Tourism Management 106 (2025) 104981

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Topic-based engagement analysis: Focusing on hotel industry Twitter accounts

Inmaculada Rabadán-Martín [a], Lucía Barcos-Redín [b], Jorge Pereira-Delgado [c], Francisco Aguado-Correa [a,*], Nuria Padilla-Garrido [d]

[a] Department of Financial Economics, Accounting and Operations Management, Faculty of Business Sciences and Tourism, University of Huelva, Plaza de la Merced, 11, 21071, Huelva, Spain
[b] Faculty of Economics and Business Administration, ICADE, Universidad Pontificia Comillas, Alberto Aguilera, 23, 28015, Madrid, Spain
[c] Telefónica Tech, Ronda de la Comunicación Oeste 1, 28050, Madrid, Spain
[d] Department of Economics, Faculty of Business Sciences and Tourism, University of Huelva, Plaza de la Merced, 11, 21071, Huelva, Spain

A R T I C L E I N F O

Keywords:
Firm-generated content
Twitter
Natural language processing
Topic modelling
Latent Dirichlet allocation
Engagement
Hotel

A B S T R A C T

In a social-media context, brands need to understand how to frame their messages, so that a topic can be quickly recognized, promoting higher levels of user engagement. However, knowledge about the link between content type and its engagement is not sufficiently studied. We first explore hotel Firm-Generated Content (FGC) and its inherent themes using topic modelling; we then use an *ad hoc* metric to investigate the engagement levels associated with each topic; thirdly, we compare the relevance attached to the topics with their engagement levels. In total, 44,448 corporate tweets from 62 hotel brands were analyzed to identify 14 topics, one of which had not previously been uncovered. Notably, there was a positive correlation with engagement for content related to hotel management activities and, among smaller groups, to sustainability. The results will expand FGC-related investigation within the hotel sector and will be of interest to firms seeking effective communication strategies.

1. Introduction

Social media have been turned into ideal platforms, so that firms may publish information on their products, launch promotions, and provide responses to their clients, targeting potential consumers more selectively (Kaplan & Haenlein, 2010; Wan & Ren, 2017). It is essentially marketing information, also referred to as Firm-Generated Content (FGC), that is gathered to try to improve knowledge of the brand, brand fidelity, client satisfaction, and purchasing intention (Brodie, Ilic, Juric, & Hollebeek, 2013; Cheng, Liu, Qi, & Wan, 2021; Colicev, Malshe, Pauwels, & O'Connor, 2018; de Vries, Gensler, & Leeflang, 2012; Kumar, Bezawada, Rishika, Janakiraman, & Kannan, 2016; Leung, Bai, & Stahura, 2015; Poulis, Rizomyliotis, & Konstantoulaki, 2019; Santiago, Borges-Tiago, & Tiago, 2022).

However, firms also seek to stimulate Customer Engagement through FGC (Brodie et al., 2013; Goh, Heng, & Lin, 2013; Hanson, Jiang, & Dahl, 2019; Harmeling, Moffett, Arnold, & Carlson, 2017; Kumar et al., 2016; Sheng, 2019), which is considered an indicator of marketing performance in social networks (Ji, Li, North, & Liu, 2017; Lee & Park, 2022). Customer Engagement has been widely studied through various lenses (Santos, Cheung, Coelho, & Rita, 2022), one of which establishes that user interaction in terms of likes, comments, and shares depends on the characteristics of the messages that each brand posts (Schultz, 2017). Among those characteristics, the volume (Meire, Hewett, Ballings, Kumar, & Van den Poel, 2019), the type of content (de Vries et al., 2012), richness (Shahbaznezhad, Dolan, & Rashidirad, 2021), vividness (Sabate, Berbegal-Mirabent, Cañabate, & Lebherz, 2014), interactivity (Schultz, 2017), the topic that is covered (Jalali & Papatla, 2019; Zhang, Moe, & Schweidel, 2017), and the congruency of the post with the identity of the brand (Montaguti, Valentini, & Vecchioni, 2023) have all been studied.

In this way and in the specific case of FGC published on Twitter (rebranded in the summer of 2023 to X) —one of the most popular microblogging platforms (Viñán-Ludeña & de Campos, 2022)—, brands are not only seeking to communicate with their followers, but are also trying to reach larger audiences more swiftly when, for example, the

* Corresponding author.
E-mail addresses: rabadan@uhu.es (I. Rabadán-Martín), lbarcos@comillas.edu (L. Barcos-Redín), jorge.pereiradelgado@telefonica.com (J. Pereira-Delgado), agucor@uhu.es (F. Aguado-Correa), padilla@uhu.es (N. Padilla-Garrido).

https://doi.org/10.1016/j.tourman.2024.104981
Received 4 April 2023; Received in revised form 6 May 2024; Accepted 10 June 2024
Available online 15 June 2024

followers are sharing or retweeting messages. However, users only retweet, if they recognize, within a few seconds, that the tweet covers a topic of interest for them (Jalali & Papatla, 2019; Zhang et al., 2017). Thus, it is crucial for a brand to be aware of how to write a tweet, so that users can quickly identify a topic. In this regard, according to the Uses and Gratifications theory -U&G- (Katz & Foulkes, 1962), if a brand wishes to attract an audience on social media, it must provide value, or gratification through its content to achieve greater customer engagement (Cvijikj & Michahelles, 2013; Dolan, Conduit, Frethey-Bentham, Fahy, & Goodman, 2019; Muntinga, Moorman, & Smit, 2011; Whiting & Williams, 2013). Nevertheless, there has been insufficient research on topic-related content (Lee & Park, 2022). The unstructured nature of the text included in the message may be added to the above, constituting one of the current challenges of modelling data on social networks (Zhang et al., 2017).

The hotel industry has not remained indifferent to the advance of social media (Luo & Zhong, 2015) and has recognized their fundamental role in both the attitudes and the booking intentions of guests (Leung et al., 2015), and both brand trust, and brand loyalty (Tatar & Eren-Erdoğmus, 2016). The use of social media has implied an adaptation in publicity, marketing, and promotions (Kim & Park, 2017; Taylor, Barber, & Deale, 2015), enabling, furthermore, a prompt response to crises in tourism, such as the COVID-19 global pandemic. In this latter case, hotel brands adjusted their social media content by generating new topics (Rabadán-Martín, Aguado-Correa, & Padilla-Garrido, 2023; Kwok, Lee, & Han, 2022; Mušanović, Dorčić, & Gregorić, 2023). However, the study of FGC in the hotel sector has been of limited scope (Kim, Park, & Kim, 2023), a majority of researchers having opted for the study of User-Generated Content (UGC) and consumer behaviour (Zarezadeh, Rastegar, & Xiang, 2022).

The topics covered within the area of hotel FGC have been sparsely studied. Most research in which those topics have been analyzed has employed manual content analysis (Aydin, 2020; Bonsón, Bednárová, & Wei, 2016; Leung, Bai, & Erdem, 2017; Minazzi & Lagrosen, 2013), proceeding to assign a single predefined topic to each message, based on the main theme or purpose of the communication: informative, entertaining, promotional, social, remunerative (Luarn et al., 2015; Menon et al., 2019), and emotional (Malhotra, Malhotra, & See, 2013).

However, it was never considered when assigning topics to messages in those studies that a single message may cover various themes. It is here that the topic modelling methodology comes into its own, situated within the field of natural language processing and machine learning, specifically employing the Latent Dirichlet Allocation (LDA) technique (Blei et al., 2003). This approach proves useful for decomposing messages into a set of topics, along with their corresponding weights, providing a detailed understanding of the underlying content structure (Calheiros, Moro, & Rita, 2017; Kim et al., 2023; Moro & Rita, 2018; Reisenbichler & Reutterer, 2019). However, despite the progress represented by topic modelling, its utilization as a technique to study FGC in hotels is very limited (Kim et al., 2023).

Added to which is the scant attention lent to the FGC effect on online consumer engagement of messages published on social media (Luarn et al., 2015; Osei-Frimpong & McLean, 2018) and, in concrete, on the engagement levels that the FGC-related topics can generate. That information is fundamental for firms searching for a better understanding of the factors that drive and that can above all improve the experience of client engagement (Aydin, 2020; Leung et al., 2017).

To the best of the authors' knowledge, no research has explored the engagement generated by the topics obtained through natural language processing in messages posted by tourism firms through social media. This is the focus of our study.

To do so, we operate under the premise that a brand message may contain more than one topic, that there is no necessity to employ predefined topics, and that topic modelling can assist with uncovering the topics. Given the exploratory nature of the study, four research questions were formulated:

RQ1. What are the topics and in what proportion are they contained in the Twitter messages that hotel chains post?

RQ2. What was the temporal evolution of the topics and the potential impact of COVID-19 on them?

RQ3. Which topics achieved greater/lesser engagement?

RQ4. To what extent is the proportion of a topic related to its level of engagement?

The contribution of this investigation is its: a) exploration of the underlying themes in the FGC of hotel chains through natural language processing; b) assessment of the weight and the temporal evolution of each topic within the messages, considering the impact of COVID-19; c) determination of the effectiveness of the content strategy through the assessment of the engagement levels that each topic generates using a new metric designed for that purpose, and d) comparison between brand-assigned topic importance and the engagement levels that each topic generates.

The rest of the paper is organized as follows. After a literature review, the research methodology and the main results of the study are described. The paper then closes with the conclusions, the theoretical and the practical implications, and the limitations of this research.

2. The importance of FGC and its use within the hotel industry

Firms continually use social media to commercialize their brands and products and to interact with consumers. In turn, consumers pore over FGC and UGC to make informed choices (Wan & Ren, 2017). FGC therefore plays a fundamental role in creating relevant and attractive content for the target public that can be understood as any form of information that the firm may communicate through its official social media platforms (Poulis et al., 2019).

When designing marketing in social networks, the difference is marked by appropriate content and its congruence with the brand, because it affects the way in which consumers will perceive the information in the posts (Montaguti et al., 2023; Wan & Ren, 2017). The idea behind FGC is therefore to provide topics for conversation with relational consequences for brands and consumers (Ceballos, Crespo, & Cousté, 2016). Throughout the literature, different classifications of the messages that firms send out can be found on the basis of message content: informative, entertaining, promotional, social, remunerative (Luarn et al., 2015; Menon et al., 2019), and emotional (Malhotra et al., 2013). Informative messages contain information on the firm, the brand and/or services, as well as important announcements for consumers. Messages for entertainment awaken interest among users, employing, for example, interesting images, photos, and videos. The promotional messages refer to various publicity campaigns, while the remunerative ones are linked to sales promotions. Social messages contain social and sporting events, open questions to fans on contemporary activities and humanitarian work. Finally, emotional posts seek emotional connections with the target audience, awakening feelings and reactions. However, in one of the most prominent works within this field, Kumar et al. (2016) summarized the types of FGC as follows: 1) promotional or informative messages; and 2) non-promotional messages.

With regard to the scarce few studies that have analyzed the type of content issued by hotels on social media (Table 1), differences can be found in relation to the categories of content type that are investigated, sample size, the social media that are studied, hotel typology, methodology in use, the measurement of engagement, and findings.

Commencing with the type of content and despite the differences that were found, the informative/promotional content related with the tourist establishment and the non-promotional content can be clearly distinguished. With the passage of time, studies have shown increasingly fewer differences between the number of promotional and non-promotional messages that hotels publish (Aydin, 2020; Leung et al., 2017).

Table 1
Summary of relevant studies on FGC in the hotel sector.

Author(s)	Content type	Sample	Methodology	Measurement of Engagement	Findings
Minazzi and Lagrosen (2013)	Promotion related, specific to hotels (29%); Promotion of destinations (28%); Contest related (16%); Factual information (15%).	15 hotel brands Facebook pages (3,981 posts)	Content analysis	Not studied	High propensity to use posts to promote hotels or the destinations where they are located and a low focus on the potential offered by interactions with users.
Cervellon and Galipienzo (2015)	Informational (68.6%); Emotional (31.4%).	7 Luxury hotel Facebook pages (379 posts)	Content analysis	Questionnaire	Informational content (compared to emotional content) positively impacts the attitude toward the brand name and the perception of quality.
Yoo and Lee (2015)	General hotel information (38%); User-generated contents (33%); Conversational message (13%); Hotel news (5%); Promotion (3%); Travel tips (3%); Entertainment (0.5%); Others (4.5%).	6 Heritage hotel Facebook pages (549 posts)	Content analysis	Likes, comments, and shares (individually)	Most hotels focus more on sending messages to their fans than engaging in conversations with them. No analysis of which type of content has higher engagement levels was offered in the study.
Bonsón et al. (2016)	Marketing (41.46%); Reply to Stakeholder (17.57%); Retweet (22.27%); General Information (16.63%); Social and Environmental Concern (1.70%); Employment (0.38%).	78 hotel groups Twitter account (3,705 tweets)	Content analysis Spearman coefficient/ Kruskal–Wallis test/ GLM	Likes, comments, and retweets	Content related to environmental, social and employment issues leads to greater audience engagement and could, therefore, positively influence the hotel's online awareness and presence.
Leung et al. (2017)	Product attributes/informational: Product (19.5%); Promotion (6.9%); Reward (16.2%)/Brand image-related/ emotional messages: Information not directly related to the hotel (19.9%); Brand (12.2%); Involvement (25.4%).	12 hotel brands Facebook pages (1,837 posts)	Content analysis Multivariate analysis of variance (MANOVA)	Likes, comments, and shares (individually)	Product, brand and involvement messages were more effective than information, reward and promotion messages. Hotel Facebook pages work best when referring to hotel branding, introducing new products, interacting with customers, sharing travel information and giving rewards.
Ferrer-Rosell, Martin-Fuentes, and Marine-Roig (2019)	Hotel services -other facilities, restaurants, rooms-; Tourism and destination -heritage, urban sites/events, nature/sport gastronomy- (ranking by weight of content).	189 hotel Facebook pages (4,725 posts)	Content analysis/ Compositional data analysis (CoDa)/ ANOVA	Not studied	Higher category hotels promote themselves by highlighting their facilities, while lower category hotels do so by promoting the destination.
Aydin (2020)	Promotion/event (22.6%); Congratulations/condolences (19%); Informative (facilities & services) (48.1%); News sharing (1.9%); Entertaining content (8.4%).	441 Five-star hotel Facebook pages (1,847 posts)	Content analysis/ Regression Analysis Results	Likes, comments, and shares	The appearance of celebratory messages (congratulations/condolences) as a type of content that leads to more interaction. This type of content is better in terms of creating interaction compared to promotions and event announcements and sharing general news or entertainment content.
Ferrer-Rosell et al. (2020)	Hotel services content (50.89%); Destination content (49.11%)/Hotel services; Tourism and destination; Emotional/personality attributes: modernity, entertainment, tradition and romanticism, relaxation (ranking in weight of content).	236 hotel Facebook pages (5,900 posts)	Content analysis Compositional data analysis (CoDa)/ ANOVA	Not studied	External (destination-level) vs. internal (hotel-level) content of the hotels is adapted to emphasize what can be stronger and what will better attract new clients. Differences are also found in the emotional attributes by content post both between hotels in different cities and between hotel classes.
Sari, Putra, Maemunah, and Dianawati (2022)	Promotion (10.28%); introducing products or services (70.15%); giving life-related advice and making simple daily conversations (16.05%); handling complaints (3.51%)	50 hotel Twitter accounts (5,438 tweets)	Content analysis	Not studied	Twitter pages were used most often for introducing and explaining a hotel's products and services, and it was rarely used for informing customers of sales promotions.
Kim et al. (2023)	One-way information giving: promotions and advertising (24.1%); Interactive communication (two-way communication): customer-relationship management (36.5%); Service recovery (two-way communication) (39.3%)	10 hotel chains Twitter accounts (175,358 tweets)	Topic modelling (Structural Topic Models, STM)	Not studied	Hotels try to share their latest news and marketing updates with the public. Luxury/high hotels focus on service recovery. High and upper-middle hotels tend to use Twitter to interact with customers interested in their hotels.
Mušanović et al. (2023)	Peak season (9 topics: hotel ambiance … hotel offer); Shoulder season (15 topics: fine dining experience … brunching); Off season (13 topics: content co-creation … special occasions)	17 Five-star hotel Facebook pages (5,032 posts)	Topic modelling (LDA)	Not studied	Hotel brands focused on promotion and offers or incentives to book. Prior to the pandemic, hotel brands communicated a lot of information on hotel facilities, ambiance, services, offers and cuisine. The important themes during the pandemic were trust, safety, and cordiality.

Most studies could be said to coincide, in so far as the same content types were identified, although new themes were also identified in some, such as "social and environmental concern" and "employment" (Bonsón et al., 2016), and "congratulations/condolences" (Aydin, 2020). Meta-dimensions or groups of content types were created before their content was analyzed in two studies (Ferrer-Rosell, Martin-Fuentes, & Marine-Roig, 2020; Leung et al., 2017), but the themes were never grouped later on in any studies, so as to conduct a more aggregated analysis, based on the topics discovered in the messages.

Regarding the methodology employed, 81.8% of the studies used manual content analysis, while the others applied topic modelling. Although both approaches presented limitations, a common criticism of manual content analysis, given that it depends on human effort, is that coding is a tedious and time-consuming task that requires experience with both content and coding. In addition, the interpretation of the results based on the qualitative analysis of content can be quite subjective and can introduce bias (Lee, Jang, Sun, Broadwell, & Yoon, 2022). In other words, manual analysis offers decontextualized results, as it reduces complex theoretical constructions to indices that are too general and over-simplified (Hannigan et al., 2019). In addition, the categories may not cover all relevant aspects of the messages. On the contrary, topic modelling has led to advances that are linked to the way that the topics are conceptualized and interpreted in the text-based data, because it facilitates the discovery of inherent or latent topics, through automatic classification, instead of imposing pre-established categories on the data, which also means that large volumes of data may be employed. In addition, topic modelling allows for assigning more than one topic to each message and determining their proportions.

The engagement of different types of content was not measured in all studies. Those that did measure it utilized likes, comments, and shares or retweets, employing various statistical techniques for that purpose. On the one hand, some studies grouped the engagement values for the entire sample (Aydin, 2020; Leung et al., 2017). Interactions between accounts with few and with many followers were equally valued with this approach. On the other hand, the study by Bonsón et al. (2016) did differentiate the brands' ability to generate engagement by weighting it based on the number of followers at a given moment. However, since the number of followers on social media can vary over time, it may impact the results, especially over an extensive study period.

The main findings can be summarized in terms of engagement, the proportion of topics, the temporal trends of changing topics, and the clustering of hotel brands.

Beginning with the measurement of engagement, the research findings indicated that promotional messages generated inadequate engagement in Facebook (Aydin, 2020; Leung et al., 2017) or Twitter (Bonsón et al., 2016). Those studies, in which a single topic was assigned to each message and all message-related engagement was attributed to that topic, revealed significantly different engagement levels according to the type of message content, suggesting a need to increase the content that was more appealing to the audience. However, no comparison between brand-assigned topic importance and the engagement levels that each topic generated were conducted in those studies. That is, they did not provide insight into the content strategy of hotels on social media, revealing whether differences existed between the importance assigned to each topic and its level of engagement.

Other findings from the studies were focused on the high proportion of posts used to promote hotels and hotel destinations through informative messages (Minazzi & Lagrosen, 2013). An approach that was considered effective a few years ago, because when potential clients were browsing the Facebook pages of the hotels, their primary motivation was to make a suitable choice (Cervellon & Galipienzo, 2015). However, it is necessary to contextualize these results and their interpretation, as the widespread use of social media over time has changed the ways that users engage with them. Sari et al. (2022) arrived at a similar conclusion, specifying that Twitter pages were rarely used for informing customers of sales promotions.

However, the temporal trends of changing topics were only found in the study of Mušanović et al. (2023). In their work, they differentiated between the content of messages posted before and after the COVID-19 pandemic; however, the proportional spread of the topics they had identified was not specified. An analysis of whether the hotels had increased or decreased the importance of each topic was not therefore feasible.

Finally, no disparities were discussed in the FGC studies on hotels, based on the size of hotel brands, which could be a significant variable in the identification of differing social media content type strategies. Nevertheless, differences relating to both content and the proportion of topics based on the hotel segment were observed (Ferrer-Rosell et al., 2019, 2020; Kim et al., 2023).

3. Topic modelling: treatment of non-structured text data

Topic modelling is one of the most valuable and powerful text-mining techniques for data extraction, the discovery of latent data, and the search for relations between data and text documents (Jelodar et al., 2019). The modelling of topics is mainly achieved through unsupervised learning algorithms that group non-labelled data by topics (Egger & Yu, 2022). A topic can therefore be seen as words within a group that share similar meanings with a high probability of being detected together. Topic modelling has been used in different fields of knowledge, covering such varied aspects as, for example, infodemiology relating to COVID-19 (Abd-Alrazaq, Alhuwail, Househ, Hamdi, & Shah, 2020), product planning (Jeong, Yoon, & Lee, 2019), human resources (De Mauro, Greco, Grimaldi, & Ritala, 2018), and climate change (Dahal, Kumar, & Li, 2019).

Among the set of tools for topic modelling, LDA (Dickinger, Lalicic, & Mazanec, 2017), an unsupervised generative probabilistic model for collections of discrete text data (Blei et al., 2003), is a basic and popular method that is especially useful for identifying hidden topics and themes based on word co-occurrences for each document in the corpus (Kim, Lee, & Park, 2021). Its principal advantage is that latent topics can be inferred from the structure with no need for pre-defined classes (Maier et al., 2018), given that the algorithm has no external information on the topics that are found and the documents are neither labelled by topic nor by key words.

"LDA assumes that each document can be represented as a probabilistic distribution over latent topics, and that topic distribution in all documents share a common Dirichlet prior. Each latent topic in the LDA model is also represented as a probabilistic distribution over words and the word distributions of topics share a common Dirichlet prior as well" (Jelodar et al., 2019, p. 15172). Unlike conventional analyses of word frequency, certain topics are extracted with this method while taking into account a broad variety of documental units, instead of considering frequency within each documental unit (Kim, Park, Barr, & Yun, 2019).

Nevertheless, LDA presents a series of limitations and challenges associated with its application in practice. Maier et al. (2018) mentioned that it was important to consider that the results of the model were not deterministic. Those authors then highlighted the need to conduct an appropriate pre-processing of the text collection, an adequate selection of several model hyperparameters, such as the number of topics, as well as the importance of evaluating model reliability, and valid interpretations of the topics that may be discovered. Besides, it has in some investigations been affirmed that generating semantic meaning from short texts can be a challenge for LDA and might not be optimal for those sorts of texts (Cheng, Yan, Lan, & Guo, 2014; Laureate, Buntine, & Linger, 2023). Nevertheless, 15 topic models were identified in the systematic review of Laureate et al. (2023) on 193 scientific papers, focused on the use of topic models for short text social media analysis. From among those 15, the most widely used was LDA, in 80.3% of all the articles under study, after which at some distance came Structural Topic Modelling (STM) and BiTerm Topic Modelling (BTM), with 6.7% and 4.7%, respectively.

However, and despite its utility, topic modelling has not yet gained widespread application in tourism research (Dickinger et al., 2017). Most of the investigation has been generated over the past five years and has in almost every way been centred on UGC. TripAdvisor predominates as a principal source of text data for analysis; fundamentally service and client satisfaction attributes. For example, topic modelling has therefore been used for studying UGC in relation to hotels and accommodation (Guo, Barnes, & Jia, 2017; Hu, Zhang, Gao, & Bose, 2019; Song, Liu, Guo, Yang, & Jin, 2022; Sutherland, Sim, Lee, Byun, &

Kiatkawsin, 2020), airlines (Lim & Lee, 2020), museums (Kirilenko, Stepchenkova, & Dai, 2021), restaurants (Jia, 2020; Liu, Yu, Mehraliyev, Hu, & Chen, 2022), tourism destinations (Wang, Li, Wu, & Wang, 2021), tourism attractions (Taecharungroj & Mathayomchan, 2019), travel itineraries (Vu et al., 2019), urban smart tourism (Brandt et al., 2017), and staycations or holistays (Muritala, Hernández-Lara, & Sánchez-Rebull, 2022).

As discussed in the previous section, the application of topic modelling to the FGC of hotel companies is testimonial, despite their utility for better understanding tendencies and patterns in the tourist market through the messages of firms posted on their respective social media accounts.

4. Online customer engagement

Customer engagement has been found to boost loyalty, trust, and brand evaluations. It is well known that in the search for greater engagement, informative content facilitates positive consumer commitment (Pansari & Kumar, 2017), in particular when it is a matter of developing positive brand-related attitudes (Colicev et al., 2018; Kumar et al., 2016) and the improvement of consumer feeling (Meire et al., 2019). In turn, emotional content seeks to improve states of mind and presents greater probabilities of being commented upon and posted between users (Akpinar & Berger, 2017; Coelho, De Oliveira, & De Almeida, 2016; Luarn et al., 2015). In addition to the type of content, the platform employed to share the information can also influence customer engagement (Shahbaznezhad et al., 2021).

In the hyper-competitive tourism sector, the way that brands can utilize social media to increase client engagement is a key question (Harrigan, Evers, Miles, & Daly, 2017). However, as Cheung, Shen, Lee, and Chan (2015) mentioned, the interpretation of customer engagement is still a rather vague and controversial matter. In that sense, the systemic literature review that Hao (2020) completed on customer engagement involving 173 articles within the field of hospitality and tourism, discovered that the concept of customer engagement was expressed in 25 different terms, twenty-seven definitions, and four sub-forms. In that review, Hao (2020) also identified online customer engagement as one of the four sub-forms of customer engagement and defined it as “customers’ engagement with firms’ online platforms, ***i.e.***, a Social Network Site (SNS), booking websites, virtual communities, online travel agent and mobile App” (Hao, 2020, p. 1845). In turn, Pino et al. (2019, p. 4) defined online engagement as “expressing favour (“like”) or disfavour (“dislike”) toward a given object (***e.g.***, a tourist event or a destination); better articulating a users’ position with respect to that object (through comments), or reaching other people (by sharing online contents)”.

In the studies centred on online customer engagement included in the review of Hao (2020), data were primarily collected through posts from online platforms, such as Facebook (Aydin, 2020; Hashim & Fadhil, 2017; Lei, Pratt, & Wang, 2017; Menon et al., 2019; Pino et al., 2019; Yang, Lin, Carlson, & Ross, 2016; Yoo & Lee, 2015), Twitter (Menon et al., 2019; Pino et al., 2019; Sashi, Brynildsen, & Bilgihan, 2019), and TripAdvisor (Torres & Singh, 2016). From among the total of 173 studies that were identified, 37 were on online customer engagement, in 33 of which online customer engagement was measured as a unidimensional behavioural construct. In particular, engagement with an SNS was mainly measured as the number of likes, following, comments and (re) posts (Hao, 2020).

The measurement of online engagement has been the object of various approaches and proposals (Buhalis & Mamalakis, 2015; Muñoz-Expósito, Oviedo-García, & Castellanos-Verdugo, 2017; Oviedo-García, Muñoz-Expósito, Castellanos-Verdugo, & Sancho-Mejías, 2014; van Doorn et al., 2010; Vivek, Beatty, & Morgan, 2012). Given the varied treatment of reactions and interactions, the interactions were separately analyzed in some studies on FGC (Bonsón & Ratkai, 2013; Coelho et al., 2016; Lee, Hosanagar, & Nair, 2018; Lei et al., 2017; Luarn et al., 2015; Menon et al., 2019), evaluating the type of content that generated more “likes”, “comments”, “shares”, “replies”, and “retweets”. In other cases, total engagement was calculated as a grouping of the values “likes”, “comments”, “shares”, “replies”, and “retweets”, either as an arithmetical sum (Aydin, 2020; Bonsón et al., 2016; Buhalis & Mamalakis, 2015; Mariani, Mura, & Di Felice, 2018; Pino et al., 2019; Yang et al., 2016), or as a weighting of each type of interaction according to the degree of implication of the followers (Mariani et al., 2018; Muñoz-Expósito et al., 2017). Despite the different approaches, the various authors coincided over the relevance of engagement and the need to measure it as an indicator of the effectiveness of the actions that formed part of the FGC strategy.

5. Methodology

The idea behind the methodology followed in this exploratory study is to identify the topics upon which the FGC of the large Spanish hotel groups is focused when their messages are shared through the Twitter social network, and to analyse the engagement generated by each topic. To do so, five stages were completed: data collection, text preprocessing, topic modelling, engagement, and content type strategy (Fig. 1).

5.1. Data collection

Twitter was selected as a data source for three fundamental reasons: firstly, because it is a very popular social network among firms for information sharing, commercial sales, and engagement (Chae, 2015); secondly, it is utilized more than any other in research focused on data generated in social media (Zachlod, Samuel, Ochsner, & Werthmüller, 2022); and thirdly, at the time of the study, Twitter provided an Application Programming Interface (API) for Academic Research, through which academic researchers could access precise and comprehensive Twitter data, both historical and real-time. The specific software package, academictwitteR, was employed to download the tweets (Barrie & Ho, 2021).

Alongside the text of the messages, the API provided the following metadata: the date on which the message was sent, the language, the typology, and the number of reactions to the message. Regarding typology, Twitter categorized messages as general tweets (a message posted on Twitter containing text, photos, a GIF, and/or video), retweets (re-posting of a tweet), replies (a reply to someone else’s tweet), and quote tweets (tweeting another person’s Tweet and adding a comment). Reactions to the message referred to the number of retweets, likes, quotes, and replies that it had received from other users.

Data were collected using purposive sampling to include the largest hotel brands in Spain, a country that is globally ranked second with regard to the number of annual visitors (UNWTO World Tourism Organization, 2023), third in the Travel & Tourism Development Index ranking (World Economic Forum, 2022), fifth in the UNESCO’s World Heritage Site list (UNESCO World Heritage Convention, n.d.), and first in Biosphere Reserves (UNESCO, n.d.). Specifically, the 50 hotel groups with the largest numbers of rooms in Spain in 2019 listed on the Hosteltur ranking of large hotel chains (Hosteltur, 2020) were selected. After inspecting the official websites of these groups, one by one, the different brands that they advertised on those websites were identified. In all, 62 brands (57.4%) were selected, which were the ones that had their own Twitter accounts and that, in addition, published tweets in Spanish. Those 62 brands comprised our initial sample, which formed part of 44 hotel groups and implied an offer close to 278,000 hotel rooms.

After collecting the 62 official Twitter accounts, the tweets exclusively published by the hotel brands themselves (FGC) in the Spanish language were downloaded. The data were collected between January 1, 2018, and May 21, 2021, resulting in the retrieval of 44,448 valid messages (average number of corporate tweets per day and brand: 0.58).

DATA COLLECTION	Tweets download
TEXT PREPROCESSING	Tokenization Cleaning and unification of text Stop words removal Part-of-speech tagging algorithm to retain only nouns and adjectives Lemmatization Filtering tokens that appear either too often or too rarely Identification of n-grams
TOPIC MODELLING	Topic modelling (LDA) Topic evolution (Time graph)
ENGAGEMENT	Analysis of engagement
CONTENT TYPE STRATEGY	Weight-Engagement matrix Statistical correlation between the weight and the engagement

Fig. 1. Data collection and analysis process.

Importantly, UGC was not included in any instance.

5.2. Text pre-processing

In second place, and with the purpose of creating a cleaner corpus, the text of all the messages underwent pre-processing, with the help of Python libraries, which included the habitual procedures of tokenization, lemmatization, n-grams identification, and stopword removal (Heydt, 2018; Kirilenko et al., 2021; Laureate et al., 2023; Maier et al., 2018; Manning, Raghavan, & Schütze, 2008; Vázquez, Pereira-Delgado, Cid-Sueiro, & Arenas-García, 2022). In concrete, the following steps were followed:

a) Tokenization with removal of short (fewer than 4 symbols) and long (more than 25 symbols) tokens.
b) Removal of mentions, hashtags, punctuation symbols, numbers, emoticons, and URLs.
c) Transformation of words to lower case and removal of accents for the purpose of term unification. Additionally, adjustment of some particularities of the Spanish language (*i.e.*, the letter ñ and its tilde accent).
d) Removal of Spanish stopwords with the NLTK library (Loper & Bird, 2002), one of the most widely used in the academic literature (Laureate et al., 2023). Additionally, we created a corpus-specific stopword list, which included such tokens as: the names of the establishments, the days of the week or month, and the names of the brands (hotel chains) comprising the sample.
e) Part of speech tagging with the spaCy library (Honnibal & Montani, 2017), retaining nouns and adjectives, which are often the most informative parts of speech in text.
f) Application of a lemmatization algorithm with the purpose of converting the tokens to their dictionary forms (or lemmas), so that the inflected words were comparable with each other.
g) Removal of tokens that appeared either too often or too rarely in the corpus, as they could not be used to discriminate between the topics. Those tokens that appeared in more than 60% of the documents or in less than 10 documents were excluded.
h) Identification of n-grams with the NLTK library (Loper & Bird, 2002) that needed to be treated as an entity (*e.g.*, Valentine's Day).

5.3. Topic modelling

Given the nature of our data, the performance of the LDA and the BTM algorithms were compared to perform the topic modelling. The first, because it is one of the most widely used in studies with short texts on social media, and the second, despite it being scarcely used in those studies, because it was specifically designed to analyse short texts such as Twitter messages (Laureate et al., 2023). The fact that LDA yielded results of greater coherence tipped the scales in favour of that algorithm.

Taking a corpus as an input, LDA provides two important outputs (Vázquez et al., 2022): 1) The uncovered topics, each defined by a weighted list of characteristic terms (the *per*-topic word distribution); 2) The proportion of the document that is assigned to each of those topics (the *per*-document topic distribution).

The LDA algorithm requires specification of the K hyper-parameter that refers to the number of topics to discover, as well as to adjust the hyper-parameters to corpus level, alpha (α), and beta (β). α corresponds to the Dirichlet distribution parameter built into the algorithm to extract topic distribution by document. This hyper-parameter controls the density of topics by document, in other words, whether a document is composed of a small or a large number of topics. The algorithm also uses another concentration parameter, β, which controls the topic-word density, in other words, whether the topics are characterized by a smaller or greater number of terms. In addition, it should be taken into account that the choice of those three hyper-parameters can affect the performance of the algorithm (Vázquez et al., 2022).

Two metrics are commonly used, in order to adjust these hyper-parameters and to compare the performance of the LDA topic models: perplexity and coherence (Kirilenko et al., 2021; Vázquez et al., 2022). However, better perplexity will not always be related with greater human interpretability of the topics (Chang, Boyd-Graber, Wang, Gerrish, & Blei, 2009). For that reason and given that it is one of the most widely used metrics for topic model evaluation, it was decided to use coherence. Topic coherence measures the semantic similarity between the words with the highest probability of belonging to a particular topic, and, according to Röder, Both, and Hinneburg (2015), the CV coherence metric outperforms all others. So, the Gensim library (Řehůřek & Sojka, 2011) was chosen, which offers several coherence measures, including the CV metric (Laureate et al., 2023).

We then used the MALLET topic modelling toolkit library (McCallum, 2002) to run the LDA algorithm, as it is a very efficient tool for that purpose. The library uses an optimised Gibbs sampling algorithm (Yao, Mimno, & McCallum, 2009). Different hyper-parameter values were explored for K and α (Table 2), and the CV coherence score was calculated for each combination. Furthermore, and given the stochastic nature of the LDA process, we executed 10 runs of the algorithm for each combination, generating a total of 720 models. The library recommended default value of 0.01 for the β hyper-parameter was used.

Fig. 2 shows the coherence score for each of the 720 generated models as a function of the values of K and α. On average, the best results were obtained for a value of $\alpha = 5$ and a value of $K = 14$. Nevertheless, following Maier et al. (2018), those quantitative results were combined with intersubjective qualitative human judgment, so as to select an appropriate combination of hyperparameters.

The 10 models with the highest coherence values were selected for further research, all of which with a coherence score higher than 0.44. Those 10 candidate models corresponded to K values within a range of 10–28 topics, thereby covering different levels of granularity. For each of the 10 models, three researchers individually reviewed the 20 most representative words (principal words) in each topic and assigned a provisional label to it that described the underlying content of each topic in the most concise possible way (Maier et al., 2018).

Through a discussion group and taking topic interpretability as a criterion, the three researchers arrived at a consensus, selecting the second repetition of the model that included 14 topics with an α of 5. The inter-rater agreement for label coding results for the selected model was 83%, a value considered to be satisfactory (Kassarjian, 1977). Discrepancies observed in the coding were resolved by the researchers through a second simultaneous review of the messages with the highest loading on each topic. The same procedure was followed to ensure that the labels of the 14 topics described the substantive content of the topic.

Having selected the model and labelled the topics, the weight (θ_{ik}) of each of the 14 established topics (k) was compiled for each of the 44,448 tweets (i) comprising the sample. It is important to note that LDA assumes that each message is a combination of contents, and the sum of their weights is equal to 1.

Subsequently, the average weights of each topic for each of the 62 hotel brands (u) in the analysis were calculated:

$$Mean\_Brand\_Topic_{uk} = W_{uk} = \frac{\sum_{i \in I_u} \theta_{ik}}{n_u}$$

where I_u refers to the set of tweets that each hotel brand posted on its Twitter account, and n_u refers to the number of tweets in that set.

Having determined W_{uk}, a measure of topic concentration could also be calculated, as with the study conducted by Ceballos et al. (2016). Specifically, the Herfindahl-Hirschman Index (HHI) was utilized, which indicates whether the brands were focused or not on a few topics using the following expression:

$$HHI_u = W_{u1}^2 + W_{u2}^2 + \cdots + W_{u14}^2$$

The maximum value of the index was 1 (all messages were centred

Table 2
LDA hyperparameter values.

Hyper-parameters	Values
Number of topics (K)	6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40
Concentration hyper-parameter α	0.5, 1, 5, 10

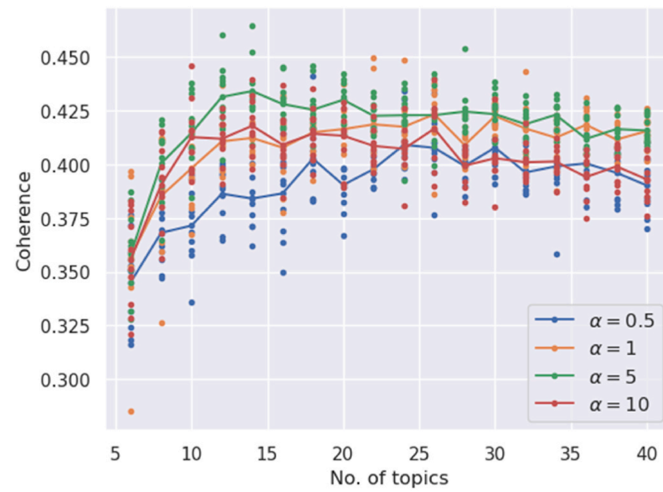

Fig. 2. Coherence of the models under analysis.

around a single topic) and the minimum was 0.0714 (the messages were evenly distributed among all 14 topics).

Subsequently, the 14 topics were then grouped, in accordance with their themes, into 6 large categories or meta-dimensions that the authors had defined, so as to conduct a more aggregated analysis (Enjoyment, Hotel management, Destination, Hotel features, Offers, and Entertainment). Finally, the evolution of the weight of each topic over time was depicted in a graph.

5.4. Engagement

The engagement analysis was based on the reactions to each message obtained through the Twitter API: number of replies, likes, retweets, quotes. We centred our attention on the messages that the brands actually composed (general tweets and replies), given that the retweets (5,453) and the quote tweets (916), which were also posted on the account of the brand, included content prepared by third parties, which might have distorted the levels of engagement relating to the original FGC. In particular, after excluding those last two, the sample for this analysis amounted to 38,079 messages.

In the first place, the engagement generated by each tweet (E_i) was quantified as the sum of the number of reactions (replies, likes, retweets, quotes) that it provoked:

$$E_i = \#replies_i + \#likes_i + \#retweets_i + \#quotes_i$$

where i refers to each tweet forming the sample. As from that point, the average level of reactions *per* tweet was calculated for each of the 62 Twitter accounts of the sample, thereby obtaining the average level of engagement for each of the hotel brand accounts:

$$Mean\_Brand\_Eng_u = \frac{\sum_{i \in I_u} E_i}{n_u}$$

where I_u refers to the set of tweets that each hotel brand posted on its Twitter account, and n_u refers to the number of tweets in that set.

The above calculation provided a global average of the reactions for each brand, without taking the topic into account. Therefore, the next step was to evaluate the impact generated by each topic for each of the 62 brands. To do so, the *per*-document topic distribution, provided by the LDA algorithm was used, *i.e.*, the proportion of the document or topic weight that was assigned to each of the k topics in that document. The engagement generated by a tweet was distributed between the k topics in a proportional way to the weight of each topic in each tweet and the average engagement associated with each topic within a brand Twitter Account was calculated as:

$$Mean\_Brand\_TopicEng_{uk} = \frac{\sum_{i \in I_u} E_i \bullet \theta_{ik}}{\sum_{i \in I_u} \theta_{ik}}$$

where θ_{ik} corresponds to the weight assigned to topic k in each one of the i tweets posted by brand u.

The Engagement Ratio, ER_{uk}, generated by each topic was calculated for each of the brands, with the purpose of evaluating the impact of each topic in relation to the average of each brand:

$$ER_{uk} = \frac{Mean\_Brand\_TopicEng_{uk}}{Mean\_Brand\_Eng_u}$$

In this way, an ER_{uk} greater than 1 meant that, for account u, topic k had generated a greater response than the global average for that account, and higher than 1, the greater the impact of the topic. Likewise, a result lower than 1 represented a topic that generated little engagement in relation to the global average of the account, and if equal to 1, it was a topic that had provoked a response equal to the average. For example, an ER value for a certain brand, u, and a topic, k, equal to 1.8, means that the engagement generated by topic k is 80% greater than the global average of the brand, whereas an ER equal to 0.3 means that the engagement generated by topic k is 70% lower than the global average.

This average in relation to the global average level of engagement was sufficient to compare the topic engagement levels for different brands with very different global engagement levels, due to such factors as having a substantially different number of followers, or a different volume of daily messages. These ER_{uk} were represented in a heatmap to facilitate their analysis.

The engagement that was globally generated by each of the k topics within the whole sample was evaluated through the calculation of the weighted average of the ER_{uk}, both for each month and throughout the whole study, taking as weights the number of tweets sent through each of the accounts.

$$Global\_Topic\_ER_k = ER_k = \frac{\sum_{u \in U} ER_{uk} \bullet n_u}{\sum_{u \in U} n_u}$$

where U is the set of the 62 brands that comprised the sample and n_u is the number of tweets published by brand u.

5.5. Content type strategy analysis

Through the analysis of the content type strategy, the aim was to establish a relationship between the weight of each topic and the level of engagement it generated. In this context, the goal was to verify whether the topics to which brands assigned more weight were the ones that, in turn, had higher engagement and *vice versa*.

To do so, having determined the global engagement ratio per topic (ER_k), the global weight ratio of each topic (WR_k) was calculated using the same sample employed to determine engagement (38,079 messages). Thus, the Weight Ratio, WR_{uk}, was calculated for each topic and brand as follows:

$$WR_{uk} = \frac{Mean\_Brand\_TopicW_{uk}}{Mean\_Brand\_TopicW_u}$$

where $Mean\_Brand\_TopicW_u$ is always equal to 0.0714 (the result of dividing 1 by 14), as the sum of the weights of the 14 topics is, by definition, always equal to 1.

Finally, the global ratio for each topic was calculated, taking into consideration the number of tweets published per brand:

$$Global\_Topic\_WR_k = WR_k = \frac{\sum_{u \in U} WR_{uk} \bullet n_u}{\sum_{u \in U} n_u}$$

where U is the set of the 62 brands that comprised the sample and n_u is the number of tweets published by brand u. In this case, a value higher than 1 signifies that the hotel brands included in the sample assigns greater importance to the topic, as it takes values above the average.

The global weight ratio of the topic (WR_k) compared with its global engagement ratio (ER_k) yielded the quadrants within a Weight-Engagement matrix:

- High WR_k and High ER_k (HH): The topics located within this quadrant were assigned above-average importance by the brands and user engagement levels with those topics were also above-average.
- Low WR_k and Low ER_k (LL): This quadrant included topics where both the weight and engagement were below average.
- High WR_k and Low ER_k (HL): Topics within this quadrant were assigned above-average weight, but users rated them with below-average engagement.
- Low WR_k and High ER_k (LH): Topics in this quadrant had below-average weights, while their engagement was above average.

5.6. Correlation between the proportion and the mean engagement value of each topic

Finally, the potential correlation between the mean engagement value of each topic, ER_{uk}, and the proportion of each topic, W_{uk}, was analyzed. Specifically, a global analysis was conducted with the 62 brands. Similarly, correlation analysis was conducted for small, medium, and large-sized brands (Annex I). Given that our data did not adhere to a normal distribution and included outliers, the Spearman's rank correlation coefficient was utilized.

6. Results

6.1. Types of messages

Table 3 shows a description of the 44,448 messages sorted by type. In all, 88.5% of the tweets published by the hotel brands implied one-way information giving (74.2% were original and 14.3% referred to other messages -retweets or quote-), whereas 11.5% reflected two-way content, with responses to users (replies).

6.2. Topic discovery and visualization

The application of LDA to the 44,448 messages was sufficient to differentiate 14 latent topics, with their respective weights (Table 4). A label was attached to each topic and was placed in one of the 6 general categories or meta-dimensions defined for that purpose.

The average for the global weights of the different topics was situated at 7.14%, Topic 12 (Customer service/Thanks) presenting the highest value at 7.96% and Topic 14 (Events), the lowest value with 6.65%. By categories, the most exploited was "Enjoyment" (28.14%) that

Table 3
Description of the messages forming the sample.

Type	Frequency (%)	Example
General tweets	32,963 (74.2)	To celebrate International Beer Day, we invite you to play a game with us. Find your favourite beer partner and give us your answer!
Retweets	5,453 (12.3)	RT @sevillaciudad: The historic centre of #Sevilla, almost four-square kilometres, the largest in Spain and one of the largest in …
Replies	5,116 (11.5)	@user Thanks for your feedback and see you soon.
Quote tweets	916 (2.1)	We are ready for the fourth edition of @MallorcaLiveFestival in #NewMagaluf! The biggest music event on the island will offer a quality experience for tourists and locals. More info here.

Table 4
List of topics, words and categories.

Topic	Label	Top 10 words	Weight	Ranking by weight	Category
1	Management	tourism, sector, manager, new, president, Spain, hotelier, CEO, today, forum	0.0695	10	Hotel management
2	Family/Festive	day, year, happy, family, today, kid, everything, activity, Christmas, festive	0.0701	8	Enjoyment
3	Discount	offer, reservation, discount, special, night, stay, breakfast, web, price, room	0.0781	2	Offers
4	Travel	destination, place, city, travel, perfect, history, getaway, natural, world, best	0.0692	11	Enjoyment
5	Sustainability	year, world, guest, team, person, part, best, hotelier, project, initiative	0.0721	5	Hotel management
6	City	city, new, Madrid, centre, great, first, via, Barcelona, door, capital	0.0685	12	Destination
7	Coast	view, beach, sea, pool, coast, spa, terrace, sun, resort, spectacular	0.0749	3	Destination
8	Experience/Safety	experience, everything, new, vacation, moment, safety, security, measure, guest, maximum	0.0698	9	Enjoyment
9	Contest	stay, night, photo, award, day, contest, new, raffle, tomorrow, edition	0.0677	13	Entertainment
10	Holiday	day, best, beach, holiday, great, Majorca, next, summer, destination	0.0723	4	Enjoyment
11	Culture	history, art, city, unique, visit, palace, work-of-art, century, experience, old	0.0703	7	Destination
12	Customer service/ Thanks	thank you, much, greetings, please, information, reservation, possible, email, comment, detail	0.0796	1	Hotel management
13	Gastronomy	gastro, restaurant, menu, kitchen, best, dish, chef, dinner, bar, list	0.0713	6	Hotel features
14	Events	day, perfect, event, best, space, golf, place, service, room, sport	0.0665	14	Hotel features

encompassed Topics 2 (Family/Festive), 4 (Travel), 8 (Experience/ Safety), and 10 (Holiday), closely followed by "Hotel management" (22.12%), and "Destination" (21.37%).

Together with the global weight of each topic for the whole sample, the weight that each topic reached for each message and brand in the study was also analyzed. Thus, for example, in the first tweet of Table 5, Topics 7 and 10 stood out, meaning that the message predominantly featured combined content of "Coast" and "Holiday". Similarly, in the second tweet, Topics 8 and 11 were prominent, relating to "Experience/ Safety" and "Culture".

Important differences were found in the weights of the topics according to the brand that was analyzed. As an example, Fig. 3 covers the brands with more than 300 tweets in the period under study, observing that the communication strategies of some brands were centred more on a few topics, while other brands to a greater or lesser extent diversified their content. Specifically, it was the Hotusa brand that achieved the highest concentration index (HHI_u), with a value of 0.11, by focusing on the topic of "Management" (0.257). A topic that includes the concept of hotel management, both internally and externally, as well as the opportunities and threats affecting the hotel sector. That is, it encompasses, among others, themes of tourist management and territorial development, given their impact on the hotel sector.

Specifically, the Hotusa group promoted forums on tourist innovation as part of its approach to hotel management, with debates and roundtable discussions on the future of tourism, publishing the main ideas and conclusions from those forums on Twitter. In particular, "Management" had a weight of over 50% in 516 out of the 981 messages analyzed for this brand, as exemplified by:

> "We celebrated for the first time at the Elite World Istanbul Hotel 5*, the Activate event, with the participation of representatives of Hotusa Hotels, trainings about hotel marketing, the latest trends in online marketing and the GDS market".

> "Jose Theotonio, CEO of @pestanahotels, highlights the importance of digitalization for hotel management and product promotion. #HotusaExplore".

> "*We begin the first debate of #HotusaExplora: Tourism as a strategic driver for economic development. With @VECI_Esp @CdEconomia @GGuevaraM @WTTC. Moderated by @bieitorubido, director of @abc_es.*"

6.3. Topic evolution and possible effects of COVID-19

Fig. 4 shows the evolution of average weight by month for each topic, thereby reflecting in which periods a topic acquired greater or lesser relevance. The significative effects of the declaration of the COVID-19 pandemic (marked with a discontinuous red line in Fig. 4) and the closure of accommodation establishments within Spain between March and May 2020 may be appreciated. Specifically, the most negatively affected topic was related with the promotion of events (Topic 14), which in 2021 had not recovered the weight that it had in the years leading up to the pandemic. Nevertheless, if attention is paid to the maximum weights, Topics 2 (Family/Festive), 3 (Discount), 5 (Sustainability), 7 (Coast), 8 (Experience/Safety), 9 (Contest), 11 (Culture), and 12 (Customer service/Thanks) rose to higher peaks after rather than before the outbreak of the pandemic in Spain.

In general, after the inflection point marked by COVID-19, the brands placed emphasis on stimulating a desire to travel, strengthening the categories of "Enjoyment" (Topics 2, 4, 8, and 10), "Destination" (Topics 6, 7, and 11), "Offers" (Topic 3), and "Entertainment" (Topic 9).

6.4. Engagement

Fig. 5 shows the engagement, calculated as $Global\_Topic\_ER_k$, generated by each of the 14 topics on a heatmap ($Global\_Topic\_ER_k$ appears between parentheses in the part below the horizontal axis that refers to each of the topics). In global terms, the topic that generated higher engagement was T5 (Sustainability), with a value of 1.162, followed by T9 (Contest; 1.137), and T1 (Management; 1.076). In the case of T5, for instance, this topic therefore generated an engagement level that was on average 16.2% higher than the average value for each brand. Topic 12 (Customer Service/Thanks) presented the lowest engagement level, with a value of 0.755, considering, in addition, that it was the topic with the highest concentration of reply-type tweets, whose likes, comments, and retweets were at very low levels.

Besides, engagement was represented by topic (calculated in terms of ER_{uk}) for the 27 brands with a higher number of messages than 300, as each Twitter profile might have followers with different characteristics, in accordance with its hotel offer. Thus, a white colour on the heatmap reflects ER_{uk} values close to 1, green reflects ER_{uk} values higher than 1, and red, values below 1. So, for example, in the case of the "Paradores" account and Topic 9 (Contest), the heatmap shows a moderately intense green colour, which corresponds to a value of $ER = 1.34$. It means that the engagement generated by Topic 9 was, in this case, 34% higher than the average engagement with "Paradores". Likewise, for Topic 3 (Discount), the heatmap shows a moderately intense red colour, which corresponds to a value of $ER = 0.82$. It means that the engagement generated by Topic 3 was 18% lower than the average engagement with "Paradores". Thus, considering the great weight that "Paradores"

Table 5
Weight of each of the 14 topics in various examples of tweets.

Brand	Tweet	Topic 1 Manag.	Topic 2 Family/ Festive	Topic 3 Discount	Topic 4 Travel	Topic 5 Sustain.	Topic 6 City	Topic 7 Coast	Topic 8 Exper./ Safety	Topic 9 Contest	Topic 10 Holiday	Topic 11 Culture	Topic 12 Cust. serv./ Thanks	Topic 13 Gastro.	Topic 14 Events	Sum
Hoteles Globales	Need some travel inspiration? This is Teresitas beach, in Tenerife #Tenerife! #islascanarias #canarias #wanderlust #instatravel	0.027	0.027	0.027	0.027	0.027	0.027	**0.258**	0.027	0.027	**0.412**	0.027	0.027	0.027	0.027	1
Paradores	A medieval 12th-century castle, set within an Arab fortress. A dream come true. And if you think about it, you can stay here! An unparalleled experience with nine centuries of history in its stones.	0.071	0.019	0.019	0.019	0.019	0.019	0.019	**0.282**	0.019	0.019	**0.440**	0.019	0.019	0.019	1
Hoteles santos	The @HotelPortaFira is the best skyscraper for hosting events! Spacious halls, diverse services, sophisticated gastronomic menus … Everything is carefully designed to achieve success in all your events. Discover their most exclusive halls!	0.018	0.018	0.018	0.068	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	**0.318**	**0.418**	1
Ilunion hotels	We also open our urban hotels! 35% discount with our Unique Club + an additional 5% with URBAN5.	0.015	0.015	**0.432**	0.015	0.015	**0.265**	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.140	1
Paradores	.@fraconalv, General Manager of the Parador de El Saler, explains alongside Javier Blasco, Communication Manager of @xalocmar, the sustainability actions they carry out: species recovery, dune regeneration, and turtle protection. An example of #SustainableParadores!	**0.335**	0.019	0.019	0.019	**0.440**	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	1

Brand	Topic 1 Manag.	Topic 2 Family/ Festive	Topic 3 Discount	Topic 4 Travel	Topic 5 Sustainab.	Topic 6 City	Topic 7 Coast	Topic 8 Experience /Safety	Topic 9 Contest	Topic 10 Holiday	Topic 11 Culture	Topic 12 Cust. serv./ Thanks	Topic 13 Gastronomy	Topic 14 Events	HHI_u
Vincci_Hoteles	0.058	0.085	0.063	0.081	0.063	0.071	0.082	0.083	0.062	0.082	0.065	0.060	0.084	0.063	*0.073*
Servigroup	0.055	0.089	0.061	0.073	0.063	0.072	0.093	0.081	0.059	0.109	0.055	0.058	0.072	0.060	*0.075*
sercotelhoteles	0.087	0.075	0.073	0.064	0.080	0.067	0.053	0.065	0.078	0.070	0.059	0.085	0.061	0.086	*0.073*
SenatorHR	0.066	0.081	0.062	0.075	0.112	0.065	0.071	0.092	0.072	0.064	0.049	0.084	0.048	0.058	*0.075*
riuplazahotels	0.057	0.056	0.067	0.063	0.058	0.152	0.062	0.065	0.067	0.060	0.059	0.051	0.072	0.112	*0.081*
RiuHoteles	0.057	0.060	0.057	0.059	0.081	0.059	0.087	0.065	0.085	0.070	0.045	0.161	0.055	0.059	*0.082*
PrincessHotels	0.051	0.102	0.056	0.061	0.052	0.060	0.085	0.094	0.065	0.096	0.046	0.054	0.096	0.081	*0.077*
paradores	0.059	0.060	0.122	0.077	0.056	0.055	0.059	0.053	0.055	0.057	0.131	0.056	0.104	0.055	*0.081*
nhhotelgroup	0.113	0.068	0.075	0.065	0.114	0.072	0.045	0.075	0.079	0.050	0.054	0.069	0.060	0.060	*0.077*
MeliaHtlResorts	0.059	0.068	0.057	0.075	0.080	0.063	0.086	0.084	0.065	0.068	0.061	0.111	0.061	0.063	*0.074*
MeliaHotelsInt	0.139	0.059	0.052	0.044	0.159	0.054	0.039	0.071	0.070	0.051	0.043	0.127	0.040	0.051	*0.092*
Ilunionhotels	0.090	0.067	0.081	0.057	0.167	0.071	0.046	0.070	0.062	0.054	0.053	0.064	0.046	0.072	*0.083*
IBEROSTAR	0.064	0.068	0.049	0.061	0.083	0.057	0.064	0.085	0.068	0.057	0.053	0.187	0.052	0.054	*0.087*
Hotusa	0.257	0.048	0.048	0.063	0.068	0.069	0.054	0.054	0.081	0.045	0.059	0.050	0.046	0.059	*0.110*
hotelessilken	0.056	0.057	0.079	0.085	0.062	0.074	0.076	0.092	0.050	0.063	0.075	0.047	0.094	0.089	*0.075*
HotelesGlobales	0.038	0.156	0.060	0.049	0.049	0.046	0.181	0.068	0.053	0.117	0.043	0.053	0.040	0.047	*0.099*
hoteles_santos	0.052	0.054	0.083	0.083	0.058	0.074	0.079	0.088	0.050	0.062	0.064	0.039	0.108	0.107	*0.077*
hipotels	0.056	0.082	0.087	0.059	0.073	0.057	0.076	0.060	0.063	0.146	0.046	0.075	0.057	0.062	*0.079*
h10hotels	0.057	0.068	0.093	0.068	0.057	0.091	0.087	0.060	0.067	0.078	0.061	0.069	0.086	0.059	*0.074*
FERGUSHotels	0.079	0.090	0.067	0.064	0.078	0.074	0.062	0.086	0.066	0.102	0.047	0.060	0.055	0.070	*0.074*
exehotels	0.043	0.061	0.085	0.090	0.046	0.095	0.082	0.074	0.056	0.076	0.081	0.077	0.061	0.074	*0.075*
Eurostars	0.061	0.060	0.076	0.079	0.049	0.079	0.064	0.059	0.137	0.055	0.090	0.074	0.057	0.060	*0.078*
Elbahoteles	0.046	0.089	0.073	0.072	0.049	0.078	0.099	0.065	0.050	0.103	0.054	0.039	0.070	0.112	*0.078*
CataloniaHotels	0.055	0.072	0.120	0.059	0.074	0.076	0.078	0.059	0.067	0.090	0.050	0.068	0.070	0.062	*0.075*
BlueSeaHotels	0.052	0.085	0.105	0.070	0.053	0.057	0.137	0.057	0.062	0.091	0.066	0.055	0.053	0.057	*0.079*
BlueBayHotels	0.078	0.078	0.073	0.082	0.072	0.085	0.071	0.075	0.071	0.077	0.060	0.050	0.068	0.061	*0.073*
barcelohoteles	0.060	0.058	0.057	0.066	0.065	0.075	0.087	0.071	0.053	0.070	0.053	0.169	0.054	0.060	*0.083*

Fig. 3. Average weight of topics (W_{uk}) and concentration index (HHI_u) of Twitter accounts with the highest levels of activity.

assigned to Topic 3 (Fig. 3), it indicates that a high number of messages had been published whose content generated few interactions among users.

6.5. Content type strategy: weight-engagement matrix

The Weight-Engagement matrix offers a static view of the content strategy of the hotel brands in the sample, revealing differences in the importance assigned to each topic and its engagement level (Fig. 6). Specifically, the weights attached to the topics “Gastronomy”, “Discount”, and “Customer Service/Thanks” were to a greater extent higher than their respective engagement levels (HL quadrant). Whereas above-average interaction levels were assigned to “Sustainability”, “Contest”, “Management”, “Experience/Safety”, “Family/Festive”, “City”, and “Culture” despite their having lower weightings (LH quadrant).

On the other hand, above-average importance was assigned to the topics “Coast” and “Holiday” for both weight and engagement (HH quadrant), in contrast to the lower values for both weight and engagement (LL quadrant) assigned to “Event” and “Travel".

6.6. Correlations

Additionally, a statistical correlation analysis was conducted. The existence of correlations was determined globally (62 brands) and by grouping hotel brands based on the number of available rooms (Table 6). Significant correlations were observed in the overall analysis, albeit with weak or moderate coefficients. Specifically, the only significant positive correlation was found for Topic 1 (Management; 0.370), indicating that a higher proportion of the topic generates a higher level of engagement, while the two negative correlations corresponded to Topic 3 (Discount; −0.290) and Topic 12 (Customer service/Thanks; −0.693).

Regarding the group analysis, in the groups labelled Small and Large, a negative correlation of Topic 12 (Customer service/Thanks) was also observed. Only in the Small group, were moderate significant positive correlations obtained between the proportion of the topic and its engagement. Specifically, in Topic 1 (Management; 0.478) and Topic 5 (Sustainability; 0.632). The group of medium-sized hotels showed no significant correlation.

7. Discussion

The use of social media as a marketing strategy is almost essential nowadays, making it necessary to control the content of the messages that are published on those media. In the specific case of the hotel industry, although a large number of works on the analysis of UGC have been published, far fewer investigations have been centred on FGC. More specifically, and as far as the authors are aware, there have been no previous investigations on tourist firm FGC in which topic modelling has been jointly employed alongside engagement analysis. Hence the focus of our exploratory work on analysing the content of the messages published on Twitter by hotel brands with establishments in Spain, a global leader in the management of mass tourism, seeking to discover the principal latent topics, their weights, their evolution over time, the engagement levels that they generated, as well as the relation between the weight and the engagement of the topics. That sort of framework can be valuable for advancing the development of content-type strategies, not only in the tourism sector but also in other industries that use social media to disseminate firm-generated content. It is worth mentioning that the approach followed in this work is aimed at linking the proportion of topics found in messages to engagement, based on the premise that a message can contain multiple topics. Working under this premise can lead to a more precise evaluation of both topic prevalence within the corpus and the engagement levels that can be attributed to each one.

Beginning with the latent topics identified in the Twitter messages that hotel chains were posting (*RQ1*), a total of 14 were identified through LDA, one of the most popular topic modelling methods. The topic weights ranged from 7.96% (Customer service/Thanks) to 6.65% (Events). In our study, previously identified topics were observed along with new ones, such as “Management”, indicating a commitment by hotel brands to highlight their achievements within this area. In addition, the “Sustainability” topic (7.21%) and its notably higher significance is worth highlighting, even though that observation is not entirely novel, as it was carefully examined by Bonsón et al. (2016) under the designation “Social and Environmental Concern” (1.70%). This heightened importance may be attributed to the mega-trend nature of sustainability and the increased awareness of hotel brands within this domain.

In turn, the individual analysis of the most active Twitter profiles revealed different topic weightings among the hotel chains, reflecting

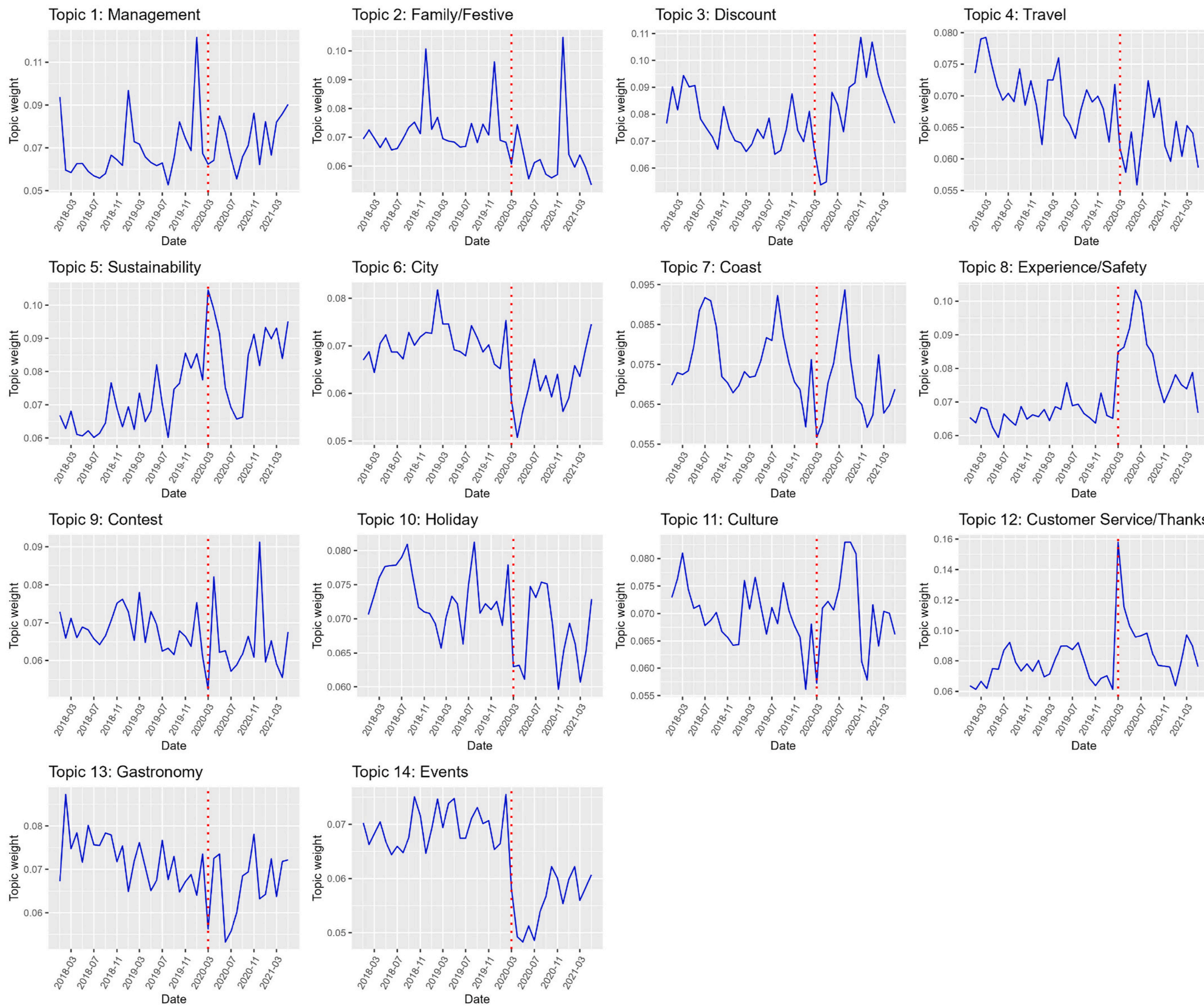

Fig. 4. Temporal evolution of the weight of each topic.

the unique characteristics of each hotel chain. This analysis was supported by the Herfindahl-Hirschman Index (HHI), a measure of market concentration applied, in our case, to the concentration of topics, allowing us to determine to what extent a brand focuses on a limited number of topics. For example, the Hotusa Group, a parent company organized into three business units—the hotel services unit, the distribution unit, and the hotel management area— achieved the highest level of topic concentration by placing significant emphasis on Topic 1 (Management) in its messages.

With regard to the 6 categories or meta-dimensions created for the purpose of conducting a more aggregated analysis, differences with other studies were once again found, perhaps motivated by the social network (mainly Facebook) and the types of hotel groups that were studied. Although no meta-dimensions were identified in studies on hotel FGC after the identification of content themes, similarities were observed with the type of content identified in previous studies. Thus, the content that was mostly emphasized in our analysis was "Enjoyment", closely followed by "Hotel management", and "Destination". More specifically, the promotion of the tourist destination had a weight of 21.37%, similar to the 28% obtained by Minazzi and Lagrosen (2013), or the 38% of the study of Yoo and Lee (2015), which reflects that this theme has been gaining strength for the past decade. In relation with the promotions conducted by the hotels, the results were in line with those shown in the content analysis of Leung et al. (2017) where this topic was not reflected in more than 7% of the messages similar to the figure of 7.81% for topic 3 "Offers" in our own work. The oldest studies centred on the hotel sector showed important differences in the proportion of messages on the topic "Contests", placed within the Entertainment category: ranging from 16% in the work of Minazzi and Lagrosen (2013) to 0.5% in the work of Yoo and Lee (2015), focused on heritage hotels. The most similar results were found in the work of Aydin (2020), who reported 8.4% as against the figure of 6.77% in our study. The differences found between our results and those of Minazzi and Lagrosen (2013) and Aydin (2020) may be attributed to the temporal evolution of social media usage, with a reduced emphasis on the importance of contests. Additionally, within the "Hotel management" category, we included the "Customer Service/Thanks" topic (7.96%), which was divided by Kim et al. (2023) into "Customer-relationship management" (two-way service recovery) and "Service recovery" (two-way communication), which yielded a weight of 75.9% when combined. This substantial difference from our study's 11.5% could be attributed to the possibility that the proportion of responses was higher.

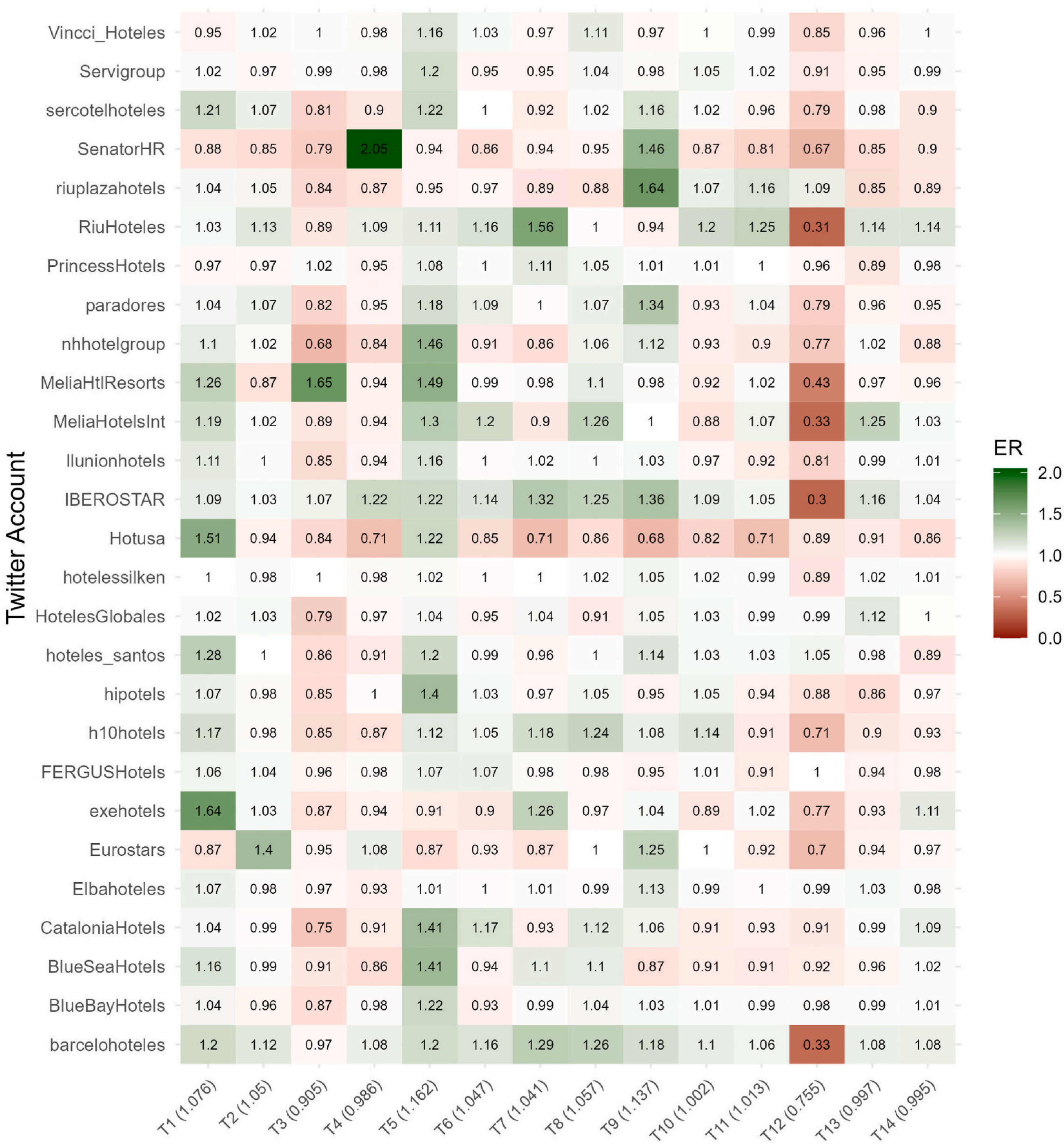

Fig. 5. Average topic engagement levels of Twitter profiles.

The time at which the World Health Organization declared the outbreak of a pandemic was taken into consideration, to approach the monthly development of the topics, so as to analyse the temporal evolution of the topics and the potential impact of COVID-19 on them (RQ2). The lock-down at home and the hotel closures that then ensued were accompanied by an emphasis on topics relating to "Sustainability", "Experience/Safety", and "Customer service/Thanks". That result was aligned with other works on the content of the social media messages sent by tourist destinations during that period (Huertas, Oliveira, & Girotto, 2020; Ketter & Avraham, 2021) and in hotel accommodation (Aguado-Correa, Rabadán-Martín, & Padilla-Garrido, 2022; Mušanović et al., 2023). However, as the pandemic advanced, the topic "Experience/Safety" started to lose weight and was overtaken by 6 topics that were trailing behind it at that time. Besides those works, following the declaration of the pandemic, hotel chains decided to enhance several additional topics, reaching some unprecedented peaks, as seen in the case of Topic 2 (Family/Festive), Topic 3 (Discount), Topic 7 (Coast), Topic 9 (Contest), and Topic 11 (Culture). In the case of Topics 2, 7, and 11, the recorded peaks were not significantly higher than those prior to the pandemic. However, both discounts and contests showed

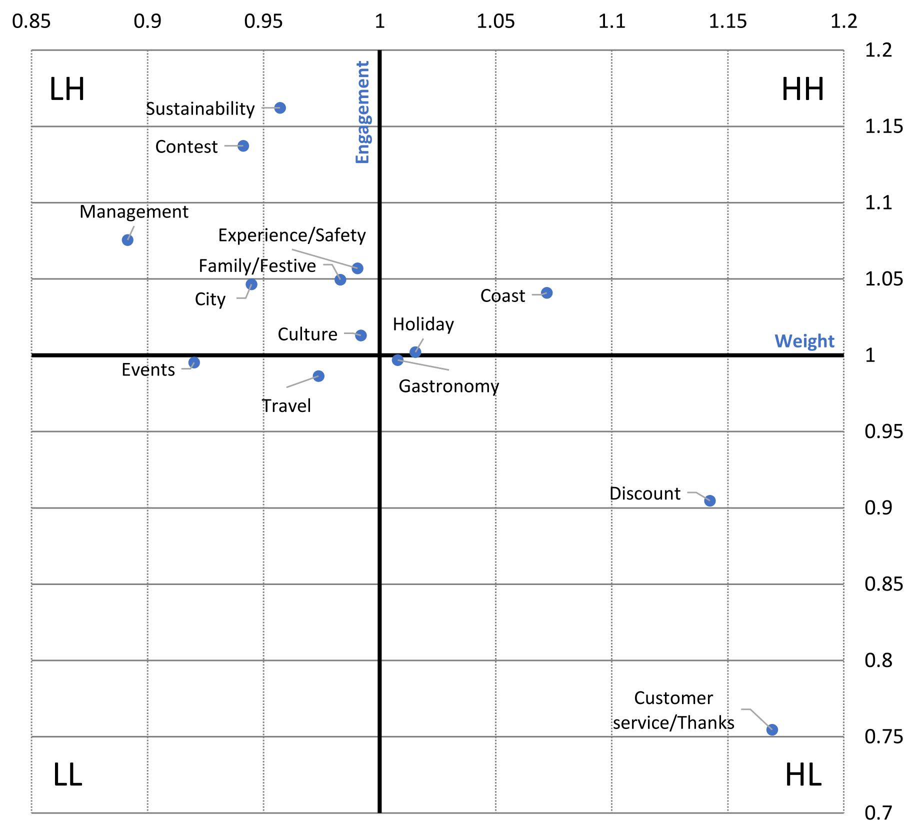

Fig. 6. Weight-Engagement matrix (WR_k vs ER_k).

substantially high levels, perhaps driven by an attempt to stimulate demand. It is essential to highlight that none of the studies on FGC in hotels conducted a dynamic analysis of the evolution of message topics. Only Mušanović et al. (2023) assessed the differences between themes discussed in 2019, 2020, and 2021, but they performed a static and overall analysis for each of these periods, without distinguishing the evolution of the importance of each topic over time, as no topic weightings were provided in their work.

With regard to the analysis of topics and their engagement levels (RQ3), it is necessary to mention that measuring the engagement of topics obtained through topic modelling presents clear differences from the measurement of engagement for the content analysis. Studies employing content analysis methodology assigned only one topic to each message, attributing all of the message's engagement to that topic. However, in our case, working under the premise that a message is composed of a set of topics with their corresponding weights, the engagement of each message can be allocated to each topic in proportion with its weight, thus allowing for greater precision in the attribution of engagement to the topics. Having said as much, it was found that sustainability-related content generated higher engagement levels among users. Those results confirmed that the environmental content continued to be more attractive for Twitter users of hotel chains within Spain, as Bonsón et al. (2016) likewise concluded in their study referring to data in 2013.

A Weight-Engagement matrix was designed to study the content type strategy of the hotel brands included in the sample. This matrix served to analyse how the hotel brands were relating the weight and engagement of each topic (RQ4). Considering that the primary reason why international tourists choose to travel to Spain is "to enjoy the sea and beaches" (Loss, 2022), hotel brands should continue to emphasize the content of "Coast" and "Holiday," as the high weight assigned to these themes is associated with high engagement ratios. On the other hand, themes such as "Sustainability," "Contest," "Management," "Experience/Safety," "Family/Festive," "City," and "Culture" may represent an opportunity for connection with followers, as their importance assigned by the brands was below average, yet they achieved higher levels of engagement. However, the different weights and engagement levels of each hotel brand should be noted, which may influence its content type strategy.

Complementarily, regarding the statistical correlation between the proportion of the topic within the messages and the average engagement value (RQ4), it was found that, when analyzing the 62 brands globally, content related to "Management" was the only one that showed a significant yet weak positive correlation. That result was noteworthy, because it is a new topic that has not been addressed in previous studies focused on the FGC of hotels. It might therefore suggest that followers of the accounts value the initiatives that hotel brands undertake from a management perspective.

Adding the above results to the study of the Weight-Engagement matrix in Fig. 6, an increase in the proportion of the "Management" topic would not only cause it to move towards the High weight-High engagement quadrant (horizontal movement in the Weight-Engagement matrix), but the positive correlation suggests that its engagement might also increase (vertical movement).

Concerning the analysis by number of rooms, it was observed that only the small chains had significant positive correlations. Specifically, the "Management" topic obtained a moderate correlation along with the "Sustainability" topic. A situation that may be due to the followers of the accounts of small chains assuming that those chains have fewer resources than large chains and appreciating, in such matters, the efforts of small chains.

With respect to significant negative correlations, Topic 12 (Customer service/Thanks) obtained a moderate value both globally and by groups (Small and Large). This value could be attributed to the fact that response messages to other users only very rarely prompt interaction with those users. Additionally, the significant negative correlation of

Table 6
Spearman's correlation matrix.

Topic/Engagement		Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14
Spearman's rho		Manag.	Family/ Festive	Discount	Travel	Sustain.	City	Coast	Exper./ Safety	Contest	Holiday	Culture	Cust. serv./ Thanks	Gastro.	Events
Global	Correlation Coefficient	0.370[b]	−0.092	−0.290[a]	0.060	0.215	0.060	0.148	0.094	−0.161	0.073	0.065	−0.693[b]	−0.202	−0.162
N = 62 (38,079 tweets)	Sig. (2-tailed)	0.003	0.477	0.022	0.640	0.093	0.644	0.250	0.467	0.210	0.570	0.613	0.000	0.116	0.207
Small (2,000–5,000 rooms)	Correlation Coefficient	0.478[a]	0.195	−0.102	0.248	0.632[b]	0.277	0.160	0.212	−0.342	0.052	0.130	−0.502[a]	−0.151	−0.323
N = 25 (12,516 tweets)	Sig. (2-tailed)	0.017	0.365	0.767	0.115	0.001	0.180	0.443	0.307	0.094	0.804	0.534	0.011	0.470	0.115
Medium (5,001–10,000 rooms)	Correlation Coefficient	0.264	−0.271	−0.386	−0.071	−0.393	−0.021	−0.061	−0.418	0.082	0.175	−0.175	−0.500	0.014	0.361
N = 15 (11,538 tweets)	Sig. (2-tailed)	0.340	0.327	0.157	0.802	0.148	0.944	0.832	0.123	0.773	0.532	0.532	0.060	0.964	0.187
Large (10,001–35,000 rooms)	Correlation Coefficient	0.274	−0.085	−0.029	−0.305	0.269	−0.010	0.286	0.371	−0.134	0.034	−0.215	−0.672[b]	−0.393	−0.342
N = 22 (14,025 tweets)	Sig. (2-tailed)	0.216	0.707	0.898	0.167	0.225	0.966	0.197	0.090	0.549	0.880	0.336	0.001	0.072	0.119

[a] . Correlation is significant at the 0.05 level (2-tailed).
[b] . Correlation is significant at the 0.01 level (2-tailed).

Topic 3 (Discount) indicates that higher discount content in the messages is associated with lower generated engagement, as observed in other studies within the hotel domain (Aydin, 2020; Leung et al., 2017). Regarding this topic, Cvijikj and Michahelles (2013) noted that posts with discount-related content were negatively correlated with the number of "likes" and had no effect on shares. On the contrary, Luarn et al. (2015) concluded that remuneration content significantly influenced consumer participation through "likes", but had no effect on comments and shares. However, Lee et al. (2018) discovered that the directly informative content that mentioned price, availability, and other offers or promotions resulted in less consumer commitment. One year later, Menon et al. (2019) showed that informative, entertaining, and promotional posts on Twitter had a positive and significative relation with "likes", comments, and shares. Finally, Montaguti et al. (2023) discovered that posts that included promotional incentives generated fewer shares, although that negative effect was reversed when the post was congruent with the brand.

Additionally, the Weight-Engagement matrix revealed the high importance assigned to the "Customer Service/Thanks" and "Discount" topics, despite having below average engagement levels. That situation may be because customer service and discounts are considered to act as levers to support sales within the sector.

8. Theoretical and practical implications

Our approach in this study has proposed a framework for a joint analysis of latent topics in FGC and the user engagement levels that those topics generate, starting from the premise that a message can contain more than one topic and expanding the literature on the application of topic modelling to FGC in the hotel industry. Thus, a more comprehensive and nuanced view of the data has been provided, revealing patterns that, had the assignment been restricted to a single topic per message, might not have been evident. In that way, the 4 research questions of our study approach have been more precisely and realistically addressed, reflecting on the multifaceted nature of natural language and the data that have been analyzed. In particular, various theoretical contributions to the research on FGC in the hotel sector have been presented.

The first theoretical contribution corresponds to the FGC topics and the categories that emerged from the analysis. Specifically, a new and previously unstudied topic — "Management" — has been introduced in our research. The topic, unexplored in prior research, was, in turn, part of a meta-dimension or category — "Hotel management" —. The category was comprised of the topics "Customer service/Thanks", "Sustainability", and "Management", and it was positioned as the third most important, following "Enjoyment" and "Destination". This positioning indicates a shift in trends in the content type strategies of hotel firms. The new taxonomy can serve as a starting point for future research in the field of FGC.

The second theoretical contribution is focused on the link between the identified topics and engagement. An aspect that has not been previously examined in works applying topic modelling techniques to the FGC of hotels. The main implications of this connection are: (a) the identification of the "Sustainability" topic as the one generating the highest engagement, suggesting an increase in global awareness of sustainability; (b) the identification of topics to which, up until now, the hotel brands included in the sample have assigned little weight, yet which could, insofar as they are topics with above-average engagement levels, provide opportunities for connecting with the target audience; (c) the positive correlation between the weight and engagement of the newly identified theme "Management"; (d) the negative correlation between the weight and engagement of the topics "Discount" and "Customer service/Thanks"; and (e) the analysis of brand groups based on the number of rooms, showing a positive correlation between the weight and engagement of "Management" and "Sustainability" in hotel brands with fewer rooms.

In that sense, the results have provided a valuable contribution to content marketing in social media within the hotel industry, highlighting how topics and engagement have been evolving in response to contemporary concerns of consumers and hotel brands, which may influence customer behavior on social media. It is aligned with the Uses and Gratifications theory, which posits that the content of messages on social media should be designed to create value or gratification for individual consumers, in order to foster a higher level of engagement. Furthermore, this approach can have a dynamic nature, making it particularly suitable for analyzing new types of content that emerge over the years (Muntinga et al., 2011), as has been the case of our study.

With regard to the practical implications of the study, if hotel brands base their marketing communications on the joint results of topic modelling and engagement analysis, they can improve their appeal when communicating with the target public. The engagement levels that are generated in relation to each topic represent information of vital importance for a brand to orient marketing strategies, social responsibility, and promotion. More specifically, it may be of use to firms, so that they can evaluate the effectiveness of their content type strategies with regard to the topics that they really wish to promote. One clear example of the need to change strategy, in accordance with the results of effectiveness reported in this study is the case of the "*Paradores*" brand, which centred its content type strategy on the topics of "Discount", "Culture", and Gastronomy" (see Fig. 5). However, none of them reached engagement values above the average. We would therefore recommend, in the case of the "*Paradores*" brand, an adjustment of the proportions of the messages with that content, increasing those that had "Contest" as a topic, because that is the one that offers a higher level of engagement on Twitter.

In addition, the spread of topics may be compared with the brands of competitors to detect whether any of the topics they are promoting present higher engagement levels, so as to analyse the content of the messages and the reasons for that greater engagement. Nevertheless, based on the findings of this study, it would be advisable for hotel brands to enhance their messages with "Management" and "Sustainability" content, as it has the potential to generate higher engagement, especially for smaller brands. They could achieve that through frequent posts that highlight the tasks of CEOs, approaches to addressing challenges in the tourism sector, and the sustainability strategies of the hotel brand, such as promoting actions aimed at achieving sustainable development goals.

On the other hand, this work has contributed to a deeper understanding of the trends and patterns in the FGC of the major hotel groups, shedding light on the topics that have gained prominence and those that have receded since the outbreak of the pandemic.

Furthermore, the Weight-Engagement matrix of topics designed in this study is proposed as a new supporting tool in the process of adapting social media content type strategy. Although the matrix has been presented for the set of studied brands, each brand can individually determine the position of its topics in relation to weight and engagement ratios to analyse themes that could be scaled back in favour of others more highly valued by their social media users. The versatility of this tool is its usability across various sectors and social media platforms, implying that each content strategy may be customized, based on the characteristics of users/customers on each social network, with the aim of enhancing the effectiveness of the message.

9. Limitations and future lines of investigation

This study presents various limitations. In the first place, the messages that the hotel brands posted on Twitter have been analyzed, although they may not be representative of all the content posted through their different social media. The reason for having a single data source to collect the FGC was due to the difficulty of downloading massive volumes of messages from other social media. In second place, tweets directed at Spanish speakers and solely written in Spanish were employed in the investigation. In third place, the sample is conditioned by the choice of hotel brands with a strong presence in Spain and that had Twitter accounts. Thus, the generalizability of the findings may be limited to the hotel groups and brands included in the sample. It is nevertheless true that Spain is the third most competitive country in the world in terms of tourism, according to the World Economic Forum (2022), which might give some indication of the solidity and representativeness of those brands. The fourth limitation refers to the size of the observations used to perform group-wise correlations, which may influence the results. The fifth and last limitation is related to the interpretation of global results from the entire sample's content type strategy, as each hotel brand should analyse the position of topics within the Weight-Engagement matrix, to identify the quadrants they fall into and make strategic decisions that are aligned with their policies and objectives.

As future lines of investigation, we can highlight the application of the proposed methodology to the analysis of the tweets that the main international hotel chains post in English in other countries. By doing so, any variation in the weight of the topics and the engagement levels that they generate could be analyzed. A second line of research would emerge from the inclusion of user opinions to understand how those opinions when tweeted interact with and respond to the FGC that may be under study.

Statements and declarations

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

The authors have no relevant financial or non-financial interests to disclose.

As an original work, the paper and its associated data have not previously been published elsewhere. All the authors can confirm their participation and their areas of work in the investigative project. The paper has not been submitted to any other publication or conference.

The authors declare that there is no conflict of interest regarding the publication of this article.

Impact statement

Social media have become an essential tool for communication and information exchange within the hotel sector. In the present study, it has been demonstrated that hotel chains use networks such as Twitter, primarily to promote content related to enjoyment, hotel management activities, and destinations. Consequently, "hotel management" emerges as a new category, encompassing sustainability, hotel management, and customer service topics. This social media usage extends beyond mere commercialization, aiming to enhance the perception of organizational transparency and foster sustainability awareness, aspects positively valued by users, according to the results of this paper. The joint use of topic modelling and engagement analysis presented in this study constitutes a tool with which key topics that are discussed in the sector may be identified and the interest that they can arouse can be assessed. They therefore constitute the basis for properly defining the content type strategies of hotel firms.

CRediT authorship contribution statement

Inmaculada Rabadán-Martín: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Lucía Barcos-Redín:** Writing – review & editing, Validation, Software, Resources, Methodology, Formal analysis. **Jorge Pereira-

Delgado: Writing – review & editing, Validation, Software, Resources, Methodology, Formal analysis. **Francisco Aguado-Correa:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Nuria Padilla-Garrido:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation.

Declaration of competing interest

None.

Acknowledgment

Funding for open access charge: Universidad de Huelva/CBUA.

Annex I. Twitter accounts of hotel brands grouped by the number of available rooms

Size (rooms)	Small (2,000–5,000)	Medium (5,001–10,000)	Large (10,001–35,000)
Tweets	12,516	11,538	14,025
Number of Twitter accounts	25	15	22
Twitter accounts	AlegriaHotels	allsunHotels	_Best_Hotels
	azuLineHotels	aluahotels	Anantara_Hotels
	BlueSeaHotels	AMResorts_Corp	barcelohoteles
	Elbahoteles	BlueBayHotels	Eurostars
	EstivalGroup	CataloniaHotels	exehotels
	EveniaHotels	DreamsResorts	GranMeliaHotels
	FERGUSHotels	grupotel	h10hotels
	HesperiaResorts	hipotels	Hotusa
	hoteles_magic	HotelesGlobales	IBEROSTAR
	hoteles_santos	Lopesan	InnsidebyMelia
	hotelessilken	PalladiumHG	MEbyMelia
	HSMHoteles	paradores	MeliaHotelsInt
	htophotels	PrincessHotels	MeliaHtlResorts
	Ilunionhotels	SecretsResorts	nhcollection
	insotelresorts	SenatorHR	nhhotelgroup
	medplaya		NHHotels
	onahotels		OCEAN_HOTELS
	PlayaSolHG		Paradisus
	proturhotels		RiuHoteles
	RocHotels		Riuplazahotels
	sercotelhoteles		solbymelia
	Servigroup		TrypHotels
	THBHOTELS		
	Vincci_Hoteles		
	ZafiroHotels		

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: A surveillance study. *Journal of Medical Internet Research, 22*(4), 1–9. https://doi.org/10.2196/19016

Aguado-Correa, F., Rabadán-Martín, I., & Padilla-Garrido, N. (2022). COVID-19 and the accommodation sector: first measures, and online communications strategies. A multiple case study in a Spanish province. *Investigaciones Turísticas, (24),*, 172–197. https://doi.org/10.14198/INTURI2022.24.9

Akpinar, E., & Berger, J. (2017). Valuable virality. *Journal of Marketing Research, 54*(2), 318–330. https://doi.org/10.1509/jmr.13.0350

Aydin, G. (2020). Social media engagement and organic post effectiveness: A roadmap for increasing the effectiveness of social media use in hospitality industry. *Journal of Hospitality Marketing & Management, 29*(1), 1–21. https://doi.org/10.1080/19368623.2019.1588824

Barrie, C., & Ho, J. (2021). academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software, 6*(62), 3272. https://doi.org/10.21105/joss.03272

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bonsón, E., Bednárová, M., & Wei, S. (2016). Corporate Twitter use and stakeholder engagement: An empirical analysis of the Spanish hotel industry. *European Journal of Tourism Research, 13*, 69–83. https://doi.org/10.54055/ejtr.v13i.232

Bonsón, E., & Ratkai, M. (2013). A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review, 37*(5), 787–803. https://doi.org/10.1108/OIR-03-2012-0054

Brandt, T., Bendler, J., & Neumann, D. (2017). Social media analytics and value creation in urban smart tourism ecosystems. *Information and Management, 54*(6), 703–713. https://doi.org/10.1016/j.im.2017.01.004

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research, 66*(1), 105–114. https://doi.org/10.1016/j.jbusres.2011.07.029

Buhalis, D., & Mamalakis, E. (2015). Social media return on investment and performance evaluation in the hotel industry context. In I. Tussyadiah, & A. Inversini (Eds.), *Information and communication Technologies in tourism 2015*. Cham: Springer. https://doi.org/10.1007/978-3-319-14343-9_18.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management, 26*(7), 675–693. https://doi.org/10.1080/19368623.2017.1310075

Ceballos, M., Crespo, Á. G., & Cousté, N. L. (2016). Impact of firm-created content on user-generated content: Using a new social media monitoring tool to explore Twitter. In L. Petruzzellis, & R. Winer (Eds.), *Rediscovering the essentiality of marketing. Developments in marketing science: Proceedings of the academy of marketing science*. Cham: Springer. https://doi.org/10.1007/978-3-319-29877-1_61.

Cervellon, M. C., & Galipienzo, D. (2015). Facebook pages content, does it really matter? Consumers' responses to luxury hotel posts with emotional and informational content. *Journal of Travel & Tourism Marketing, 32*(4), 428–437. https://doi.org/10.1080/10548408.2014.904260

Chae, B. (2015). Insights from hashtag #supplychain and Twitter analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics, 165*, 247–259. https://doi.org/10.1016/j.ijpe.2014.12.037

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Neural Information Processing Systems, 22*, 288–296.

Cheng, M., Liu, J., Qi, J., & Wan, F. (2021). Differential effects of firm generated content on consumer digital engagement and firm performance: An outside-in perspective. *Industrial Marketing Management, 98*(July), 41–58. https://doi.org/10.1016/j.indmarman.2021.07.001

Cheng, X., Yan, X., Lan, Y., & Guo, J. (2014). BTM: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering, 26*(12), 2928–2941. https://10.1109/TKDE.2014.2313872.

Cheung, C. M. K., Shen, X. L., Lee, Z. W. Y., & Chan, T. K. H. (2015). Promoting sales of online games through customer engagement. *Electronic Commerce Research and Applications, 14*(4), 241–250. https://doi.org/10.1016/j.elerap.2015.03.001

Coelho, R. L. F., De Oliveira, D. S., & De Almeida, M. I. S. (2016). Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics. *Online Information Review, 40*(4), 458–471. https://doi.org/10.1108/OIR-06-2015-0176

Colicev, A., Malshe, A., Pauwels, K., & O'Connor, P. (2018). Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media. *Journal of Marketing, 82*(1), 37–56. https://doi.org/10.1509/jm.16.0055

Cvijikj, I. P., & Michahelles, F. (2013). Online engagement factors on Facebook brand pages. *Social Network Analysis and Mining, 3*(4), 843–861. https://doi.org/10.1007/s13278-013-0098-8

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining, 9*(1), 1–20. https://doi.org/10.1007/s13278-019-0568-8

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management, 54*(5), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004

de Vries, L., Gensler, S., & Leeflang, P. S. H. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing, 26*(2), 83–91. https://doi.org/10.1016/j.intmar.2012.01.003

Dickinger, A., Lalicic, L., & Mazanec, J. (2017). Exploring the generalizability of discriminant word items and latent topics in online tourist reviews. *International Journal of Contemporary Hospitality Management, 29*(2), 803–816. https://doi.org/10.1108/IJCHM-10-2015-0597

Dolan, R., Conduit, J., Frethey-Bentham, C., Fahy, J., & Goodman, S. (2019). Social media engagement behavior: A framework for engaging customers through social media content. *European Journal of Marketing, 53*(10), 2213–2243. https://doi.org/10.1108/EJM-03-2017-0182

Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology, 7*(May), 1–16. https://doi.org/10.3389/fsoc.2022.886498

Ferrer-Rosell, B., Martin-Fuentes, E., & Marine-Roig, E. (2019). Information and communication Technologies in tourism 2019. In *Information and communication Technologies in tourism 2019*. Springer International Publishing. https://doi.org/10.1007/978-3-030-05940-8.

Ferrer-Rosell, B., Martin-Fuentes, E., & Marine-Roig, E. (2020). Diverse and emotional: Facebook content strategies by Spanish hotels. *Information Technology & Tourism, 22*(1), 53–74. https://doi.org/10.1007/s40558-019-00164-z

Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Information Systems Research, 24*(1), 88–107. https://doi.org/10.1287/isre.1120.0469

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., et al. (2019). Topic modeling in management research: Rendering new theory from textual data. *The Academy of Management Annals, 13*(2), 586–632. https://doi.org/10.5465/annals.2017.0099

Hanson, S., Jiang, L., & Dahl, D. (2019). Enhancing consumer engagement in an online brand community via user reputation signals: A multi-method analysis. *Journal of the Academy of Marketing Science, 47*(2), 349–367. https://doi.org/10.1007/s11747-018-0617-2

Hao, F. (2020). The landscape of customer engagement in hospitality and tourism: A systematic review. *International Journal of Contemporary Hospitality Management, 32*(5), 1837–1860. https://doi.org/10.1108/IJCHM-09-2019-0765

Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science, 45*(3), 312–335. https://doi.org/10.1007/s11747-016-0509-2

Harrigan, P., Evers, U., Miles, M., & Daly, T. (2017). Customer engagement with tourism social media brands. *Tourism Management, 59*, 597–609. https://doi.org/10.1016/j.tourman.2016.09.015

Hashim, K. F., & Fadhil, N. A. (2017). Engaging with customer using social media platform: A case study of Malaysia hotels. *Procedia Computer Science, 124*, 4–11. https://doi.org/10.1016/j.procs.2017.12.123

Heydt, M. (2018). *Python web scraping cookbook: Over 90 proven recipes to get you scraping with Python, microservices. Docker, and AWS* (p. 136). Birmingham, UK: Packt Publishing.

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.

Hosteltur. (2020). *Ranking Hosteltur de grandes cadenas hoteleras 2020*. Retrieved from https://www.hosteltur.com/139934_ranking-hosteltur-de-grandes-cadenas-hoteleras-2020.html.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management, 72*(March 2018), 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Huertas, A., Oliveira, A., & Girotto, M. (2020). Gestión comunicativa de crisis de las oficinas nacionales de turismo de España e Italia ante la Covid-19. *El Profesional de la Información*, 1–18. https://doi.org/10.3145/epi.2020. jul.10.

Jalali, N. Y., & Papatla, P. (2019). Composing tweets to increase retweets. *International Journal of Research in Marketing, 36*(4), 647–668. https://doi.org/10.1016/j.ijresmar.2019.05.001

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., et al. (2019). Latent dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jeong, B., Yoon, J., & Lee, J. M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management, 48*(April 2017), 280–290. https://doi.org/10.1016/j.ijinfomgt.2017.09.009

Ji, Y. G., Li, C., North, M., & Liu, J. (2017). Staking reputation on stakeholders: How does stakeholders' Facebook engagement help or ruin a company's reputation? *Public Relations Review, 43*(1), 201–210. https://doi.org/10.1016/j.pubrev.2016.12.004

Jia, S. S. (2020). Motivation and satisfaction of Chinese and U.S. Tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management, 78*(October 2019), Article 104071. https://doi.org/10.1016/j.tourman.2019.104071

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons, 53*(1), 59–68. https://doi:10.1016/j.bushor.2009.09.003.

Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research, 4*(1), 8–18. https://doi.org/10.1086/208674

Katz, E., & Foulkes, D. (1962). On the use of the mass media as 'escape': Clarification of a concept. *Public Opinion Quarterly, 26*, 277–388.

Ketter, E., & Avraham, E. (2021). #StayHome today so we can #TravelTomorrow: Tourism destinations' digital marketing strategies during the Covid-19 pandemic. *Journal of Travel & Tourism Marketing, 38*(8), 819–832. https://doi.org/10.1080/10548408.2021.1921670

Kim, D. S., Lee, B. C., & Park, K. H. (2021). Determination of motivating factors of urban forest visitors through latent Dirichlet allocation topic modeling. *International Journal of Environmental Research and Public Health, 18*(18). https://doi.org/10.3390/ijerph18189649

Kim, W. G., & Park, S. A. (2017). Social media review rating versus traditional customer satisfaction: Which one has more incremental predictive power in explaining hotel performance? *International Journal of Contemporary Hospitality Management, 29*(2), 784–802. https://doi.org/10.1108/IJCHM-11-2015-0627

Kim, K., Park, O., Barr, J., & Yun, H. (2019). Tourists' shifting perceptions of UNESCO heritage sites: Lessons from Jeju Island-South Korea. *Tourism Review, 74*(1), 20–29. https://doi.org/10.1108/TR-09-2017-0140

Kim, W. H., Park, E., & Kim, S. B. (2023). Understanding the role of firm-generated content by hotel segment: The case of Twitter. *Current Issues in Tourism, 0*(0), 1–15. https://doi.org/10.1080/13683500.2021.2003759

Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tourism Management, 83*(June 2020), Article 104241. https://doi.org/10.1016/j.tourman.2020.104241

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2016). From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing, 80*(1), 7–25. https://doi.org/10.1509/jm.14.0249

Kwok, L., Lee, J., & Han, S. H. (2022). Crisis communication on social media: What types of COVID-19 messages get the attention? *Cornell Hospitality Quarterly, 63*(4), 528–543. https://doi.org/10.1177/19389655211028143

Laureate, C. D. P., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-023-10471-x

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science, 64*(11), 5105–5131. https://doi.org/10.1287/mnsc.2017.2902

Lee, H., Jang, S. J., Sun, F. F., Broadwell, P., & Yoon, S. (2022). Comparing themes extracted via topic modeling and manual content analysis: Korean-language discussions of dementia on Twitter. *Studies in Health Technology and Informatics, 295*, 230–233. https://doi.org/10.3233/SHTI220704

Lee, J., & Park, C. (2022). Customer engagement on social media, brand equity and financial performance: A comparison of the US and Korea. *Asia Pacific Journal of Marketing and Logistics, 34*(3), 454–474. https://doi.org/10.1108/APJML-09-2020-0689

Lei, S. S. I., Pratt, S., & Wang, D. (2017). Factors influencing customer engagement with branded content in the social network sites of integrated resorts. *Asia Pacific Journal of Tourism Research, 22*(3), 316–328. https://doi.org/10.1080/10941665.2016.1250792

Leung, X. Y., Bai, B., & Erdem, M. (2017). Hotel social media marketing: A study on message strategy and its effectiveness. *Journal of Hospitality and Tourism Technology, 8*(2), 239–255. https://doi.org/10.1108/JHTT-02-2017-0012

Leung, X. Y., Bai, B., & Stahura, K. A. (2015). The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter. *Journal of Hospitality & Tourism Research, 39*(2), 147–169. https://doi.org/10.1177/1096348012471381

Lim, J., & Lee, H. C. (2020). Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel. *Current Issues in Tourism, 23*(10), 1261–1276. https://doi.org/10.1080/13683500.2019.1604638

Liu, J., Yu, Y., Mehraliyev, F., Hu, S., & Chen, J. (2022). What affects the online ratings of restaurant consumers: A research perspective on text-mining big data analysis. *International Journal of Contemporary Hospitality Management, 34*(10), 3607–3633. https://doi.org/10.1108/IJCHM-06-2021-0749

Loper, E., & Bird, S. (2002). *Nltk: The Natural Language toolkit arXiv*.

Loss, L. (2022). Survey: European tourists traveling to Spain. *TourismReview Media*. Retrieved from https://www.tourism-review.com/people-traveling-to-spain-mostly-to-enjoy-sun-and-beach-news12727.

Luarn, P., Lin, Y. F., & Chiu, Y. P. (2015). Influence of Facebook brand-page posts on online engagement. *Online Information Review, 39*(4), 505–519. https://doi.org/10.1108/OIR-01-2015-0029

Luo, Q., & Zhong, D. (2015). Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management, 46*, 274–282. https://doi.org/10.1016/j.tourman.2014.07.007
Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754
Malhotra, A., Malhotra, C., & See, A. (2013). How to create brand engagement on Facebook. *MIT Sloan Management Review, 54*(2), 18–20.
Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge, England: Cambridge University Press.
Mariani, M. M., Mura, M., & Di Felice, M. (2018). The determinants of Facebook social engagement for national tourism organizations' Facebook pages: A quantitative approach. *Journal of Destination Marketing & Management, 8*(July 2016), 312–325. https://doi.org/10.1016/j.jdmm.2017.06.003
McCallum, A. K. (2002). Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.
Meire, M., Hewett, K., Ballings, M., Kumar, V., & Van den Poel, D. (2019). The role of marketer-generated content in customer engagement marketing. *Journal of Marketing, 83*(6), 21–42. https://doi.org/10.1177/0022242919873903
Menon, R. G. V., Sigurdsson, V., Larsen, N. M., Fagerstrøm, A., Sørensen, H., Marteinsdottir, H. G., et al. (2019). How to grow brand post engagement on Facebook and Twitter for airlines? An empirical investigation of design and content factors. *Journal of Air Transport Management, 79*(May), Article 101678. https://doi.org/10.1016/j.jairtraman.2019.05.002
Minazzi, R., & Lagrosen, S. (2013). Investigating social media marketing in the hospitality industry: Facebook and European hotels. In Z. Xiang, & I. Tussyadiah (Eds.), *Information and communication Technologies in tourism 2014*. Cham: Springer. https://doi.org/10.1007/978-3-319-03973-2_11.
Montaguti, E., Valentini, S., & Vecchioni, F. (2023). Content that engages your customers: The role of brand congruity and promotions in social media. *Journal of Interactive Marketing, 58*(1), 16–33. https://doi.org/10.1177/10949968221129817
Moro, S., & Rita, P. (2018). Brand strategies in social media in hospitality and tourism. *International Journal of Contemporary Hospitality Management, 30*(1), 343–364. https://doi.org/10.1108/IJCHM-07-2016-0340
Muñoz-Expósito, M., Oviedo-García, M.Á., & Castellanos-Verdugo, M. (2017). How to measure engagement in Twitter: Advancing a metric. *Internet Research, 27*(5), 1122–1148. https://doi.org/10.1108/IntR-06-2016-0170
Muntinga, D. G., Moorman, M., & Smit, E. G. (2011). Introducing COBRAs. *International Journal of Advertising, 30*(1), 13–46. https://doi.org/10.2501/ija-30-1-013-046
Muritala, B. A., Hernández-Lara, A. B., & Sánchez-Rebull, M. V. (2022). COVID-19 staycations and the implications for leisure travel. *Heliyon, 8*(10), Article e10867. https://doi.org/10.1016/j.heliyon.2022.e10867
Mušanović, J., Dorčić, J., & Gregorić, M. (2023). Corporate communication on social media: A case study before and during pandemic COVID-19. *Corporate Communications*. https://doi.org/10.1108/CCIJ-07-2022-0085
Osei-Frimpong, K., & McLean, G. (2018). Examining online social brand engagement: A social presence theory perspective. *Technological Forecasting and Social Change, 128*, 10–21. https://doi.org/10.1016/j.techfore.2017.10.010
Oviedo-García, M. A., Muñoz-Expósito, M., Castellanos-Verdugo, M., & Sancho-Mejías, M. (2014). Metric proposal for customer engagement in Facebook. *The Journal of Research in Indian Medicine, 8*(4), 327–344. https://doi.org/10.1108/JRIM-05-2014-0028
Pansari, A., & Kumar, V. (2017). Customer engagement: The construct, antecedents, and consequences. *Journal of the Academy of Marketing Science, 45*(3), 294–311. https://doi.org/10.1007/s11747-016-0485-6
Pino, G., Peluso, A. M., Del Vecchio, P., Ndou, V., Passiante, G., & Guido, G. (2019). A methodological framework to assess social media strategies of event and destination management organizations. *Journal of Hospitality Marketing & Management, 28*(2), 189–216. https://doi.org/10.1080/19368623.2018.1516590
Poulis, A., Rizomyliotis, I., & Konstantoulaki, K. (2019). Do firms still need to be social? Firm generated content in social media. *Information Technology and People, 32*(2), 387–404. https://doi.org/10.1108/ITP-03-2018-0134
Rabadán-Martín, I., Aguado-Correa, F., & Padilla-Garrido, N. (2023). Exploring hidden topics on hotel Facebook fan pages before and during the COVID-19 crisis. *Review of Business Management, 25*(3), 291–314. https://doi.org/10.7819/rbgn.v25i3.4227
Řehůřek, R., & Sojka, P. (2011). *Gensim–python framework for vector space modelling* (Vol. 3). Brno, Czech Republic: NLP Centre, Faculty of Informatics, Masaryk University, 2.
Reisenbichler, M., & Reutterer, T. (2019). Topic modeling in marketing: Recent advances and research opportunities. In *Journal of business economics* (Vol. 89)Springer Berlin Heidelberg. https://doi.org/10.1007/s11573-018-0915-7.
Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Wsdm 2015 - proceedings of the 8th ACM international conference on web search and data mining* (pp. 399–408). https://doi.org/10.1145/2684822.2685324
Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Lebherz, P. R. (2014). Factors influencing popularity of branded content in Facebook fan pages. *European Management Journal, 32*(6), 1001–1011. https://doi.org/10.1016/j.emj.2014.05.001
Santiago, J., Borges-Tiago, M. T., & Tiago, F. (2022). Is firm-generated content a lost cause? *Journal of Business Research, 139*(October 2021), 945–953. https://doi.org/10.1016/j.jbusres.2021.10.022
Santos, Z. R., Cheung, C., Coelho, P. S., & Rita, P. (2022). Consumer engagement in social media brand communities: A literature review. *International Journal of Information Management, 63*. https://doi.org/10.1016/j.ijinfomgt.2021.102457. August 2021.
Sari, R., Putra, F., Maemunah, I., & Dianawati, N. (2022). The use of Twitter in social media marketing: Evidence from hotels in Asia. *APMBA (Asia Pacific Management and Business Application), 11*(1), 59–74. https://doi.org/10.21776/ub.apmba.2022.011.01.4
Sashi, C. M., Brynildsen, G., & Bilgihan, A. (2019). Social media, customer engagement and advocacy: An empirical investigation using Twitter data for quick service restaurants. *International Journal of Contemporary Hospitality Management, 31*(3), 1247–1272. https://doi.org/10.1108/IJCHM-02-2018-0108
Schultz, C. D. (2017). Proposing to your fans: Which brand post characteristics drive consumer engagement activities on social media brand pages? *Electronic Commerce Research and Applications, 26*, 23–34. https://doi.org/10.1016/j.elerap.2017.09.005
Shahbaznezhad, H., Dolan, R., & Rashidirad, M. (2021). The role of social media content format and platform in users' engagement behavior. *Journal of Interactive Marketing, 53*, 47–65. https://doi.org/10.1016/j.intmar.2020.05.001
Sheng, J. (2019). Being active in online communications: Firm responsiveness and customer engagement behaviour. *Journal of Interactive Marketing, 46*, 40–51. https://doi.org/10.1016/j.intmar.2018.11.004
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*(June 2021), 132–138. https://doi.org/10.1016/j.jhtm.2022.02.027
Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability, 12*(5), 1–15. https://doi.org/10.3390/su12051821
Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management, 75*(June), 550–568. https://doi.org/10.1016/j.tourman.2019.06.020
Tatar, S., & Eren-Erdoğmus, I. (2016). The effect of social media marketing on brand trust and brand loyalty for hotels. *Information Technology & Tourism, 16*(3), 249–263. https://doi.org/10.1007/s40558-015-0048-6
Taylor, D. C., Barber, N. A., & Deale, C. (2015). To tweet or not to tweet: That is the question for hoteliers: A preliminary study. *Information Technology & Tourism, 15*, 71–99. https://doi.org/10.1007/s40558-015-0022-3
Torres, E. N., & Singh, D. (2016). Towards a model of electronic word-of-mouth and its impact on the hotel industry. *International Journal of Hospitality & Tourism Administration, 17*(4), 472–489. https://doi.org/10.1080/15256480.2016.1226155
UNESCO (n.d.). Biosphere Reserves. Retrieved from https://en.unesco.org/biosphere.
UNESCO World Heritage Convention (n.d.). World Heritage List Statistics. Retrieved from https://whc.unesco.org/en/list/stat#d1.
UNWTO World Tourism Organization. (2023). World tourism barometer excerpt. *UNWTO, 21*(2). Retrieved from https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2023-05/UNWTO_Barom23_02_May_EXCERPT_final.pdf?VersionId=gGmuSXlwfM1yoemsRrBI9ZJf.Vmc9gYD.
van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., et al. (2010). Customer engagement behavior: Theoretical foundations and research directions. *Journal of Service Research, 13*(3), 253–266. https://doi.org/10.1177/1094670510375599
Vázquez, M. A., Pereira-Delgado, J., Cid-Sueiro, J., & Arenas-García, J. (2022). Validation of scientific topic models using graph analysis and corpus metadata. *Scientometrics, 127*(9), 5441–5458. https://doi.org/10.1007/s11192-022-04318-5
Viñán-Ludeña, M. S., & de Campos, L. M. (2022). Analyzing tourist data on Twitter: A case study in the province of Granada at Spain. *Journal of Hospitality and Tourism Insights, 5*(2), 435–464. https://doi.org/10.1108/JHTI-11-2020-0209
Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer engagement: Exploring customer relationships beyond purchase. *Journal of Marketing Theory and Practice, 20*(2), 122–146. https://doi.org/10.2753/MTP1069-6679200201
Vu, H. Q., Li, G., & Law, R. (2019). Discovering implicit activity preferences in travel itineraries by topic modeling. *Tourism Management, 75*(June), 435–446. https://doi.org/10.1016/j.tourman.2019.06.011
Wan, F., & Ren, F. (2017). The effect of firm marketing content on product sales: Evidence from a mobile social media platform. *Journal of Electronic Commerce Research, 18*(4), 288–302.
Wang, J., Li, Y., Wu, B., & Wang, Y. (2021). Tourism destination image based on tourism user generated content on internet. *Tourism Review, 76*(1), 125–137. https://doi.org/10.1108/TR-04-2019-0132
Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal, 16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041
World Economic Forum. (2022). Travel & tourism development index 2021. *Rebuilding for a Sustainable and Resilient Future*. Retrieved from https://www3.weforum.org/docs/WEF_Travel_Tourism_Development_2021.pdf.
Yang, S., Lin, S., Carlson, J. R., & Ross, W. T. (2016). Brand engagement on social media: Will firms' social media efforts influence search engine advertising effectiveness? *Journal of Marketing Management, 32*(5–6), 526–557. https://doi.org/10.1080/0267257X.2016.1143863
Yao, L., Mimno, D., & McCallum, A. (2009). Efficient methods for topic model inference on streaming document collections. *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 937–945). https://doi.org/10.1145/1557019.1557121
Yoo, K. H., & Lee, W. (2015). Use of Facebook in the US heritage accommodations sector: An exploratory study. *Journal of Heritage Tourism, 10*(2), 191–201. https://doi.org/10.1080/1743873X.2014.985228

Zachlod, C., Samuel, O., Ochsner, A., & Werthmüller, S. (2022). Analytics of social media data – state of characteristics and application. *Journal of Business Research, 144*, 1064–1076. https://doi.org/10.1016/j.jbusres.2022.02.016

Zarezadeh, Z. Z., Rastegar, R., & Xiang, Z. (2022). Big data analytics and hotel guest experience: A critical analysis of the literature. *International Journal of Contemporary Hospitality Management, 34*(6), 2320–2336. https://doi.org/10.1108/IJCHM-10-2021-1293

Zhang, Y., Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing, 34*(1), 100–119. https://doi.org/10.1016/j.ijresmar.2016.07.003

Francisco Aguado-Correa, Ph.D., is an Associate Professor in the Department of Financial Economics, Accounting and Operations Management, at the Faculty of Business Sciences and Tourism, University of Huelva. He has experience teaching in the Business Administration degree course at the University of Seville, Carlos III University of Madrid and the University of Huelva. His research interests include natural language processing, tourism big data, and hotels' presence in social media.

Lucía Barcos-Redín, Ph.D., is an Associate Professor in the Business Management Department, Universidad Pontificia Comillas. She also collaborates with the Quantitative Methods Department. Lucía holds a Ph.D. in Engineering, and her main lines of research are Logistics, Operations Management and Business Analytics.

Nuria Padilla-Garrido, Ph.D., is an Associate Professor in the Department of Economics, at the Faculty of Business Sciences and Tourism, University of Huelva. She has experience teaching in the Business Administration degree course at the University of Seville and the University of Huelva. Her research interests include multi-criteria decision-making, natural language processing, tourism big data, and hotels' presence in social media.

Jorge Pereira-Delgado is currently a Data Scientist at Telefónica Tech and a former Machine Learning (ML) Researcher at Carlos III University of Madrid. During his career, his researched has mainly focused on Natural Language Processing (NLP) and, more specially, in Topic Modelling (TM), although he has some work on parallel computing for ML. With a Bachelor in Telecommunication Technologies Engineering and a Masters in ML and Signal Processing, he is currently working in Big Data related consulting projects, developing solutions such as aquaculture fish demand predictions or automatic job offer classification algorithms.

Inmaculada Rabadán-Martín, Ph.D., is an Associate Professor in the Department of Financial Economics, Accounting and Operations Management, at the Faculty of Business Sciences and Tourism, University of Huelva. She has experience teaching in the Business Administration degree course at the University of Huelva. Her research interests include operations management, natural language processing, tourism big data, and hotels' presence in social media.