Methodological proposal for the identification of tourist routes in a particular region through clustering techniques

Juan B. Duarte-Duarte, Leonardo H. Talero-Sarmiento, Diana C. Rodríguez-Padilla*

Universidad Industrial de Santander, Facultad de Ingenierías Físico Mecánicas, Escuela de Estudios Industriales y Empresariales, Bucaramanga, Colombia

ARTICLE INFO

Keywords:
Clustering techniques
Sustainable development
Tourism factors
Tourist clusters
Tourist routes
Tourism

ABSTRACT

The tourism sector has an essential role in the sustainable development of a country. Therefore, in this research we propose a methodology to identify tourist routes that integrate the most important Points Of Interest in a region taking up as criteria profile characteristics in common between the sites evaluated using clustering techniques. To attain this goal, firstly, a literature review focused on compiled information used for location selection and evaluation in attraction potential sites. Then, clustering techniques are applied to identify similarities between sites, and finally, a layout of tourist routes is presented. We applied this methodology using data from a Region in Colombia. As a result, eight factors are proposed: Natural, Cultural, Tourist Plant, Infrastructure, Superstructure, Accessibility, Human and Tourist Capital and Security. From the second phase, three tourist groups were identified with three tourism factors for each of them; and then, two examples of tourist routes are proposed.

1. Introduction

Tourism site integration using curated routes is a strategy that could improve sustainable tourism development and, at the same time, increase the quality of life of local people. Different countries often regard tourism as the preferred alternative to development (Briedenhann and Wickens, 2004) because of its economic and social value at the global scale. Tourism plays a key role in sustainable development and impacts the climate (UNWTO, 2008); tourism has also been identified as a tool to raise the economic growth of underdeveloped regions and to improve the standard of living of local communities (Ritz and Pueckló, 1998) (Kombo, 2000). Additionally, tourism presents sociocultural benefits, improves education, and reinforces the maintenance of culture and heritage (Jafari, 2005).

Taking advantage of the benefits of tourism, over the years, its economic impact has been studied using different methodologies. For example, Deng et al. (2002) have used multicriteria analysis techniques to categorize Victoria Park in Australia as a tourism resource, and the same technique have been used by Triantaphyllou (2000) in Bosnia, Laguna Marín-Yaseli and Nogues Bravo (2003) in Iberian Mountain Ranges, Infante Sánchez (2014) in Colombia, Egil & Ik (2014) in Turkey, Rubio et al. (2016) in Argentina, Camarena (2016) in México, and Morteza et al. (2016) in Qeshm Island, Iran, among others. On the other hand, Zimmer and Grassman (1996), Padin and Pardellas (2004), Blanco (2008), Andrés Cabello and Pasqual Bellido (2015) have used the LEADER method, a guide to evaluate the tourism potential of a territory, in different regions of Spain and México. Furthermore, Eagles et al. (2001) and Mishra (2009) in different areas of Asia; Mart et al. (2013) in Mexico; and Alam et al. (2015) in Malaysia, have used statistical methods to evaluate tourism sites. Many other authors have employed other methodologies.

Most methodologies usually use qualitative techniques that require expert analysis (Lahtamaki et al., 2016). However, there are different quantitative techniques based on the structural information of tourist sites, such as clustering techniques (CTs). There is evidence that the application of CTs can create cooperative relations in high-tech sectors, where the most valuable capital is knowledge and cooperation occurs between companies and academia (Porter, 2000)(Rocha, 2004). However, it is possible to apply these techniques in more traditional sectors, such as tourism: for instance, Pitchayadejanant and Nakpathom (2018) clustered the agrotourism activities in an orchard using machine learning, Gosal et al. (2019) processed and clustered the attractiveness of different tourist locations, Nilashi et al. (2019) used clusters to predict touristic choice preferences in ecofriendly tourism, and Guo et al. (2020) clustered tourism attractions to predict the competitiveness of a city.

* Corresponding author.

E-mail address: diana.rodriguez15@correo.uis.edu.co (D.C. Rodríguez-Padilla).

https://doi.org/10.1016/j.heliyon.2021.e06655

Received 24 September 2019; Received in revised form 26 February 2020; Accepted 29 March 2021

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
among others. These clustering techniques can create a promising new development direction for the tourism industry (Esteveio et al., 2009).

We propose the use of CTs to design tourist routes (TRs) in a region to maximize the potential benefits of tourism. These routes are important because TRs are tools that highlight the cultural heritage of sites, promoting and benefiting those who are part of it (Jeambey, 2016) (Briedenham and Wickens, 2004). Additionally, TRs integrate several attractions that have not been studied independently, allowing an equitable distribution of their benefits (Meyer, 2004). Consequently, tourist route implementation is a strategy to optimize job creation and tourism and conserve the historical, cultural, and folkloric heritage of any site (Open Africa, 2002).

Therefore, the research question is, “How can we identify a methodology to propose tourist routes in a particular region?” In this paper, we propose a methodology to identify theoretical tourist routes in such a way that different tourist attractions can be integrated into a region. In Phase 1, a literature review is performed to collect characteristics or attributes to describe the tourism cluster of a site. In Phase 2, we apply a clustering technique among sites to find patterns and similarities among them, identifying specific clusters. Finally, Phase 3 designs tourist routes by integrating these clusters.

2. Literature review

To allow tourism on a site, different organizations and researchers have proposed diverse methodologies to identify and categorize touristic factors (UNWTO, 2013). These factors are an elemental guide in the tourism process (Knezevic, 2008). Consequently, this investigation uses predefined characteristics and proposes to group them into a hierarchical structure, with eight factors, covering 27 subfactors (Figure 1) described in Table 1.

Next, we describe the methodologies used in different studies due to their essential contributions to the classification of tourist factors.

2.1. Pierre Defert

Defert (1972) proposes the categorization of touristic resources as follows: Hidrom (water); Litom (elements built by man); Antropíom (socioeconomic structure); Phitom (Natural terrestrial elements); and Mnémonome (elements that generate memory). This categorization of resources has been used by authors such as Camara and Morcate Labrada (2014) to evaluate the city of Fort de France. Additionally, the consulting company Auren Consultores (2014), directed by the School of Industrial organization of Spain, compared Defert’s classification to those of other authors to choose the most suitable classification for Andalusia, Spain. This comparison was also performed by researchers at the University of San Luis Posotis in Mexico (2015) and by Rubio et al. (2016), a Ph.D. in Geography of the “Universidad Nacional del Sur” in Argentina.

2.2. SECTUR

The Secretary of Tourism of Mexico designs and implements policies aimed at strengthening tourism activity (SECTUR, 1975). These policies have been applied by various authors, such as Reyez and Sánchez (2005). Reyes has used three variables to calculate the potentiality index of natural tourism: 1) natural variables (geomorphologic, vegetal associations, and natural elements); 2) accessibility variables (terrestrial, maritime, and air transport, gas station, and road density); and 3) equipment variables (hotels, food establishments, tourist promotion units, banks, and commercial establishments). Other authors who have used the SECTUR information are Martínez (2010), Mikery Gutiérrez and Pérez-Vázquez (2014), Camara and Morcate Labrada (2014) and Cortés (2015); these authors used SECTUR information to analyze the tourism potential of their respective case studies.

2.3. OAS

The Organization of American States (1987) proposed a resource categorization divided into the following categories: 1) natural sites (landscape-based resources); 2) museums and cultural traditions (art, history, and monuments); 3) folklore (related to the traditions of the resident population); 4) contemporary technical, scientific and artistic realizations (elements that by their singularity have tourist interest and are more current than historical characters); and 5) programmed events (organized events, current or traditional). This categorization is easy to replicate; several authors such as Blanco (2008), Conti et al. (2012), Solis et al. (2013), Mikery Gutiérrez and Pérez-Vázquez (2014), Castellanos

![Figure 1. Tourism Factores and sub-factors.](image-url)
Menjura and Ariza Cortes (2015), and Navarro (2015) have applied this methodology to their tourist resources inventory.

### 2.4. LEADER guide

Zimmer and Grassman (1996) created the LEADER guide to evaluate the tourist potential of a territory. He took into account the offer (natural factors, socioeconomic factors, infrastructures and services available, cultural factors, sports and leisure offerings, health offerings, accommodation, restaurants, and sites with the possibility of organizing conferences and seminars) and the demand (defined from customer surveys). Then, Peter carried out an analysis of the competition and, finally, performed a study of the market trends to anticipate the opportunities and risks linked to the new expectations of clients. Authors such as Blanco (2008), Vázquez (2014), Andrés Cabello and Pascual Bellido (2015), and Vanegas (2017) have used the guide to apply tourism research methodologies in different parts of Europe and Latin America.

### 2.5. Gunn

Gunn (2002) has proposed factors such as market preferences, advocacy, information systems, restrictions and benefits (social environment, governmental disposition, among others), and physical factors, composed of potential destination areas, by geographic and competitive contexts, to identify new and improved attractions and threats to the environment. The authors who have used this categorization are Chá and Uysal (1995), Smith (1987) and Smith (2013).

### 3. Methodology

#### 3.1. Data

First, a region is selected (taking into account a region could be divided into micro regions or sites to apply the following methodology where the identified characteristics in the literature review are consulted from primary or secondary sources such as: interviews from the inhabitants of the region, going to each site of interest, looking for information in the regional official website where you can find its general tourist characteristics; official website of general and specific statistics of the sites of the region, newsletters, articles, housing census, citizenship, development plans, tourism development plans, inventories of the region’s assets, among others.

The data must be considered in terms of different types of variables (e.g., the scale of measurement, reason variables, quantity variables, and dichotomous variables). These variables have to be stored and classified; furthermore, each evaluated site should have a numerical value for each of these variables; that is, no missing data are allowed.

#### 3.2. Touristic clusters

Clustering techniques can recognize patterns and similarities between the characteristics of entities (Instituto Valenciano de Tecnologías Turísticas, 2015). CTs organize studied entities from the information provided by the decision maker; in this case, the information represents the characteristics collected in the literature review section. Moreover, the hierarchical clustering technique has excellent performance (Everitt et al., 2011), allowing classifications of greater accuracy than permutation methods such as K-means (Karypis et al., 2000) (Li et al., 2002). Nevertheless, to apply this technique, there are two requirements: 1) the selection of a link method and 2) the implementation of a distance function (Wilks, 2011).

However, considering that, the proposed factors and their corresponding characteristics have different types of variables, such as the scale of measurement, according to reason, quantity and dichotomous, it is necessary to transform them to the same scale. To do this, we use the Gower coefficient of similarity because it allows the integration of quantitative and qualitative variables (López and Villa, 2011) (Podani, 2012). Furthermore, we can estimate the Euclidean distance $D_{ij}$ between each pair of townships $ij$ using the Gower similarity ($S_{ij}$) according to the following relation $D_{ij}^2 = 2(1 - S_{ij})$. The coefficient of similarity between the townships $ij$ is explained by Eq. (1):
where \( W_{jk} = 0 \) if entities \( j \) and \( k \) cannot be compared for variable \( i \) because \( x_{ij} \) or \( x_{ik} \) is unknown.

Additionally, Gower defines:

For binary variables: \( W_{jk} = 1 \) if \( y_j \neq y_k \); \( W_{jk} = G_{jk} = 1 \) if \( x_j = x_k = 1 \) or if \( x_j = x_k = 0 \) and double zeros (mutual absences) are included; \( W_{jk} = G_{jk} = 0 \) if \( x_j = x_k = 0 \) and double zeros are excluded from the comparison.

For nominal variables: \( W_{jk} = 1 \) if \( x_j \) and \( x_k \) are known; then let, \( G_{jk} = 0 \) if \( x_j \neq x_k \); \( G_{jk} = 1 \) if \( x_j = x_k \).

Furthermore, for variables measured on the Interval and ratio scale: \( W_{jk} = 1 \) if \( x_j \) and \( x_k \) are both known and \( G_{jk} = 1 - (|x_j - x_k| / \text{range of variable } i) \) (Gower and Legendre, 1986).

Finally, we study the link method in each case; however, considering the distance functions estimated in this investigation and its characteristics, we suggest applying methods that use squared distances.

3.3. Tourist routes: related sites integration

For the design of the routes, we consider the recommendations of Aravena Paillalef et al. (2013), Cadena et al. (2013) and Garrido et al. (2015). The design process begins by identifying tourist points of interest to integrate using routes. These points are subsequently located on a map to determine their geographical distribution and to analyze the possible integrations to make between them.

The design of these routes must consider the location of the sites and their geographical proximity to one another, without taking into account the road infrastructure between them. That is, routes can have two kinds of sections: real and hypothetical (a.k.a. theoretical) roads. The real routes link the sites without taking into account the road type (i.e., whether the road is paved or not). The hypothetical routes, on the other hand, are unbuilt connections that will be required. We use this methodology to identify possible infrastructure improvements that could be helpful in connecting tourist sites.

4. Results

The methodology proposed was applied in Santander, which is one of the 32 departments in Colombia. Santander is in the northwest part of the country (Figure 2). Santander is composed politically of 87 townships and is considered the fourth most important economy in Colombia, contributing 7.8% of the total national gross domestic product (GDP). We select Santander in this study due to its diversified economy, which causes heterogeneity in the distribution of wealth among its inhabitants (de Santander, 2016), as represented in the indices of monetary poverty, the Gini coefficient (indicator of the distribution of income of individuals within an economy) and the index of unmet basic needs (Departamento Administrativo Nacional de estadística, 2018). By applying this methodology and taking into account the benefits of tourism, the economic disparity in Santander may be able to be reduced.

4.1. Data

According to the information available from secondary sources (the official website of Santander governorate, the official website of the National Administrative Department of Statistic (DANE), among others), we collect 79 characteristics for every township (Annex 1. 79 characteristics for each Townships) 12 belong to the Accessibility factor, 10 belong to the Human and Tourist Capital factor, 11 belong to the Cultural factor, 11 belong to the Natural factor, 11 belong to the Infrastructure factor, 16 belong to Tourist Plant, 5 belong to Security and 3 belong to Superstructure. Applying the hierarchical clustering technique, we identified that 6 of the 79 characteristics are dichotomous variables, 50 are variables of quantity, 21 are expressed as a percentage, and one is a reason variable (Annex 1. 79 characteristics for each Townships).

4.2. Touristic clusters

According to the characteristics identified, to apply the clustering techniques, we selected the Ward link method and the squared Euclidean distance due to the ease of identifying atypical values between entities in the same group; as a result, three tourist groups are proposed considering the potential touristic identification. The Natural, Cultural, and Human and Tourist Capital factor data are used to form the clusters; additional data (Accessibility, Infrastructure, Tourist Plant, Security, and Superstructure) are used to analyze the general infrastructure of each township. The Similarity Matrix and the Distance Matrix are in Annex 2. Similarity Matrix and Annex 3. Distance Matrix, respectively.

We choose three groups because fewer groups would lead to negative similarity (atypical relationships) and more groups would not significantly increase the similarity. We highlight that the groups are heterogeneous in size, as indicated in the dendrogram, a diagram showing the similarity of attributes between each pair of entities grouped sequentially (Minitab 18, 2018). We use yellow, blue, and red (Figure 5) to differentiate groups; the geographical distribution of each group is given on Santander’s political map (Figure 4).

Touristic Cluster 1 (Blue): This cluster has the most significant “rural” population (the native population of San Andrés, Providencia and Santa Catalina due to crossbreeding between indigenous, Spanish, French, English, Dutch and Africans (Ministerio de Cultura de Colombia, 2004)) located in the El Peñón, Coromoro, Guadalupe and Oiba townships. This cluster also includes some of the most representative cultural events in the region, such as “Las Ferias Bonitas” (the beautiful fairs), the Tiple national week celebrating the traditional musical instrument, the Luis A. Calvo festival in Bucaramanga, the Guabina and Tiple festival in
Vélez, the music festival in San Vicente de Chucurí, and the Guascarrilera festival in Matanza. Additionally, most of the townships around the Chicamocha canyon (Aratoca, Cepitá, Los Santos, and Villanueva) are in this group.

**Touristic Cluster 2 (Red):** This group of townships stand out because they have large indigenous populations, mulattos and Afro-Colombians (referring to Colombian citizens of African descent), mostly distributed in the towns of Cimitarra (with 3326 inhabitants of this type of population), Puerto Parra (with 1454), Palmas de Socorro (with 1369) and Puerto Wilches (with 1337) according to DANE's census and projections for the population of each township in 2015. This group also has the fewest cultural components compared to the other groups, with only 5 of the 40 townships having representative aspects in this area. These components are found in Socorro, widely known due to the insurrection of the “comuneros” community, where the basilica “Nuestra Señora del Socorro” is located; in Suratá, where the Bambuco National Festival (traditional dance) is celebrated; in Pinchote, where the Trios National Festival is celebrated; in Güepa, where the “Panela” festival is held, featuring traditional brown sugar food; and in Chipatá, the location of the song festival. Finally, this group includes the Páramo of Santurbán, which includes the townships of California, Santa Barbara, Suratá, and Vetas.

**Touristic Cluster 3 (Yellow):** Represented by the most significant number of the “ROM” population, or Gypsy ethnic communities, established in Colombia (Ministerio de Cultura, 2010). Located in Girón, Floridablanca and Lebrija, these townships have several museums, such as the Guane Archaeological Museum located in Floridablanca, “Mansión del Fraile” museum in Girón, the “Quijote de la Mancha” house museum in Zapatoca, and Petroleum National Museum in Barrancabermeja. This group is also recognized for its fairs and festivities and its handicrafts as well as because there are some renowned typical food delicacies in these townships, such as “Obleas” in Floridablanca, “Culonas” ants in Barichara, artisan wine in Zapatoca and “Panuchas” (traditional candy) in Malaga. These townships have several well-known caves, including “La Pintada” cave in Lebrija, Caverne in Malaga, Nitro cave in Zapatoca, and Guane cave in Girón, among others.

### 4.3. Tourist routes: related sites integration

The routes are shown on a political map of Santander and have been added to Google Earth, a web-based mapping service platform. The routes can be designed considering each cluster; for example, from Cluster 1, a route was designed considering the proximity of each site of interest to the townships without taking into account the existing routes (Figure 5 and Figure 6) (this route is in Google Earth in Annex 4. Cluster 1 Route 1).

“Route: Chicamocha canyon and more.”

This route focuses on a scenic tourism tour, and it begins at Los Santos Township, Cepitá, Aratoca; continues to Curití, San Gil; and ends at Villanueva. These locations contain the extensive Chicamocha canyon, and it is also possible to appreciate other touristic sites, such
as “el Salto del Duende” in Los Santos, “Salto del Mono Aullador” in Aratoca, the “Yeso” cave and the “La Vaca” cave in Curití. In addition, tourists can enjoy the famous Fonce River in San Gil, where foreigners can practice sports such as rafting or canoeing and, finally, the “Indio” cave in Villanueva.

Analyzing the general infrastructure of these townships, considering their geographical information and other data factors (Accessibility, Infrastructure, Tourist Plant, Security, Superstructure), the canyon of Chicamocha provides opportunities for landscape and scenery sightseeing, extreme sports, and historic tours. Accordingly, it will be
necessary to build access roads between Los Santos and Aratoca due to the lack of existing connections between them. Furthermore, most of the townships on this route must improve and expand their internet coverage. Additionally, the pavement of the roads in this region should increase the number of tourist agencies as well as the options to practice different kinds of sports.

Another tourist route we can find in Cluster 2 is route 1 (Figure 7 and Figure 8) named:

“Route: Biodiversity of the Santurbán Páramo”

This route focuses on landscape sightseeing tourism that begins with Suratá, continues to California, and ends in Vetas Township to develop tourism around the Santurbán Páramo. Here, tourists can see the Cattleya Mendelii and Befaria flowers, the Frailejon, 15 species of mosses, and several species of animals, including the Chirradora, the “Pato Zambullidor,” and the Curi and the Condor, among others. Additionally, tourists can visit “La Laguna” and buffaloes in Suratá; the Páez, Pico, Quelpa and Toro lagoons; hot springs; the San Antonio de Padua sanctuary, which is located in the foothills of the eastern Colombian Cordillera in the Andean Mountain system in California; and the La negra Lagoon in Vetas.

Analyzing the infrastructures of these three townships, we note that although California is geographically next to Vetas, these townships are not connected in terms of roadway infrastructure; therefore, it is proposed that these townships be connected to take advantage of the tourist activities provided by the Páramo.

5. Discussion

This article proposed to answer the following research question: How can a methodology be identified to proposed touristic routes in a particular region? To accomplish this aim, artificial intelligence techniques that allow us to systematically analyze all the characteristics that can define a possible tourist site are proposed. However, considering the literature review, we found that there are multiple variables to determine potential tourists of a site that can vary depending on the site analyzed and the methodology used to evaluate the site. For example, there are discrepancies in SECTUR (1975) methods; Zimmer and Grassman (1996), among others. Additionally, it should be noted that not all sites have the same characteristics, and therefore, as a preliminary step, the fields must be filtered. In this study, since we evaluated the department of Santander, we eliminated some characteristics considering its geographical location (e.g., number of beaches, and seasons).

This process is essential, as the characteristics found can also change according to the data available at the time of the study. Variable selection is a critical process because if there are different variables, it could be possible to develop different tourist profiles than those found in the present study. However, if in the future, different characteristics are considered, these will not invalidate the proposed routes; instead, complementary routes could be designed.

On the other hand, in the data collection, we used secondary sources; more specifically, the official national, departmental, and township websites, on which it is possible to find reliable information. However, according to the publication and updating policies, some information required the prior permission of the owner, and some had outdated publications, which means that although the characteristics selected are appropriate for the site, their figures may vary.

In the last methodological section, the purpose of designing tourist routes was to integrate potential tourist sites considering the position and geographical proximity between townships. In Santander, some of this integration was made taking into account the roadway infrastructure available; however, we found nearby townships that do not have a direct connection; for these cases, we propose theoretical roads so that, in the future, stakeholders could propose infrastructure projects.

The results of this study are aligned with those of Pitchayadejanant & Nakpathom (2018), who use data mining to identify associations and patterns of activities in tourism, as we take advantage of the association of patterns to create touristic clusters that are linked by routes. On the other hand, Gosai et al. (2019) perform clustering that includes tourist factors, including such as nature, ornithology, and religious pilgrimage, among others, that complement our proposed factors; also, Nilashi et al. (2019) conclude that big data and machine learning are essential factors with enormous growth potential in the tourism industry. Finally, Guo et al. (2020) compared different methods to identify economic clusters that are essential to improve the tourism competitiveness of a city.

Taking into account the proposed data hierarchy, as well as the proposed method, for future work, it could be necessary to update the information with data in situ, using deep techniques to analyze the construction of building roads, bridges, and other infrastructures. This research could be proposed to government entities as an alternative method to growth and to improve the social, cultural, and economic situation of each township of Santander. The proposed clusters could be part of the departmental development plans around Colombia, taking into account that this methodology could be replicable anywhere. In addition, the tourism agencies could offer different attractions in Santander using the proposed routes that include the preference of tourists in terms of nature, culture, history or adventure; finally, these routes can be promoted on platforms such as social networks and official websites of the department's cities.

6. Conclusions

In this study, we introduce a novel approach to speed up the decision-making process related to designing and creating routes by identifying tourism clusters that consider the similarity between their characteristics.
based on a hierarchical structure. These clusters are geographical sites that offer a particular type of tourism; however, the cluster groups depend on the attributes that describe each location. That is, if we use different characteristics, it is possible to find a diverse cluster group. For this, it is essential to justify the selection of attributes, the data collection process and the sources. We also highlight that this methodology is based on quantitative methods. At the route creation phase, we need to use an expert criterion to speed up the interest site connection, meaning that when we choose among feasible sites, we decide which ones to integrate with the main route based on the expert criterion. In conclusion, clustering techniques reduce the use of resources such as time, money and expertise. Additionally, the application of this technique can be used in future tourism studies in other places or sites where a tourism manager can identify touristic activities, as well as natural, cultural and other resources. Thus, this technique can be used to determine and design adventurous, religious, gastronomic, and other exciting and attractive tourism routes. Finally, this methodology allows government staff and entrepreneurs to highlight and prioritize future investments, such as transportation structure (e.g., new roads, bridges, cableways, among others), which can improve tourism development.

**Declarations**

**Author contribution statement**

Diana Carolina Rodriguez-Padilla: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Juan Benjamín Duarte-Duarte: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Leonardo Hernán Talero-Sarmiento: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

**Funding statement**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data availability statement**

Data will be made available on request.

**Declaration of interests statement**

The authors declare no conflict of interest.

**Additional information**

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e06655.

**References**

Alam, A.S.A.F., Er, A.C., Begum, H., Alam, M.M., 2015. The factors of selecting Malaysia as tourist destination. Mediter. J. Soc. Sci. 6 (3), 491–498.

Almeda, M.V., 2009. Matriz de Avaliação do potencial turístico de localidades receptoras. Revista Turismo Em Análise 20 (3), 541–563.

Andrés Cabello, S., Pacsal Bellido, N., 2015. La construcción del turismo en nuevos destinos: luces y sombras. El caso de la Rioja (España). Revista de Ciencias Sociales y Humanidades 24, 30–49.

Aravena Paillafet, G., Espinosa Sepúlveda, A., Flores Chávez, J., 2013. Manual para el desarrollo de circuitos de turismo de intereses especiales 1. Auren Consultores, 2014. Recursos Turísticos, 30.

Blanco, M., 2008. Guía para la elaboración del plan de desarrollo turístico de un territorio, 45.

Boull, R.C., 1997. Planificación del espacio turístico. 3° ed. México, Trillas. Retrieved from. http://www.apae.pe/archivos_up/0107-planificacion-del-espacio-turistico-r-obero-c-ballon.pdf.

Briemhann, J., Wicken, E., 2004. Tourism routes as a tool for the economic development of rural areas-vibrant hope or impossible dream? Tourism Manag. 25 (1), 56–79.

Cadena, A., Christian, V., Ing, F., 2013. Nuevos atractivos turísticos tangibles en los cantones: Tulcán, Espejo y Mira para la creación de circuitos turísticos. Camarena, C., 2016. Evaluación de los recursos turísticos naturales del municipio de San Pedro Lagunillas, Nayantza, México, a partir de la metodología multicriterio, pp. 43–69.

Castellanos Menjura, Claudia Patricia, 2015. Evaluación de los recursos turísticos con vocación ecoturística y caracterización de la demanda turística en las zonas de uso público de la Reserva Forestal Protectora del Cerro Quinín (Tibucay–Cundinamarca).

University Distrital. Retrieved from. http://repository.udea.edu.co/biblioterra m/1/1349/2383/1/CastellanosMenjuraClaudiaPatricia2015.pdf.

Chá, S., Uyul, M., 1995. Regional analysis of tourism resources for marketing purposes. J. Hospit. Leisure Market. 2, 61–74.

Conti, A., Charne, U., Moscovo, F.V., Comparato, G.J., Cassani, M.J., Avalís, V.S., Rucci, A.C., 2012. Evaluación de atractivos para la identificación de nuevos productos turísticos. In: Caso de estudio: región Capitanía de la provincia de Buenos Aires, 18.

Gobernación de Santander. 2016. Plan Desarrollo Departamental. Gobernación de Santander.

Defert, P., 1972. Essai de formulation d’une typologie intégrée des ressources et activités touristiques. In: Métodos de Recreación Turística y su Aplicación en Pueblos y Regiones En Voe do Desenvolvimento, pp. 64–75.

Deng, J., King, B., Bauer, T., 2002. Evaluating natural attractions for tourism. Ann. Tourism Res. 29 (2), 422–438.

Departamento Administrativo Nacional de estadística. 2018. Boletín Técnico Pobreza Monetaria Santander Boletín Técnico. Departamento Administrativo Nacional de estadística, pp. 1–13.

Eagles, F.F.J., Bowman, Margaret E., Chang-Hung Tao, Teresa. 2001. Guidelines for planning in Parks and Protected Areas of East Asia.

Egul, A.Y.S., B., T.U.L.S.I., 2014. Tourist hotel location selection with analytic hierarchy process. Market. Tourism Manag. Conf. 47–58.

Estevao, C., Ferreira, J., 2009. The tourism clusters role in regional development: presenting a competitiveness conceptual model. Tourism Destination. Develop. Brand. 2013, 127–139. Retrieved from. https://repository.ipch.it/p/bitstream/10400.11 /1479/1/tourism.pdf.

Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis. In: Quality and Quantity, 14.

Garrido, M.A., Sánchez, J.A.L., Enríquez, A.F., 2015. Rutas turísticas culturales e itinerarios culturales como productos turísticos: reflexiones sobre una metodología para su diseño y evaluación. Análisis Espacial y Representación Geográfica: Innovación y Aplicación (Octubre), 463–471.

Gosal, A.S., Geijzendorffer, I.R., Václavík, T., Poulin, B., Ziv, G., 2019. Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosyst. Serv. 38 (June), 100958.

Gover, J.C., Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classif. 3 (1), 5–48.

Gunn, C.A., 2002. Tourism Planning: Basics, Concepts, Cases/Clare A. Gunn with Turgut Karypis, M., Kumar, V., Steinbach, M., 2000. A comparison of document clustering techniques. TextMining Workshop At.

Instituto Valenciano de Tecnologías Turísticas, 2015. BIG DATA: retos y oportunidades.

Infante Sánchez, E. del P., 2014. Elementos determinantes en Cundinamarca para el desarrollo del turismo como actividad estratégica regional. Sama de Negocios 5 (10), 40–48.

Instituto Valenciano de Tecnologías Turísticas, 2015. BIG DATA: retos y oportunidades para el turismo. Instituto Valenciano de Tecnologías Turísticas, p. 64.

Jafari, J., 2005. El turismo como disciplina científica. Politic. Soc. 42 (1), 39–56.

Jeambey, Z., 2016. Rutas turísticas en nuevos cantones: Tulcán, Espejo y Mira para la creación de circuitos turísticos. Arquit. Urban. 35 (1), 48.

Kombol, T.P., 2000. Rural tourism on the Croatian islands - Sustainable development and regenerative strategies, 102, 425–437.

Knezevic, R., 2008. Contents and assessment of basic tourism resources, 14 (1), 79–118.

Kropf, M., 2002. Evaluating natural attractions for tourism. Ann.

Kunc, P., 2007. Tourism planning: a conceptual model. Tourism in Parks and Protected Areas of East Asia. Tourism in Parks and Protected Areas of East Asia. Tourism in Parks and Protected Areas of East Asia. Tourism in Parks and Protected Areas of East Asia. Tourism in Parks and Protected Areas of East Asia.

La Marín- Yusaf, M., Nogues Bravo, D., 2003. La potencialidad turística del medio natural en el lic de las Sierras Ibericas Riojanas mediante evaluacióan del turismo en nuevos cantones: Tulcán, Espejo y Mira para la creación de circuitos turísticos. Arquit. Urban. 35 (1), 48.

Laitamaki, J., Tada, M., Liu, S., Setadya, N., Vatcharaoonorn, N., Zheng, F., 2016. Tourism Industry, 14 (1), 7–29.

Laino, M., 2011. Análisis Factorial Testamento De Tablas Mista. Nuevas equivalencias entre el ACP normado y el ACM, 1st21. QUESTII O, pp. 99–108.

Li, C., Hung, T., Ng, Y., 2016. Urban tourism competitiveness: a revealed preference approach. J. Travel Res. 55 (2), 214–228.

Li, C., Chu, C.-H., Wang, Y., Yan, W., 2002. An improved fuzzy c-means algorithm for manufacturing cell formation. 1. In: 2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE’02 Proceedings (Cat. No.02CH379211), 2 pp. 1505–1510.
Lillo, Adelaida, Ramón Rodríguez, Ana, Sevilla-Jimenez, Martín. 2006. Un marco de análisis del capital humano en turismo. Papeles de Turismo 39, 45–59.

Lopez, Fernando. 2002. Libro Verde de la Accesibilidad en España. Diagnostico y bases para un plan integral de supresión de barreras. Madrid: Instituto de Migraciones y Servicios Sociales.

Lucas, Robert C. 1968. Clawson and Knetsch, economics of outdoor recreation. Nat. Resources 3, 738–743. Available from: http://digitalrepository.unm.edu/njr/vol11/iss4/13.

Maass, S.F., Luz, I., De, R., 2010. Evaluación multicriterio de los recursos turísticos del Parque Estatal Multicriteria evaluation of tourism resources of the park sierra de. Martínez, Mario Alberto, Osorio García, Maribel, Franco Maass, Sergio, Ramírez de la Parque Estatal. El periplo Sustentable 18, 1–11. Retrieved from: http://www.mincultura.gov.co/areas/poblaciones/noticias/Documents/Caratrerizacin%20del%20pueblo%20Awa.pdf.

Meyer, D., 2004. Tourism Routes and Gateways: Key Issues for the Development of Tourism Routes and Gateways and Their Potential for Pro-poor Tourism, pp. 1–31 (April).

Mikery Gutiérrez, Mildred Joselyn, Pérez-Vázquez, Arturo. 2014. Métodos para el análisis del potential turístico del territorio rural. Revista Mexicana de Ciencias Agrícolas (9), 1729–1740. https://doi.org/10.20931/remexca.v09.i09.1066.

Ministerio de Cultura de Colombia. 2004. Raíces, isleños descendientes de europeos y africanos. In: Awa Kualiker, Gente de La Montana, (018000). Ministerio de Cultura de Colombia, pp. 1–11. Retrieved from: http://www.mincultura.gov.co/areas/poblaciones/noticias/Documents/Caracterizac%20de%20pueblos%20Awa.pdf.

Minitab 18, 2018. Dendrograma. Retrieved from: https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/.

MINCIT, 2003. Seguridad Turística: Reto competitivo de Colombia. Ministerio de comercio de Colombia.

Misra, P.G.M.S., 2009. Asia Pacific Journal of Marketing and Logistics. Retrieved from: http://www.emeraldinsight.com/doi/pdfplus/10.1108/13555850910928281.

Mortezá, R., Reza, F.M., Seddiqu, M.M., Shararveh, P., Jamal, G., 2016. Selection of the optimal tourism site using the ANP and fuzzy TOPIS in the framework of Integrated Coastal Zone Management: a case of Qeshm Island. Ocean Coast Manag. 130, 179–187.

Navarro, D., 2015. Recursos turísticos y atractivos turísticos: conceptualización, clasificación y valoración. Cuad. Tur. (25), 335.

Nilashi, M., Abani, A., Dalvi, M., Vedegardhåkordi, E., Samad, S., Ibrahim, O., Akbari, E., 2019. Preference learning for eco-friendly hotels recommendation: a multi-criteria collaborative filtering approach. J. Clean. Prod. 215, 767–783.

OEA. 1987. Touristics Resources. CICATUR-OEA.

Open Africa, 2002. Information Relating to the African Dream Project. Open Africa, Cape Town.

Papin Fabrío, C., Paredes de Blas, X.X. 2004. Una propuesta de turismo sostenible para el municipio de Caldas de Reis (Pontevedra). Cuad. Tur. 13 (13), 107–126. Retrieved from: http://dialnet.unirioja.es/servlet/articulo?codigo=1033354&info=resumen

Pichayadejanant, K., Nakaphoem, P., 2018. Data mining approach for arranging and clustering the agro-tourism activities in orchard. Kasetsart J. Soc. Sci. 39 (3), 407–413.

Podani, J., 2012. Extending Gower’s General Coefficient of Similarity to Ordinal Characters, 48 (2), 331–346.

Porter, M.E., 2000. Location, competition, and economic development: local clusters in A global economy. Econ. Dev. Q. 14 (1), 15–21.

Querida Castro, Renato. 2006. Elementos Del Turismo. EUNED, pp. 1–216.

Ratiz, T., Puczkó, L., 1998. Rural Turpism and Sustainable Development in Hungary. ‘Rural Tourism Management: Sustainable Options’ International Conference, Conference Proceedings. SAC, Auchincruive, Ayr, pp. 450–464.

Real Academia Española (RAE). (Spanish Royal Academy). 2005. Spanish dictionary. Second edition. Available at: https://dle.rae.es/infraestructura?m–form (consulted in octubre, 2018).

Reyes, O., Sánchez, C., 2005. Metodología para determinar el potential de los recursos turísticos naturales en el Estado de Oaxaca, México. Cuad. Tur. 1 (16), 153–173.

Rocha, H., 2004. Entrepreneurship and development: the role of clusters. Small Bus. Econ. 23, 363–400.

Rubio, M.L., Duval, V.S., Pezzola, A., 2016. Análisis de Localización de Emprendimientos Turísticos en El Sector Norte del Partido de Villarino (Argentina). InterEspacio: Revista de Geografía e Interdisciplinaridad 2 (5), 9–35.

SECTOR, 1975. ¿Qué Hacemos? Retrieved from: http://www.gob.mx/sectur/es/#192.

Smith, S.L.J., 1987. Regional tourism analysis of tourism resources. Tourism 14, 254–273.

Smith, S., 2013. Identification of funtional tourism in North America. Am. Pharmaceut. Rev. 16 (5), 13–20.

Solís, M.J., Errazuriz, M.J., Caradeuc, C., Seabra, G., 2013. Evaluación multicriterio de la potencialidad turística de un territorio. caso de estudio parque nacional pan de ázcar, region de Atacama. Chile 1–12.

Triantaphylloiou, E., 2000. Multi-Criteria Decision Making Methods. (March), pp. 5–21.

UNWTO, 2008. In: Responding to Global Challenges Climate Change and Tourism Responding to Global Challenges (Copyright; World Tourism Organization and United Nations Environment, World Tourism Organization and the United Nations Environment Programme, Madrid, Spain. UNWTO, 2013. Sustainable Tourism for Development Guidebook. Enhancing Capacities for Sustainable Tourism for Development in Developing Countries, 228.

Vanegas, Juan Gabriel. 2017. Evaluación Multicriterio e inventario de atractivos turísticos. Espacios 38, 1–6.

Wilks, D.S., 2011. Cluster Analysis. Int. Geophys. 100, 603–616.

Zimmer, P., Grassman, S., 1996. Eauat el potential turístico de un territorio. In: Seminario Leader,