TOURISTS' ARCHETYPES AND SUSTAINABILITY AT THE DESTINATION: A DIGITAL BEHAVIORAL SEGMENTATION APPROACH

Authors:

Sofía Blanco-Moreno – University of León, Faculty of Economics and Business Administration, Av. de la Facultad nº 25 24004 León – Assistant professor – sblanm@unileon.es - +34 647718163

Ana M. González-Fernández – University of León, Faculty of Economics and Business Administration, Av. de la Facultad nº 25 24004 León – Full professor – amgonf@unileon.es

Roman Egger – Modul University - Am Kahlenberg 1 1190 Vienna | Austria - Full professor – roman.egger@smartvisions.at

Luis V. Casaló – University of Zaragoza, Faculty of Economy and Business, Gran Vía 2, 50005 Zaragoza – Full professor – lcasalo@unizar.es

Abstract:

Traditional segmentation in tourism has mainly focused on individual socio-demographic characteristics or trip motive, which sometimes prove to be a poor reflection of real differences in behaviour. To move on this topic, this work-in-progress employs artificial intelligence and big data analysis to: (1) identify various categories of tourists based on their posting behaviour in Instagram and (2) evaluate whether the level of engagement with sustainability at the destination differs among these segments or not. Based on data collected via web scraping from 245 Instagram locations within León (Spain), 139,273 posts from 43,000 users were analysed. Specifically, six archetypes based on their spatial movement, posting and engagement behaviours were identified, having differences in their sustainability engagement. Gastronomy and cultural segments exhibit the greatest sustainability engagement in their Instagram posts, and luxury and shopping segments the less.

Keywords: Behavioural segmentation, sustainable tourism, tourist archetypes, artificial intelligence.

1. INTRODUCTION

The development of tourism in last decades seems unstoppable, becoming a major economic activity globally, with a relevant contribution to world GDP and millions of jobs related to these activities everywhere (UNWTO, 2025). However, it is also a cause of devastation of the environment and causing strife to host communities (Rehman et al., 2024). Thus, with an increase in world tourism, the demand for sustainable tourism practices has increased, and it is necessary to examine tourist behaviour at tourist destinations and how they engage with sustainability practices (Agrawal et al., 2022).

In this respect, we need to consider that several profiles of tourists visit a touristic destination. Traditional tourism segmentation generally aims at settled categories like demographics and travel motivations, among others (Song et al., 2025). However, nowadays we have plenty of information about tourists' behaviours available in social media platforms, especially on Instagram (Blanco-Moreno et al., 2024b). To move in to this topic, through the use of artificial intelligence (AI) and big data analysis, the objective of this work is twofold: First, we seek to identify various categories of tourists based on their posting behaviour in Instagram; second, we aim to evaluate whether the level of engagement with sustainability at the destination differs among tourists' archetypes or not.

To do that, this research analyses Instagram data as this is the main social media platform where travellers share their photo experiences (Blanco-Moreno et al., 2024b). In addition, we employ AI-tools that allow for the post classification of tourists based on their spatial behaviours and the segmentation of profiles based on their preferences displayed in posts.

2. METHODOLOGY

The methodology involved a multi-stage analysis of Instagram data:

- The first phase involved collecting data through web scraping from 245 Instagram locations within León (Spain), resulting in 139,273 posts from 43,000 users. The data included photographs, text, and metadata such as likes, anonymized user information, and timestamps.

- In the second phase, users were classified as tourists or residents, excluding residents (e.g., users were classified as tourists if their recorded stay was less than 30 days within the destination (Gunter and Önder, 2021) from the analysis.

- The third phase involved text analysis using AI and machine learning to extract relevant information. To examine sustainability engagement, posts were filtered based on textual content using natural language processing techniques (Blanco-Moreno et al., 2024a). Posts were identified as sustainability-related if they contained predefined sustainability-related keywords, including terms related to eco-tourism, conservation, responsible travel, carbon footprint, and waste reduction, among others. In addition, the sentiment polarity of the post was calculated. Additionally, posting behaviours is analysed by examining the frequency of posts and the level of

engagement they generate, as measured through likes and comments. Finally, content analysis is performed, considering the length of the text, the number of hashtags used, and the presence of sustainability-related hashtags.

- The fourth phase applied deep learning techniques for photo analysis. Image analysis utilized convolutional neural networks, specifically Deepface, an open-source model integrating VGG-Face and Google FaceNet with 97.53% reliability (Serengil, 2024). This model detected the number of people in each image and classified their emotions into seven categories.

- The fifth phase involved the development of spatial behaviour archetypes by clustering tourists' geolocated posts. In other words, tourist archetypes are derived from the geospatial movement patterns of users; geolocation data is considered, focusing on the specific locations visited within the destination.

- Finally, the sixth phase involves the relationship between tourist archetypes and sustainability engagement.

Then, tourist segmentation was conducted on Orange Data Mining software, an open-source data visualization and machine learning environment. Clustering analysis module, specifically the K-Means algorithm, was used in order to segment tourists based on their geospatial movement as well as posting behaviour (Arefieva et al., 2021). Clustering was a process involving several steps: (1) Geospatial data were initially normalized for consistency within users' movement behaviour as distance and visit frequency variations required standardization to offer meaningful comparisons; (2) the most suitable behavioural metrics were selected (frequency of destination visits, frequency of posting within the destination, and the level of interaction in terms of likes and comments); (3) the most number of clusters that provided optimal clusters was found based on the Silhouette Score, which provided intra-cluster cohesion and inter-cluster separation to increase clustering.

3. PRELIMINARY RESULTS

3.1. Tourists' archetypes identification

The segmentation process properly captures various behaviour patterns and led to the identification of six archetypes based on their spatial movement, posting and engagement behaviours. Specifically:

- Archetype 1 - Event Tourists – These tourists focus on attending events and festivals and show high engagement in urban nightlife locations. They exhibit a moderate level of geospatial movement, primarily visiting high-density event venues, modern bars, and cultural spaces. People who are interested in events and prefer trendy bars.

- Archetype 2 - Classic Tourists – These visitors show a balanced interest between cultural sites and gastronomy, frequently visiting historic landmarks such as the Cathedral, Parador, and

MUSAC while engaging with well-known local restaurants, people who have a balanced interest in museums and restaurants. Its main locations include the Cathedral, the Parador and the MUSAC, with a preference for historical and cultural sites and traditional tourist experiences.

- Archetype 3 - Cultural Tourists – Characterized by their preference for museums and historical sites, this group has the highest visitation rate for landmarks like Casa Botines, the Cathedral, and MUSAC, with a lower emphasis on gastronomy, people who prefer museums to restaurants and have a keen interest in cultural and historical sites. Its main locations include the Cathedral, the Botines building and the MUSAC, i.e. tourists deeply committed to cultural heritage.

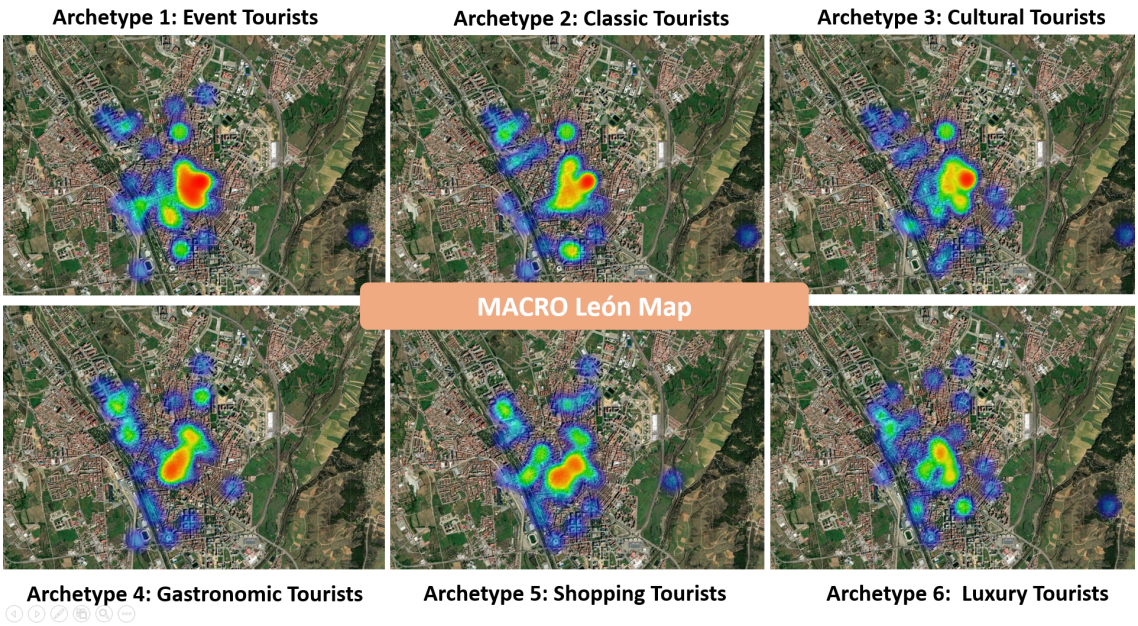
- Archetype 4 - Gastronomic Experiential Tourists – This segment prioritizes culinary experiences over traditional sightseeing, frequently posting about local dishes, restaurants, and street food. They focus on experiential travel, engaging deeply with local dining culture, people who focus on restaurants and bars and are more interested in the overall experience than specific cultural sites. They prefer places like the Barrio Húmedo and the Plaza Mayor, with a strong inclination towards gastronomic experiences.

- Archetype 5 - Shopping-Oriented Tourists – These tourists prefer commercial areas and retail experiences, frequently posting from shopping streets such as Calle Ancha and Ordoño II, with minimal engagement in cultural attractions, people who have a strong preference for shopping streets and often visit places such as Calle Ordoño II and Calle Ancha. They show minimal interest in museums and focus more on retail and leisure experiences.

- Archetype 6 - Luxury Tourists – Visitors in this segment seek exclusive and premium experiences, favouring high-end restaurants, boutique hotels, and fine dining establishments such as Cocinandos (Michelin star) and Hotel Real Colegiata, people who prefer exclusive and high-level experiences, visiting places such as the Hotel Real Colegiata and the Michelin-starred restaurant Cocinandos. These tourists are characterized by a preference for luxury and exclusivity.

Figure 1 provides a visual overview of the spatial tourist activity of the different digital tourist archetypes in León. The map indicates where the tourist activity is most concentrated, e.g., the most visited places in the destination. Tourists archetypes differ in posting (e.g., higher number of photos for luxury and event segments) and engagement behaviours (e.g., higher engagement rates for gastronomy segment).

Figure 1. Heatmap of tourists' archetypes in León related to their movements



3.1. Tourists' archetypes identification

Finally, we analysed the relationship between tourist archetypes and their environmental engagement, an approach to guarantee sustainable tourism. Different travellers may exhibit distinct behaviours, preferences, and priorities in their digital tourism experiences, which may shape how they interact with sustainability topics. Specifically, following the aforementioned methodology, we compare the percentage of posts identified as sustainability-related in each archetype.

Initial analyses confirm significant differences in environmental engagement among tourists' archetypes (χ^2=17.89; $p<0.01$); the gastronomy (61%) and cultural (45%) segments exhibit the greatest percentage of pro-environmental posts and luxury (23%) and shopping (8%) segments the less. These results are consistent with recent proposals relating types of tourism and sustainability (e.g., luxury hotels might hesitate to operate in a more environmentally friendly way because they worry such practices will harm their performance as their customers may not value that (Peng & Chen, 2019).

4. CONCLUSIONS

This work-in-progress aims to contribute to tourism and sustainability research in two ways. First, different from traditional tourist segmentation that has mainly focused on individual demographic characteristics and travel motivations, we use actual travellers' posting behaviour in Instagram and combine AI-based segmentation with behavioural analysis to better reflect real differences in behaviour among segments. Second, our results confirm that tourists' archetypes

exhibit different levels of pro-environmental engagement in their posts. These results offer interesting implications for destination marketing organizations. For example, as cultural and gastronomy tourists care more about sustainability at the destination, these tourists could be approached by initiating niche campaigns in social media where sustainability messages are emphasized. In turn, luxury and shopping tourists, who are less environmentally engaged, could be treated differently (e.g., framing sustainability as an exclusive, high-end feature as opposed to an ethical obligation). In addition to these interesting initial results, new analyses are going to be performed to evaluate whether posts identified as sustainability-related exhibit a more positive polarity of sentiment and have greater interaction levels (in terms of likes and comments) than those posts not related to sustainability.

References

Agrawal, R., Wankhede, V. A., Kumar, A., Luthra, S., & Huisingh, D. (2022). Big data analytics and sustainable tourism: A comprehensive review and network based analysis for potential future research. International Journal of Information Management Data Insights, 2(2), 100122. https://doi.org/10.1016/j.jjimei.2022.100122

Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. Tourism Management, 85, 104318. https://doi.org/10.1016/j.tourman.2021.104318

Blanco-Moreno, S., González-Fernández, A. M., Muñoz-Gallego, P. A., & Egger, R. (2024a). What do you do or with whom? Understanding happiness with the tourism experience: an AI approach applied to Instagram. Humanities and Social Sciences Communications, 11(1), 1–16. https://doi.org/10.1057/s41599-024-02859-z

Blanco-Moreno, S., González-Fernández, A. M., Muñoz-Gallego, P. A., & Casaló, L. V. (2024b). Understanding engagement with Instagram posts about tourism destinations. Journal of Destination Marketing & Management, 34, 100948. https://doi.org/10.1016/j.jdmm.2024.100948

Gunter, U., & Önder, I. (2021). An exploratory analysis of geotagged photos from Instagram for residents of and visitors to Vienna. Journal of Hospitality & Tourism Research, 45(2), 373-398. https://doi.org/10.1177/1096348020963689

Peng, N., & Chen, A. (2019). Luxury hotels going green–the antecedents and consequences of consumer hesitation. Journal of Sustainable Tourism, 27(9), 1374-1392 https://doi.org/10.1080/09669582.2019.1622710

Rehman, S., Khan, S. N., Antohi, V. M., Bashir, S., Fareed, M., Fortea, C., & Cristian, N. P. (2024). Open innovation big data analytics and its influence on sustainable tourism development: A multi-dimensional assessment of economic, policy, and behavioral factors. Journal of Open Innovation: Technology, Market, and Complexity, 10(2), 100254. https://doi.org/10.1016/j.joitmc.2024.100254

Serengil, S. I. (2024). Deepface: A lightweight face recognition and facial attribute analysis (age, gender, emotion and race) library for python. GitHub. Retrieved from: https://github.com/serengil/deepface

Song, H., Chen, J., & Li, P. (2025). Decoding the cultural heritage tourism landscape and visitor crowding behavior from the multidimensional embodied perspective: Insights from Chinese classical gardens. Tourism Management, 110, 105180. https://doi.org/10.1016/j.tourman.2025.105180

UNWTO (2025). UNWTO World Tourism Barometer, January 2025 – Excerpt. World Tourism Organization, 23(1), January 2025 (version 24/01/25). ISSN: 1728-9246. Retrieved from: https://pre-webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2025-01/UNWTO_Barom25_01_January_EXCERPT_v3.pdf