>
- Cheng, I. C. C., Ma, J., Ye, H., & Law, R. (2021). A Comparison of Hotel Guest Experience Before and During Pandemic: Evidence from Online Reviews. *Information and Communication Technologies in Tourism*, 549–555. <https://doi.org/10.1007/978-3-030-65785-7>
- Cibere, G. U., Basaran, M. A., & Kantarci, K. (2020). Evaluation of Hotel Performance Attributes Through Consumer Generated Reviews: The Case of Bratislava. *Advances in Hospitality and Tourism Research*, 8(1), 48–75. <https://doi.org/10.30519/ahtr.592312>
- Emir, O., & Bayer, R. U. (2017). Examination of Urban Hotels Online Review Site Performance with a Comparative Approach based on TripAdvisor and Booking. *Journal of Business Research - Turk*, 3(9), 72–85. <https://doi.org/10.20491/isarder.2017.288>
- Ferreira, A., Correia, M. B., & Renda, A. I. (2024). Systematic Literature Review on the Profile of Tourists in Four- and Five-star Hotels Based on Online Reviews. *European Journal of Tourism, Hospitality and Recreation*, 14(1), 20–34. <https://doi.org/10.2478/ejthr-2024-0002>
- Hu, F., Teichert, T., Deng, S., Liu, Y., & Zhou, G. (2021). Dealing with pandemics: An investigation of the effects of COVID-19 on customers' evaluations of hospitality services. *Tourism Management*, 85, 104320. <https://doi.org/10.1016/j.tourman.2021.104320>
- Instituto Nacional de Estatística. (2020). *Estatísticas do turismo - 2019*. INE. <https://www.ine.pt/xurl/pub/133574>

- Instituto Nacional de Estatística. (2021). *Estatísticas do turismo - 2020*. INE. <https://www.ine.pt/xurl/pub/280866098>
- Instituto Nacional de Estatística. (2022). *Estatísticas do turismo - 2021*. INE. <https://www.ine.pt/xurl/pub/22122921>
- Instituto Nacional de Estatística. (2023). *Estatísticas do turismo - 2022*. INE. <https://www.ine.pt/xurl/pub/72384938>
- Instituto Nacional de Estatística. (2024). *Estatísticas do Turismo - 2023*. INE. <https://www.ine.pt/xurl/pub/439494299>
- Instituto Nacional de Estatística. (2025). *Estatísticas do Turismo - 2024*. INE. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=66197834&PUBLICACOESmodo=2
- Kathunia, A., Delhi, I., Kaif, M., Arora, N., & Narotam, N. (2024). Sentiment Analysis Across Languages: Evaluation Before and After Machine Translation to English. *ArXiv*. <https://doi.org/10.48550/arXiv.2405.02887>
- Kim, B., Kim, S., & Heo, C. Y. (2016). Analysis of satisfiers and dissatisfiers in online hotel reviews on social media. *International Journal of Contemporary Hospitality Management*, 28(9), 1915–1936. <https://doi.org/10.1108/IJCHM-04-2015-0177>
- Kitsios, F., Kamariotou, M., Karanikolas, P., & Grigoroudis, E. (2021). Digital Marketing Platforms and Customer Satisfaction: Identifying eWOM Using Big Data and Text Mining. *Applied Sciences*, 11, 8032. <https://doi.org/10.3390/app11178032>
- Lee, H. (2024). Comparative analysis of data augmentation on sentiment classification. *Cornerstone: A Collection of Scholarly and Creative Works*. <https://cornerstone.lib.mnsu.edu/etds>
- Li, H., Ye, Q., & Law, R. (2013). Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784–802. <https://doi.org/10.1080/10941665.2012.708351>
- Liang, S., Schuckert, M., & Law, R. (2016). Multilevel Analysis of the Relationship Between Type of Travel, Online Ratings, and Management Response: Empirical Evidence from International Upscale Hotels. *Journal of Travel & Tourism Marketing*, 34, 239–256. <https://doi.org/10.1080/10548408.2016.1156613>
- Limberger, P. F., Anjos, F. A. dos, Meira, J. V. de S., & Anjos, S. J. G. dos. (2014). Satisfaction in hospitality on TripAdvisor.com: An analysis of the correlation between evaluation criteria and overall satisfaction. *Tourism & Management Studies*, 10(1). <https://www.tmstudies.net/index.php/ectms/article/view/648>
- Mozilla. (2025). *GeckoDriver*. Mozilla Developer Network. <https://firefox-source-docs.mozilla.org/testing/geckodriver/>
- Oliveira, A. S., Renda, A. I., Correia, M. B., & Antonio, N. (2022). Hotel customer segmentation and sentiment analysis through online reviews : an analysis of selected European markets. *Tourism & Management Studies*, 18(1), 29–40.
- Oliveira, Anderson S., Renda, A. I., & Correia, M. B. (2020). Online reviews: A pathway to improve hotel management. *Dos Algarves: A Multidisciplinary e-Journal*, 36, 108–126. <https://doi.org/10.18089/DAMeJ.2020.36.7>

- Organização Pan-Americana da Saúde. (2023). *Histórico da pandemia de COVID-19*. Organização Pan-Americana Da Saúde. <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>
- Pacheco, L. (2016). An analysis of online reviews by language groups: the case of hotels in Porto, Portugal. *European Journal of Tourism Research*, 14, 66–74. <https://doi.org/10.54055/ejtr.v14i.243>
- Pacheco, L. (2017). Customer Satisfaction in Portuguese Hotels: Evidence for Different Regions and Hotel Segments. *Tourism Analysis*, 22, 337–347. <https://doi.org/10.3727/108354217X14955605216087>
- Padma, P., & Ahn, J. (2020). Guest satisfaction & dissatisfaction in luxury hotels: An application of big data. *International Journal of Hospitality Management*, 84, 102318. <https://doi.org/10.1016/j.ijhm.2019.102318>
- Raedts, M., Roozen, I., & Dothee, C. (2023). The influence of the Covid-19 pandemic on guests' evaluations of hotel services. *Current Issues in Tourism*, 26(23), 3751–3756. <https://doi.org/10.1080/13683500.2022.2162374>
- Registo Nacional de Turismo. (2023). *Registo Nacional de Empreendimentos Turísticos*. Turismo de Portugal. https://rnt.turismodeportugal.pt/RNT/Pesquisa_ET.aspx?Filtro-Visivel=True
- Rhee, H. T., & Yang, S.-B. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior*, 50, 576–587. <https://doi.org/10.1016/j.chb.2015.02.069>
- Ríos-Martín, M. Á., Folgado-Fernández, J. A., Palos-Sánchez, P. R., & Castejón-Jiménez, P. (2019). The Impact of the Environmental Quality of Online Feedback and Satisfaction When Exploring the Critical Factors for Luxury Hotels. *Sustainability*, 12, 299. <https://doi.org/10.3390/su12010299>
- Schuckert, M., Liu, X., & Law, R. (2015). A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *International Journal of Hospitality Management*, 48, 143–149. <https://doi.org/10.1016/j.ijhm.2014.12.007>
- Selenium. (2025). *Selenium WebDriver*. Selenium. <https://www.selenium.dev/>
- Song, Y., Liu, K., Guo, L., Yang, Z., & Jin, M. (2022). Does hotel customer satisfaction change during the COVID-19? A perspective from online reviews. *Journal of Hospitality and Tourism Management*, 51(January), 132–138. <https://doi.org/10.1016/j.jhtm.2022.02.027>
- Srivastava, A., & Kumar, V. (2021). Hotel attributes and overall customer satisfaction: What did COVID-19 change? *Tourism Management Perspectives*, 40. <https://doi.org/10.1016/j.tmp.2021.100867>
- Tontini, G., Bento, G. dos S., Milbratz, T. C., Volles, B. K., & Ferrari, D. (2017). Exploring the nonlinear impact of critical incidents on customers' general evaluation of hospitality services. *International Journal of Hospitality Management*, 66, 106–116. <https://doi.org/10.1016/j.ijhm.2017.07.011>
- TravelBI. (2025a). *Dormidas*. TravelBI. <https://travelbi.turismodeportugal.pt/alojamento/dormidas-dashboard/>

- TravelBI. (2025b). *Emprego no Turismo por situação profissional e duração do trabalho*. TravelBI. <https://travelbi.turismodeportugal.pt/emprego-formacao/emprego-turismo-situacao-profissional-duracao-trabalho/>
- TravelBI. (2025c). *Oferta no Alojamento Turístico*. TravelBI. <https://travelbi.turismodeportugal.pt/alojamento/oferta-no-alojamento-turistico-dashboard/>
- TravelBI. (2025d). *Receitas do Turismo*. TravelBI. <https://travelbi.turismodeportugal.pt/turismo-em-portugal/receitas-do-turismo-dashboard/>
- TripAdvisor. (2025). *About TripAdvisor*. TripAdvisor. <https://tripadvisor.mediaroom.com/pt-about-us>
- Tsiotsou, R. H. (2019). Rate my firm: cultural differences in service evaluations. *Journal of Services Marketing*, 33(7), 815–836. <https://doi.org/10.1108/JSM-12-2018-0358>
- Tsiotsou, R. H. (2022). Identifying value-creating aspects in luxury hotel services via third-party online reviews: a cross-cultural study. *International Journal of Retail & Distribution Management*, 50(2), 183–205. <https://doi.org/10.1108/IJRDM-04-2021-0207>
- UNWTO. (2020). *Impact assessment of the COVID-19 outbreak on International Tourism*. <https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism>
- Wong, E., Rasoolimanesh, S. M., & Pahlevan Sharif, S. (2020). Using online travel agent platforms to determine factors influencing hotel guest satisfaction. *Journal of Hospitality and Tourism Technology*, 11(3), 425–445. <https://doi.org/10.1108/JHTT-07-2019-0099>
- World Economic Forum. (2024). *Travel & Tourism Development Index 2024*. <https://www.weforum.org/publications/travel-tourism-development-index-2024/>
- World Health Organization. (2023). *Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic*. World Health Organization. [https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic)
- Wu, M.-Y., Pearce, P., & Dong, W. (2017). How satisfying are Shanghai's superior hotels? The views of international tourists. *International Journal of Contemporary Hospitality Management*, 29(4), 1096–1115. <https://doi.org/10.1108/IJCHM-01-2015-0014>
- Xu, Y., Li, S., Law, R., Jin, Y., & Lyu, Z. (2023). How does the COVID-19 pandemic influence tourist rating behaviour? An empirical exploration based on expectation theory. *Current Issues in Tourism*, 26(24), 4052–4068. <https://doi.org/10.1080/13683500.2022.2157707>
- Ye, B. H., Luo, J. M., & Vu, H. Q. (2018). Spatial and temporal analysis of accommodation preference based on online reviews. *JOURNAL OF DESTINATION MARKETING & MANAGEMENT*, 9, 288–299. <https://doi.org/10.1016/j.jdmm.2018.03.001>
- Zhou, L., Ye, S., Pearce, P. L., & Wu, M.-Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1–10. <https://doi.org/10.1016/j.ijhm.2013.12.004>