1

Big Data Analytics and Instagram: an exploratory study on Italian hotel accounts[1]

Tommasina Pianese[*], Giulio Rossetti[+], Virginia Morini[±]

Abstract:

A wealth of tourism-related data is available on the Internet, particularly on social networking sites (SNSs) like Facebook and Instagram. Big data analytics (BDA) allows this large quantity of data to be processed, supported by machine learning and artificial intelligence, and gain an in-depth understanding of traveller preferences and behaviours. With regard to hotels, the analysis of data from SNSs provides countless actionable insights into customers' socio-demographic features, habits, daily trends and brand attitudes. This enables communication to be perfectly targeted, besides supplying valuable information to improve customer satisfaction. Nevertheless, the study of the implications of the automatic processing of data from SNSs in the hotel industry is still in its embryonic state. In order to demonstrate the utility of BDA to understand

[1] **Acknowledgment** This work is supported by the European Union – Horizon 2020 Programme under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement no. 871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

[*] CNR-ISMed, e-mail: tommasina.pianese@ismed.cnr.it - corresponding author.

[+] CNR-ISTI, e-mail: giulio.rossetti@isti.cnr.it

[±] Computer Science Department, University of Pisa (Italy) – CNR-ISTI, e-mail: virginia.morini@phd.unipi.it

how hotels leverage SNSs, we conducted an exploratory study on the Instagram accounts – the photo-sharing SNS known worldwide – of eleven Italian hotels. To this end, the average sentiment score, the average length, lexical diversity and word clouds were calculated on textual data, collected with the instagrapi python package and pre-processed leveraging a standard NLP pipeline. These evidenced different stages of implementation of digital communication on SNSs, shorter text-based messages written on Instagram compared to other SNSs, and specific patterns of user engagement in hotel accounts. BDA also provides information about the online self-promotion process: hotel digital communication is clearly connected to destination, and hashtags are chosen to reach the desired community of travellers.

Keywords: big data, tourism, hotel, social network, Instagram.

1.1 Introduction

Tourism 4.0 is used to refer to the application of key enabling technologies (i.e., Internet of Things, big data, artificial intelligence, virtual realities, etc.) in the tourism industry (Atzeni et al., 2021; Stankov and Gretzel, 2020). Among such technologies, big data analytics (BDA), i.e. processing the information that revolves around tourism, supported by machine learning and artificial intelligence, is a powerful tool for triggering a smart revolution in the tourism industry (Vaid and Kesharwani, 2018). The Internet is a highly suitable resource for collecting such big data. Before choosing a destination, tourists search for information and inspiration for experiences on websites, blogs and social networking sites (SNSs). When planning their travels, individuals use metasearch (e.g. Skyscanner, Trivago) to compare flights and hotel rates, they read comments and reviews on peer-to-peer travel applications (e.g. tourism-related forums and blogs), and make reservations on online travel agencies (OTAs, e.g. Booking.com, Expedia). During their visit, tourists post real-time on SNSs and continue to share comments, photos and/or videos about their stay when back home. They also provide feedback for other travellers on OTAs and platforms like TripAdvisor (Alaei et al., 2019). However, not all travellers trust the reviews uploaded in the different peer-to-peer applications to the same extent (Del Chiappa et al., 2018).

BDA is fundamental for tourism companies to gain, from these data, insights into preferences, expectations and behaviours of travellers. Such information is impossible to obtain via traditional channels and are likely to positively affect their

performance (Ben Youssef and Zeqiri, 2022). The mining of this large set of data also reveals hotel brand positioning and the main attributes differentiating key competitors (Hu and Trivedi, 2020).

An enormous amount of tourism-related data comes from SNSs, particularly Facebook and Instagram, that have progressively reshaped the tourism landscape (Jimenez-Marquez et al., 2019). Travellers have started to use SNSs to share online travel experiences with virtual communities. As most travellers engage with SNSs, tourist destinations have realized the opportunities to exploit them to enhance their attractiveness (Nixon, 2022). In the last few years, the hotel industry has also been compelled to engage with SNSs, with two mutually connected functions. They are an effective communication channel to reach customers at any time at any place and reinforce the hotel brand. At the same time, analysis of data from SNSs (i.e. BDA) provides hotels with countless actionable insights about customers' socio-demographic features, habits, daily trends and brand attitudes. This enables perfectly targeted communication to be implemented, besides supplying valuable information to improve customer satisfaction (Stankov and Gretzel, 2020).

While SNSs have attracted academic interest as a tool to be leveraged for hotel branding (Leung et al., 2015), few studies have so far exploited big data collected from hotel SNSs (Li et al., 2018). Of the latter, some have used BDA to classify posts on Facebook hotel accounts (e.g. O'Connor, 2011). However, there is a paucity of studies on data uploaded on Instagram hotel accounts (e.g. Kapoor et al., 2022). Nonetheless, Instagram offers many more business opportunities for hotels. It is an online travel album, where hotels and travellers upload millions of photos every day and interact by putting a like, by commenting, and using hashtags and mentions (Distel et al., 2022). Instagram allows the visibility of unknown locations to be rapidly raised and potential customers to be reached by paying for sponsorship and/or influencers, or through travellers' e-word-of-mouth. Even more important is that Instagram users are more engaged with business profiles, i.e., they visit hotel profiles, often suggested by the platform. In this regard, Instagram monitors search users' preferences, and exploits AI-driven search tools to offer every user contents that they will most likely appreciate.

With this in mind, this chapter aims to further deepening the scientific debate related to this somehow overlooked research area. Specifically, our aim was to demonstrate the utility of big data analytics to obtain insights into how hotels leverage Instagram in their digital communication. To this end, an exploratory study was conducted on the Instagram accounts of eleven appropriately selected Italian hotels. The word-based contents (i.e. captions) of the 11,095 Instagram posts published since the creation of the hotel accounts on the popular SNSs until April 2022 were analysed to extract descriptive features for quali/quantitative analysis. The ability of the hotels to engage their virtual community (i.e. with likes and comments) was also analysed.

The chapter enriches the literature on BDA in the tourism industry by showing the insights offered by big data analytics in understanding how hotels strive to communicate on Instagram (Go and You, 2016; Xiang et al., 2015). In so doing, we focus on the under-explored perspective of hotel managers, while most studies to date have focused on the users of SNSs. Our multidisciplinary approach to the phenomenon is also original, as we have combined data mining with managerial competences.

This chapter is structured as follows. First, we present a critical review of the literature addressing big data and SNSs in the tourism industry, including the perspectives of tourist destinations and hotels. After describing the research method, we present the findings obtained from the analysis of Instagram hotel accounts with some of the most widely used tools and techniques. In the last section, we discuss how this study enriches the existing literature, and then present the practical implications, limitations and future research directions.

1.2 Literature review

To date, the literature addressing big data in the tourism industry has focused on data classification (Li et al., 2018) and data analytics (Li and Law, 2020). Based on their sources, the massive volume of tourism data available on the Web has been narrowed down to three categories, namely 1) UGC data, generated by users, including online textual and photo data; 2) device data, generated by devic-

es, including e.g. GPS data, mobile roaming data; 3) transaction data, generated by operations, including such fields as web search data and webpage visiting data (Li et al., 2018).

A double-stage framework intended to inform analysis of tourism-related data was provided by Jimenez-Marquez et al. (2019). The first stage involved the sequencing of data, while the second aimed to create a big data architecture for automatically processing unstructured heterogeneous data. As part of BDA, sentiment analysis has gained momentum in tourism research, especially to determine the contextual positive, neutral or negative polarity of a text document, an opinion, or an emotion expressed in online UGC reviews and blogs (Alaei et al., 2019).

Other studies have focused on the opportunities that BDA offers tourism destinations and hotels (e.g., Mariani, 2019). BDA assists destination management organizations (DMOs) in obtaining real-time knowledge on tourist behavioural patterns. Thus, knowing when and how many photos were taken in a place plays a role in destination management strategic plans (Miah et al., 2017). Likewise, analysis of multiple data with BDA (e.g. website traffic, weather information) has been shown to be valuable to forecast hotel occupancy at destinations (Pan and Yang, 2017).

BDA may well also support performance by enabling hotels to forecast booking cancellations and profile clients (cfr. Zarezadeh et al., 2022). Likewise, analysis of reviews on platforms such as TripAdvisor (Alaei et al., 2019) allows customer satisfaction to be predicted (Zhao et al., 2019), reveal brand positioning (Hu and Trivedi, 2020), identify the factors that most influence the hotel guest experience (Xiang et al., 2015) and detect differences in guest preferences based on their culture and language (Liu et al., 2017).

Social networking sites constitute a large resource for collecting big data on tourism supply and demand (Jimenez-Marquez et al., 2019). DMOs use SNSs to promote a destination brand (Fatanti and Suyadnya, 2015). The hospitality industry leverages SNSs to build online reputations. Travellers use SNSs to portray, reconstruct and relive their trip, which is often instrumental to self-promote a desired image (Stankov and Gretzel, 2020).

In the literature analysing tourism-related data obtained from SNSs, there is a prevalence of studies on the Instagram accounts of tourism destinations (e.g. Mükisch and Surkic, 2022). By analysing photos, *mise en scène* and captions, Mele et al. (2021) found that the communication of DMOs was strongly influenced by national culture. BDA can also prove that the congruence between the destination "projected image" and "perceived image" on Instagram is relevant to travellers' intention to visit a place (Egger et al., 2022; Nixon, 2022). The use of emoji allows DMOs to express an emotional mood in Instagram posts, help them convey the message effectively, reduce ambiguity and solve misunderstandings in intercultural communication (Distel et al., 2022). Both face and non-face emoji add positive emotions to a destination post, even though face emoji, associated with a beautiful picture, are found to have a more positive effect on the desire to visit a destination. Finally, BDA reveals that sustainably-oriented influencers – that is third-party endorsers who shape audience attitudes based on their experiences and opinions- support a destination communication strategy aimed at developing sustainable tourism (Palazzo et al., 2021).

As for the hotel perspective, in-depth knowledge of travellers' needs through the automatic analysis of the data from SNSs is likely to trigger a dynamic of innovation, thus keeping abreast of the latest tourism trends (Garrido-Moreno and Lockett, 2016). Despite the variety of SNSs, researchers have largely focused on Facebook usage in the hotel business (e.g. Leung and Baloglu, 2015) and mostly limited the use of BDA to classifying posts. Phelan et al. (2013) distinguish posts related to presenting content features (e.g. upcoming promotions), providing property information (e.g. hotel contact details), and favouring interactions with customers (e.g. answers to customers' comments). Minazzi and Lagrosen (2013) further introduced the category "call to action", implying an explicit invitation to users to engage in conversations.

Other scholars have used BDA to understand the features to make Facebook posts effective, revealing the importance of using simple words in a concise text to ensure that users read the entire posts (Lo and Fang, 2018). Leveraging humour, posing questions and showing hotel pictures are appreciated by Facebook users as well (Su et al., 2015).

As for travellers' engagement, researchers distinguish between passive users (i.e., they take knowledge but do not contribute to the community) and active users (i.e., they engage in community activities such as sharing information) (Su et al., 2015). Further measurements of user engagement are: "likes", "comments" and/or "share" the contents or the "check-in" function (Lo and Fang, 2018; Phelan et al., 2013).

Although Instagram provides relevant business-oriented data, the study of the implications of the automatic processing of such data in the hotel industry is still in an embryonic state (Asanbekova and Maksudunov, 2018; Kapoor et al., 2022; Kurniawan et al., 2021). An exception is Kapoor et al. (2022) who used BDA to demonstrate that "objective" environmental-oriented messages on Instagram posted by influencers were more effective than "emotional" messages in demonstrating to travellers a hotel's commitment to sustainability, thus influencing their intention to stay at the eco-friendly hotel.

1.3 Research method

For the purposes of our research, an exploratory study was conducted to demonstrate the utility of big data analytics to obtain insights into how hotels leverage Instagram in their digital communication. Instagram is one of the most popular SNSs worldwide, with roughly one billion monthly active users. Italy represents one of the most active countries with 27.5 million users. At the time of this study, more than 639 millions Instagram posts from governments, tourism bodies, local businesses, tour guides, travellers over the world used the hashtag #travel, along with thousands of secondary travel-related hashtags. In 2021, the World Tourism Organization established a partnership with Instagram considered to be a powerful platform for digitally reconnecting with travellers around the world during and after the Covid-19 pandemic (UNWTO, 2021). Its main features are the sharing of contents (i.e. posts) with virtual communities (i.e. followers). Each post includes a photo (i.e. max 10 as an album) or videos (i.e. lasting 15, 30 or 60 seconds), which are accompanied with a caption, and usually by hashtags (i.e. labels to catalogue the content on certain themes), tag (i.e. aimed at associating the content with another account) and/or geographic tag (i.e.

where content was created). Another popular feature are the Stories, that is photos and videos that are live for others to view for 24 hours, except for those that are kept in the profile account as "featured stories". Indicators of user engagement are the following functions: like, comment and share (Su et al., 2015). As for hotel selection, we adopted the method of convenience sampling, which collects data from a conveniently selected pool of respondents. We considered the eleven Italian hotels included in the 2020 Gold List of top hotels in Europe drawn up by Condé Nast Travellers, i.e an annual list of best European hotels selected by the CNT's global crew. This list enabled to select a sample of hotels that are homogeneous for their qualitative level as judged by a large group of experts, but heterogeneous for their social presence. In Table 1.1, we report the hotel and Instagram name, the number of followers and posts at April 2022 (Asanbekova and Maksudunov, 2018). This sample encompassed a variety in terms of Instagram activation (between 2013 and 2021), number of followers (ranging from 168,000 to 3,700) and posts (ranging from around 2,300 to 422).

Content analysis was conducted on the hotel Instagram accounts (Krippendorff, 2018), which is a method frequently adopted in tourism-related studies (Mele et al., 2021). Our focus was on text messages, that is word-based contents (i.e. captions) published in their accounts.

Table 1.1 Italian hotel Instagram accounts included in the sample.

Hotel name	*Instagram account*	*Starting from mm/year*	*Number of followers*	*Number of posts*	*Language*
Aman, Venice	*@aman_venice*	*11/2016*	*52,500*	*382*	*Eng.*
Belmond Hotel Cipriani, Venice	*@belmondhotelcipriani*	*09/2016*	*70,400*	*1,490*	*Eng. & Ital.*
Belmond Hotel Splendido, Portofino	*@belmondhotelsplendid*	*11/2015*	*143,000*	*1,299*	*Eng. & Ital.*
Borgo Santo Pietro, Siena	*@borgosantopietro*	*04/2018*	*159,000*	*1,249*	*Eng. & Ital.*

Grand Hotel Tremezzo, Lake Como	*@ghtlakecomo*	*07/2018*	*168,000*	*2,283*	*English*
Hotel De Russie, Rome	*@hotelderussie*	*09/2015*	*52,800*	*975*	*English*
Hotel Vilòn, Rome	*@hotelvilon*	*07/2017*	*10,400*	*422*	*English*
San Pietro di Positano, Amalfi Coast	*@ ilsanpietrodipositano*	*07/2013*	*93,500*	*882*	*Eng. & Ital.*
Palazzo Venart, Venice	*@palazzovenart*	*01/2017*	*3,700*	*653*	*Eng. & Ital.*
The Gritti Palace, Venice	*@thegrittipalace*	*06/2016*	49,000	*1,347*	*English*
The Place, Florence	*@theplacefirenze*	*01/2021*	14,800	*113*	*English*

Source: authors' calculations.

As for data collection, in March 2022 we crawled the Instagram accounts of the eleven Italian hotels included in our sample. Leveraging the instagrapi python package, a wrapper around the official Instagram API (application programming interface), we collected all hotels' shared contents focusing on: their date of publication, type (photos, videos, albums), captions (including hashtags and mentions) and generated engagement (i.e. number of likes and number of comments).

The collected textual data were pre-processed leveraging a standard NLP pipeline (i.e., removing non-printable characters and punctuation) and analysed to extract descriptive features for a quanti/qualitative analysis. We computed, for each hotel, the average sentiment score (e.g., negative, positive, neutral and compound scores according to VADER[2] (Hutto et al., 2014), the average length

[2] Valence Aware Dictionary and sEntiment Reasoner (VADER) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. A sentiment lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative. VADER not only gives the positivity and negativity score but also tells us about how positive or negative a sentiment is.

(i.e., number of words) and lexical diversity (i.e., the ratio of unique words to the total number of words (McCarthy et al., 2005) of its posts. We focused on these textual measures since they allowed us to assess whether the vocabulary used, and the sentiments expressed by each hotel reveal different narrative styles. Moreover, we qualitatively summarised the employed narratives by means of word cloud visualization (Halvey, 2007), thus detecting the most relevant terms used by each hotel for online ads. We filtered terms, taking into account both their observed absolute frequency and their uniqueness, i.e., according to the TF-IDF measure (Luhn, 1957; Jones, 1972), including hashtags and emoji.

Finally, leveraging each hotel posts' metadata, we analysed online presence and activity in terms of: average number of published posts per day, content type preferences, observed distribution of likes/comments, correlation of received comments and likes for posts, use of hashtags for brand building and relative frequency of users' mentions.

1.4 Understanding hotel usage of Instagram through BDA

BDA showed that the Instagram hotel accounts were managed continuously and, on average, at least one content was published per day on this photo-sharing SN. This daily update, which is not linked to certain events, was important to ensure a "social presence", in other words to maintain an active and continuous interaction with their followers, thereby boosting the perceptions of being engaged in a relationship (Kietzmann et al., 2011; Minazzi and Lagrosen, 2014).

Moving on to type preferences, by automatically analysing the 11,095 posts published on Instagram, we found that all hotels preferred to publish a single photo which resembled virtual postcards. Slight differences in the communication style emerged for albums and videos (see Figure 1.1), with some hotels, like @borgosantopietro, making greater use of videos to share with followers the large property and narrate the available travel experiences like cookery classes, cheese making, wine fermentation, olive oil extraction and also river swimming.

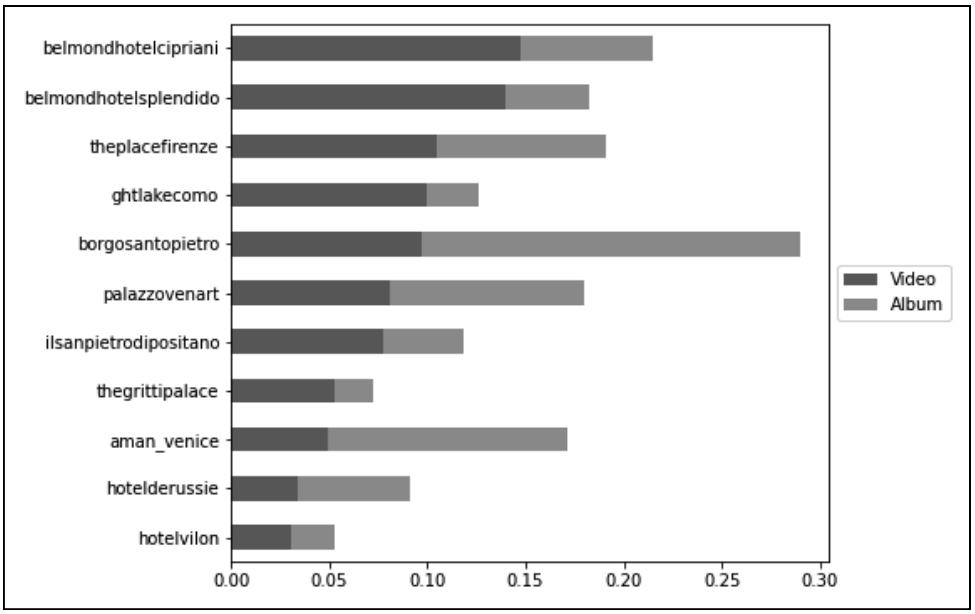

Figure 1.1 Content type preferences

Source: authors' calculations.

Beyond the type used, posts were always accompanied by a text-based caption which, according to the Avg Text Length, was between 50 and 60 words long for all hotels, including those written in two languages (i.e. English and Italian). In contrast, analysis of the lexical diversity in all the text-based captions written on the hotel Instagram accounts revealed a different level of variety in terms of vocabulary, such that the terminology was richer for some hotels (e.g. @theplacefirenze) and less for others (e.g. @belmondhotelsplendido) (see Figure 1.2).

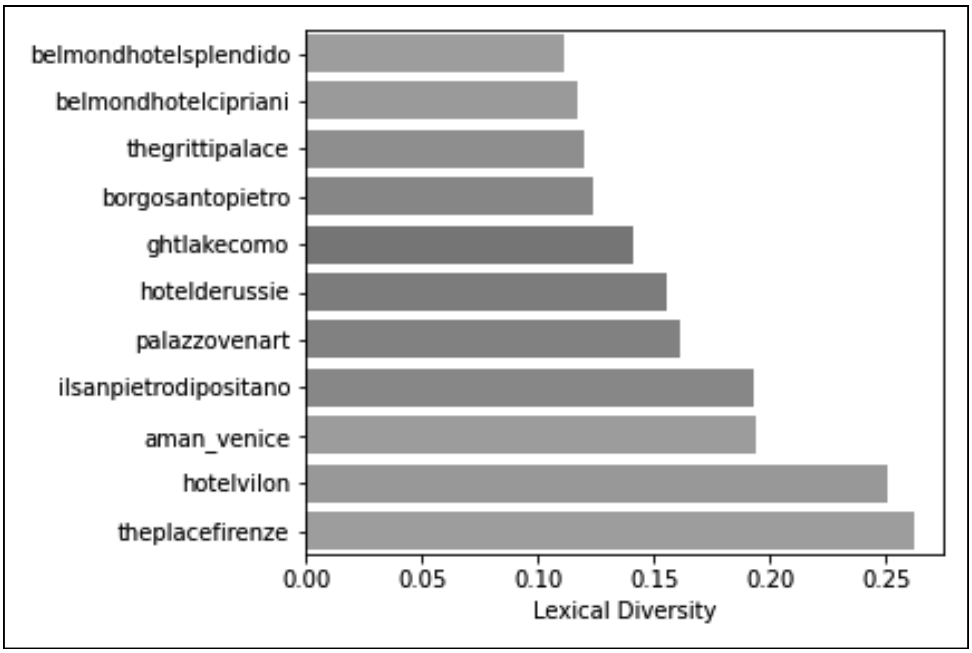

Figure 1.2 Lexical diversity

Source: authors' calculations.

Captions usually also included "hashtags" aimed at connecting the post to broader online themes and discourses, as well as "mentions" to engage a wider audience and increase the message relevance (see Table 1.2). Both Instagram functions aim to rapidly create a community of connected individuals (i.e. users with similar interests) which is larger than the hotel community.

Table 1.2 Distribution of hashtags and mentions.

Hotel brand	Hashtags			Mentions		
	Total	Unique	Avg. per Post	Total	Unique	Avg. per Post
aman_venice	1503	271	3.9	140	117	0.4

belmondhotelcipriani	2876	225	2.9	788	476	0.8
belmondhotelsplendido	2974	176	3.0	934	586	0.9
borgosanpietro	17679	16119	17.7	225	118	0.2
ghtlakecomo	9610	1767	9.7	716	406	0.7
hotelderussie	7465	1494	7.6	907	603	0.9
hotelvilon	918	322	2.3	77	56	0.2
ilsanpietrodipositano	7642	1549	8.7	142	92	0.2
palazzovenart	7150	947	11.0	309	197	0.5
thegrittipalace	12367	1275	12.4	1068	533	0.5
theplacefirenze	233	31	2.0	165	72	1.4

Source: authors' calculations.

Interestingly, looking at the @thegrittipalace, we observed their extensive (average per post) use of hashtags (i.e. 12.4) aiming essentially to connect user-generated contents to the hotel account, thus enriching their online narration, amplifying the hotel's visibility and engaging more with its community. In this regard, they did not limit use of hashtags to the hotel (e.g. #thegrittipalace, #ohsogritti), but extended the hashtags to the wider tourist destination (i.e. #venezia, #venice) and generally to luxurious living (e.g. #theluxurycollection). As for the mentioning function, hotels belonging to international chains were more prone to connect their post to the brand chain (e.g. Belmond) and to sister properties in other destinations (Garrido-Moreno and Lockett, 2016; O'Connor, 2011).

The aim of sentiment analysis was to use an automatic tool to extract relevant information concerning hotel communication on discourses from text-based captions on Instagram. As for polarity, the language used in the text-based captions of Instagram posts was mainly neutral, resembling a showcase where someone is describing to the virtual community the spaces and experiences which guests may enjoy at the hotel. In contrast, we found slight differences, as some hotels deliberately opted for a more "positive" communication language for their hedonic online promotion. This neutral (or slightly positive) polarisation is quite surprising given that the period of time under analysis included approximately two years of the pandemic, during which there were entire months of lockdown and restrictions on travelling and commuting.

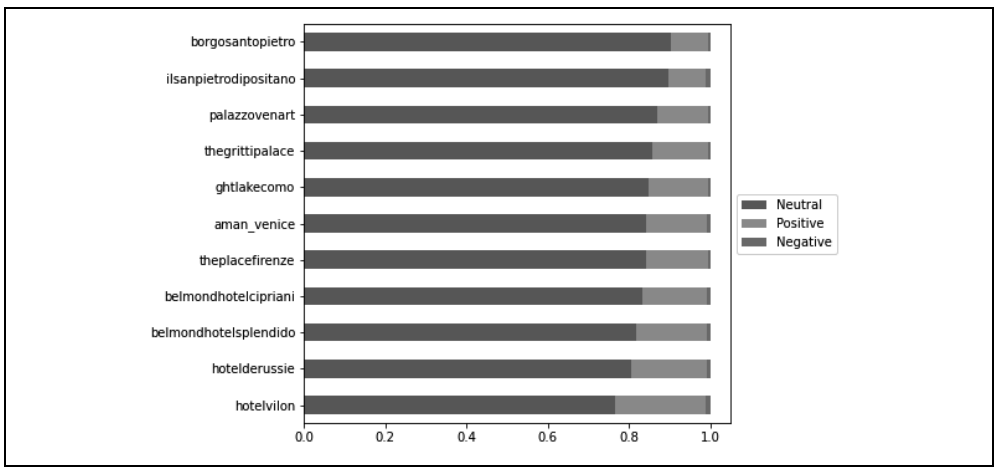

Figure 1.3 Sentiment analysis

Source: authors' calculations.

Word cloud visualization (Halvey, 2007), generated by leveraging the "wordcloud" python package (https://pypi.org/project/wordcloud/) starting from a TF-IDF analysis of hotel texts, enabled us to detect the most relevant terms used by each hotel on their Instagram account, with the inclusion, or otherwise, of hashtags and emoji used to reinforce the message.

In this regard, a common feature that emerged for most hotels was that the recurrent words included the hotel name and, most importantly, the tourist destination where the hotel was based. Indeed, all hotels considered in this study are located in world-renowned destinations and their brand names have been extensively exploited in the hotel online communication. From figure 5 below which reports the findings for the Grand Hotel Tremezzo, Lake Como, Lombardy (Instagram account @ghtlakecomo), it emerges that "Lake" and "Como" are the most frequent words, followed by "grand tremezzo". "Gualtiero Marchese", an internationally acclaimed chef based in the hotel restaurant, and "Bellaggio", the residence of a famous American actor, were recurrent words as well.

Figure 1.4 Word clouds for the @ghtlakecomo without (left) and with hashtags (right)

Source: authors' calculations.

Finally, as for likes and comments, which are indicators of user engagement (Asanbekova and Maksudunov, 2018), the selected hotels underlined different patterns. In figure 1.5, for each hotel, succinct representations of likes (left) and comments (right) over the posted contents are reported using boxplots, a standardised way to summarise the interquartile range (i.e., the box containing data points between the 25th and 75th percentile of the distribution), the median (i.e., the vertical line within the box), and the minimum and maximum observed distribution values (i.e., the whiskers on the left and right of the box), having removed outliers. The absence of symmetricity in hotel boxplots underlines the fact that the observed data follow a Gaussian, rather than a right-skewed distri-

bution (long-tailed in the case of likes). Most of the observed hotels register a median value of likes within the range [500, 800] (with a few notable exceptions reporting a lower median rate of user likes - e.g., palazzovenart, hotelvilon). However, the different sizes of the interquartile range, as well as the length of the right whisker, suggest significant differences on the ability to share contents able to reach high user endorsement (e.g., thegrittipalace is able to attract 5000 likes to some of its contents, while, on the other extreme, palazzovenart attracts at most 30 likes per post). The distribution of comments per post (right) underlines a similar behaviour, despite there being a few notable exceptions. We can observe how different hotels engage users differently by comparing the two figures. As an example, consider thegrittipalace: its contents are those attracting more likes (left) while, at the same time, being able to generate only a modest volume of users' comments.

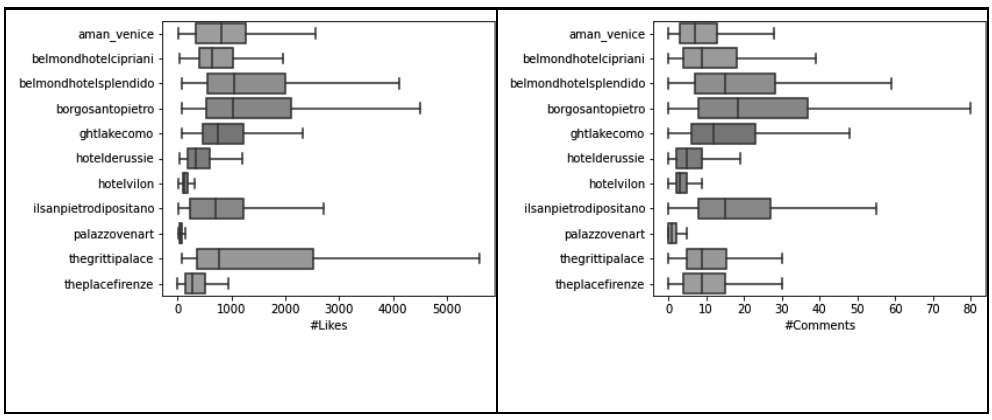

Figure 1.5 Boxplot of distributions of likes (left) and comments (right)

Source: authors' calculations.

When analysing the Pearson correlation (Benesty, 2009) among hotels' received likes and comments, there was usually a positive trend (individual values ranging between [0.45, 0.8]): the more frequently a hotel post is commented, the more likely it has also received a large number of likes. However, depending on the hotel, the engagement variable with the highest growth rate may vary (e.g., the number of likes for thegrittipalace, the number of comments for theplacefiren-

ze). Although all hotels reported a level of user engagement, none of them was able to optimise both liking and commenting functions.

1.4 Discussion and conclusions

By demonstrating the utility of big data analytics for hotel digital communication, this study enriches the surprising paucity of research that has exploited big data from SNSs to overcome limited survey samples and sampling bias, and gain insight into tourism-related issues (e.g. Li et al., 2018; Xiang et al., 2015). At the aggregate level, BDA allows us to detect different stages in implementing digital communication by hotels (Effing and Spil, 2016), intended as the process by which hotels employ Instagram strategically, "for customer-facing purposes, by producing content regularly, engaging customers in an ongoing relationship and generating analytics and customer insights to drive strategic marketing actions" (Tafesse and Wien, 2018). It assumes a social media strategy, active presence, customer engagement initiatives and social media analytics (Tafesse and Wien, 2018). For example, @palazzovenart, despite activating its profile in 2017, has published few posts, has few followers and little interaction with them.

Use of concise and essential language confirms previous studies on other SNSs that have shown that short messages effectively raise the interest of the virtual community: users read only parts of the text messages if they are long (Lo and Fang, 2018). However, our study shows that hotels, on average, use fewer words in captions of Instagram posts than Facebook, and this may be connected to its intrinsic nature as a social networking site that attracts more via images rather than text-based posts (Lo and Fang, 2018).

BDA, specifically word cloud visualization, by detecting the most recurrent terms used by hotels in their text-based messages including hashtags on Instagram, provides information about the online self-promotion process, with hotel digital communication strongly connected to destination character, which reflects its tangible along with intangible, experiential and symbolic nature (Chi et al., 2018). "Lifestyle experiences" are emphasised as well through the deliberate choice of

hashtags that clearly position the hotel in the community of travellers paying attention to hyper-personalised services and experiences (Han and Lee, 2021).

BDA also reveals specific patterns of user engagement in hotel Instagram accounts. We show that "like" and "comments", considered the main metrics for follower participation, are actually not related to each other: having many likes is not the same as having many comments (Phelan et al., 2013). However, there is still the need to understand how to detect the activity and hence engagement of passive users (Su et al., 2015).

Digital presence has become essential, and hotels are now forced to widen the scope of their traditional marketing and branding strategies to include virtual communities. One of the challenges concerns the extent to which hotel managers are able to collect and systematically analyse the vast amount of data available on the Internet, including SNSs, to implement effective digital communication. Our findings strongly suggest that hotel managers invest in infrastructure (i.e., software for BDA) and human resources, including data scientists working closely with social media managers, to interpret data insights and translate them into concrete strategies and actions. This would also enable new tourist trends to be defined and help determine the hotel's position in relation to that of others at the destination in question.

This study is not without its limitations. We considered a limited number of hotels in popular Italian destinations, but are aware that there are other types of hotels (e.g. in rural areas) with a successful presence on social networking sites that are worth studying. Further, Instagram data, whose access was partially limited by the company owner Meta, would need to be enriched with context analysis and triangulated with data collected from other sources to reduce potential bias and increase data reliability. Finally, the techniques used for data analysis allowed to understand "how" (e.g. posts length), but not "what" hotels communicated in their posts on Instagram. Nonetheless, going into details about what hotels are saying would be extremely important to know.

Despite the above limitations, this study opens up avenues for future research. Importantly, researchers could use big data analytics to obtain insights into such aspects as the hotel brand from the large amount of data available on the Web,

including tourist and non-tourist related applications. Another major avenue to pursue would be to conduct research including hotel websites (i.e. the hub of digital communication) and different SNSs exploited with different purposes in order to arrive at an understanding of the overall social media strategy.

References

Alaei, A. R., Becken, S., Stantic, B. "Sentiment analysis in tourism: capitalizing on big data", *Journal of Travel Research*, 2019, n. 58 (2), pp. 175-191.

Asanbekova, M., Maksudunov, A. "The Marketing Power of Instagram: A Content Analysis of Top Hotel Enterprises in Kyrgyzstan", *Uluslararası Türk Dünyası Turizm Araştırmaları Dergisi*, 2018, n. 3 (2), pp. 141-152.

Atzeni, M., Del Chiappa, G., Mei Pung, J. "Enhancing visit intention in heritage tourism: The role of object-based and existential authenticity in non-immersive virtual reality heritage experiences", *International Journal of Tourism Research*, 2021, pp. 1-16.

Ben Youssef, A., Zeqiri, A. "Hospitality industry 4.0 and climate change", Circular Economy and Sustainability, 2022, pp. 1-21.

Benesty, J., Chen, J., Huang, Y., Cohen, I. "Pearson correlation coefficient", In Noise reduction in speech processing, Springer, Berlin, Heidelberg, 2009, pp. 1-4.

Chi, C. G. Q., Pan, L., Del Chiappa, G. "Examining destination personality: Its antecedents and outcomes" *Journal of Destination Marketing & Management*, 2018, n. 9, pp. 149-159.

Del Chiappa, G., Lorenzo-Romero, C., Alarcon-del-Amo, M. D. C. "Profiling tourists based on their perceptions of the trustworthiness of different types of peer-to-peer applications", *Current Issues in Tourism*, 2018, n. 21 (3), pp. 259-276.

Distel, V., Egger, R., Petrovic, U., Phan, V. L., Wiesinger, S. "The Usage of Emoji in Tourism-Related Instagram Posts: Suggestions from a Marketing Perspective", in

J.L. Stienmetz, B. Ferrer-Rosell, D. Massimo (eds), ENTER22 Information and Communication Technologies in Tourism, Cham: Springer, 2022, pp. 134-145.

Effing, R., Spil, T. A. "The social strategy cone: Towards a framework for evaluating social media strategies", *International Journal of Information Management*, 2016, n. 36 (1), pp. 1-8.

Egger, R., Gumus, O., Kaiumova, E., Mükisch, R., Surkic, V. "Destination Image of DMO and UGC on Instagram: A Machine-Learning Approach", in J.L. Stienmetz, B. Ferrer-Rosell, D. Massimo (eds), ENTER22 Information and Communication Technologies in Tourism, Cham: Springer, 2022.

Fatanti, M.N., Suyadnya, I. W. "Beyond user gaze: How Instagram creates tourism destination brand?", *Procedia-Social and Behavioral Sciences*, 2015, n. 211, pp. 1089-1095.

Garrido-Moreno, A., Lockett, N. "Social media use in European hotels: benefits and main challenges", *Tourism & Management Studies*, 2016, n. 12 (1), pp. 172-179.

Halvey, M. J., Keane, M.T. "An assessment of tag presentation techniques". *Proceedings of the 16th international conference on World Wide Web*, 2007.

Han, Y., Lee, H. "Lifestyle Experiences: Exploring Key Attributes of Lifestyle Hotels Using Instagram User-Created Contents in South Korea", *Sustainability*, 2021, 13 (5), p. 2591.

Hu, F., Trivedi, R. H., "Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content", *International Journal of Hospitality Management*, 2020, n. 84, 102317.

Hutto, C., Gilbert E. "Vader: A parsimonious rule-based model for sentiment analysis of social media text". *Proceedings of the international AAAI conference on web and social media*, 2014, n. 8 (1).

Jimenez-Marquez, J. L., Gonzalez-Carrasco, I., Lopez-Cuadrado, J. L., Ruiz-Mezcua, B. "Towards a big data framework for analyzing social media content", *International Journal of Information Management*, 2019, n. 44, pp. 1-12.

Jones, K. S. "A statistical interpretation of term specificity and its application in retrieval". Journal of documentation, 1972.

Kapoor, P. S., Balaji, M. S., Jiang, Y., & Jebarajakirthy, C. "Effectiveness of travel social media influencers: a case of eco-friendly hotels", *Journal of Travel Research*, 2022, n. 61(5), pp. 1138-1155.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., Silvestre, B. S. "Social media? Get serious! Understanding the functional building blocks of social media", *Business horizons*, 2011, n. 54 (3), pp. 241-251.

Krippendorff, K. Content analysis: An introduction to its methodology. Sage publications, 2018.

Kurniawan, Y., Caesar, G. C., Maliki, W. P., Christabel, K., Anwar, N., "Analysis of Social Media Instagram Effectivity in Hotel Industries", paper presented at the 2021 International Conference on Information Management and Technology (ICIMTech), 2021, n. 1, pp. 458-463.

Leung, X. Y., Bai, B., Stahura, K. A. "The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter", *Journal of Hospitality & Tourism Research*, 2015, n. 39 (2), pp. 147-169.

Leung, X.Y., Baloglu, S. "Hotel Facebook marketing: an integrated model", *Worldwide Hospitality and Tourism Themes*, 2015, n. 7(3), pp. 266-282.

Li, J., Xu, L., Tang, L., Wang, S., Li, L. "Big data in tourism research: A literature review", *Tourism Management*, 2018, n. 68, pp. 301-323.

Li, X., Law, R. "Network analysis of big data research in tourism", *Tourism Management Perspectives*, 2020, 33, 100608.

Liu, Y., Teichert, T., Rossi, M., Li, H., Hu, F. "Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews", *Tourism Management*, 2017, n. 59, pp. 554-563.

Lo, Y. C., Fang, C. Y. "Facebook marketing campaign benchmarking for a franchised hotel", *International Journal of Contemporary Hospitality Management*, 2018, 30 (3), pp. 1705-1723.

Luhn, H. P. "A statistical approach to mechanized encoding and searching of literary information", *IBM Journal of research and development 1.4*, 1957, pp. 309-317.

McCarthy, P.M., "An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)", Ph. Diss. The University of Memphis, 2005.

Mele, E., Kerkhof, P., & Cantoni, L. "Analyzing cultural tourism promotion on Instagram: a cross-cultural perspective", *Journal of Travel & Tourism Marketing*, 2021, n. 38 (3), pp. 326-340.

Nixon, L. "Do DMOs Promote the Right Aspects of the Destination? A Study of Instagram Photography with a Visual Classifier", in J. L. Stienmetz, B. Ferrer-Rosell, D. Massimo (eds), ENTER22 Information and Communication Technologies in Tourism, Cham: Springer, 2022, pp. 174-186.

O'Connor, P. "An analysis of the use of Facebook by international hotel chains", paper presented at the ICHRIE Conference, 2011, July 27th, University of Massachusetts Amherst.

Ong, Y. X., Sun, T., Ito, N. "Beyond Influencer Credibility: The Power of Content and Parasocial Relationship on Processing Social Media Influencer Destination Marketing Campaigns", in J. L. Stienmetz, B. Ferrer-Rosell, D. Massimo (eds), ENTER22 Information and Communication Technologies in Tourism, Cham: Springer, 2022.

Palazzo, M., Vollero, A., Vitale, P., Siano, A. “Urban and rural destinations on Instagram: Exploring the influencers’ role in #sustainabletourism”, *Land use policy*, 2021, n. 100, 104915.

Pan, B., Yang, Y. “Forecasting destination weekly hotel occupancy with big data”. Journal of Travel Research, 2017, n. 56 (7), pp. 957-970.

Phan, V. L., Wiesinger, S. “The Usage of Emoji in Tourism-Related Instagram Posts: Suggestions from a Marketing Perspective”, in J.L. Stienmetz, B. Ferrer-Rosell, D. Massimo (eds), ENTER22 Information and Communication Technologies in Tourism, Cham: Springer, 2022.

Phelan, K. V., Chen, H. T., Haney, M. ““Like” and “Check-in”: how hotels utilize Facebook as an effective marketing tool*”, Journal of hospitality and Tourism Technology*, 2013, n. 4 (2), pp. 134-154.

Stankov, U., Gretzel, U. “Tourism 4.0 technologies and tourist experiences: a human-centered design perspective”, *Information Technology & Tourism*, 2020, n. 22 (3), pp. 477-488.

Su, N., Reynolds, D., Sun, B. “How to make your Facebook posts attractive: A case study of a leading budget hotel brand fan page”, *International Journal of Contemporary Hospitality Management*, 2015, n. 27 (8), pp. 1772-1790.

Tafesse, W., Wien, A. “Implementing social media marketing strategically: an empirical assessment”, *Journal of Marketing Management*, 2018, n. 34 (9-10), pp. 732-749.

UNWTO, Recover and Rediscover. An Instagram Guide to Tourism Recovery for Governments and Small Businesses, 2021, available at https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2021-11/recover-rediscover-an-instagram-guide-to-tourism-recovery-for-governments-and-small-businesses-b.pdf?VersionId=Ruv2mo8YJTOcJ4JqZ2N0MvlnAaJHrXN4

Vaid, J., Kesharwani, S. “Role of big data analytics in social media marketing of MICE tourism”, *Global Journal of Enterprise Information System*, 2018, n. 10 (1), pp. 55-61.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., Uysal, M. “What can big data and text analytics tell us about hotel guest experience and satisfaction?”, *International Journal of Hospitality Management*, 2015, n. 44, pp. 120-130.

Zarezadeh, Z. Z., Rastegar, R., Xiang, Z. “Big data analytics and hotel guest experience: a critical analysis of the literature”, *International Journal of Contemporary Hospitality Management*, 2022.

Zhao, Y., Xu, X., Wang, M. “Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews”, *International Journal of Hospitality Management*, 2019, n. 76, pp. 111-121.