Image of a Tourist Attraction and Psychographic Profile of the Tourist: Data Mining Approach

Nancy Guillen Rojas [1], Aldo Medina Gamero [1], Klinge Villalba Condori [1], Lizbet Herrera Quinto [1] and Beatriz Zevallos Allcca [1]

[1] *Tourism Management Department, Universidad San Ignacio de Loyola, Lima - Peru*

Abstract

The goal of this research is to understand the image of an attraction from the national tourist point of view and its psychographic profile. The study was made in relation to the Archeological Sanctuary of Pachacamac in Lima, Peru. For this purpose, unstructured data mining—a cutting-edge method—was used. In order to get relevant data, 209 comments from TripAdvisor were collected, while for the Big Five personality analysis, five comments per visitor about different places were processed to determine their psychographic profile (a total of 1045 comments). Through qualitative research method, terms that describe the touristic image of this place were identified, which were then analyzed using the Lexalytics Software to determine the related keywords, their emotional valuation, and frequency. The applied research design was phenomenological, as we sought to know about the visitor experience of the tourist. Results indicated a positive perspective on the image of the Archeological Sanctuary of Pachacamac by visitors due to how close they are to the location—the site's museum, the historical weight that precedes this place, cultural tourism, and the services provided there. Moreover, it was noted that most visitors were men (75.60%) from the X Generation and Baby Boomers and they are basically liberal, artists, organized, hard-working people, contemplative, and kind. This data contributes to better decision-making, generating use full information that could be used by tourism professionals and those in charge of the destination's management.

Keywords

Tourist image, Data mining, Psychographic profile

1. Introduction

The image of the tourist attraction is viewed as the whole idea represented in the mind of the traveler, which could be a result of the perception of the characteristics associated with the place they are visiting (Groen, 2012). The image arises from the very same visitors that go to the destination (Baloglu, Henthorne & Sahin, 2014). Beerli and Martín (2004) assert that the gathered information after a visit or a personal experience in the destination contributes to the creation of the place's touristic image. Thus, studying the touristic image allows for the identification of differences between what it is offered and what the visitor perceives (Kaur, Chauhan & Medury, 2016). This facilitates decision-making, the design of marketing strategies, and the management of the destination. Furthermore, through the psychographic study, detailed information on the lifestyles, activities, interests, and opinions of the visitors can be obtained in order to find out the profile of the people who visit a destination (Moutinho, 2020).

Due to the coronavirus disease (COVID-19) and with the purpose to reactivate the sector's economy, many countries are carrying out advertising campaigns via digital platforms, TV, radio, and other more innovative media, using tools such as big data, where structured and unstructured data can be found. It is important to mention that all these campaigns focus on national tourism due to the traveling restrictions that some countries are still implementing and also because of the national tourism volume, which is six times greater than international tourism (OMT, 2020). It must be stressed that unstructured data mining allows for the transformation of unstructured information into systematic patterns or behaviors that can be used by tourism professionals through marketing strategies, easy access to information for an effective tour-operation, management of questions and complaints, etc. (Bucur, 2015). The right projection of a touristic image is a key factor to establish the competitiveness, recognition, and positioning of a place. This way, marketing strategies can be established as well as the delivery of a satisfactory touristic product in keeping with the projected touristic image (Camprubí, Guia & Comas, 2008)

This research aims at proving that the use of free availability unstructured data in text format can be used in academic research (Groen, 2012). The web has become a wide resource to obtain this kind of information, whose content is a valuable contribution to organizations as well as for decision-making (Gupta & Rathore, 2013). Plenty of information on different topics can be found online, such as processed documents, e-mails, audio, videos, and texts (Anderson, 2016). This includes free-access platforms such as TripAvisor, where travelers exchange information about their experiences (Godnov & Redeck, 2016). The use of free-text format data help in the carrying out of more research in tourism given its unobstructed availability and easy accessibility (Gretzel & Yoo, 2008).

Table 1
Arrival of national tourists to the Archeological Sanctuary of Pachacamac. Source: Ministry of External Trade and Tourism, 2020.

YEAR	2015	2016	2017	2018	2019	2020
Number of visitors	86028	112627	105237	108680	118726	12226
Growth percentage	5.85%	30.92%	−6.56%	3.27%	9.24%	−89.70%

The development of domestic tourism contributes to the knowledge and valuation of the culture, and fosters a sense of belonging, and pride in what belongs to the region. According to World Tourism Organization (UNWTO - 2020), domestic tourism is being used as a tool to reactivate the sector. To this, it must be added that the proportion of domestic tourism in Peru is 92% higher than foreign tourism. Moreover, studying and understanding the image of this attraction and linking it to the psychographic profile of the visitors becomes useful to improve web contents and the promotion provided by the management of the Archeological Sanctuary of Pachacamac, the Ministry of Culture, and the Ministry of External Trade and Tourism.

This research proposes to integrate the development of the psychographic study through data mining to be executed for specific research in the tourism field.

2. Literature review
2.1. Factors of a touristic destination image

Beerli and Martín (2004), after going over the attractions and characteristics in the existing scales, propose the inclusion and classification of all factors that influence the appreciation of the image created by the individuals in nine dimensions, namely, natural resources, political and economic factors, natural environment, social environment, and overall environment of the place. The researchers also highlight the selection of characteristics that are used in the design of a scale that will greatly depend on the attractions that each destination has, their positioning in the market, and on assessing the goals of the image that the individual may perceive.

Conversely, Baloglu and McClaery (1999) present a model that allows for understanding the factors that influence the creation of the touristic image of a destination. According to the proposed model, it can be observed that in the creation of the touristic image of a destination, there are two core factors of influence. On the one hand, we have the stimulus factors, which are secondary sources of information (can be induced, organic, or autonomous), and primary sources of information (life experience and intensity of the visit). On the other hand, we find personal factors, which can be determined by the motivations, psychographic profile, and sociodemographic characteristics of the individual. These factors directly influence the destination image perceived in the perceiving/cognitive image, and in the affective image, as well as, consequently, in the overall complex image of the destination (Beerli & Martín, 2004; Baloglu & McCleary, 1999; Gartner, 1993).

2.2. Unstructured data mining

These are information-gathering processes based on statistical analysis, database methodology, and mathematical algorithms. They are built to discover useful information within unstructured data that have patterns or systematic behaviors that present the existent relations between the data in order to predict trends and behaviors (Olmeda & Sheldon, 2001). In other words, unstructured data mining does not have a predefined format, is not ordered, is hard to categorize, and can be found in many forms, such as social media and web pages (Gretzel & Yoo, 2008; Kaur, Chauhan & Medury, 2016). However, that information can be turned into a format that allows being worked on in order to identify patterns. In other words, it is possible to give it a structure that allows for the collection of information and carrying out a content analysis (Tsujii, Takahashi, Fujita & Tsuda, 2014). A clear example is what Jalbani, Memon, Memon, Depar, and Koondhar (2018) point out; they performed the processing of comments from different social media in Cloud Natural Language to identify the feelings of the users about opinions on social media.

2.3. Psychographic profile

Demby (1994) defines psychography as "the use of psychological, sociological, anthropological factors as desired benefits from the desired behavior, self-concept and lifestyle" (p. 1). Psychography is developed through the discovery of several consumer typologies, becoming the most relevant factor due to its psychological diversity (Sarli & Hon, 2011). Additionally, Hsu, Kang, and Wolfe (2002) point out that "psychography is the development of consumer psychological profiles, being psychologically based on the measures of different life models or lifestyles" (p. 4). By applying psychography, diverse aspects of consumers can be analyzed, such as thoughts, reflections, lifestyles, personality traits, demography, socioeconomic level, and expenditure level (Sarli & Hon, 2011; Tintaya, 2018).

2.4. Big Five personality model

For a long time, a large number of researchers have studied the validation of personality measurement; many agree that there are five personality factors (Barrick & Mount, 1991). According to Zhao and Seibert (2006), the Big Five model provides a complete classification of the personality, each of the dimensions broadly describe psychological functioning, which is composed of more specific traits.

In this sense, authors Costa and McCrae have posed the most developed operational model of the Big Five (Zhao & Seibert, 2006). First, openness to experience: characterizes people that are intellectually curious and tend to seek new experiences, and explore ideas (Zhao & Seibert, 2006). It also describes the curiosity and the artistic side of people (Verma, Kumar & Chandra, 2017). Individuals with high levels of openness to experience are creative, innovative, imaginative, reflexive, and not so traditional. On the contrary, individuals with low levels of openness to experience are characterized as being conventional, of limited interest, and not so analytical (Costa, Busch, Zonderman & McCrae, 1986; McCrae & John, 1992; Zhao & Seibert, 2006). Second, conscientiousness: indicates the level of

organization, persistence, hard-work, and motivation of an individual when seeking to achieve a goal (Costa, Busch, Zonderman & McCrae, 1986 and Zhao & Seibert, 2006). Unlike the other dimensions, it is composed of two sides: motivation for achievement and reliability (Barrick & Mount, 1991). It is also defined as the tendency to self-discipline, responsibility, and sticking to rules (Verma, Kumar & Chandra, 2017). Third, extraversion: allows understanding up to where a person is assertive, dominant, energetic, active, communicative, and enthusiastic (Zhao & Seibert, 2006). Individuals with a high level of extraversion tend to be joyful, very sociable, love to have fun, and are constantly looking for external emotions or stimuli (Costa, Busch, Zonderman & McCrae, 1986; McCrae & John, 1992; Zhao & Seibert, 2006). However, individuals with a low level of extraversion prefer to spend more time alone, and are very reserved, calm and independent (Costa, Busch, Zonderman & McCrae, 1986; McCrae & John, 1992; Zhao & Seibert, 2006). Fourth, kindness: evaluates the interpersonal orientation of each individual (Zhao & Seibert, 2006).

3. Methodology

This research addresses the image of an attraction using unstructured data mining, which is a cutting-edge method that can be used by tourism professionals (Godnov & Redeck, 2016). It is worth mentioning that this method benefits from the amount of content in free-text format (Gassiot, 2012). Moreover, data was collected by means of verbatim opinions (comments about own experiences), and they were analyzed with the purpose of understanding the most eye-catching characteristics of the attraction.

Based on the above, this is descriptive and qualitative research, as components related to the image were identified in order to submit them to in-depth analysis, for that purpose, the software Lexalytics was used to determine keywords related to the attraction's image as well as their emotional valuation and Apply Magic Sauce to learn about the psychographic profile of the visitors (Hernández, Fernández & Baptista, 2014). To determine the touristic image, 209 assessments were employed made by travelers that visited the Archeological Sanctuary of Pachacamac from January 2015 to March 2020, who left their comments on TripAdvisor. Five comments from different places were processed in order to know about the Big Five personalities of each user that takes part in the sampling. A total of 1,045 comments were processed.

The collection process was made in October 2020. All comments were captured in English (original comment in Spanish) and subsequently evaluated with Grammarly in order to correct grammatical, contextual, and spelling mistakes.

4. Findings and debate

Table 2
Concepts related to the Areas–Services within the Archeological Sanctuary of Pachacamac

Keywords	Frequency	Emotional score
Museum	High	+3.32 (positive)
Well-trained guides	Medium	+1.83 (positive)
Visitor center	Medium	+1.61 (positive)
Toilet facilities	Medium	+1.49 (positive)
Wide areas	Medium	+1.48 (positive)
Well-signaled paths	Medium	+0.78 (positive)
Several sections	Medium	+0.66 (positive)
Interesting gift shops	Medium	+0.65 (positive)
Good staff	Medium	+0.61 (positive)
Cozy cafeteria	Low	+0.48 (neutral)
Specific stops	Low	−0.82 (negative)

The term "museum" is the one with higher frequency among the terms related to the complementary areas and services that visitors can enjoy within the Archeological Sanctuary of Pachacamac. It gets a positive connotation (+3.32), and that is because the museum was remodeled and adjusted to the attraction's characteristics, thereby causing a positive impact on the visitors. The museum has to be visited to understand the history and chronology of the cultures that are developed in the sanctuary. Moreover, guides are provided in the attraction, among other services. The national tourist uses the term "well-trained guides" (+1.83) to refer to the work that the guides of the place do in assisting and informing visitors. Among other terms of medium frequency, there are: "visitor center" (+1.61), "toilet installations" (+1.49), "wide areas" within the attraction (+1.48), "well-signaled paths" (+0.78), "several sections" or divisions within the sanctuary (+0.66), an "interesting gift shops" (+0.65), and "good staff" (+0.61). Visitors express a positive emotion toward the services that are offered within the attraction. Among the low-frequency terms, we can find "cozy cafeteria" (+0.48), even if it highlighted the good service offered by the cafeteria, the circuit per se did not exceed the visitor's expectations, that is why the term was scored as neutral, and "specific stops" (−0.82), which make reference to the touristic bus stops in order to listen to the guide's explanations, were negatively evaluated because some visitors would rather walk, expressing that the tour would be more enjoyable if done by foot. However, that is not an option as there are people who do not respect the heritage and damage the place.

Table 3
Concepts relates to the Touristic Activity- Sustainability

Keywords	Frequency	Emotional Score
School trip	High	+0.02 (neutral)
Restoration works	Medium	+2.39 (positive)
Cultural tour	Medium	+0.74 (positive)
Touristic resource	Low	+0.35 (neutral)
Receptive tourism	Low	+0.35 (neutral)
Low investment	Low	+0.01 (neutral)

The national tourist considers that the term "school trip" is highly associated to the touristic activity that is carried out in the attraction. However, it gets a neutral valuation (+0.02), taking into account that the Archeological Sanctuary of Pachacamac is one the most visited attractions by students, with more than 50% of X Generation visitors. Likewise, "restoration works" being done in the sanctuary are very valuable to visitors, since the sanctuary is being preserved and new national archeology findings can be discovered. It also gets a positive connotation (+2.39). It is also thought that, in the past, there were "low investment" in the maintenance and research works, for being one the most important touristic centers in Lima. However, currently, the biggest concerns are maintenance and preservation of the attraction. On the other hand, regarding the touristic activity, the "cultural tour" is the most common due to the sanctuary characteristics. It gets a medium frequency and a positive connotation (+0.74). Pachacamac hosts the history of different civilizations. It is the perfect place for cultural tourism.

Table 4
Concepts related to the Emotions that the visit to the Archeological Sanctuary of Pachacamac generates

Keywords	Frequency	Emotional score
Magical setting	Alta	+0.44 (positive)
Wonderful sight	Alta	+2.71 (positive)
It is definitely worthy	Alta	−0.10 (negative)
Fascinating place	Media	+2.81 (positive)
Great experience	Media	+1.82 (positive)

Must-see spot	Media	+1.00 (positive)
Amazing history	Media	+0.52 (positive)
Bad feeling	Baja	+0.35 (neutral)
Full discovery	Low	−0.75 (negative)

The emotions generated by the visit to the Archaeological Sanctuary of Pachacamac have been diverse, with valuations ranging from positive to negative. In this sense, the term that obtained the highest frequency is that of Pachacamac, a "magical setting," the visitors have in mind all the history that it houses due to the passing of different cultures, this term receives a positive connotation (+0.44). Moreover, visitors consider that the sanctuary provides a "wonderful sight" (+2.71), and that could be because the attraction is located very close to the Pacific Ocean and, from Templo del Sol, an impressive view of Lurín and the ocean can be enjoyed, especially if the visit takes place during the afternoon to watch the sunset, as many visitors' stress: "it is definitely worthy" to visit the attraction and to experience what it has to offer. Even though it is scored with a high frequency, this term receives a negative connotation (−0.10), because many visitors consider that, in spite of being close to the city and presenting cultural and historical richness, not everyone is willing to visit and enjoy its wonder. Terms of medium frequency such as "fascinating place" (+2.81), "great experience" (+1.82), "must-see spot" (+1.00), and "amazing history" (+0.52) are perfectly complemented with the high-frequency terms, and they generated a positive feeling in the visitors. Finally, the national tourist considers that the ruins of Pachacamac are in "full discovery," that archeologically speaking it is a place with a lot to offer and that it is still being discovered, which is why it receives a negative connotation (−0.75).

Table 5
Concepts related to the scores Museum–Historical Aspects

Keywords	Frequency	Emotional score
Pachacamac God	High	+0.27 (neutral)
Religious center	High	+0.81 (positive)
Main sanctuary	High	+0.55 (positive)
Religious authorities	Medium	+3.32 (positive)
Ritual objects	Medium	+3.03 (positive)
Interesting museum	Medium	+1.92 (positive)
Gold plates	Medium	+1.54 (positive)
Nice collection	Medium	+1.49 (positive)
Excellent design	Medium	+1.49 (positive)
Artistic sense	Medium	+0.66 (positive)

This cluster makes reference to the scores that the museum receives as well as the historical aspects. In this sense, about historical aspects, it must be mentioned that the term "Pachacamac God" (+0.27) has a high frequency, but at the same time it gets a neutral valuation. This is because the Pachacamac God was de most asked Oracle of the Andean world, able to predict the future and control the movements of the Earth and constantly mentioned in the tours, and therefore, visitors remember him. Additionally, the terms "religious center" (+0.81) and "main sanctuary" (+0.55), are mentioned, which present a high frequency and positive valuation, since Pachacamac has been the main religious sanctuary of the central coast for thousands of years. The term "religious authorities" (+3.32) is also used with medium frequency and positive valuation to refer to the cult that priests of the Pachacamac God managed. Moreover, tourists assess the museum as an "interesting museum" (+1.92), of "excellent design" (+1.49), with "artistic sense" (+0.66), and have a "modern room" (+0.52), all these items with medium frequency and positive valuation, since the museum has a contemporary design, which construction lays in keeping with the archeological site with modern exhibition areas, with good lighting

and distributed following a proper museography script that seeks to enrich the experience of the visitors. On the other hand, regarding the objects shown to the tourists, it was mentioned that the museography counts on "ritual objects" (+3.03), "gold plates" (+1.54) and a "nice collection" (+1.49), since the goal is to show the richness of the collections that this museum houses through a series of sections. These last terms were mentioned at a medium frequency and had a positive valuation.

Xu, Z., Zhang, H., Zhang, C., Xu, M., and Dong (2019), in their research, point out that museums promote the economy and serve as a point of attraction for a tourist destination so it helps residents and visitors to understand the local culture, as they can represent the image of a destination. The image of a museum is a factor that greatly influences the positive emotions of a tourist so that the interpretation service, exhibition, guide service, and the museum environment have a relevant role in their behavior. Therefore, an adequate image and service allow attracting more tourists. Regarding the Pachacamac museum, it should be mentioned that most of the keywords found reflect positive emotions and scores about the services offered. This results in an adequate image of the sanctuary, as this place serves as a prelude to the visit to the ruins and is a welcoming space for visitors, showing the main findings of the sanctuary and facilitating the understanding of the history that preceded it.

First, regarding openness to the experience, the results of the Apply Magic Sauce show that 86.12% of the analyzed individuals are liberal and artistic, in other words, creative, innovative and imaginative people, unlike the 13.88% of the analyzed individuals, who are described as conventional and of limited interest, that is to say, conservative and very traditional people (Zhao & Seibert, 2006; Verma, Kumar & Chandra, 2017). Second, regarding conscientiousness, the results present that 83.25% of the individuals are organized and hard-working people, who are persistent and are always in pursuit of achievement. On the other hand, 16.75% of the individuals are spontaneous and impulsive (Costa, Busch, Zonderman & McCrae, 1986, Zhao & Seibert, 2006). According to Barrick and Mount (1991), this behavior trait is highly related to the work performance, as it refers to the way in which we control, regulate, and manage our impulse. Third, about extraversion, 29.19% of the analyzed users are individuals committed to the outside world, meaning they are joyful, very sociable people, who love to have fun and being in a permanent search for emotions and external stimulus (Costa, Busch, Zonderman & McCrae, 1986; McCrae & John, 1992 and Zhao & Seibert, 2006). However, 70.81% of the users are contemplative, which indicates that this group prefers to spend more time on their own, they are very reserved, calm and independent. Fourth, about kindness, 66.99% of the users are team players and confident, they are capable of forgiving, they are loving, altruistic and gullible, which represent people who are kind to others (Costa, Busch, Zonderman & McCrae, 1986; McCrae & John, 1992).

Also, according to Apply Magic Sauce software, most visitors are men (75.60%). The X Generation (50.24%) is the most representing group, followed by Baby Boomers (29.67%). Both Generation X and Baby Boomers seek to learn and understand (cultural tourism), while the Millennials or Centennials seek adventure, fun, and entertainment. Baby Boomers get the highest score of leadership (63%), indicating that age can be a relevant factor to be a leader.

5. Conclusion

Based on the data above, it can be inferred that there is such a difference due to the interests and/or motivations of the traveler; that is, both Generation X and the Baby Boomers seek to learn and understand (cultural tourism), while the Millennials or Centennials seek adventure, fun, and entertainment, something they will obviously not find in the Archaeological Sanctuary of Pachacamac. Moreover, by analyzing the Big Five personality model, it was shown that Baby Boomers are the generation that stands out in most factors, that is, they are characterized as being liberal, artistic, creative, innovative, and imaginative (openness to experience); organized and hard-working people always look for achievement (conscientiousness); contemplative people, seek to spend time by themselves, they are reserved and calm (extraversion); people who seek teamwork, who are trustworthy, capable of forgiveness and altruistic (kindness); relaxed, confident, clam (neuroticism). Regarding the potential for leadership, once again, Baby Boomers get the higher score (63%), indicating that age can be a relevant factor to be a leader.

In addition, it should be noted that the image of the Archaeological Sanctuary of Pachacamac and the National Tourist Profile are highly related, since the results obtained through the frequently

mentioned key terms reflect the generational characteristics of the Baby Boomers as predominant, followed by Generation X, all obtained by means of the Big Five personality model. In this sense, it is useful to analyze these two variables in the same study, as the results will allow for appropriate decision-making in the design of marketing strategies for each generation. We should also work on the positive projection of the image of the sanctuary since it has been observed that predominant groups have leading characteristics and their comments can be used as references for other visitors.

According to the consulted background, it should be taken into account that, at the international level, the methodology and model used in this research have been applied only to touristic destinations. However, at the national level it has been proven that it is possible to apply them to the study of the image of a tourist attraction, even with certain limitations. By means of this research, it was possible to confirm that it is possible to apply them in an attraction's image study. Even if it is true that the Archeological Sanctuary of Pachacamac as such does not have restaurants nor accommodation as part of its touristic infrastructure, there are rural restaurants and haciendas in the district Pachacamac or Valle de Lurín that offer shows of Paso Horses and complement the visit very well.

6. References

[1] I. Anderson.: Integration and Analysis of Unstructured Data for Decision Making: Text Analytics Approach. International Journal of Open Information Technologies, 4(10), (2016): 82–88. URL: https://www.researchgate.net/publication/307919142_Integration_and_Analysis_of_Unstructured_Data_for_Decision_Making_Text_Analytics_Approach.

[2] S. Baloglu, T. Henthorne, S. Sahin. Destination Image and Brand Personality of Jamaica: A Model of Tourist Behavior. Journal of Tourism and Marketing, 31(8), (2014): 1057–1070. DOI: 10.1080/10548408.2014.892468.

[3] S. Baloglu, K. McCleary. A model of destination image formation. Annals of Tourism Research, 26(4), (1999): 868–897. URL: https://doi.org/10.1016/S0160- 7383(99)00030-4.

[4] M. Barrick, M. Mount. The big five personality dimensions and jo bperformance: A Meta-analysis. Personnel Psychology, 44(1), (1991) 1–44. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.744-6570.1991.tb00688.x.

[5] A. Beerli, J. Martín. Factors influencing destination image. Annals of Tourism Research, 31(3), (2004) 657–681. DOI: 10.1016/j.annals.2004.01.010.

[6] C. Bucur. Using Opinion Mining Techniques in Tourism. Procedia Economics and Finance, 23, (2015) 1666 – 1673. URL: https://www.sciencedirect.com/science/article/pii/S2212567115004712.

[7] R. Camprubí, J. Guía, J. Comas. Destination networks and induced tourism image. Tourism Review, 63(2), (2008) 47–58. DOI:10.1108/16605370810883941.

[8] P. Costa, K. Busch, A. Zonderman, R. McCrae. Correlations of MMPI Factor Scales with Measures of the Five Factor Model of Personality. Journal of Personality Assessment, 50(4), (1986) 640–650 (1986). DOI: 10.1207/s15327752jpa5004_10.

[9] E. Demby. Psychographics revisited: The birth of a technique. Marketing Research, 6(2), 1-26 (1984). URL: fromhttps://search.proquest.com/docview/202674861/fulltext/A71324D578964C23PQ/1?accountid=43847.

[10] W. Gartner. Image Formation Process. Journal of Travel & Tourism Marketing, 2(2-3), (1993) 191-216 (1993). DOI: 10.1300/J073v02n02_12.

[11] A. Gassiot. Destination image of Girona: An online text- mining approach (Tesis de Maestría, Universidad de Girona, Girona, España) (2012).

[12] U. Godnov, T. Redeck. Application of text mining in tourism: Case of Croatia. Annals of Tourism Research, 58, 162–166 (2016).

[13] U. Gretzel, K. Yoo. Use and Impact of Online Travel Reviews. Information and Communication Technologies in Tourism. 35–46 (2008).

[14] M. Groen. Where do we go from here? Visualising Amsterdam for tourists, based on geotagged user generated content (Tesis de maestría, Universidad de Van Amsterdam, Amsterdam, Holanda) (2012). Retrieved from http://dare.uva.nl/cgi/arno/show.cgi?fid=452530.

[15] V. Gupta, N. Rathore. Deriving Business Intelligence from Unstructured Data. International Journal of Information and Computation Technology, 3(9), (2013): 971– 976. URL: https://www.ripublication.com/irph/ijict_spl/16_ijictv3n9spl.pdf.

[16] R. Hernández, C. Fernández, P. Baptista. Metodología de la Investigación [Research Methodology]. México: McGraw-Hill / Interamericana Editores S.A. (2014).

[17] C. Hsu, S. Kang, K. Wolfe. Psychographic and demographic profiles of niche markets leisure travelers. Journal of hospitality & tourism research, 26(1), (2002): 3–22. DOI: 10.1177/1096348002026001001.

[18] A. Jalbani, M. Memon, M. Memon, S. Depar, M. Koondhar. A Study of News Recommender System using Natural Language Cloud Computing Services. Sinh University Research Journal, 50(2), 249–254. URL: https://sujo-old.usindh.edu.pk/index.php/SURJ/article/view/447.

[19] A. Kaur, A. Chauhan, Y. Medury. Destination image of Indian tourism destinations: an evaluation using correspondence analysis. Asia Pacific Journal of Marketing and Logistics, 28(3), 499–524 (2016). URL: https://doi.org/10.1108/APJML-05-2015-0074.

[20] R. McCrae, O. John. An Introduction to the Five-Factor Model and Its Application. Journal of Personality, 60(2), (1992): 175–215. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.1992.tb00970.x.

[21] Ministerio de Comercio Exterior y Turismo. Llegada de Visitantes a Sitios Arqueológicos y Áreas Naturales Protegidas por el Estado [Arrival of visitors to archeological sites and natural areas protected by the State]. (2020). URL: http://datosturismo.mincetur.gob.pe/appdatosTurismo/Content2.html.

[22] L. Moutinho. Segmentation, targeting, positioning and strategic marketing. En L. Moutinho (Eds.), Strategic Management in Tourism (pp. 121–166). Reino Unido: CAB International. (2020).

[23] L. Olmeda, P. Sheldon. Data Mining Techniques and Applications for Tourism Internet Marketing. Journal of Travel and Tourism Management, 11(2-3),1–20 (2001). URL: https://www.tandfonline.com/doi/abs/10.1300/J073v11n02_01.

[24] World Tourism Organization. UNWTO Briefing Note–Tourism and COVID-19 Understanding Domestic Tourism and Seizing its Opportunities. UNWTO, (3), 1–40 (2020). Doi: https://doi.org/10.18111/9789284422111.

[25] A. Sarli, H. Hon. The role of psychographic for distinguish main categories of consumers based on lifestyle, personality and value variables. IJER, 2(4), (2011): 29–34. URL: https://pdfs.semanticscholar.org/934d/52cdc85c1df3b13e74cdeaa40ba4c4816b8d.pdf.

[26] P. Tintaya. Pictografía: una estrategia de estudio de los sentidos personales del ser humano [Pictography: a study strategy on the human beings' personal senses]. Revista de Psicología, (20), (2018): 97–109. URL: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2223-30322018000200008.

[27] K. Tsujii, M. Takahashi, Y. Fujita, K. Tsuda. Extraction of Accommodation Evaluation by Foreign Tourists with Text Mining. International Journal of Trade, Economics and Finance, 5(2), (2014): 130–135 (2014). URL: http://www.ijtef.org/papers/355-C00005.pdf.

[28] V. Verma, S. Kumar, B. Chandra. Big five personality and tourist's intention to visit green hotels. Indian j. Sci.Res, 15(2), (2017): 79–87. URL: https://www.researchgate.net/publication/317674176

[29] Z. Xu, H. Zhang, C. Zhang, M. Xu, N. Dong. Exploring the Role of Emotion in the Relationship between Museum Image and Tourists' Behavioral Intention: The Case of Three Museums in Xi'an. Sustainability, 11(3), (2019): 2–20. DOI:10.3390/su11030559.

[30] H. Zhao, S. Seibert. The Big Five Personality Dimensions and Entrepreneurial Status: A Meta-Analytical Review. Journal of Applied Psychology, 91(2), (2006): 259–271. DOI: 10.1037/0021-9010.91.2.