Measuring destination image: a novel approach based on visual data mining

A methodological proposal and an application to European islands

Anastasia Arabadzhyan[1]

Paolo Figini[2*]

Laura Vici[3]

Abstract

Availability of User Generated Content and the development of Big Data and machine learning algorithms have paved the way to collecting and analysing great volumes of data. We scan imagery data from travelling-related posts on Instagram to identify the key features of the destination image and of its dynamics. Specifically, we exploit a newly introduced Visual Object Recognition tool (Google Cloud Vision) to convert into textual labels the content of about 860,000 travel-related pictures posted on Instagram in Summer 2019 for several European islands. The output, a vector of labels' frequencies on a very fine-grained scale, is used to proxy the destination image at different points in time. We then introduce the Index of Distance in Destination Image, a metric built on the pictures' labels ranking, and aimed at providing a quantitative measure of (dis)similarity between destination images. We show that the analysis of labels and the index are fit to compare destinations cross-sectionally and over time, providing a useful tool for researchers, marketers and DMOs. We also deliver evidence on how external shocks (like extreme events linked to climate change) or the organization of events modify the cognitive sphere of the destination image, with repercussions on activities undertaken by tourists and relevant implications for local policies.

Keywords

destination image; data mining; image recognition; user generated content; external events.

Published on the Journal of Destination Marketing and Management. Please cite it as:

Arabadzhyan, A., Figini, P., Vici, L., 2021. Measuring destination image: a novel approach based on visual data mining. A methodological proposal and an application to European islands, *Journal of Destination Marketing and Management*, DOI: https://doi.org/10.1016/j.jdmm.2021.100611

---

[1] CAST – Centre for Advanced Studies in Tourism, University of Bologna.

[2] Department of Economics and CAST – Centre for Advanced Studies in Tourism, University of Bologna.

[*] Corresponding author: Paolo Figini, Department of Economics, Piazza Scaravilli 2, I-40126, Bologna, Italy. E-mail: paolo.figini@unibo.it. Research for this paper has been supported by European Union's Horizon 2020 research and innovation program under grant agreement No 776661, project "SOCLIMPACT - DownScaling CLImate imPACTs and decarbonisation pathways in EU islands and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and beyond". We thank Ali Gholami for data scraping and data organization and participants of the 2019 CBTS held in Bruneck (Italy) and of the 2020 CAST days held in Rimini (Italy) for their comments. The usual disclaimers apply.

[3] Department of Statistics and CAST – Centre for Advanced Studies in Tourism, University of Bologna.

Measuring destination image: a novel approach based on visual data mining

A methodological proposal and an application to European islands

1. Introduction

Traditionally, destination image was mainly thought of as the outcome of marketing and branding strategies, centrally decided by Destination Management Organizations (DMOs), tourism stakeholders or policy makers. Communication was a one-way process, where DMOs and private tourism businesses were transmitting predetermined information and images to consumers (through brochures, websites, travel books, promotional material, etc.), shaping tourist gaze (Urry, 2002). In this framework, tourists were considered as passive consumers who could just receive and process this type of information: the very act of taking pictures and accumulating images was merely seen as a way of certifying an experience without really living it (Sontag, 2002). Photography during trips was predetermined by social roles, with sites, monuments, and experiences to be documented (Bourdieu and Bourdieu, 2004; Harrison, 2004).

Web 2.0 has revolutionized communication and information flows in many sectors, including tourism. Social media are having a huge impact on how tourists choose the destination to visit (Buhalis and Law, 2008; Lo et al., 2011; Morosan and Jeong, 2008) and the way visitors share and search for information about destinations (Kim et al., 2017; Xiang and Gretzel, 2010; Zeng and Gerritsen, 2014). Destination image is also affected: the use of social media and User-Generated Content (UGC) have been diminishing the market power of the DMOs in promoting the destination and shaping its image (Akehurst, 2009). Tourists become active consumers, spreading information, sentiments, and perceptions through online reviews. They upload pictures, videos, and comments from their tourism experiences, thus contributing to the destination image (Stylianou-Lambert, 2012).

The changing structure of travel information is also reshaping scientific investigation. Leung et al. (2013), in their literature review on social media in the tourism sector, highlight a common result: tourists mainly search for information in the travel-planning phase through social media. Hence, trustworthiness of UGC is a crucial factor which can strongly affect the destination choice (Xiang and Gretzel, 2010; Kim et al. 2017). Tourists consider UGC as more reliable and unbiased than traditional communication sources (Akehurst, 2009) and information posted by DMOs or by private tourism organizations (Mack et al., 2008; Zheng and Gretzel, 2010). More recently, User Generated Images (UGI) have been capturing the attention of this stream of literature, as posted pictures have become the most important way tourists use to share their experiences (Deng and Li, 2018; Lu et al., 2017; Stepchenkova and Zhan, 2013).

The present paper contributes to the line of research investigating the role of tourists in shaping destination image in several ways: first, we make use of Big Data tools and machine learning algorithms to analyse UGI. To the best of our knowledge, our paper is the first to apply a Visual Object Recognition (VOR) tool, recently developed by Google (Google Cloud Vision – GCV), to scan images posted on Instagram by visitors. This way we can identify, classify, and investigate the content of pictures, and the resulting time-varying vector of frequencies of features that appear in the UGI is then used to proxy the destination image.

Second, this proxy is used to feed a new metric (the Index of Distance in Destination Image – IDDI) which measures the degree of destination uniqueness / substitutability in the tourists' eyes. This index might become a simple but powerful metric to guide DMOs and destinations in their marketing strategies and in pinpointing mismatches between projections and perceptions of the destination image.

Third, we investigate the dynamics of the destination image by showing how key features of the pictures, as well as the distance between destination images, can change with time. This contributes from a novel perspective to the debate on how time, in general, and events, specifically, affect the destination image. Indeed, we test how exogenous events (specifically, shocks induced by climate change, arguably one of the most important drivers of changes in the future tourism flows) and endogenous events (organized by the destination) impact the destination image.

In a nutshell, both the Big Data tool and the metric herein proposed can be used to compare destinations (across units) and against the time dimension (over time), in a dynamic framework where the destination image continuously adapts, consistently with the Arrow-Debreu framework where the economic value of a good depends on space, time and contingency.

The focus of the paper is a set of European islands (four of the Canary Islands, Cyprus, Crete, Malta, and Sicily) monitored during the Summer of 2019. The choice of the summer season is self-explaining, as this is the peak season, when much of tourism demand for those destinations occurs. The investigation of islands also allows us to precisely define the destinations, eliminating confounding effects stemming from blurring boundaries of mainland sites. They are also areas where inflows and outflows are more regulated and where events are likely to hit the territory in a more homogeneous way. The social media analysed is Instagram, one of the fast-growing platforms and arguably the most popular UGI application worldwide.

The novelty of our contribution is threefold: first, the application of GCV, a VOR tool, to scan images posted by visitors on Instagram allows integrating textual and visual analysis. To the best of our knowledge, this powerful tool is applied for the first time in tourism studies, where most of the literature on social media focuses instead on textual content. Second, we develop and propose a destination image proxy and an index (IDDI) built from pictures posted on Instagram, which is fit to compare destinations cross-sectionally and over time. Three, via econometric analysis, we investigate how external factors influence destination image.

The paper is organized as follows: Section 2 reviews the vast literature on destination image, focusing on the contributions more closely related to this paper's approach (the one linking destination image to tourists' activity on social media); Section 3 presents our conceptual framework and the hypotheses to be tested in the empirical part; Section 4 introduces and describes the data and the methodology; Section 5 presents the results of our empirical investigation. Finally, Section 6 discusses the main conclusions, the policy implications, and the limitations of the present work.

2. A literature review on the destination image and its measurement in the era of Big Data

The notion of destination image was introduced by Gunn (1972), Mayo (1973), and Hunt (1975) and has been largely analysed by the tourism literature since then (Pike, 2002). It plays a key role, affecting destination choice, on-site experience, satisfaction, loyalty, and intention to revisit or recommend the destination (Baloglu and McLeary, 1999; Chen et al, 2014; Chi and Qu, 2008; Gallarza et al., 2002; Kozak and Baloglu, 2011; Li et al., 2015; Stylidis et al., 2018; Pike et al., 2018; Wang and Hsu, 2010; Zhang et al., 2014). Destination image is a composite multidimensional concept represented by a set of image-forming characteristics of the destination which visitors find the most valuable and/or intrinsic to a given location. As Pearce (1988) highlights, the term has "vague and shifting meanings", which has led to proliferation of definitions over time. Gallarza et al. (2002), Martin and Bosque (2008), Tasci et al. (2007) and Zhang et al. (2014) list dozens of alternative definitions, the most common and accepted ones identifying the destination image as a blend of "impressions, perceptions, feelings, beliefs that people have about a destination" (Crompton, 1979).

It is generally accepted to split the destination image into three dimensions: cognitive, affective, and conative (Agapito et al., 2013; Baloglu and McCleary, 1999; Pike and Ryan, 2004). The cognitive dimension refers to knowledge and beliefs related to the attractions to be seen, expected experiences to remember, and to the general environment of the destination (weather conditions, accommodation structures, attractions, health conditions, accessibility, etc.). The affective dimension is more related to feelings and emotions that can be triggered by different characteristics of the destination (Beerli and Martin, 2004). The conative image is consequential to the previous two spheres and refers to behavioural intentions of tourists regarding future activities. Destination image formation is a complex process, and the three dimensions are highly interconnected. While cognitive and affective spheres are the subject of many studies (Chew and

Jahari, 2014; Wang and Hsu, 2010), the overall image (a holistic impression of the destination – Echtner and Ritchie, 1991) is more difficult to capture.

Several factors affect image formation (Beerli and Martin, 2004; Pike, 2002; Tasci et al., 2007); Jenkins (1999) identifies and classifies demand factors (including motivation, perceptions, psychological characteristics, socio-economic features, hearsay) and supply factors (comprising destination promotion and tourism marketing). The advent of social media and UGC affects in various ways the cognitive, affective, and conative dimensions of the destination image (Kim et al. 2017). Whereas conventional media (such as broadcasting, newspapers, or magazines) have now a limited or nil impact on image formation, word of mouth and independent sources of information are becoming much more important influencers (Hanlan and Kelly, 2005; Tham et al., 2013). The advent of Web 2.0 has, in a way, democratized the process of creation and dissemination of the destination image (Lo et al. 2011), sometimes confirming, sometimes challenging the existing images imposed by DMOs (Schmallegger and Carson, 2008). Kim et al. (2017) highlight that destination image formation via social media is still under-investigated, especially from an empirical perspective.

Within this framework, the complexity of the concept and the multitude of definitions have led to various and inconsistent ways to measure the destination image (Stepchenkova and Mills, 2010). Most of the studies focus on measuring the cognitive dimension, since it is easier to assess it via quantitative analyses and using Likert scales or semantic differential scales. Factor analysis, multidimensional scaling, conjoint analysis, correspondence analysis have also been used. Particularly, data reduction techniques allow identifying latent dimensions of destination image (Gallarza et al., 2002; Pike, 2002; Tasci et al., 2007). Qualitative analyses are instead more suitable to assess the affective dimension, but these methodologies are time consuming and costly. Thus, it is rare to find papers that assess both dimensions of the image (Stepchenkova and Mills, 2010) and there is no consensus on which metrics should be used for providing a quantitative measure of the destination image.

Current research on the impact of social media on destination image has focused on textual messages by using content analysis (Chua et al., 2016; Költringer and Dickinger, 2015; Liang et al., 2009; Stepchenkova et al., 2009; Xiang et al. 2015). On the contrary, just a few works analyse the content of UGI (Deng and Li, 2018; Jiang et al., 2013; Miah et al., 2017; Stepchenkova and Zhan, 2013). This is disappointing, as "a picture is worth a thousand words" (Deng and Li, 2018; Pittman and Reich, 2016) and images are well known to have a more pronounced impact on people's attitudes, perceptions, and preferences than words (Hirschman, 1986), thus playing an important role during holidays and in shaping tourism experiences (Garlick, 2002). In fact, text in posts and comments is often short, especially on Instagram, recalling only a few aspects of the tourist's perspective. On the contrary, pictures spotlight in one shot several features, some of them are iconic representations of the holiday (the ocean in seaside destinations, a monument in cultural destinations, landscapes or plants for nature-based holidays, etc.) while others might play a crucial role in transmitting the quality of the experience and in the process of destination image formation. These secondary aspects (colours, human expressions, sunshine, lifestyle, surrounding buildings, etc.) are often not explicitly mentioned in the caption but refine the tourists' perception of a destination, thus conveying tangible, intangible, and complex concepts. Moreover, images are considered more reliable (being cues within the "realism heuristic", Sundar, 2008), more immediate and intimate than words, and easy to be retained in memory (the so-called "picture superiority effect", Childers and Houston, 1984; Whitehouse et al., 2006; Hockley, 2008; Pittman and Reich, 2016).

Nonetheless, the phenomenon of taking and sharing pictures is currently under-investigated (Caton and Santos, 2008; Lo et al., 2011, Lo and McKercher, 2015; Volo and Irimiás, 2020), and the analysis of pictures posted online has mainly detected tourists' behaviour, recommendations, perceptions, travel routes and trip duration (Kurashima et al. 2013; Lee et al. 2005; Okuyama and Yanai, 2013), but the destination image has been the research focus of only a few studies. As Volo and Irimiás (2020) underline, a systematic research on the meanings of photos can enhance the understanding of tourists' visuals from both theoretical and empirical perspectives.

Li et al. (2018) review the use of Big Data in tourism research and, among the different types of methods, sources, and types of data to be considered, they also describe the analysis of photos, which is triggered by the development of Application Programming Interface (API) of photo-sharing websites. Many details of posts, including user's demographics, picture's timestamp, and geographical position can be downloaded, thus allowing spatially correlating photo density and distinguish between residents and tourists (Kádár and Gede, 2013). Many studies use geographical data to focus on spatial aspects of tourists position or mobility (Girardin et al., 2008; Kádár, 2014; Koerbitz and Önder, 2014; Miah et al., 2017; Önder, 2017; Vu et al., 2015).

However, UGIs can provide more than the mere description of the picture (through hashtags) and of the tourist position (through geotags): methodologically, Pullman and Robson (2007) state that photos can be investigated analogously to text. In this regard, very few studies analyse the actual content of pictures (Zhou et al., 2015). Like text, the analytical treatment of photographs includes two broad groups of methodologies: content analysis and semiotic analysis. Content analysis is attribute-based and is mainly concerned with describing the appearance of certain features in the pool of images, allowing for quantitative analysis of their frequencies, co-occurrence, and other related evidence. Differently, the semiotic approach is highly interpretive and deals with the latent content of the photographs (Stepchenkova and Zhan, 2013).

Content analysis is applied to pictures in a few cases: Stepchenkova and Zan (2013) manually analyse around 1000 photos uploaded to Flickr by DMOs and by tourists, highlighting the emerging differences between projected (via DMOs) and perceived (by tourists' posts) images of destinations in Peru. They locate the destination image at the convergence between DMO-projected and user-generated images in terms of frequencies and confirm findings of other studies adopting text-mining approaches (e.g., Költringer and Dickinger, 2015; Stepchenkova et al, 2009). Miah et al. (2017) use photos uploaded by tourists on Flickr to analyse and predict tourists' behavioural patterns in Melbourne. They develop a method for extracting information and insights from 238,900 geotagged photos and analyse unstructured big data by combining four computational techniques: text processing, geographical data clustering, visual content processing (or representative photo identification) and time series modelling, translating information into a set of visual words. Jiang et al. (2013) study a sample of 10,357 pictures and 9537 texts uploaded on Panoramio and related to 10 cities in Europe, to assess attraction popularity and fluctuations over time. In their paper, visual features are extracted, and each picture is associated with a set of labels describing it. Visual features clustering is then applied to classify pictures which are similar in their content.

To the best of our knowledge, only the last two papers use VOR tools to analyse pictures content in tourism literature, but none of them exploits such elaboration to investigate the issue of destination image. Therefore, our paper aims at filling a gap in the literature, by proposing a novel methodology (based on VOR tools) to associate posted pictures' attributes with destination image, and by developing a metric based on this Big Data approach. Finally, unlike most of the recent literature, which analyses pictures from Flickr and Panoramio (Li et al., 2018), we contribute to the strand of studies that use Instagram as the source of imagery content, which are very few at present (Lu et al., 2017). As Volo and Irimiás (2020) highlight, the visual data analysis of pictures posted on Instagram allows capturing the physical and societal changes in tourism destinations and in tourists' behaviours.

3. Conceptual Framework

Our conceptual framework, summarised in Figure 1, is coherent with the literature on destination image. It is widely accepted, since Jenkins (1999), that there are two forces contributing to destination image formation. The traditional one pushes from the supply side (through DMOs, tourism stakeholders, and conventional media), positioning the destination on the market and promoting a target "projected" image. At the same time, specifically with the advent of Web 2.0, the image stemming from the community of tourists, independent media and word-of-mouth becomes more important, pushing from the demand side (Akehurst, 2009; Stylianou-Lambert, 2012).

The perceived destination image emerges from these two forces, which are interconnected, affecting each other in a hermeneutic circle of representation (Urry, 1990): on the one hand, DMOs are active on social media, interacting with other users and promoting the target image. On the other hand, the image perceived by tourists emerges and can push the destination to implement policies to adapt or to mitigate this image (Deng and Li, 2018; Hanlan and Kelly, 2005; Schmalleger and Carson, 2008; Tham et al., 2013). The distinction between projected and perceived image is well recognised in the literature (Andreu et al., 2000; Volo and Irimiás, 2020) and, conceptually, it is fairly easy to disentangle it. Empirically, it is a much harder task because a picture represents both a perception of the place and, at the same time, a projection (if we consider that the observer is different from the one who took the photo). Our approach follows Schmalleger et al. (2010) and Stepchenkova and Zhan (2013), among others, in considering users' pictures as a representation of the aggregated perceived destination image.

[Insert Figure 1 about here]

Within this framework, the destination image can be disaggregated into three components: the cognitive, affective, and conative spheres (Agapito et al., 2013; Baloglu and McCleary, 1999; Beerli and Martin, 2004; Pike and Ryan, 2004), which are interrelated. The knowledge and beliefs, the quality of the experience and of the attractions to be seen (proper to the cognitive component) trigger attitudes, feelings, and emotions of tourists (the affective component). Both dimensions are then important for the conative component (Kim et al., 2017; Prayag et al., 2013), which refers to behavioural intentions regarding future activities, choices, and decisions (e.g., the decision to repeat the visit or to use word-of-mouth to promote the destination).

Having the baseline model of Figure 1 in mind, the paper focusses on the demand side, thus investigating how the destination image builds itself through the activity of tourists captured on a UGI platform, which is still an overlooked research topic (Lo et al., 2011; Kim et al., 2017). In line with Deng and Li (2018), Lu et al., (2017), Stepchenkova and Zhan (2013), social media based on imagery are considered a powerful force of destination image formation: UGI can easily communicate the primary tourism activities and experiences but, better than textual UGC, images highlight complementary aspects of a holiday that importantly affect tourist gaze and that are often ignored in posts or comments. The contribution of UGI can thereby be represented by the red arrows in Figure 1. In fact, labels returning image content are generally associated to nouns and to objective features of each picture (Deng and Li, 2018), allowing the identification of what Stepchenkova and Zhan (2013) call the aggregated perceived destination image.

Within this framework, we then focus on the upper red arrow of Figure 1, as the image content analysis of UGI can be used to study the cognitive component of the destination image. We then investigate differences and similarities across destination images by looking at how labels associated to pictures posted by tourists differ. To validate this perspective, we then propose hypothesis 1 to be tested:

H1: Differences across destinations in the cognitive component of the aggregated perceived destination image are captured by differences in labels associated to the content of pictures posted on social media.

Destination image is a dynamic concept, evolving over time (Gallarza et al., 2002; Gkritzali et al., 2018; Kim et al., 2019; Lo and McKercher, 2015). Accordingly, the continuous flow of new pictures, posts, information, events, unceasingly forges the destination image, making it a sort of time-lapse photography. The impact of time can be better understood by separating its main components. First, mutation of tourist gaze is affected by long-term structural aspects, including urban and cultural development of the destination, changes in its promotional strategy, and the socio-economic and political evolution. Second, tourist gaze is also affected by seasonal aspects, since different tourism segments visit the destination in different periods of the year, each one with specific consumption patterns, preferences, and priorities. In this respect, some specific features of the destination image can recur with seasonal frequency.

Given the data at our disposal, which only span over a few months, we cannot empirically investigate these two aspects. However, a mutable destination image could be detected by looking at the variability of pictures' labels in different periods of the year, namely different months of the summer season (Sarma, 2007; Tasci, 2007). To verify whether destination images change in this respect we then propose hypothesis 2.

H2: The cognitive component of the aggregated perceived destination image is dynamic.

Finally, time affects destination image through one shot or recurrent events happening at the destination (sport competitions, cultural festivals, climate events, natural disasters, terrorist attacks, etc.) In this respect, it is important to separate organized from exogenous events. In line with the systematic literature reviews (Lai and Li, 2014; Dragin-Jensen and Kwiatkowski, 2018), events are increasingly acquiring a key role in urban and regional development strategies. They also expand awareness and familiarity with the destination, affect tourists' loyalty, reinforce local identity and, especially, they become important image builders for the destination. Accordingly, we intend to test whether endogenous events (i.e., events organized and hosted by the destination) affect the image by analysing how frequently aspects connected with the events appear in tourists' posted pictures. Thus, we propose hypothesis 3:

H3: The cognitive component of the aggregated perceived destination image is affected by endogenous factors, such as events organized at the destination.

On the other hand, destination image is not free from unpredicted shocks like climate change, terrorist attacks, or natural disasters (Arana and Leon, 2008; Alvarez and Campo, 2014; Avraham, 2015; Becken et al., 2016; Wu and Shimizu, 2020). Some of these effects (e.g., climate change, political instability) can be partly reduced by adopting responsible behaviours and policies, but others are completely out-of-control (e.g., earthquakes, terrorist attacks, etc.). Consistently with this approach, we test how frequently exogenous events (i.e., external events not under control of the destination), especially the ones connected with climatic and weather conditions, appear in posted pictures, thus impacting the destination image. We therefore propose hypothesis 4:

H4: The cognitive component of the aggregated perceived destination image is affected by exogenous factors, like weather conditions and extreme climate events.

These four hypotheses, which breakdown the upper red arrow depicted in Figure 1, are the core of our analysis and can be graphically summarised in Figure 2, where the images of two tourism destinations, *i* and *j*, are represented as bubbles. The overlapping (non-overlapping) area of the bubbles graphically represents the degree of similarity (uniqueness) between destination images, according to the characteristics of the pictures posted by tourists on the social media (H1). These characteristics can vary with time because they are highly dependent on contingent situations occurring in specific moments (changing the size of the overlapping area): for example, destinations host tourism segments with (even if slightly) different gazes and behaviours at different times (H2); the flow of events organized at the destination (H3) and other external factors (H4) impact the characteristics of tourists' posts, continuously forging the destination image. For the sake of simplicity, the cognitive component of the aggregated perceived destination image is often shortened to destination image in the rest of the paper.

[Insert Figure 2 about here]

4. Data and Methodology

The test of H1-H4 is undertaken through a novel methodology that takes advantage of recently developed tools to handle Big Data which, differently from traditional data, are often unstructured and need new analytic tools, storage systems, and methods (Gandomi and Haider, 2015; Tsai et al., 2015). Big Data have been analysed in different fields such as technology security and safety, healthcare, services, and tourism (Költringer and Dickinger, 2015; Lu and Stepchenkova, 2015; Önder, 2017; Xiang et al., 2015), but many aspects are still unexplored, especially in destination management (Li et al, 2018).

The data for this paper come from Instagram, one of the most rapidly growing social network platforms. While Snapchat's daily user count and Facebook's monthly count grew by 2.13% and 3.14% in Q1 2018 respectively, Instagram was growing by about 5% per quarter and from 300 million monthly active users recorded at the beginning of 2015 it has reached 1 billion by June 2018 (https://techcrunch.com/2018/06/20/instagram-1-billion-users/). Unlike social networks that have text as the main means of connection, Instagram has imagery at its core, and this is the main reason for such a booming popularity. Data available for retrieval from Instagram posts are pictures, image caption, hashtags, comments, and geotag (if the user opted for attaching it). Geotagging is a feasible option also in other networks, but users are much more prone to indicate their location when posting on Instagram, thus making it a precise tool for geospatial investigation, for example to study visitors' activity. The data were collected using an in-house programmed scraper.

From a purely technical point of view, it is preferable to use Instagram because its data are less of a black box than, for example, Twitter: while the Twitter API returns a sample of tweets that match the criteria outlined in the query (without giving an indication of the total number of tweets of interest), data obtained from Instagram are the universe of posts from open accounts which match specific search criteria. Panoramio, another image-based social media, is like Twitter in filtering posts in an undocumented way. Flickr shares the same characteristics of Instagram but is far less popular and the main reason why the great majority of recent studies on photographs use Flickr is because its API easily allows getting metadata and pictures (Li et al., 2018). Finally, given that the primary aim of this paper is to develop a time-varying vector of the features composing the image of the destination as perceived by travellers, Instagram data appear to fit the purpose better than online review platforms, since reviews are mostly tied to specific hotels, attractions, restaurants, etc.

Among the many destinations theoretically of interest, we focus on the Canary Islands archipelago, specifically on the four largest islands with highest tourism inflows: Tenerife, Gran Canaria, Fuerteventura, and Lanzarote. Additionally, posts related to four Mediterranean islands (Cyprus, Crete, Malta, and Sicily) are explored for comparison purposes. Although the latter islands are like the Canaries in the sense of being mainly leisure (sea & sun) destinations, they also have a more pronounced cultural component, which we expect to find in the pictures posted by visitors. Islands allow a more precise identification of the territory under investigation, eliminating confounding effects stemming from blurring boundaries of mainland destinations.

The pool of data consists of all retrievable (public) posts that contain a hashtag with the island's name and one of the travel-related keywords (*travel, visit, vacation, holiday, trip* and cognate words) in the caption: this way we target posts that focus on the selected islands as tourism destinations. The timespan used in the analysis runs from the 8th of June 2019 to the 28th of September 2019, hence covering the whole 2019 Summer season. Table 1 reports some descriptive statistics of posts, from which we see that islands are similar in terms of average number of comments, likes and shared geotags.

[Insert Table 1 about here]

Operationally, the content of each picture must be identified and translated into plain text (labels) before proceeding with statistical analysis. When facing a similar task, Stepchenkova and

Zhan (2013) selected a small sample of pictures and performed it manually. However, since we aim at obtaining a vector of feature occurrences for each of the destinations on a very fine-grained scale, we need to process several hundred thousand photos (about 860,000 in our study, shared through 500,000 posts), which, of course, cannot be done manually. We then make use of VOR tools which have recently been developed and gave life to a flourishing research field, particularly in engineering and computer science literature (Bello-Orgaza et al., 2016; Li et al., 2018; Miah et al., 2017; Qiu et al., 2016).

Specifically, we resort to Google's Cloud Vision API (GCV), a tool based on powerful machine learning algorithms such as convolutional neural networks, which perform various image processing tasks. From the available GCV options (label or facial detection, object localization, others), we selected label detection, which returns a broad set of categories associated to the pictures: image labels can identify objects, locations, phenomena, activities, specific plant and animal species, products, and even more generic and abstract categories (e.g., love, friendship, tourism). GCV processes each single image and returns an output with a list of up to ten labels describing the content of the picture. Each label comes with a related score indicating the degree of accuracy (taking values from 0.5 to 1) in matching the given image. In general, the accuracy of GCV is estimated to be high (overall, more than 0.8), and it scores considerably higher than similar tools of other big providers (Enge, 2019). This is confirmed in our case, where the level of accuracy is over 0.8 for all islands, ranging from 0.818 (sd = 0.086) in Cyprus to 0.834 (sd = 0.087) in Fuerteventura.

The output of the VOR processing is then aggregated and used to define the image of a given destination *i* as a time-varying vector (with a time span from 1 to *T*) of label frequencies for a fixed set of *K* labels returned by GCV. In matrix notation, we write a *K* x *T* matrix as in [1]:

$$Image_i = \begin{pmatrix} freq\_lab_1^1 & \cdots & freq\_lab_T^1 \\ \vdots & \ddots & \vdots \\ freq\_lab_1^K & \cdots & freq\_lab_T^K \end{pmatrix} \qquad [1]$$

This notation has some degree of flexibility. In the most general case, *K* may include all possible labels that have ever appeared in the imagery during the given time span. Alternatively, labels can also be grouped in broader categories. Frequencies may refer to the ratio of the number of occurrences of label *k* to either the total number of labels *K*, or to the number of pictures, or to the number of posts at time *t*. Finally, the degree of time disaggregation determines whether time periods from 1 to *T* are hours, days, weeks, or months. Researchers are required to tailor these parameters in accordance with the question of interest.

One of the aims of this paper is to exploit characteristics of the pictures to build a quantitative metric proxying the cognitive component of the destination image; this way, it is possible to compare destinations and determine the uniqueness/dissimilarity of their images, thereby testing H1. To proceed in this direction, we use the labels frequencies to rank labels for each island. The ranking (which is shown in the Appendix, Table A.1) allows constructing a measure of "destination image distance", that is, to which extent the image of one destination differs from another, hence being a relative measure of its distinctiveness or uniqueness. Although many different concepts of distance can be used, the simplest measure is proposed in this paper and reported in [2]: the average absolute rank distances of the set of top *K* labels (*K*=20 is the chosen benchmark in our analysis). The Index of Distance in Destination Image (IDDI) can hence be formulated in [2a]:

$$IDDI_{ij} = \frac{\sum_{k_i}\left|rank_{ik_i} - rank_{jk_i}\right| + \sum_{k_j}\left|rank_{jk_j} - rank_{ik_j}\right|}{2} \qquad [2a]$$

The index measures the distance in the image of destinations *i* and *j*, based on the first *K* labels sorted by the number of occurrences, so that $k_i = \{label_{1i}, \dots, label_{Ki}\}$. Alternatively, squared rather than absolute distances could be considered in [2a], although this latter approach is more sensitive to the presence of outliers (single labels with very large rank distance across destinations). Therefore, we suggest using absolute distances as a baseline, and check whether results are qualitatively similar if squared distances are considered.

As IDDI is a measure of relative distance, when the time dimension is not present, it becomes meaningful only if there are three or more destinations in comparison. We further suggest normalizing the absolute distance between destinations to the smallest distance in the sample, so that:

$$\widetilde{IDDI}_{ij} = \frac{IDDI_{ij}}{min\,\{IDDI\}} \qquad [2b]$$

IDDI can be computed for each destination at different points in time, to evaluate the dynamics of the perceived image of the destination and hence test H2. The absolute and relative frequency of labels in Instagram posts can also be used to determine how specific factors (for example organized events or extreme weather conditions stemming from climate change) affect behaviours and activities undertaken by tourists, possibly influencing the destination image (in line with H3 and H4). Exploratory analysis addressing this question is carried out through econometric estimations of models which are presented throughout the next section.

Before moving to the empirical section, it is important to mention that several steps of data cleaning have been carried out to obtain the sample used in the analysis. First, travellers often post pictures with a time-lag, for example from their previous trips. This may create distortion because the picture posted does not correspond to the time of the post, but could have been taken weeks, months or even years before. While in cross-sectional analysis this is less of a concern, an adjustment is necessary when investigating the time dimension. To mitigate possible biases and to lower the amount of unrelated content, we only use posts that are geographically tagged to the destination.

Second, a critique may arise as DMOs also have Instagram accounts and they post on Instagram, thus influencing tourists' perceptions. We argue that this concern does not undermine our approach, since the number of such posts is negligible. To have an idea of the frequency of posts of DMOs, hotels, shops, and other commercial entities, we randomly selected 500 posts from the sample and manually checked into which category they fall; we found that only 8.5% of posts come from DMOs and similar accounts, allowing us to conclude that the images we investigate are coming from tourists.

Third, as shown in Table 1, many posts contain more than one picture, and some users attach several similar pictures to a single post. Therefore, we drop repetitive photos within a post when they have the same first label. In these cases, only one of the pictures with the same first label is randomly selected and kept in the sample for subsequent analysis.

Fourth, we process all labels attached to each post in a similar fashion as text is processed in text-mining exercises, i.e., using the "bag of words" approach. This implies also stemming and eliminating stop-words and yields a list of variables representing the frequency of each word in the text corpus (or in the label set, in our case).

Finally, these data may be transformed in various ways depending on the question to be addressed. In the next section, we demonstrate possible applications and always apply different transformations to the original data. When testing H1, a cross-sectional analysis of the destinations under investigation is carried out and labels are pooled over time to measure similarities and differences in their images. Next, we study how these differences converge or diverge on a seasonal (T=2) and monthly (T=4) basis to test H2. However, given that our data only cover the summer of 2019, we are not able to disentangle specific recurrent seasonal patterns, for which multiannual data would be needed, but only differences across periods. Finally, we further disaggregate the time dimension and use daily label frequencies; this way, it is possible to investigate how various events impact the perceived destination image, thus testing H3 and H4.

5. Results

5.1. Cross-sectional analysis of destination image

For each island under investigation, Figure 3 presents a descriptive analysis of label frequencies through word clouds (composed with wordart.com). Word clouds are widely used as a straightforward visual tool representing the relative frequencies of words appearing in a text corpus (Steel and Iliinsky, 2010; Chandrapaul et al., 2019). The larger the size of the word in the cloud, the higher its frequency within the text. In this paper we apply the logic of word clouds to compare the frequency of labels attached to images, considering the ten labels provided by GCV for each scanned picture. In Figure 3, the full range of labels in their raw form (i.e., without any grouping of labels) is used: this way, a massive number of features is analysed but, on the other hand, aggregating characteristics to more general topics (e.g., pooling together all food-related labels into a "Food" category) would likely smoothen differences and lead to overlooking specific characteristics.

[Insert Figure 3 about here]

A first glance at Figure 3 shows that all destinations look extremely similar which, perhaps, is of little surprise given that they could all be considered European sea & sun destinations: hence, labels like *Sky, Sea, Vacation, Tree, Beach* are among the most frequent for all islands (Table A.1 in the Appendix shows the ranking of the top-20 labels for each island). Nonetheless, some differences can be spotted: *Mountain* appears relatively more frequently in Tenerife than in other islands; *Sea* and *Ocean* have relatively more weight in Fuerteventura; *Architecture* and *Building* are of more importance in Cyprus, Crete, Malta and Sicily than in the Canary Islands, something that is clearly linked to the density of cultural heritage in the Mediterranean islands: in fact, all the labels representing architectural, religious and historical sites (*History, Historic, Ruins, Site, Ancient, Building,, Dome, Mosque, Holy, Medieval*, etc.) have higher ranks in these islands than in the Canaries. The islands of this archipelago have more similar images, but also have distinct features: for example, Gran Canaria appears the most urban, Tenerife is characterized by a higher frequency of labels related to partying and nightlife but also for wildlife spotting, Lanzarote stands out for its arid landscapes and Fuerteventura for the vast sandy seashores and turquoise waters as the frequencies of labels such *Beach, Shore, Sand, Coast, Turquoise, Ocean* show.

Once the frequency ranks of labels are available, and differences are spotted through the word cloud descriptive tool, a more precise quantitative metric can be applied. The proposed index (IDDI) measures how much the image of a destination differs from another, allowing a more precise evaluation of the H1 validity. Table 2 is a symmetric matrix presenting the average absolute rank distances (computed according to [2a] between the islands under investigation), based on the set of top 20 labels. Distances are normalized according to [2b] to the value of the two closest destinations (in our case Tenerife and Gran Canaria), the distance between which is hence set to unity. It appears that Sicily and Malta, being a mix of leisure and cultural destination, lie on one side of the spectrum, whereas the Canary archipelago resides on the other, with Crete and Cyprus being somewhere in the middle and close to each other. In addition, it is interesting to highlight the heterogeneity among the Canary Islands. Despite belonging to the same institutional and cultural background and being geographically close, Fuerteventura is the most distant from the other three islands, and Tenerife and Gran Canaria's images seem to be closer to that of Cyprus, than to Fuerteventura's. We also highlight that while the cells of Table 2 measure distances between pairs of islands, the average distance row at the bottom of the same table captures the degree of distinctiveness/peculiarity of each destination in the whole pool. In Table A.2 of the Appendix we report the IDDI built using quadratic rather than absolute distances: this approach provides very similar qualitative results.

[Insert Table 2 about here]

To test the validity of IDDI, we compare results to the correspondence analysis undertaken on the same set of top 20 labels for the 8 islands. Correspondence analysis is a statistical method used to describe associations between two or more categorical variables, widely adopted in marketing research and also used in tourism literature (Beldona et al., 2005; Chen, 2001; Marcussen, 2014; Richards and van der Ark, 2013). A plot of the two dimensions derived from this analysis (see the Appendix for the whole statistical procedure) highlights the (dis)similarities among destination images (Figure 4). Distances between islands based on Chi-square measures are then elicited in Table 3, where distances have been normalized to the lowest distance (the one between Tenerife and Gran Canaria), allowing a direct comparison with IDDI. The comparison between values obtained through IDDI (Table 2) and correspondence analysis (Table 3) strongly supports the validity of IDDI.

We conclude this section stating that the evidence stemming from labels of pictures posted on Instagram supports H1: differences in the destination images are captured by differences in labels associated to the content of pictures. Moreover, the proposed index (IDDI) returns results that are consistent with the ones of correspondence analysis, and robust to different definitions of distance.

[Insert Figure 4 about here]

[Insert Table 3 about here]

5.2. Looking into dynamics

IDDI can also be used to assess the dynamics of destination images, thus investigating, for example, whether two destinations become more similar, and hence substitutes, in the eyes of tourists. This might have relevant policy implications in terms of market competition. The data at hand do not allow undertaking a long-term analysis, but we can investigate the short-term dynamics to see whether image distances change on a monthly or seasonal basis. Tables 4 and 5 report the IDDI constructed by splitting the sample into shoulder (June and September) and peak (July and August) season (for comparability, the numeraire distance used for normalization is the same as in Table 2).

[Insert Tables 4 and 5 about here]

Looking at the average distance (last row of Tables 4 and 5), differences in IDDI appear to be higher in the shoulder than in the peak season (only Crete does not change, with the average distance being equal to 2.6 in both periods): images hence appear to be more similar in the peak season and more distant in the shoulder season. This interesting finding, which tends to support H2 (the destination image is dynamic, changing with time), is arguably due to the higher number of "representative" mass leisure tourists arriving in July and August. In the shoulder season, on the contrary, visitors are likely to be more diverse and enjoy a wider range of activities (cultural, nature-based, etc.) which differ across islands. Tables A.5-A.8 in the Appendix provide a more detailed (monthly) split of IDDI, which is however consistent with the peak/shoulder breakdown herein presented.

Our data can also pinpoint changes in the destination image on a more fine-grained scale, for example analysing if the frequency of specific labels is triggered in reaction to specific events. We illustrate this point with two examples, one related to an endogenous event (organized by the destination) and one related to an exogenous event (not depending on the destination). To start with, Figure 5 plots the relative frequency (calculated as the ratio of daily occurrences of labels of interest to the number of posts) of *windsurfing* and *kitesurfing* labels appearing in pictures of Fuerteventura. Given that surfing is one of the core activities among tourists visiting this island, *windsurfing* and *kitesurfing* appear in tourists' pictures almost every day (on average, in 10% of the pictures). However, Figure 5 shows two peaks when the relative frequency is significantly higher. These peaks coincide with the windsurfing and kitesurfing World Cup that took place on the island from July 19th to August 3rd, 2019. Interestingly, related posts appear more frequently in the opening and in the closing of the championship, whereas the remaining days of the competition are comparable with other dates before and after the World Cup. Such findings, which support H3 (the destination image is affected by events organized within the destination) may be of interest for DMOs assessing the impact of specific events on the popularity of the destination and on its image.

[Insert Figure 5 about here]

The second example investigates the tourists' response to extreme climate events. On the one hand, it is well known that tourists are not generally posting anything that makes themselves and others feel negative, and traveling-related posts typically exhibit very little negativity (Deng and Li, 2018). On the other hand, when the event is perceived of high importance, they might decide to share related images and sentiment despite the associated negative mental cost. Figure 6 plots the relative frequency of *wildfire* and *explosion* labels for Gran Canaria in the month of August 2019, compared to all the other islands of the Archipelago. Kinks appear precisely at the time of the two major fires having occurred in Gran Canaria during the period under observation, supporting H4. The relative frequency of labels, although remaining low (around 1.5 - 2% of posts) indicates a change in the image perception.

[Insert Figure 6 about here]

Continuing in this line of investigation, another novel application of our approach is exploring the substitutability of different activities undertaken by tourists at a destination when weather conditions change. Destinations might be interested in knowing to what degree visitors can easily substitute one type of activity with another when weather conditions force them to do so. Considering the pressing issue of climate change, which urges adaptation and mitigation strategies development, such analysis may shed light on the resilience of a destination to extreme weather conditions.

To illustrate this point, we collected daily data on daytime weather in the Canary Islands (from openweathermap.org) and analysed how frequencies of labels associated to different types of activity performed by tourists changed in response to shifts in weather conditions (we focused on Canary Islands to avoid too much dispersion in meteorological conditions compared to Mediterranean islands). To keep things simple, only distinction between beach-related and non-beach-related activities is considered, and we focus on the list of top-20 labels for posted pictures to find those that appear in all the four islands and are representative of one of the two activities. Accordingly, we selected *beach, sea,* and *ocean* to represent beach-related leisure activities, while *tree*, *plant, mountain,* and *landform* were chosen to represent nature-based activities undertaken away from the beach. The following equation was then estimated through OLS:

$$freq_t = \alpha + \beta temp_{10d} + trend_t + trend^2{}_t + \varepsilon_t \qquad [3]$$

where $freq_t$ is the relative frequency at day t of, alternatively, beach-related or nature-based activity labels as described above (daily observations entering the sample were required to have at least 50 posts per day). The main variable of interest $temp_{10d}$ represents the weather conditions and stands for 10-day averages of daytime temperature recorded at most visited locations within islands (the last chunk of September 2019 is a 13-day average). This way, the possibility that users post related content a few days later is taken into consideration; standard errors are clustered on 10-day chunks level for correct inference. Since warmer temperatures favour leisure activities on the beach, while lower temperatures make them less comfortable hence driving tourists to find alternatives (e.g., visit natural parks), we expect that the sign of coefficient β is positive for beach-related and negative for non-beach related activities. Possible seasonality effects are controlled through the inclusion of linear and quadratic trends in equation [3], which is estimated for each individual island.

As emerges from results presented in Table 6, beach-related labels in Fuerteventura turn out to be in a weak negative relationship with temperature, which could be explained through the fact that lower temperatures are often accompanied by stronger winds. Given that Fuerteventura attracts tourists who prefer active sports (such as various types of surfing) to sunbathing, windier (and colder) days may, in fact, attract more visitors to the seaside. As regards nature-based activities, higher temperatures lead to lower frequencies of non-beach related labels in Tenerife and Gran Canaria, which is in line with our expectations.

[Insert Table 6 about here]

This exercise indirectly confirms the validity of the cross-destination image distance index IDDI proposed in the previous subsection. Recall from Table 2 that Gran Canaria and Tenerife have the most similar images, followed by Lanzarote, whereas Fuerteventura appears to be the most distant. When using complementary data on temperature, we see from Table 6 that the most similar islands in their association between temperature and activities are again Tenerife and Gran Canaria, followed by Lanzarote, whereas Fuerteventura stands out.

Since it appears from both Tables 2 and 6 that Tenerife, Gran Canaria and Lanzarote can be grouped together, we run eq. [3] on the pooled sample to improve estimation efficiency and obtain more robust results, which are presented in Table 7. Columns (1) and (3) include destination fixed effects as additional control variables, while in Columns (2) and (4) we add destination-specific linear and quadratic time trends (as opposed to common trends in (1) and (3)), as an additional robustness check.

[Insert Table 7 about here]

In line with expectations, pooling the sample allows estimating the parameters of interest more precisely, providing more robust evidence that warmer temperatures favour beach-based activities. Given that on average the temperatures observed on islands were not extremely high, the positive coefficients in columns (1) and (2) and the negative coefficients in columns (3) and (4) are pointing at weather-driven substitution between beach and nature-based activities.

Finally, we used seemingly unrelated regression (SUR) analysis to investigate whether extreme climatic events have an impact on the cognitive dimension of the destination image, as another test of H4. The nature-based and beach-related label frequencies are now modelled jointly, and the impact of wildfire outbreak on Gran Canaria, that occurred in August, is evaluated in a difference-in-difference framework, according to eq. [4]:

$$\begin{aligned} LnFreq_{nature_{it}} &= \alpha_n + \beta_n temp_{10d} + \delta_{n1} post_t + \delta_{n2} treat_{it} + trend_t + trend^2{}_t + \gamma_i \\ &+ \varepsilon_{nit} \\ LnFreq_{seaside_{it}} &= \alpha_s + \beta_s temp_{10d} + \delta_{s1} post_t + \delta_{s2} treat_{it} + trend_t + trend^2{}_t + \gamma_i \\ &+ \varepsilon_{sit} \end{aligned} \qquad [4]$$

where $LnFreq_{j_{it}}$ is the natural logarithm of nature-related and seaside-related labels frequency observed on day *t* on island *i*, γ_i stands for the island fixed effect, $post_t$ is an indicator of the post-treatment period, which is equal to 1 starting from the date of the first fire (10^{th} of August), whereas $treat_{it}$ is an interaction between $post_t$ and Gran Canaria fixed effect, therefore being equal to 1 if an observation belongs to Gran Canaria and is observed in the post-fire period. Results of the estimation are presented in Table 8.

[Insert Table 8 about here]

Assuming that changes of label frequencies are appropriate proxies for demand, we can conclude that a 1 ℃ higher temperature implies a 2.4% decrease in demand for nature-based activities, at the same time increasing demand for beach-based activities by about 3.6%. The wildfires which took place on Gran Canaria led to a decrease in demand for nature-related activities in the island of almost 7%. Importantly, in contrast to the impact of temperature, there is no evidence that this negative effect was substituted by higher demand for beach-based leisure. Results of Table 8 have important implications for the tourism policy of the destination since they show to which extent external conditions affect behaviours and activities of tourists.

6. Discussion and Conclusions

6.1 Conclusions and main contribution

This paper exploits the novel Big Data and Artificial Intelligence techniques to investigate: (i) how the analysis of pictures posted online by tourists can enlighten the process of destination image formation; (ii) how the introduction of an index of distance in destination image built on the labels of pictures can measure the uniqueness/dissimilarity of destinations; (iii) how the individual destination image is dynamic, also responding to specific events organized or hitting the destination. By applying the Visual Object Recognition tool developed by Google (Google Cloud Vision) to about 860,000 Instagram pictures posted by tourists in 8 selected European islands, we contribute to the literature in several methodological and empirical aspects.

Methodologically, our approach highlights the potential of Big Data, especially UGI, in advancing knowledge about destination image and tourists' perceptions and behaviours. Arguably, nothing proxies the perception of destination image better than the content of pictures posted by visitors on social media. As recalled by Deng and Li (2018), an English adage says: "a picture is worth a thousand words", thereby communicating the experience at the destination and the cognitive sphere of the destination image in a precise way. VOR software allows investigating the key features of posted images, which are recognised and translated into labels by artificial intelligence. Next, methods of text mining, quantitative analysis and econometric estimation can be applied to image labels exactly as they would be to written text. Given the massive amount of visual data on social networks, artificial intelligence and machine learning tools are needed to efficiently exploit available information. To the best of our knowledge, this is the first time that VOR tools are applied to study the destination image, and the first time that these tools are applied to Instagram, the most popular image-based social media.

The analysis of image label frequencies allows comparing destinations and, accordingly, we developed and proposed an index measuring the distance between the image of different destinations. This index (IDDI – Index of Distance in Destination Image) might open relevant research avenues and policy implications regarding positioning of a destination against similar or competing tourism areas. We conducted several different tests to show how this approach can be applied to different research questions arising from a conceptual framework that is coherent with recent literature on destination image. We found support for all our hypotheses: differences in destination images can be captured by the different labels associated to the content of posted pictures (H1); the destination image is dynamic, changing with time (H2), being affected by endogenous factors (like events organized at the destination, H3) and by exogenous factors (like weather conditions and climate events hitting the destination, H4). Further econometric analysis also allowed us to estimate the degree to which tourists substitute between different types of activities according to weather conditions and climate events.

Empirically, we applied this methodology to investigate some of the Canary (Gran Canaria, Fuerteventura, Lanzarote, Tenerife) and Mediterranean Islands (Crete, Cyprus, Malta, Sicily). In a spectrum delineating the nature of destination images, Sicily and Malta are perceived as a mix of leisure and cultural destinations, the Canary archipelago emerges as a pure leisure destination, while Crete and Cyprus lay in between, and are close to each other. Fuerteventura is the most unique island of the Canary archipelago, whereas Gran Canaria and Tenerife are the two closest islands within the pool.

In the peak season, when destinations mainly host mass leisure tourists, destination images are found to be more like each other than in the shoulder season, where islands show a higher degree of dissimilarity. This is arguably because in off-peak season islands host more heterogeneous tourist segments, which enjoy a wider range of activities (cultural, nature-based, etc.), differing across islands. Images are also sensitive to events organized at the destination. Therefore, well-designed events are a useful tool to push the perceived toward the projected image.

Finally, whereas traveling-related posts typically exhibit very little negativity, external events, like bad weather conditions and extreme climatic events may impact the tourists' experience and the destination image. Assuming a direct relationship between the content of posted pictures and the activities carried out by tourists, econometric analysis suggests that 1 ℃ higher temperature implies a 2.4% decrease in demand for nature-based activities and a 3.6% increase in demand for beach-based activities. Moreover, the wildfires which took place in Gran Canaria in summer 2019 led to a decrease in demand for nature-related activities in the island of almost 7%, which was not substituted by beach-based activities.

6.2. Policy and managerial implications

The approach proposed in this paper presents a precise metric to investigate the cognitive dimension of the destination image by directly analysing the pictures' content rather than elaborating on posters' words. The introduction of an index measuring the distance in destination images also has important policy and managerial implications. Projection and perception of destination uniqueness should be one of the main goals of DMOs: when destinations are perceived as similar, they are likely to be substitutable, hence decreasing their market power. Thus, the measurement of the degree of similarity between competing destinations allows on the one hand assessing the effectiveness of specific marketing and positioning strategies and, on the other hand, better identifying the characteristics of competing destinations. For example, if IDDI shows a strong similarity with another destination, this might indicate that branding strategies have been ineffective, an issue to be tackled by the DMO. Moreover, the flexibility of our metric may guide allocation of investments in those factors identified to be the drivers of dissimilarity from other islands. Those factors might be the ones mainly affecting the segments of tourists visiting the destinations during the shoulder seasons, also helping the goal of de-seasoning tourist flows.

Moreover, as destination images are significantly affected by major non-tourism related events such as climate change (in line with Gkritzali et al., 2018), our findings could help DMOs evaluate how adaptation and mitigation policies applied in similar destinations are affecting their image. In the current tourism world, which is characterized by the three key external factors shaping the performance of destinations (geopolitical instability, climate change and health crises) this approach will become more and more relevant in the future. Similarly, careful investigation of how the organized sport or cultural events impact the destination image is of paramount importance in assessing the effectiveness of these investments. For example, as Tenerife and Gran Canaria have very similar images, the hosting and the promotion of specific events can become a way to increase product differentiation and to diversify their own images.

6.3. Limitations and further research

Our approach shares the general strength of Big Data analyses compared to surveys in terms of volume of available information; it hence achieves a better representativity of the universe of reference (tourists, in our case). That said, the present approach is obviously not free from limitations, within both the data source and the methodology applied. Instagram data, together with the advantages described in the introduction, also have several drawbacks, which could be considered as more or less severe depending on the question of interest.

First, when it comes to investigating sentiment, pictures are not fit to exhibit negativity, as opposed to textual media and online reviews (Lup et al., 2015). Indeed, especially when traveling, Instagram users prefer to post images that represent themselves as active and confident individuals who are enjoying life, choosing to post carefully selected picturesque sceneries, possibly enhanced with filters. These considerations support the conclusion that the proposed approach of investigating pictures is particularly effective when dealing with the cognitive dimension of destination image. If the research goal is to deepen the knowledge of the affective dimension (hence focusing on the lower red arrow represented in Figure 1), textual, sentiment and statistical analysis carried out on captions and comments of Instagram pictures would provide more meaningful insights. This further work goes beyond the scope of the present paper and is left to future research.

Second, a criticism may arise regarding the accuracy of the labels generated by the VOR tool, which may be prone to measurement errors, given that the algorithm used by GCV is new, in evolution and not known to the research community. A manual random inspection that has been carried out is quite reassuring in this respect, but the possibility of having similar content labelled with different keywords cannot be excluded, and the task of stemming and clustering cognate words can only partially tackle this issue.

Third, another serious concern is the distribution of the users' demographic characteristics, and in fact, this applies to the great majority of social platforms. In this respect, we bear in mind that the destination image derived from Instagram data is, to a certain extent, biased towards perceptions of younger age groups (Li et al., 2018).

Given the novelty of our approach, there are numerous directions for future research. The most straightforward one would be to investigate captions and comments of pictures posted on Instagram to learn about the affective sphere of the destination image. In addition, one could combine imagery labels and text from captions and comments to assess which labels are associated with negative sentiment, opening interesting avenues to study the destination image. Given the large datasets and the great number of unique labels, using a dimensionality reduction method such as LASSO would be necessary. Additionally, with a sufficiently long time span it would be possible to study the long-term dynamics of IDDI and investigate how distance between destinations changes over the years, and which factors affect this change. A longer time span would also allow capturing recurrent traits of destination images, highlighting seasonal aspects that characterize the perception of tourists visiting the destination in different periods of the year.

Finally, another line of research can further explore the impact of exogenous events. So far, we have only shown the existence of a relationship between temperature, wildfires, and beach / non-beach related activities at the destination. A step further would be to aggregate labels in several fixed categories representing various types of activity to adopt a fractional multinomial response model to assess weather-driven substitution. This application is not limited to climate change, but can also be of interest for health emergencies, such as the Covid-19 pandemic. A rigorous assessment of how the health crisis impacts the destination image and the preferences in terms of tourism activities could then be undertaken.

References

Agapito D., Mendes J., Valle P., 2013, Exploring the conceptualization of the sensory dimension of tourist experience, *Journal of Destination Marketing & Management*, 2(2), 62-73.

Akehurst G., 2009, User generated content: the use of blogs for tourism organisations and tourism consumers, *Service Business*, 3(1), 51-61.

Alvarez M., Campo S., 2014, The influence of political conflicts on country image and intention to visit: A study of Israel's image, *Tourism Management*, 40, 70–78.

Avraham E., 2015, Destination image repair during crisis: Attracting tourism during the Arab Spring uprisings, *Tourism Management*, 47, 224–232.

Andreu L., Bigne J.E., Cooper C., 2000. Projected and perceived image of Spain as a tourist destination for British travellers. *Journal of Travel and Tourism Marketing*, 9 (4), 47–67.

Araña J.E., León C.J., 2008, The impact of terrorism on tourism demand, *Annals of Tourism Research*, 35 (2), 299-315.

Baloglu S., McCleary K., 1999, A model of destination image formation, *Annals of Tourism Research*, 26(4), 868-897.

Becken S., Jin X., Zhang C., Gao J., 2017, Urban air pollution in China: destination image and risk perceptions, *Journal of Sustainable Tourism*, 25 (1), 130-147.

Beerli A., Martin J.D., 2004, Factors influencing destination image, *Annals of Tourism Research*, 31(3), 657-681.

Beldona S., Morrison A.M., O'Leary J., 2005, Online shopping motivations and pleasure travel products: A correspondence analysis, *Tourism Management*, 26, 561–570.

Bello-Orgaz G., Jung J.J., Camacho D., 2016, Social big data: Recent achievements and new challenges, *Information Fusion*, 28, 1-15.

Bourdieu P., Bourdieu M., 2004, The peasant and photography, *Ethnography*, 5(4), 601–616.

Buhalis D., Law R., 2008, Progress in information technology and tourism management: 20 years on and 10 years after the internet the state of eTourism research, *Tourism Management*, 29(4), 609–623.

Caton K., Santos C. A., 2008, Closing the hermeneutic circle? Photographic encounters with the others, *Annals of Tourism Research*, 35(1), 7-26.

Chandrapaul, Soni R., Sharma S., Fagna H., Mittal S., 2019, News Analysis Using Word Cloud. In: Rawat B., Trivedi A., Manhas S., Karwal V. (eds), *Advances in Signal Processing and Communication*. Springer, Singapore.

Chen J.S., 2001, A case study of Korean outbound travelers' destination images by using correspondence analysis, *Tourism Management*, 22(4), 345–350.

Chen Y.C., Shang R.A., Li M.J., 2014, The effects of perceived relevance of travel blogs' content on the behavioral intention to visit a tourist destination, *Computers in Human Behavior*, 30, 787–799.

Chew E.Y.T., Jahari S.A., 2014, Destination image as a mediator between perceived risks and revisit intention: a case of post-disaster Japan, *Tourism Management*, 40, 382–393.

Chi C.G.Q., Qu H., 2008, Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach, *Tourism Management*, 29(4), 624–636.

Childers T.L, Houston M.J., 1984. Conditions for a picture-superiority effect on consumer memory. *Journal of Consumer Research*, 11(2), 643.

Chua A., Servillo L., Marcheggiani E., Moere A.V., 2016, Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy, *Tourism Management*, 57, 295-310.

Crompton, J.L., 1979, Motivations for Pleasure Vacations, *Annals of Tourism Research*, 6(4), 408-424.

Deng N., Li X.R., 2018, Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection, *Tourism Management*, 65, 267-278.

Dragin-Jensen C., Kwiatkowski G., 2019, Image interplay between events and destinations. *Growth and Change*, 50 (1), 446–469.

Echtner C.M., Ritchie J.R.B., 1991, The meaning and measurement of destination image, *Journal of Travel Studies*, 2(2), 2-12.

Enge E., 2019. Image Recognition Accuracy Study, Perficient, July 2019. https://www.perficient.com/insights/research-hub/image-recognition-accuracy-study (last checked: September 2020).

Gallarza M.G., Saura I.G., García H.C., 2002, Destination image: towards a conceptual framework, *Annals of Tourism Research*, 29(1), 56-78.

Gandomi A., Haider M., 2015, Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35(2), 137-144.

Garlick S., 2002, Revealing the unseen: tourism, art and photography, *Cultural Studies*, 16(2), 289-305.

Girardin F., Fiore F.D., Ratti C., Blat J., 2008, Leveraging Explicitly Disclosed Location Information to Understand Tourist Dynamics: A Case Study, *Journal of Location Based Services*, 2(1), 41-56.

Gkritzali, A., Gritzalis, D., Stavrou, D., 2018, Is Xenious Zeus still alive? Destination image of Athens in the years of recession, *Journal of Travel Research*, 57(4), 540-554.

Gunn C., 1972, *Vacationscape: Designing tourist regions*. Austin, Texas: University of Texas Press.

Hanlan J., Kelly S., 2005, Image formation, information sources and an iconic Australian tourist destination, *Journal of Vacation Marketing*, 11(2), 163–177.

Harrison B., 2004, Snap happy: Towards a sociology of "everyday" photography, *Studies of Qualitative Methodology*, 7, 23–39.

Hirschman E.C., 1986. The effect of verbal and pictorial advertising stimuli on aesthetic, utilitarian and familiarity perceptions. *Journal of Advertising*, 15 (2): 27–34.

Hockley W.E., 2008. The picture superiority effect in associative recognition. *Memory & Cognition*, 36 (7), 1351–1359.

Hunt J.D., 1975, Image as a factor in tourism development, *Journal of Travel Research*, 13(3), 1-17.

Jenkins O.H., 1999, Understanding and Measuring Tourist Destination Images, *International Journal of Tourism Research*, 1, 1-15.

Jiang K., Yin H., Wang P., Yu N., 2013, Learning from contextual information of geo-tagged web photos to rank personalized tourism attractions. *Neurocomputing*, 119, 17-25.

Kádár B., 2014, Measuring tourist activities in cities using geotagged photography, *Tourism Geographies*, 16(1), 88–104.

Kádár B., Gede M., 2013, Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography, *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), 78-88.

Kim S.E., Lee K.Y., Shin S.I., Yang S.B., 2017, Effects of tourism information quality in social media on destination image formation: The case of Sina Weibo, *Information & Management*, 54(6), 687-702.

Kim S., Stylidis D., Oh M., 2019, Is perception of destination image stable or does it fluctuate? A measurement of three points in time, *International Journal of Tourism Research*, 21, 447-461.

Koerbitz W., Önder I., 2014, Destination Benchmarking with Geotagged Photographs, *Information and Communication Technologies in Tourism*, 201-211.

Költringer C., Dickinger A., 2015, Analyzing destination branding and image from online sources: A web content mining approach, *Journal of Business Research*, 68, 1836-1843.

Kozak M., Baloglu S., 2011, *Managing and marketing tourist destinations: Strategies to gain a competitive edge*. New York, London: Routledge.

Kurashima T., Iwata T., Irie G., Fujimura K., 2013, Travel route recommendation using geotagged photos, *Knowledge and Information Systems*, 37(1), 37-60.

Lai K., Li Y., 2014, Image impacts of planned special events: Literature review and research agenda. *Event Management*, 18, 111–126.

Lee C.K., Lee Y.K., Lee B., 2005, Korea's destination image formed by the 2002 World Cup, *Annals of Tourism Research*, 32(4), 839–858.

Leung D., Law R., Van Hoof H., Buhalis D., 2013, Social media in tourism and hospitality: a literature review, *Journal of Travel & Tourism Marketing*, 30(1–2), 3–22.

Li Y.R., Lin Y.C., Tsai P.H., Wang Y.Y., 2015, Traveller-generated contents for destination image formation: mainland China travellers to Taiwan as a case study, *Journal of Travel & Tourism Marketing*, 32(5), 518–533.

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

Liang T., Soojin C., Alastair M.M., Xinran Y.L., 2009, The many faces of Macau: A correspondence analysis of the images communicated by online tourism information sources in English and Chinese, *Journal of Vacation Marketing*, 15, 79-94.

Lo I.S., McKercher B., Lo A., Cheung C., Law R., 2011, Tourism and online photography, *Tourism Management*, 32(4), 725-731.

Lo, I.S., McKercher, B., 2015, Ideal image in process: Online tourist photography and impression management, *Annals of Tourism Research*, 52, 104-116.

Lu W., Stepchenkova S., 2015, User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software, *Journal of Hospitality Marketing & Management*, 24(2), 119-154.

Lu D., Wu R., Sang J., 2017, Overlapped user-based comparative study on photosharing websites, *Information Sciences*, 376, 54-70.

Lup K., Trub L., Rosenthal L., 2015. Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed, *Cyberpsychology, Behavior, and Social Networking*.May, 18, 5.

Mack R.W., Blose J. E., Pan B., 2008, Believe it or not: credibility of blogs in tourism, *Journal of Vacation Marketing*, 14(2), 133-144.

Marcussen C., 2014, Multidimensional scaling in tourism literature, *Tourism Management Perspectives*, 12, 31-40.

Martin H.S., Bosque I.R.D., 2008, Exploring the cognitive–affective nature of destination image and the role of psychological factors in its formation, *Tourism Management,* 29(2), 263-277.

Mayo E.J., 1973, Regional images and regional travel behaviour, in The fourth annual conference proceedings of the Travel Research Association, *Research for Changing Travel Patterns: Interpretation and Utilisation,* 211-218.

Miah S.J., Vu H.Q., Gammack J., McGrath M., 2017, A Big Data Analytics Method for Tourist Behaviour Analysis, *Information & Management*, 54, 771-785.

Morosan C., Jeong M., 2008, The role of the internet in the process of travel information search, *Information Technology in Hospitality*, 51, 13-23.

Okuyama K., Yanai K., 2013, A travel planning system based on travel trajectories extracted from a large number of geotagged photos on the web. In: *The Era of interactive media* (pp. 657e670). New York: Springer.

Önder I., 2017, Classifying multi-destination trips in Austria with big data, *Tourism Management Perspectives*, 21, 54-58.

Pearce P.L., 1988, *The Ulysses Factor: Evaluating Visitors in Tourist Settings*. New York: Springer-Verlag.
Phillips-Wren G., Hoskisson A., 2015, An analytical journey towards big data, *Journal of Decision Systems*, 24(1), 87-102.

Pike S., 2002, Destination image analysis – a review of 142 papers from 1973 to 2000, *Tourism Management*, 23(5), 541-549.

Pike S., Gentle J., Kelly L., Beatson A., 2018, Tracking brand positioning for an emerging destination during the advent of the social media era: 2003 to 2015, *Tourism and Hospitality Research*, 18(3), 286–296.

Pike S., Ryan C., 2004, Destination positioning analysis through a comparison of cognitive, affective, and conative perceptions, *Journal of Travel Research*, 42, 333–342.

Pittman, M., & Reich, B., 2016, Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155-167.

Prayag, G., Hosany, S., & Odeh, K., 2013, The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions. *Journal of Destination Marketing & Management*, *2*(2), 118-127.

Pullman M.E., Robson S.K.A., 2007, Visual methods: using photography to capture customers' experiences with design, *Cornell Hotel and Restaurant Administration Quarterly*, 48(2), 121-144.

Qiu J., Wu Q., Ding G., Xu Y., Feng S., 2016, A survey of machine learning for big data processing, *EURSASIP Journal on Advances in Signal Processing*, 67, 1-16.

Richards, G., van der Ark L.A., 2013, Dimensions of cultural consumption among tourists: Multiple correspondence analysis, *Tourism Management*, 37, 71–76.

Sarma, M.K., 2007, Influence of Information Sources on Tourists: A Segment-Wise Analysis with Special Focus on Destination Image, *Vision*, 11 (1), 35-45.

Schmallegger D., Carson, D., 2008, Blogs in tourism: changing approaches to information exchange, *Journal of Vacation Marketing*, 14(2), 99-110.

Schmallegger D., Carson D., Jacobsen D., 2010, The use of photographs on consumer generated content websites: practical implications for destination image analysis. In N. Sharda (Ed.), *Tourism informatics: Visual travel recommender systems, social communities, and user interface design* (pp. 243-260). Hershey, PA: Information Science Reference.

Sontag S., 2002, *On photography*. London: Penguin Books.

Steele, J., & Iliinsky, N., 2010, *Beautiful visualization: Looking at data through the eyes of experts*. " O'Reilly Media, Inc.".

Stepchenkova S., Kirilenko A.P., Morrison A.M., 2009, Facilitating content analysis in tourism research, *Journal of Travel Research*, 47, 454-469.

Stepchenkova S., Mills J.E., 2010, Destination image: a meta-analysis of 2000-2007 research, *Journal of Hospitality Marketing and Management*, 19(6), 575-609.

Stepchenkova S., Zhan F., 2013, Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography, *Tourism Management*, 36, 590-601.

Stylianou-Lambert T., 2012, Tourists with cameras: Reproducing or producing? *Annals of Tourism Research*, 39, 1817–1838.

Stylidis D., Cherifi B., 2018, Characteristics of destination image: Visitors and non-visitors' images of London, *Tourism Review*, 73(1), 55–67.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M.J. Metzger, A.J. Flanagin (Eds.), *Digital media, youth, and credibility*, The MIT Press, Cambridge (MA), pp. 72-100.

Tasci, A.D.A., 2007, Assessment of factors influencing destination image using a multiple regression model, *Tourism Review*, 62 (2), 23-30.

Tasci A.D.A., Gartner W., Cavusgil S., 2007, Conceptualization and operationalization of destination image, *Journal of Hospitality and Tourism Research*, 31, 194-223.

Tham A., Croy G., Mair J., 2013, Social media in destination choice: distinctive electronic word-of-mouth dimensions, *Journal of Travel & Tourism Marketing*, 30(1–2), 144–155.

Tsai C.W., Lai C.F., Chao H.C., Vasilakos A.V., 2015, Big data analytics: a survey, *Journal of Big Data*, 2, 21, 1-32.

Urry J., 1990, *The Tourist Gaze*. Sage, London.

Urry J., 2002, *The Tourist Gaze*. Sage, London (second edition).

Volo S., Irimiás A., 2020, Instagram: Visual methods in tourism research, *Annals of Tourism Research*, forthcoming.

Vu H.Q., Li G., Law R., Ye, B.H., 2015, Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos, *Tourism Management*, 46, 222–232.

Wang C.Y., Hsu M.K., 2010, The relationships of destination image, satisfaction, and behavioral intentions: an integrated model, *Journal of Travel & Tourism Marketing*, 27(8), 829–843.

Whitehouse A.J., Maybery M.T., Durkin K., 2006. The development of the picture-superiority effect. *British Journal of Developmental Psychology*, 24 (4), 767–773.

Wu L., Shimizu T., 2020, Analyzing dynamic change of tourism destination image under the occurrence of a natural disaster: evidence from Japan, *Current Issues in Tourism*, 23 (16), 2042-2058.

Xiang Z., Gretzel U., 2010, Role of social media in online travel information search, *Tourism Management*, 31 (2), 179–188.

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

Zhang H., Fu X., Cai L.A., Lu L., 2014, Destination image and tourist loyalty: a meta-analysis, *Tourism Management*, 40, 213–223.

Zeng B., Gerritsen R., 2014, What do we know about social media in tourism? A review, *Tourism Management* Perspectives, 10, 27–36.

Zheng X., Gretzel U., 2010, Role of social media in online travel information search, *Tourism Management*, 31, 179-188.

Zhou X., Xu C., Kimmons B., 2015, Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform, *Computers, Environment and Urban Systems*, 54, 144-153.

Tables

Table 1. Descriptive statistics of Instagram posts, by destination (full sample).

	Island							
Indicator	*Tenerife*	*Gran Canaria*	*Fuerteventura*	*Lanzarote*	*Cyprus*	*Crete*	*Malta*	*Sicily*
Num. of posts (total)	49,234	33,145	38,452	25,471	63,561	93,752	74,925	119,896
Avg. num. of pictures per post	1.77	1.67	1.56	1.8	1.76	1.74	1.81	1.68
Avg. num. of comments per post	2.23	2.65	2.39	2.1	2.32	2.44	2.24	1.84
Avg. num. of likes per post	68.9	72.05	78.92	74.82	79.89	84.54	81.7	72.28
Share of geotagged posts	67%	67%	67%	65%	70%	74%	76%	73%

Table 2. Index of Distance in Destination Image based on top-20 labels ranks.

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.8	1.4	2.0	2.1	2.0	7.4	3.8
Fuerteventura	2.8		3.7	3.1	3.3	2.3	9.5	5.0
Tenerife	1.4	3.7		1	1.8	2.7	5.7	3.3
Gran Canaria	2.0	3.1	1		2.1	3.1	6.9	4.0
Cyprus	2.1	3.3	1.8	2.1		1.5	4.5	3.4
Crete	2.0	2.3	2.7	3.1	1.5		3.5	3.1
Malta	7.4	9.5	5.7	6.9	4.5	3.5		2.8
Sicily	3.8	5.0	3.3	4.0	3.4	3.1	2.8	
Avg. Distance	**3.1**	**4.3**	**2.8**	**3.2**	**2.7**	**2.6**	**5.8**	**3.6**

Table 3. Correspondence analysis on islands based on top-20 labels frequencies.

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		6.8	3.1	2.1	5.8	7.6	15.4	12.8
Fuerteventura	6.8		9.8	8.8	6.1	7.0	16.1	15.9
Tenerife	3.1	9.8		1	7.6	9.3	15.7	12.0
Gran Canaria	2.1	8.8	1		6.9	8.7	15.5	12.2
Cyprus	5.8	6.1	7.6	6.9		1.9	10.5	9.9
Crete	7.6	7.0	9.3	8.7	1.9		9.1	9.4
Malta	15.4	16.1	15.7	15.5	10.5	9.1		5.9
Sicily	12.8	15.9	12.0	12.2	9.9	9.4	5.91	

Table 4. IDDI computed on peak season (July, August).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.6	1.2	2.0	2.1	2.1	7.5	3.4
Fuerteventura	2.6		3.5	2.9	3.2	2.5	9.3	4.5
Tenerife	1.2	3.5		0.97	1.9	2.7	5.6	3.1
Gran Canaria	2.0	2.9	0.97		2.0	3.1	7.0	4.0
Cyprus	2.1	3.2	1.9	2.0		1.6	4.5	3.1
Crete	2.1	2.5	2.7	3.1	1.6		3.4	2.8
Malta	7.5	9.3	5.6	7.0	4.5	3.4		2.4
Sicily	3.4	4.5	3.1	4.0	3.1	2.8	2.4	
Avg. Distance	**3.0**	**4.1**	**2.7**	**3.1**	**2.6**	**2.6**	**5.7**	**3.3**

Table 5. IDDI computed on shoulder season (June, September).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.7	1.3	2.0	2.1	2.0	7.7	5.2
Fuerteventura	2.7		3.7	3.1	3.2	2.4	9.7	7.4
Tenerife	1.3	3.7		1.1	1.8	2.6	5.8	4.4
Gran Canaria	2.0	3.1	1.1		2.1	2.9	6.9	4.6
Cyprus	2.1	3.2	1.8	2.1		1.6	4.4	4.9
Crete	2.0	2.4	2.6	2.9	1.6		3.4	3.5
Malta	7.7	9.7	5.8	6.9	4.4	3.4		3.0
Sicily	5.2	7.4	4.4	4.6	4.9	3.5	3.0	
Avg. Distance	**3.3**	**4.6**	**2.9**	**3.2**	**2.9**	**2.6**	**5.8**	**4.7**

Table 6. Weather conditions and different types of posted activity, clustered s.e.

	Gran Canaria	Tenerife	Fuerteventura	Lanzarote
	b/p	b /p	b /p	b /p
		Beach-related labels		
10-day avg temp	0.012	0.019	-0.057*	0.022
	[0.313]	[0.145]	[0.072]	[0.227]
Linear trend	✓	✓	✓	✓
Quadratic trend	✓	✓	✓	✓
constant	0.332	0.317	2.749***	0.474
	[0.203]	[0.190]	[0.001]	[0.190]
		Nature-based labels		
10-day avg temp	-0.038**	-0.021***	0.015	0.019
	[0.039]	[0.010]	[0.696]	[0.527]
Linear trend	✓	✓	✓	✓
Quadratic trend	✓	✓	✓	✓
constant	1.758***	1.380***	0.376	0.494
	[0.000]	[0.000]	[0.653]	[0.420]
N	113	113	113	104

p values in squared brackets; significance levels: * p<0.1, ** p<0.05, *** p<0.01

Table 7. Weather conditions and different types of posted activity, pooled sample (Gran Canaria, Tenerife, Lanzarote), clustered s.e.

	Beach-related labels		Nature-based labels	
	(1)	(2)	(3)	(4)
	b/p	b/p	b/p	b/p
10-day avg temp	0.018**	0.018**	-0.018**	-0.018**
	[0.031]	[0.028]	[0.017]	[0.030]
Linear trend	✓	✓	✓	✓
Quadratic trend	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓
Destination-specific linear trend		✓		✓
Destination-specific quadratic trend		✓		✓
constant	0.253	0.216	1.320***	1.342***
	[0.118]	[0.150]	[0.000]	[0.000]
N	330	330	330	330

p values in squared brackets; significance levels: * p<0.1, ** p<0.05, *** p<0.01

Table 8. SUR: impact of temperatures and wildfires on different types of activities in the Canary Islands

	Log (Nature-based labels)	Log (Beach-related labels)
	b/p	b/p
10-day avg temp	-0.024*	0.036**
	[0.098]	[0.012]
post-fire	0.008	-0.064
	[0.860]	[0.149]
treat	-0.071*	0.035
	[0.063]	[0.359]
Linear trend	✓	✓
Quadratic trend	✓	✓
Destination FE	✓	✓
constant	5.052***	3.393***
	(0.000)	(0.000)
N	330	330

p-values in parentheses; significance levels: * p<0.1, ** p<0.05, *** p<0.01.
Covariance matrix estimated using small sample adjustment.

Figures

Figure 1. The destination image: a conceptual framework

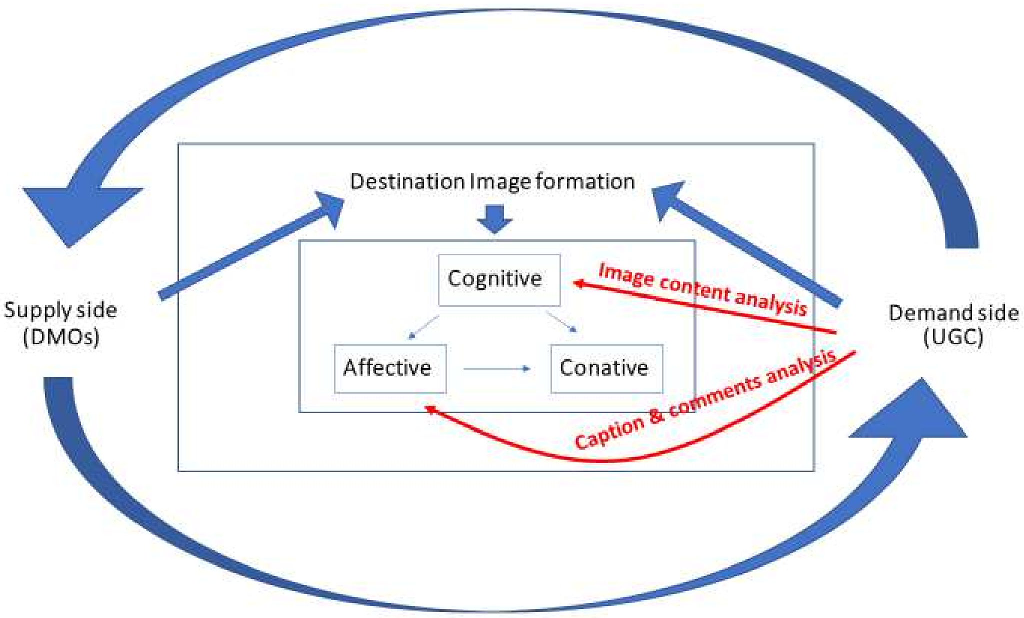
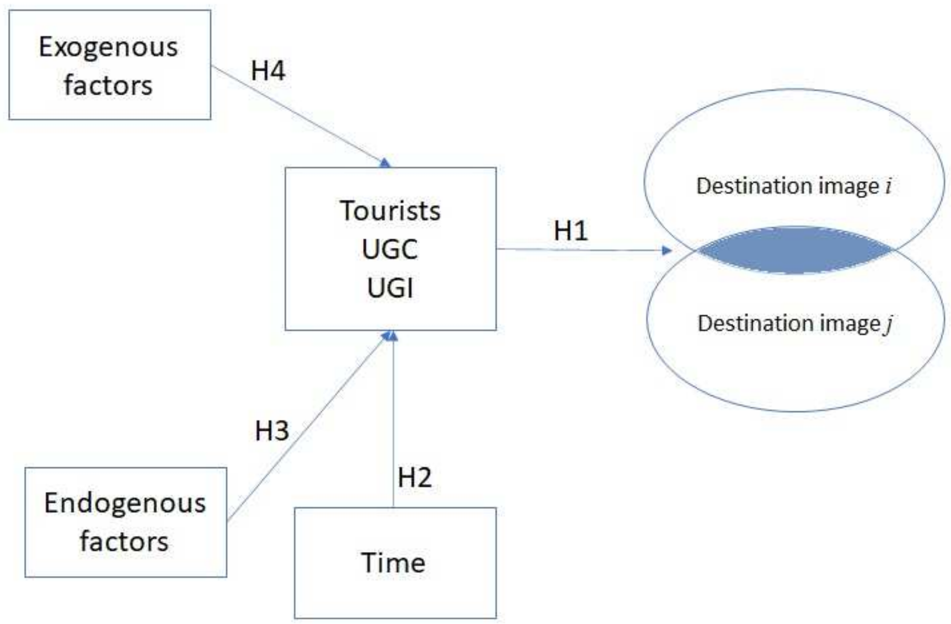


Figure 2. Different destination images and hypotheses



Notes: Destinations *i* and *j*'s images are represented by bubbles of the same size. The overlapping (non-overlapping) area of the bubbles graphically represents the degree of similarity (uniqueness) between destination images, according to the characteristics of the pictures posted by tourists on the social media (H1). These characteristics vary with time (H2, seasonal factors), endogenous factors (H3, events organized at the destination), and exogenous factors (H4, climate events hitting the destination).

Figure 3. Word clouds based on image labels.

Crete

Cyprus

Malta

Sicily

Figure 4. Correspondent biplot of destination images of different islands.

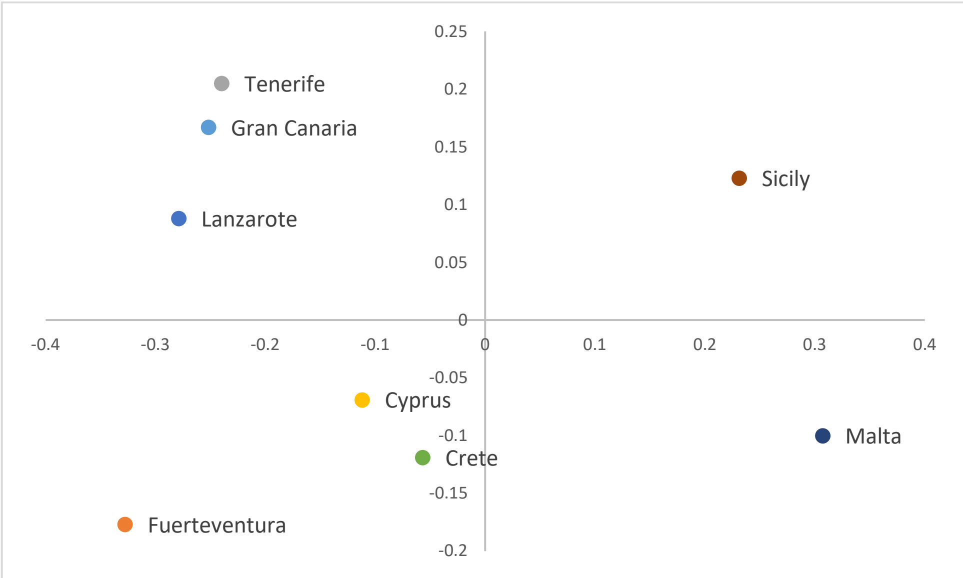


Figure 5. Daily relative frequencies of windsurfing and kitesurfing-related labels, Fuerteventura.

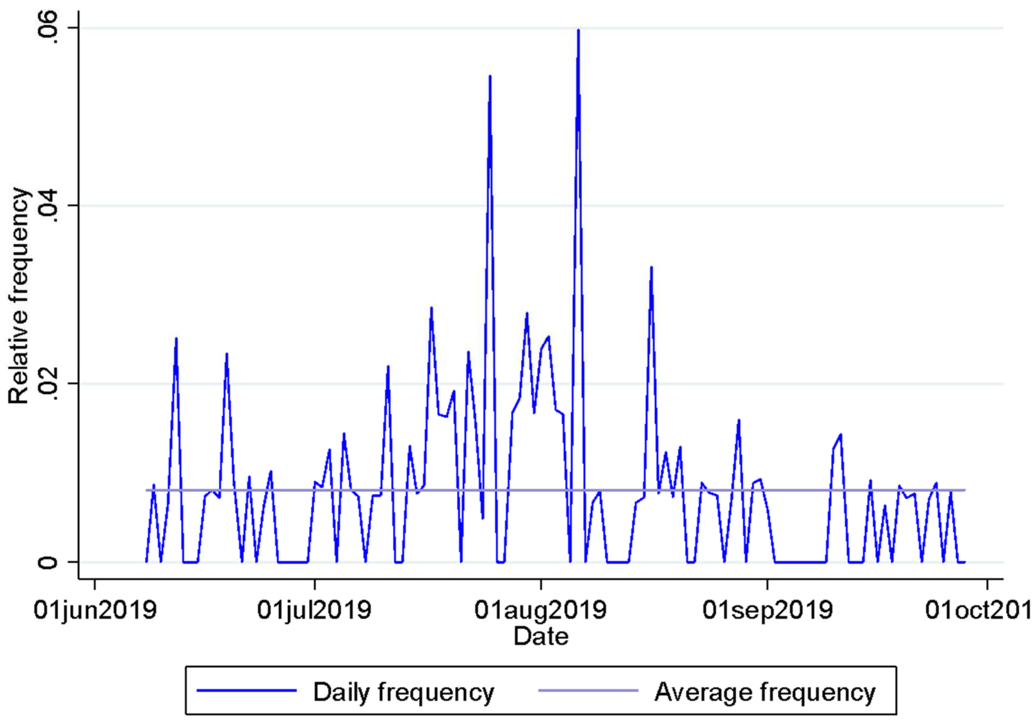
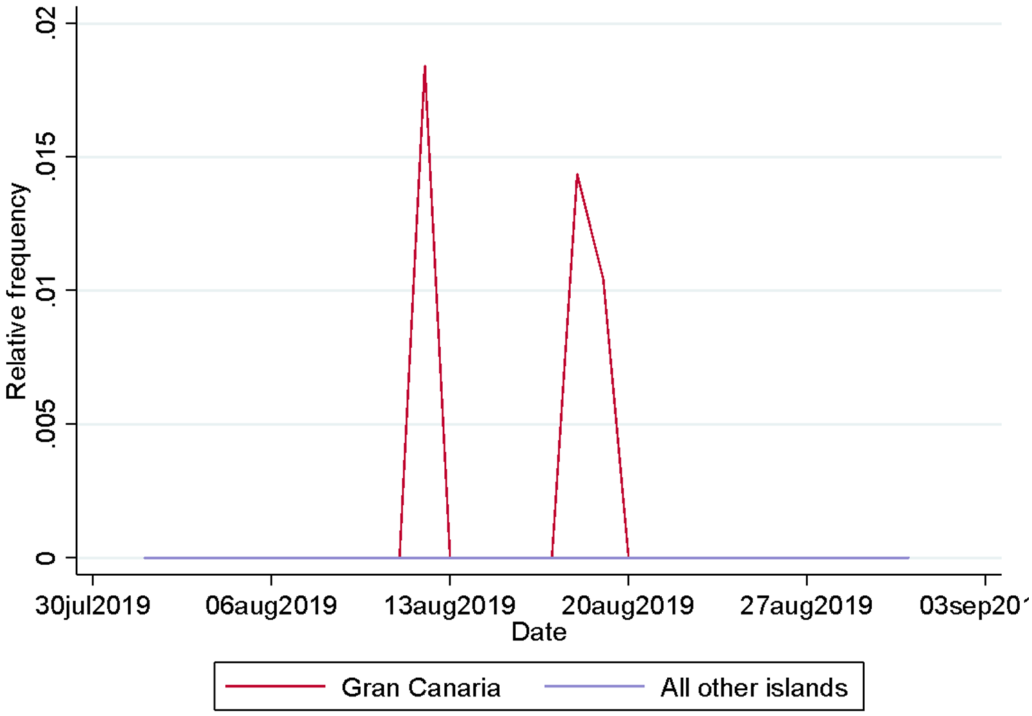


Figure 6. Daily relative frequencies of wildfire-specific labels in August (Gran Canaria vs all other islands).



Appendix

Table A.1. Top-20 labels by frequency.

	Tenerife		Gran Canaria		Lanzarote		Fuerteventura	
rank	**label**	**# of occur.**	**label**	**# of occur.**	**label**	**# of occur.**	**label**	**# of occur.**
1	Sky	30177	Sky	19309	Sky	17366	Sky	25652
2	Vacation	21544	Vacation	15070	Vacation	12353	Sea	18920
3	Tree	19964	Tree	13986	Water	11051	Vacation	18686
4	Mountain	18511	Water	10458	Tree	10246	Ocean	18145
5	Water	17875	Plant	9312	Sea	9883	Water	17784
6	Sea	15808	Sea	9140	Ocean	9609	Beach	15584
7	Ocean	14770	Summer	8057	Mountain	8386	Tree	11442
8	Plant	13683	Mountain	7760	Plant	7378	Summer	10332
9	Fun	10766	Ocean	7620	Beach	6046	Coast	10096
10	Summer	10602	Fun	7304	Summer	5865	Shore	8422
11	Tourism	10162	Tourism	6518	Landform	5653	Fun	8028
12	Landform	9920	Nature	6201	Coast	5564	Sand	7984
13	Photography	8817	Photography	6079	Fun	5552	Cloud	7095
14	Cloud	8711	Landform	5850	Cloud	5531	Plant	6915
15	Nature	8428	Architecture	5620	Tourism	5304	Blue	6470
16	Beauty	8058	Beach	5586	Landscape	5060	Landform	6360
17	Coast	7961	Hair	5235	Rock	4792	Nature	6278
18	Beach	7493	Palm	4974	Photography	4715	Body	6110
19	Leisure	7404	Beauty	4882	Blue	3889	Photography	6069
20	Hair	7227	Landscape	4757	Shore	3733	Mountain	6016
Total # of occur.		835306		541522		441131		585693
Total # of pictures		74537		48337		39381		52577

	Cyprus		Crete		Malta		Sicily	
rank	**label**	**# of occur.**	**label**	**# of occur.**	**label**	**# of occur.**	**label**	**# of occur.**
1	Sky	33698	Sky	61724	Sky	44910	Sky	76751
2	Vacation	31342	Water	55406	Water	44586	Architecture	60202
3	Water	27804	Sea	49728	Sea	38516	Water	50100
4	Sea	25210	Vacation	47536	Architecture	37323	Sea	44155
5	Ocean	22905	Ocean	43118	Vacation	34444	Ocean	39546
6	Tree	22315	Tree	28879	Ocean	30878	Vacation	36878
7	Summer	18629	Summer	27730	Tourism	27531	Building	36393
8	Fun	14970	Coast	26198	Building	24139	Tourism	31170
9	Plant	14502	Beach	25686	Coast	20417	Tree	30631
10	Architecture	14282	Tourism	23952	Summer	18759	Coast	24449
11	Beach	14054	Architecture	21059	Town	17786	Mountain	23962
12	Tourism	13287	Plant	20482	Landform	14014	Plant	22906
13	Coast	12141	Blue	18839	Blue	13937	Town	22423
14	Building	10928	Fun	17860	Tree	13875	Photography	20798
15	Photography	10836	Mountain	17719	Street	13739	Landform	20541
16	Beauty	10604	Building	17335	Photography	13652	Cloud	19231
17	Leisure	9825	Landform	17267	Coastal	13322	Summer	18640
18	Nature	9514	Body	16659	Rock	12649	Nature	17938
19	Blue	9133	Beauty	15139	Fun	12614	Beach	17363
20	Hair	9051	Shore	13900	City	12492	Food	16782
Total # of occur.		1064845		1587398		1318794		1991061
Total # of pictures		95808		141538		117576		175481

Table A.2. Index of Distance in Destination Image based on top-20 labels ranks (quadratic distances).

	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		8.8	2	4.1	6.9	4.4	113.4	27.0
Fuerteventura	8.8		32.2	10.0	22.6	12.6	187.1	49.6
Tenerife	2	32.2		1	5.4	5.6	42.6	14.0
Gran Canaria	4.1	10.0	1		5.3	13	64.9	29
Cyprus	6.9	22.6	5.4	5.3		2.8	29.0	15.6
Crete	4.4	12.6	5.6	13	2.8		19.7	13.6
Malta	113.4	187.1	42.6	64.9	29.0	19.7		13.9
Sicily	27.0	49.6	14.0	29	15.6	13.6	13.9	
avg. distance	**23.8**	**46.1**	**14.7**	**18.2**	**12.5**	**10.2**	**67.2**	**23.3**

Correspondence Analysis

The correspondence analysis run on the labels of pictures for the islands under investigation reveals a strong association between the categorical variables and a two-dimensional solution (a principal normalization was adopted), explaining 81.41% of total inertia (Table A.3). Table A.4 provides decomposition by island (decomposition by label is available upon request) and contains total inertia (squared of singular value) explained, and the quality of the approximation for each island. Additionally, Table A.4 shows, for each dimension, the coordinates, the squared correlation, and the contribution of each category to the dimensions. Figure A.2 is similar to Figure 4, but also highlights the (dis)similarities between destination images and the labels that best characterize them.

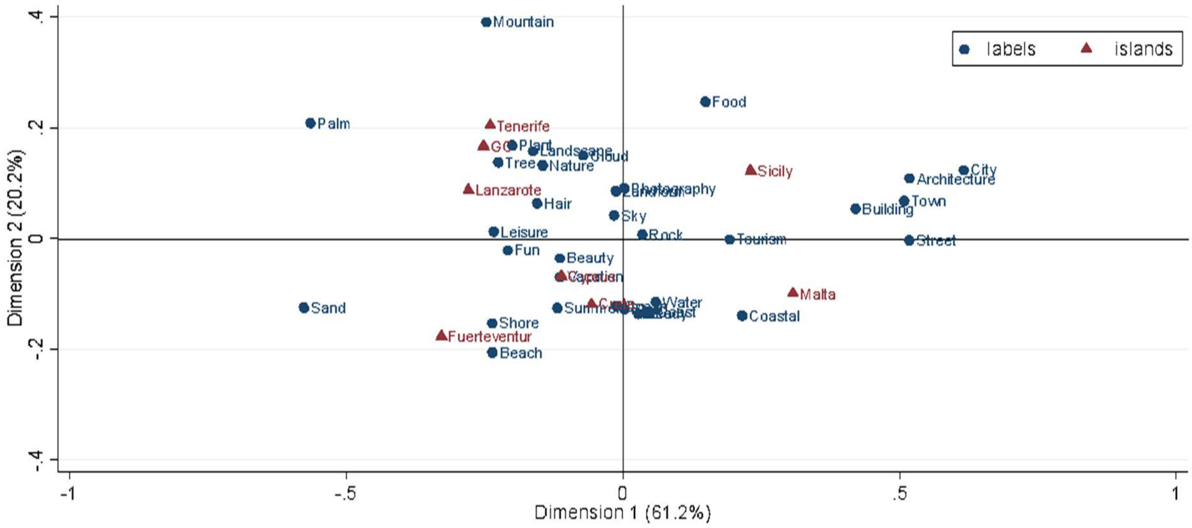
Table A.3. Correspondence analysis results.

Dimension	Singular Value	Inertia	Chi square	Proportion explained	Cumulative proportion
1	0.226767	0.051423	177623.2	61.18	61.18
2	0.130402	0.017005	58736.94	20.23	81.41
3	0.087788	0.007707	26620.01	9.17	90.58
4	0.064849	0.004206	14526.27	5	95.59
5	0.051745	0.002678	9248.7	3.19	98.77
6	0.02767	0.000766	2644.65	0.91	99.68
7	0.016305	0.000266	918.32	0.32	100
Total		**0.084049**	**290318.1**	**100**	

Table A.4. Contribution of dimensions to the inertia of each destination image.

		Total		Coordinates		Squared Correlation		Contribution to inertia	
Islands	**Mass**	**quality**	**%inert**	**dim 1**	**dim 2**	**dim 1**	**dim 2**	**dim 1**	**dim 2**
Lanzarote	0.053	0.857	0.063	-0.279	0.088	0.077	0.78	0.024	0.081
Fuerteventura	0.078	0.755	0.171	-0.328	-0.177	0.17	0.585	0.143	0.163
Tenerife	0.093	0.868	0.127	-0.24	0.205	0.366	0.502	0.23	0.104
Cyprus	0.119	0.385	0.063	-0.112	-0.069	0.106	0.28	0.033	0.029
Gran Canaria	0.06	0.756	0.087	-0.252	0.167	0.23	0.526	0.099	0.075
Crete	0.203	0.7	0.06	-0.057	-0.119	0.571	0.129	0.169	0.013
Malta	0.162	0.918	0.218	0.307	-0.1	0.088	0.83	0.094	0.296
Sicily	0.231	0.896	0.21	0.231	0.123	0.199	0.697	0.206	0.239

Figure A.2: Correspondence plot of islands and labels showing the (dis)similarities of islands and their relationship with labels.



Monthly disaggregation of the IDDI index

Table A.5. IDDI for the month of June 2019.

June	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.6	1.7	2.2	2.4	2.1	7.4	5.4
Fuerteventura	2.6		3.5	2.9	3.4	2.4	9.6	6.9
Tenerife	1.7	3.5		1.2	1.9	2.6	6.1	5.0
Gran Canaria	2.2	2.9	1.2		2.4	3.1	7.1	4.6
Cyprus	2.4	3.4	1.9	2.4		1.5	4.1	4.3
Crete	2.1	2.4	2.6	3.1	1.5		3.4	3.3
Malta	7.4	9.6	6.1	7.1	4.1	3.4		2.9
Sicily	5.4	6.9	5.0	4.6	4.3	3.3	2.9	
Avg. Distance	**3.4**	**4.5**	**3.1**	**3.4**	**2.8**	**2.6**	**5.8**	**4.6**

Table A.6. IDDI for the month of July 2019.

July	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.4	1.2	1.7	1.9	2.1	7.2	3.6
Fuerteventura	2.4		3.3	2.9	3.2	2.4	9.1	5.4
Tenerife	1.2	3.3		0.95	1.9	2.7	5.8	3.6
Gran Canaria	1.7	2.9	0.95		2.1	3.4	7.1	4.4
Cyprus	1.9	3.2	1.9	2.1		1.5	4.3	3.4
Crete	2.1	2.4	2.7	3.4	1.5		3.5	3.0
Malta	7.2	9.1	5.8	7.1	4.3	3.5		2.7
Sicily	3.6	5.4	3.6	4.4	3.4	3.0	2.7	
Avg. Distance	**2.9**	**4.1**	**2.8**	**3.2**	**2.6**	**2.7**	**5.7**	**3.7**

Table A.7. IDDI for the month of August 2019.

August	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.4	1.1	1.8	2.1	2.1	8.0	3.3
Fuerteventura	2.4		3.4	2.6	3.3	2.4	9.9	4.5
Tenerife	1.1	3.4		0.99	1.9	2.5	5.6	2.9
Gran Canaria	1.8	2.6	0.99		2.0	2.9	6.9	3.6
Cyprus	2.1	3.3	1.9	2.0		1.6	4.7	3.6
Crete	2.1	2.4	2.5	2.9	1.6		3.4	2.8
Malta	8.0	9.9	5.6	6.9	4.7	3.4		2.8
Sicily	3.3	4.5	2.9	3.6	3.6	2.8	2.8	
Avg. Distance	**3.0**	**4.1**	**2.6**	**3.0**	**2.7**	**2.5**	**5.9**	**3.4**

Table A.8. IDDI for the month of September 2019.

September	Lanzarote	Fuerteventura	Tenerife	Gran Canaria	Cyprus	Crete	Malta	Sicily
Lanzarote		2.4	1.1	1.8	2.2	1.9	7.8	5.5
Fuerteventura	2.4		3.5	2.7	3.1	2.6	10.2	8.1
Tenerife	1.1	3.5		2.3	2.1	2.5	5.8	4.0
Gran Canaria	1.8	2.7	2.3		2.5	2.9	7.6	5.5
Cyprus	2.2	3.1	2.1	2.5		1.4	4.4	5.0
Crete	1.9	2.6	2.5	2.9	1.4		3.5	3.7
Malta	7.8	10.2	5.8	7.6	4.4	3.5		3.2
Sicily	5.5	8.1	4.0	5.5	5.0	3.7	3.2	
Avg. Distance	**3.3**	**4.6**	**3.0**	**3.6**	**2.9**	**2.6**	**6.1**	**5.0**