Analysis of Medical Tourism and the Effect of Using Digital Tools to Profile Travelers in Mexico

Edmundo Arrioja-Castrejón(✉) and Andrée Marie López-Fernández

Escuela de Ciencias Económicas y Empresariales, Universidad Panamericana, Augusto Rodin 498, Insurgentes Mixcoac, Benito Juárez, 03920 Ciudad de México, Mexico
earrioja@up.edu.mx

Abstract. The increase of internet services and the availability of new cutting-edge technology, as well as the use of mobile apps and devices, set a major challenge for medical tourism service providers. E-commerce sites and social networks are an important source for consumers' awareness and knowledge regarding the medical tourism business model in México which has grown at accelerated rates in recent years; therefore their use should be enhanced by formalizing this activity, and creating integrated commercial strategies [1].

Word-of-mouth and the Internet are consumers' preferred sources of information for medical tourism [2], where trust is elemental for the perception of quality online information [3]. The use of social networks as a means of reference and information based on other users' experiences and recommendations, as well as the use of mobile applications, have contributed to making electronic commerce increasingly profitable for companies that incorporate them in their business model.

This research analyzes the preferences of consumption of medical tourism services in Mexico to determine the effect of formalizing the offer of services with the help of a digital platform on the consumers' experiences.

The study seeks to determine, through the integration of different cluster methodologies, which are the profiles of medical tourism travelers to better predict their consumption habits and channel the various options available through the use of digital tools and social networks to promote the most attractive alternatives for consumers' benefit, which translates into higher sales impacting the entire value chain of this sector.

Keywords: Medical tourism · Digital platforms · Traveler profile · Consumer behavior

1 Introduction

Improving the experience of medical tourism consumers represents a major challenge for service providers in this sector, especially when considering consumer preferences and habits, ensuring confidentiality, safety, budget, promotions, and integrated packages

J. A. Marmolejo-Saucedo et al. (Eds.): COMPSE 2021, LNICST 393, pp. 143–161, 2021.
https://doi.org/10.1007/978-3-030-87495-7_10

that are currently available, with the use of Artificial Intelligence (AI) tools and routines to contemplate other alternatives that the customer has available according to their integrated profile as a patient and traveler.

This research analyzes the consumption habits and preferences of medical tourism services in Mexico to determine the effect on customer experience when formalizing the offer of services with the help of a digital platform. The overall objective of this research was to examine the consumer preferences of the medical tourism industry in Mexico and the impact of digital tools on their choice of a service provider.

The lack of use of digital tools, transparency and promotion in the medical tourism industry in Mexico, and its possible effects, led to the formulation of the following hypotheses: an increase in the transparency of medical tourism services in Mexico through the use of digital tools, has an effect on consumer purchase decision-making; and, the greater use of digital tools to make more information available about medical tourism, the greater the patient's willingness to consume services in Mexico.

1.1 Theoretical Framework

Considering the needs and preferences of medical tourism consumers using various tools to approach the relationship among consumers and service providers should be a priority [4]; the latter should occur while still contemplating additional factors, such as supporting the patient throughout their treatment and their companions to remain in a comfortable environment throughout their stay, and providing activities to make their trip more pleasant.

Medical Tourism

Tourism is a strategic activity for Mexico and contributes 8.4% to the national Gross Domestic Product (GDP). The hotel sector in Mexico represents the main source of income of this industry and is the source of work for millions of Mexican families who directly or indirectly depend on this activity [5].

Medical services in Mexico represent another important source of direct employment and social welfare by integrating many specialized professionals such as doctors, paramedics, and nurses, as well as other related service providers, such as scientists, administrators, assistants, stretcher-bearers, among many others. According to ProMexico's data for 2011 [6], private health institutions totaled nearly 13,500 practices (of which 62% were a specialty) and 34,807 beds.

According to ProMexico's 2013 health tourism report [6], the proportion of medical specialists among general doctors in Mexico is 63.4% above the average of the countries of the Organization for Economic Cooperation and Development (OECD); therefore, this sector represents a large source of income for many medical specialists and is an important generator of indirect jobs through its supply chain.

The private health sector has the greatest probability of benefiting from medical tourism in Mexico since it has been the only provider of medical tourism services, until now, relying solely on the support of local and federal governments [7].

The medical tourism industry in Mexico has had significant growth over the last few years until 2020. According to data obtained from Euromonitor International [8], as shown in Fig. 1 below, Mexico maintained sustained growth except for the years 2009, 2015, and 2016.

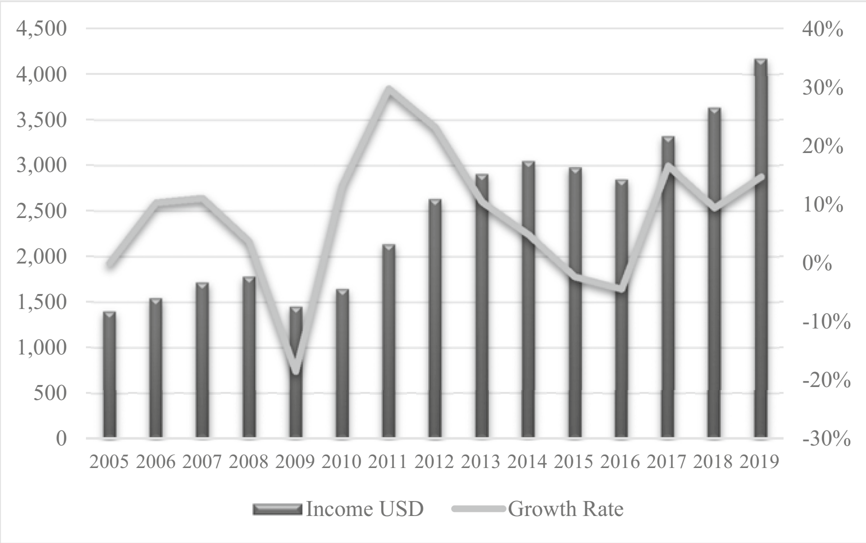

Fig. 1. Medical tourism revenue in Mexico in millions of dollars and annual growth rate.

In the period from 2005 to 2019, service providers and consumers, mainly in the United States and Canada, saw in Mexico an alternative to treat the 21 types of medical conditions (see Table 1) that can be programmed for care in clinics and hospitals, reaching growth figures close to 30% in just 2011 [9].

This is mainly due to the differential in price and the quality of the facilities and medical care that the country possesses. Mexico, together with Israel and Thailand, is one of Canada's preferred destinations for medical conditions (such as vision problems, dental implants, and gastric and aesthetic treatments) with sustained growth for Canadian consumers between 2013 and 2018 [10].

Medical tourism revenues in Mexico, according to 2018 data provided by the National Institute of Statistics and Geography (INEGI) [11], hovered around 7.1% of total revenue from international tourism, equivalent to approximately $1.6 billion dollars. If national foreign medical tourism is also considered, the figure increases in 2018 to more than $3,637 million and, at the end of 2019, to more than $4.175 million [8].

The North American market is an unbeatable opportunity for medical tourism in Mexico [10] due to proximity and prices of medical treatments in the United States and Canada. If the political and economic conditions of the country allow, it will be an important source of jobs, social welfare, and growth for many Mexicans looking for new opportunities because of the number of businesses and supply chains involved in these sectors. These conditions may only be given when medical tourism services are technically provided with the use of digital tools to bring the supply and demand of these services into contact [9].

Table 1. Major countries and medical tourism treatments in the world, data from the Medical Tourism Association – August 2020 (Figures in US dollars).

Treatment / Country	Colombia	Costa Rica	India	Israel	Jordan	Malaysia	Mexico	Poland	Singapore	Thailand	Turkey	USA
Angioplasty	7,100	13,800	5,700	7,500	5,000	8,000	10,400	5,300	13,400	4,200	4,800	28,200
Heart Bypass	14,800	27,000	7,900	28,000	14,400	12,100	27,000	14,000	17,200	15,000	13,900	123,000
Valve Replacement	10,450	30,000	9,500	28,500	14,400	13,500	28,200	19,000	16,900	17,200	17,200	170,000
Hip Replacement	8,400	13,600	7,200	36,000	8,000	8,000	13,500	5,500	13,900	17,000	13,900	40,364
Hip Resurfacing	10,500	13,200	9,700	20,100	9,000	12,500	12,500	9,200	16,350	13,500	10,100	28,000
Knee Replacement	7,200	12,500	6,600	25,000	9,500	7,700	12,900	8,200	16,000	14,000	10,400	35,000
Spinal Fusion	14,500	15,700	10,300	33,500	10,000	6,000	15,400	6,200	12,800	9,500	16,800	110,000
Dental Implant	1,200	800	900	1,200	1,000	1,500	900	925	2,700	1,720	1,100	2,500
Lap Band	8,500	9,450	7,300	17,300	7,000	8,150	6,500	6,700	9,200	11,500	8,600	14,000
Gastric Sleeve	11,200	11,500	7,300	20,000	7,500	8,400	8,900	9,400	11,500	9,900	12,900	16,500
Gastric Bypass	12,200	12,900	7,000	24,000	7,500	9,900	11,500	9,750	13,700	16,800	13,800	25,000
Hysterectomy	2,900	6,900	3,200	14,500	6,600	4,200	4,500	2,200	10,400	3,650	7,000	15,400
Breast Implants	2,500	3,500	3,000	3,800	4,000	3,800	4,500	3,900	8,400	3,500	4,500	6,400
Rhinoplasty	4,500	3,800	2,400	4,600	2,900	2,200	3,800	2,500	2,200	3,300	3,100	6,500
Face Lift	4,000	4,500	3,500	6,800	3,950	3,550	4,900	4,000	440	3,950	6,700	11,000
Liposuction	2,500	2,800	2,800	2,500	1,400	2,500	3,000	1,800	2,900	2,500	3,000	5,500
Tummy Tuck	3,500	5,000	3,500	10,900	4,200	3,900	4,500	3,550	4,650	5,300	4,000	8,000
Lasik (both eyes)	2,400	2,400	1,000	3,800	4,900	3,450	1,900	1,850	3,800	2,310	1,700	4,000
Cornea (per eye)	ND	9,800	2,800	ND	5,000	ND	ND	ND	9,000	3,600	7,000	17,500
Cataract surgery	1,600	1,700	1,500	3,700	2,400	3,000	2,100	750	3,250	1,800	1,600	3,500
IVF Treatment	5,450	ND	2,500	5,500	5,000	6,900	5,000	4,900	14,900	4,100	5,200	12,400

The Covid-19 pandemic during 2020 and 2021 has boosted the technification of many previously only face-to-face services and has accelerated the adoption of digital solutions and tools by companies and professionals [12].

It has been identified that the most widely used media by the offer of medical services are the Internet with 26%, radio and television with 14% each, the yellow pages directories with 12% and, finally, newspapers and magazines with 10% each [13].

Digital Tools

The literature review highlights the need for digital tools based on Artificial Intelligence (AI) for the analysis of human behavior in the fields of understanding, perception, problem-solving and decision-making to reproduce them with the help of a computer [14].

AI is a concept that has been studied for several decades but has been enhanced mainly in recent years because of the number of applications and the impact it is having on modern society, as well as the consequences of this technology on people's behavior and decision-making [15]. However, projects that integrate technology should focus more on information and less on technology [16], as the goal of any system is to provide information or execute a data-based action.

Big Data

Multiple companies are investing large amounts of resources in information technologies and data experts to extract valuable information from the large volumes of data available on networks. Those companies and individuals who respond more efficiently and recognize that competing in the world of analytics will provide them with greater competitive advantages will be the main beneficiaries of this technological revolution [17].

Information may come from social networks, images, and sensors on devices, among many other sources and, because of this, all service providers must base decision-making on evidence and data. Organizations also need to process and identify patterns in very large datasets and translate them into useful business information [18].

The use of data is very important, particularly in regards to the exploitation given to them through advanced analytics for the personal care industry [19], which is closely related to some treatments sought-after under the model of medical tourism such as dental and breast implants, gastric bypass and liposuction. A study on healthcare application showcases a variety of data sources, their constant interrelationship, and the applications available for processing and leveraging personal care [20].

There are several key elements to consider in handling large volumes of data: veracity, variability or complexity, and value [21]. Therefore, to the extent that data is classified through the use of technology, service providers such as hospitals, clinics, airlines, hotels, and restaurants, all of them related to medical tourism, will be able to have information available in real-time to boost their business; thus, they will be able to offer goods and services focused on customers with a greater degree of acceptance.

Currently, internet marketing and social media companies use their members' large volumes of data to analyze them for commercial gain, which is particularly "evidenced by the burgeoning popularity of many social networks such as Twitter, Facebook, and LinkedIn" [22].

Social Networks and Mobility

Social media and the internet have helped to obtain simultaneous, real-time information, and segment different types of consumers by their interests, education, sex, location, and marital status [23]. This helps companies like Amazon, Google, and Facebook anticipate people's needs and consumption habits so they can produce and deliver products and services closer to everyone's preferences.

The use of mobile devices has intensified the use of the internet for multiple personal applications which provide a means to develop programs capable of distinguishing locations and displacements; they also allow the location of consumers at all times and determine their mobility and consumption habits, which helps companies offer real-time services to customers approaching the provider's location [24]. The availability of mobile devices and the COVID-19 pandemic effect is revolutionizing the use of such technologies globally.

A study on group recommender systems describes a social media-based approach to referral systems for the tourism industry, where a group profile is built analyzing not only user preferences but also social relationships between members of the whole group [25]. This is highly useful if we consider groups with common health problems such as cornea disease or heart valve transplants.

Regarding social media, consumers' trust positively impacts brand assessment and their emotional response to advertisement [26]. As such, websites and social media are important platforms for healthcare provides as well as medical tourists to both search for and provide information on experiences, services, destinations, firms, costs, amongst others [27]. Therefore, data and information retrieved online, and particularly social media, have a significant impact on medical tourism consumers' decision-making.

Recommendation Systems

Recommendation systems have been developed in parallel with the Internet, initially built with demographic filters, based on content [28]; and, these systems are currently incorporating social media information and using device data through the Internet of Things (IoT).

A study which examined the landscape of hybrid recommendation systems introduced a system that combines knowledge-based recommendations and collaborative filters to suggest restaurants [29], could be the basis for the development and optimization of a system like this useful for medical tourism since they are services with common characteristics.

There are important achievements in the tourism industry that have been particularly benefited by the new technological advances that arise day by day. For instance, a study showed the development of an Artificial Intelligence (AI) software agent, called Traveler, which helps users select their trip [30]. This solution combines collaboration methods with content-based recommendations and demographic data about customers to suggest tailor-made vacation packages. It is important to analyze this development in greater depth to determine whether it is possible to adapt it to consider all the categories and subcategories that have arisen from interviews and future medical tourism consumer surveys.

Artificial Intelligence analytical-based recommendation systems and advanced routines for determining customer behavior in tourism are a phenomenon of increasing relevance to the economic impact it represents for AI companies. As such, it is important to study the role played by information technologies in the competitiveness of the service providers [31].

2 Methodology

The objective of this research is to segment the population of medical tourists and understand their profile. We worked with mixed data using a special metric (Gower distance) to achieve better results. This document analyzes data from international medical tourism associations: Patients Beyond Borders and Medical Tourism Association. It also analyzes data from specialized reports from Euromonitor International in August 2020 and other data sources detailed below.

2.1 Cluster Segmentation Analysis

Clustering segmentation refers to a broad set of techniques for finding groups, or subgroups, within a data set. When we carry out this grouping, we seek to divide the observations so that those belonging to the same group are quite like each other, while those belonging to other groups are different.

The model used in this study began with a correlation analysis to determine if the variables have a significant relationship. Subsequently, a cluster segmentation analysis for mixed data was performed using hierarchical agglomerative, k-means, and k-medoids to identify the relationship between the study's variables and the groups of traveler profiles, considering the sample under study and the effect of the selection of reservation tools on those profiles.

The objective of comparing a series of methodologies is to be able to select an ideal methodology for the treatment of the analyzed data. The following steps were considered in the development of the segmentation model:

1. Validation of variable correlation and trends in clustering using Hopkins statistic
2. Selection of the number of partitions to be made
3. Selection of the best cluster segmentation algorithm using Dunn index
4. Validation of the generated clusters using Silhouette coefficient.

Hierarchical Agglomerative
Hierarchical agglomerative grouping classifies objects into groups according to their similarity and, through a repetitive process, pairs of groups are successively merged until all groups merge into one large group containing all objects [32].

Clustering by K-means
Clustering by K-means is an effective algorithm to divide a data set into K groups according to the similarity of the observations [33].

Let X1,…, Xn a sample of observations. Where a partition of K groups of observations is defined by C1,…, Ck, where $C1 \cup C1 \cup \cdots \cup Ck = \{X1, \ldots, Xn\}$ and $Ci \cap Cj = \emptyset$; $i \neq j$.

The objective is solving the minimization problem (Eq. 1).

$$\begin{matrix} min \\ C1, \ldots, Ck \end{matrix} = \left\{ \sum\nolimits_{k=1}^{k} \mathrm{W(Ck)} \right\} \tag{1}$$

Where, W(Ck) is a measure of the intra-cluster variation Ck, which must solve the quadratic Euclidean distance (Eq. 2).

$$\mathrm{W(Ck)} = \frac{1}{|Ck|} \sum\nolimits_{i,i' \in Ck}^{.} \sum\nolimits_{j=1}^{p} \left(Xij - Xi'j \right)^2 \tag{2}$$

Where |Ck| denotes the number of observations in cluster K, and X_{ij} is the value of observation i in variable j.

Clustering by K-medoids

K-medoids is a partition clustering technique where each cluster is represented by an element of the cluster, these are points known as the medoids. The term medoid refers to an object within a cluster for which the average dissimilarity between it and all other members of the cluster is minimal [33]. K-medoids is less sensitive to noise and outliers because it uses the medoids as cluster centers instead of the centroids (used in k-means).

The K-medoids algorithm is a robust alternative to K-means to partition a data set into groups. In the K-medoids method, each group is represented by a selected object within the cluster. The selected objects are called medoids and correspond to the most central points located within the cluster.

The PAM (Partitioning Around Medoids) algorithm is the most common grouping method [34], that requires that the user know the data and indicate the appropriate data and number of clusters to be produced.

Hopkins Statistic

For validation of randomness, we use the Hopkins statistic (Eq. 3) that tells us whether the study database behaves uniformly or not. This allows us to define if the data are subject to clustering. This statistic measures the probability that a given data set is generated by data of a uniform distribution [35].

$$H = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i} \tag{3}$$

Where $X_i = dist(p_i, p_j)$ is the distance for each observation from its closest neighbor and $Y_i = dist\left(q, q_j\right)$ is the distance for each point from its closest neighbor in a data set simulated using a uniform distribution [36].

Gower Distance for Mixed Data

A popular option for clustering is to use the Euclidean distance. However, this metric is only valid for continuous variables. Thus, for a clustering algorithm to produce reasonable results, we must use a distance metric that can handle mixed data types such as those contemplated in this study [37]. Therefore, we will use the Gower distance (Eq. 4):

$$GOW_{jk} = \frac{\sum_{i=1}^{n} W_{ijk} S_{ijk}}{\sum_{i=1}^{n} W_{ijk}} \tag{4}$$

Where $W_{ijk} = 0$ if the objects j and k cannot be compared for the variable i if X_{ij} or X_{jk} are not known. Gower's distance concept indicates that for each type of variable, a particular measurement scale is used that works well for that type and scaled between 0 and 1. A linear join is then calculated using user-specified weights to create the final distance matrix [38].

Dunn Index

The Dunn index (Eq. 5) is an internal cluster validation measure that allows you to calculate the distance between each of the objects in the group and the objects in the other groups [39]. This index calculates the minimum of this pairwise distance as the separation between clusters (min.separation) and uses the maximum distance within the cluster as the intracluster compactness.

$$D = \frac{\text{min.separation}}{max.diameter} \tag{5}$$

Silhouette Coefficient

The Silhouette coefficient (Eq. 6) was used as an optimization criterion to determine the number of partitions in a cluster using a measure of the quality of the cluster classification [40].

$$Silhouette = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{d_i - s_i}{\max\{d_i - s_i\}} \tag{6}$$

2.2 Definition of Variables

Data were first obtained from the main medical treatments that are programmed under the medical tourism modality and from the main countries where they are marketed. Then a survey was carried out with a sample of 100 tourists and consumers of medical tourism incorporating some variables defined in a previous study carried out in Malaysia [41], with the purpose of identifying travel preferences to better establish the traveler profile.

From the questions asked, the following variables were obtained: type of trip (business, pleasure, or both), gender (male or female), age, number of trips made per year (1, 2, 3, 4, 5, 6, 7 or +), average expense per trip in Mexican Pesos, preferred type of transportation (air, land, or sea), preferred place of travel (between beach and city) and preferred reservation medium (internet/mobile app, social media or travel agency/phone). For this exercise, we considered a sample of 51 men and 49 women.

From that database, the selection variables were converted by assigning values 0 and 1 to convert them into dichotomic and trichotomic variables (see Table 2), and focus on determining which factors should be contemplated when designing a digital tool based on the user preferences; the latter resulted in the following variable relationship:

Table 2. Variables analyzed according to a sample of 100 medical tourists surveyed.

<table>
<tr><th colspan="2">Variables</th><th>Data type</th><th>Description of variables</th></tr>
<tr><td colspan="2">Expenditure</td><td>Weights figure</td><td>Average travel expense</td></tr>
<tr><td colspan="2">Age</td><td>Figure in Years</td><td>Traveler's age</td></tr>
<tr><td colspan="2">Travel</td><td>Whole number</td><td>Number of trips per year</td></tr>
<tr><td>Gender</td><td>1 – Male
0 - Female</td><td>Dichotomic variable</td><td>Gender (male or female)</td></tr>
<tr><td rowspan="3">Type of Transport</td><td>Aerial</td><td rowspan="3">Trichotomic variable</td><td>Air mean preference</td></tr>
<tr><td>Terrestrial</td><td>Land mean preference</td></tr>
<tr><td>Maritime</td><td>Maritime mean preference</td></tr>
<tr><td>Place</td><td>1 – City
0 - Beach</td><td>Dichotomic variable</td><td>Travel place preference (city or beach)</td></tr>
<tr><td rowspan="3">Type of Trip</td><td>Business</td><td rowspan="3">Trichotomic variable</td><td>Higher business travel frequency</td></tr>
<tr><td>Pleasure</td><td>Increased frequency of pleasure travel</td></tr>
<tr><td>Both</td><td>Mixed travel frequency</td></tr>
<tr><td rowspan="3">Reservation medium</td><td>Internet_App</td><td rowspan="3">Trichotomic variable</td><td>Online booking or App preference</td></tr>
<tr><td>Social_Network</td><td>Social media booking preference</td></tr>
<tr><td>Agency_Phone</td><td>Booking preference by agency / phone</td></tr>
</table>

3 Results

3.1 Correlation Analysis

From the above data, we can see that there are 3 variables associated with the reservation medium, 3 variables associated with the type of transport, 4 variables associated with the traveler profile, and 1 variable associated with the place. Considering a sample of 100 travelers of medical tourism, we use the correlation analysis to study the relationship between the set of variables associated with the traveler profile with the variables linked to the type of reservation and quantify the number of existing independent dimensions.

The variables associated with the traveler's profile are Expense, Age, Travel, and Gender. The variables linked to the reservation type are Internet_App, Social_Network, and Agency_Phone. In addition, the variable Gender is an indicator variable from 0 to 1, where 1 indicates the consumer is male and 0 indicates the consumer is female.

Table 3 below shows the descriptive analysis of the variables studied, where we can see that we have 3 quantitative (Expense, Travel, Age) and 7 qualitative variables (Gender, Air, Terrestrial, Maritime, Internet_App, Social_Network, Agency_Phone).

Below are the 3 sets of grouped variables used in this study, of which we will only use in this analysis the reservation group (reservation type, composed of the variables: Internet_App, Social_Network, and Agency_Phone), and traveler information (traveler's profile, consisting of the variables Expense, Travel, Age, Gender), to be able to determine if there is a correlation among these variables, and see if the use of digital tools such as the Internet, mobile applications or social networks have a significant relationship with the type of user profile. We can see that the correlations within and between the 2 selected sets of variables in the database are as follows (see Table 4):

Table 3. Descriptive analysis of variables, using data from 100 surveys conducted on medical tourism travelers in 2019

Descriptive	Expenditure	Travel	Age	Gender	Place	Air	Terrestrial	Maritime	Internet_App	Social_Network	Agency_Phone
Min.:	5362	1.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1st Qu.:	24659	2.00	34.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mediate:	45471	4.00	45.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00
Mean :	44022	3.82	46.87	0.51	0.55	0.60	0.33	0.07	0.55	0.17	0.28
3rd Qu.:	62596	6.00	62.25	1.00	1.00	1.00	1.00	0.00	1.00	0.00	1.00
Max.:	79857	7.00	75.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Correlation of variables, data from 100 surveys on medical tourism travelers in 2019

Correlation	Expense	Travel	Age	Gender	Internet_App	Social_Network	Agency_Phone
Expense	1.0000	0.0113	0.0302	-0.0200	0.1029	-0.0523	-0.0703
Travel	0.0113	1.0000	-0.0103	0.0214	0.3535	-0.0644	-0.3378
Age	0.0302	-0.0103	1.0000	-0.1026	-0.0270	-0.3616	0.3324
Gender	-0.0200	0.0214	-0.1026	1.0000	-0.0020	0.1241	-0.1016
Internet_App	0.1029	0.3535	-0.0270	-0.0020	1.0000	-0.5003	-0.6894
Social_Network	-0.0523	-0.0644	-0.3616	0.1241	-0.5003	1.0000	-0.2822
Agency_Phone	-0.0703	-0.3378	0.3324	-0.1016	-0.6894	-0.2822	1.0000

As we can see in Table 4, there is a clear correlation between the age and the type of reservation; this correlation is positive at older ages in relation to the telephone or travel agency reservation method, and negative in relation to the Internet and social networks. There is also a positive correlation between average expenses and travel frequency.

3.2 Hopkins Statistic and Gower Distances for Mixed Data

First, we can see that Hopkins statistic shows the probability that a given data set is generated by data of a uniform distribution, using R tool is: H = 0.5006383, this implies that the grouping to be identified is not random and that there is an underlying pattern.

Next, we see that Gower's distance can be calculated on a line using the daisy function in R. Note that due to a positive result in the variable enrolled, a record transformation is performed internally through the type of argument. Instructions for performing additional transformations, such as factors might be considered asymmetrical binaries (such as rare events).

As a sanity control, we can print the most similar (see Table 5) and different (see Table 6) pair in the data to see if it makes sense. In this case, the person aged 32, whose average expense is 36,835, and the person of age 31 whose average expense is 41,743, are the most similar given the 11 characteristics used in the distance calculation:

Table 5. Generating similar pairs of data using Gower distance

Expense	Travel	Age	Gender	Place	Air	Terrestrial	Maritime	Internet_App	Social_Network
36835	4	32	0	1	1	0	0	1	0
41743	4	31	0	1	1	0	0	1	0

While on the contrary, the person aged 71, whose average travel expense is 17,675 and the 22-year-old whose average spend is 75,504 are the most different:

Table 6. Generating different pairs of data using Gower distance

Expense	Travel	Age	Gender	Place	Air	Terrestrial	Maritime	Internet_App	Social_Network
17675	6	71	0	1	0	1	0	0	0
75504	1	22	1	0	0	0	1	0	1

3.3 Clustering Methods Comparison

Below is the comparative (see Table 7) of the 3 cluster segmentation methods proposed, where we can see that the hierarchical agglomerative model for 3 partitions is the one that reports the highest parameters using Dunn index; however, after applying the final validation using Silhouette coefficient, we can see that the difference between hierarchical agglomerative and K-medoids method is minimal, so we will select the K-medoids method for the final analysis as it is a less sensitive model to noise and outliers.

Table 7. Comparison of the proposed clustering segmentation methods

Clustering segmentation	Clusters:	2	3	4	5	6
Hierarchical	Connectivity	2.2151	8.2044	8.8405	9.4627	11.0516
	Dunn	0.445	0.5309	0.5552	0.6109	0.6109
	Silhouette	0.3591	0.3292	0.3679	0.4034	0.3949
K-means	Connectivity	4.5262	6.1825	8.8405	9.4627	11.0516
	Dunn	0.4525	0.4776	0.5552	0.6109	0.6109
	Silhouette	0.3669	0.3444	0.3679	0.4034	0.3949
K-medoids (PAM)	Connectivity	5.9893	17.9833	28.2437	35.9425	29.8905
	Dunn	0.4525	0.3068	0.2028	0.2028	0.2028
	Silhouette	0.356	0.3268	0.3314	0.2725	0.3093

3.4 Cluster with Mixed Data Using K-medoids

For an algorithm yet to choose to group observations, we must first define some notion of (dis)similarity between observations.

There are a variety of metrics to help choose the number of clusters to extract in cluster analysis. We will use silhouette width, an internal validation metric that is an aggregate measure of how similar an observation is to a cluster compared to a nearest neighbor cluster. The metric can range from -1 to 1, where the highest values are better. After calculating the Silhouette width for clusters ranging from 2 to 10 for the PAM algorithm (which uses a partition around medoids), for the analyzed data, we can see in Fig. 2, that 3 clusters yield the highest initial value:

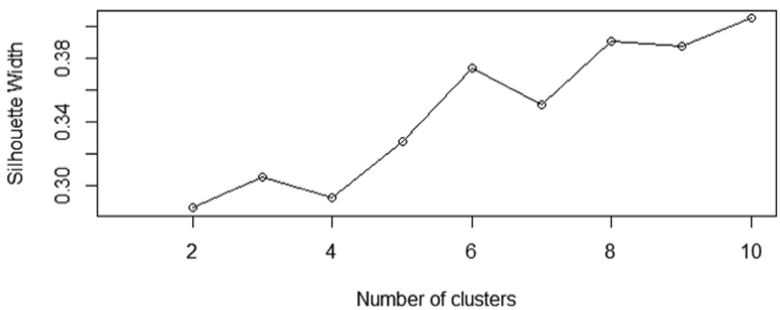

Fig. 2. Silhouette coefficient width chart to validate the number of clusters

One way to visualize many variables in a space of lower dimensions is with the graph t-SNE. This method is a dimension reduction technique that attempts to preserve the local structure to make clusters visible in a 2D or 3D plane (see Fig. 3). In this case, the graph shows the three well-separated clusters that PAM could detect:

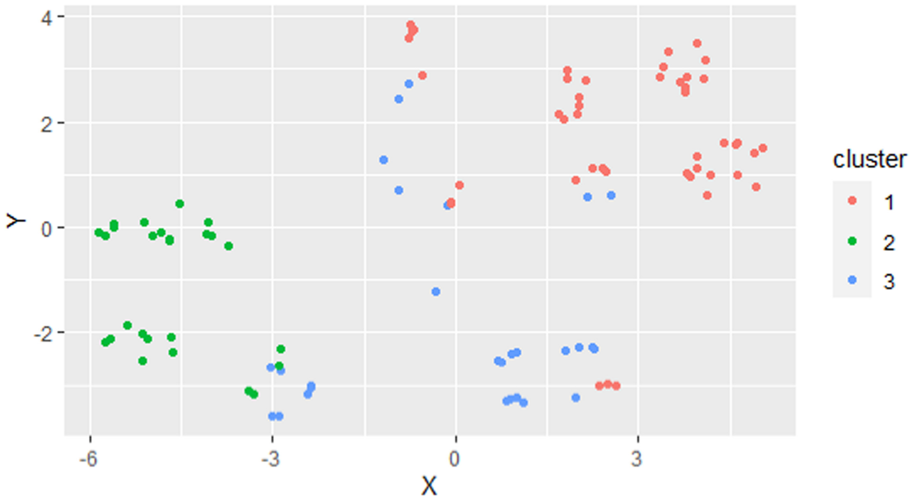

Fig. 3. Cluster scatter chart using the R tool with survey data conducted on medical tourism travelers in 2019

3.5 Cluster Validation Using Average Silhouette Width

The Silhouette width coefficient is used to validate the quality of the cluster and determine how well the observations were classified; higher values imply good classification (see Table 8), negative values mean badly classified observations and values of 0 mean that there is a possible overlap.

Table 8. Cluster widths using Silhouette coefficient

Cluster	Size	Average Silhouette width
1	41	0.42
2	34	0.3
3	25	0.21

To validate the number of clusters selected, we use the average Silhouette width. In Fig. 4 below, we can see the quality of the cluster created with K-medoids, in fact a Silhouette coefficient is generated for each observation, and the graph shows how

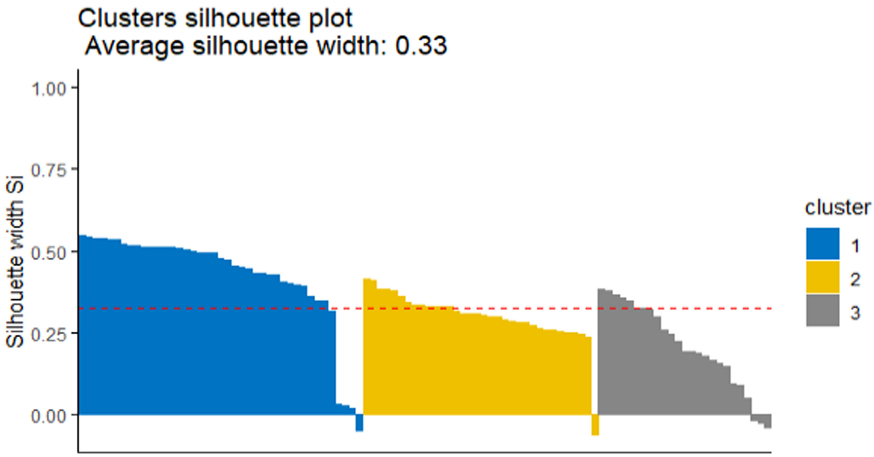

Fig. 4. Cluster validation using Silhouette width coefficient using R tool with survey data conducted on medical tourism travelers in 2019

many are correctly classified, how many can be overlaps and how many would be in the inappropriate cluster (the negative ones).

After running the algorithm and selecting 3 clusters, we can interpret the clusters by running a summary on each one, based on the following results:

Cluster 1 is mainly the profile of the Frequent/Air/Modern Traveler with average spending levels and frequent travel in the year, likes to book by electronic means and prefers to travel by plane, is the typical traveler who is looking for airlines and seeks comfort first.

Cluster 2, on the other hand, is mainly the profile of the Saver/Land/Modern Traveler with low spending levels and moderate travel in the year, that normally likes to book by electronic means and prefers to travel by land, and is the typical globetrotter.

Cluster 3 is mainly the profile of the Occasional/Air/Traditional Traveler, with the lowest levels of travel per year, average expenditure and that since it is not a frequent flyer prefers traditional means of reservation and the least possible complications.

Another benefit of the PAM algorithm, concerning interpretation, is that medoids serve as examples of each group (Table 9).

It is noteworthy that the male person, age 43 and average travel expense of 56,483, is the medoid of the Frequent/Air/Modern Traveler cluster, the male person, age 42 and average travel expense of 18,543, is the medoid of the Saving/Land/Modern Traveler cluster, and the female person, age 47 and average travel expense of 67,968, is the medoid of the Occasional/Air/Traditional Traveler cluster.

Table 9. Generation of medoids using the PAM algorithm

Cluster medoid	Expense	Travel	Age	Gender	Place	Air	Terrestrial	Maritime	Internet_App	Social_Network	Agency_Phone
1	56483	5	43	1	1	1	0	0	1	0	0
2	18543	6	42	1	1	0	1	0	1	0	0
3	67968	1	47	0	0	1	0	0	0	0	1

4 Conclusions

Consumers mostly travel from the United States, Canada and Western Europe to Asia and Latin America for medical tourism [42]. Common reasons may include insufficient insurance coverage, high costs of procedures [43] and/or even high deductibles. A study on medical tourism in Malaysia, found that demographics, such as age and sex, motivations, such as value for money, medical service quality and secondary services, as well as the type of procedure, were determining factors in the profile of medical tourists [41]. Another regarding medical tourism in India, found that it is common for medical tourists to incorporate travel plans during their medical procedure trip [44]. And, a study on expectation confirmation perspective suggested that consumer satisfaction is well associated with their expectation confirmation which may be derived from the actual performance of the medical service [45]. Thus, the study's findings are significantly related to results from previous studies in relation to medical tourism in other countries.

A key factor today to increase the competitiveness of medical tourism service providers is to take advantage of the exponential increase in internet purchases. The use of social networks as a means of reference and knowledge of recommendations based on other users' experiences, as well as the use of mobile applications have contributed to make e-commerce increasingly profitable for companies and those who incorporate it within their operating model. From the doctors who use social networks to advertise, the following social networks are identified as the most widely used: Facebook and YouTube with 32% each, Google with 18%, and Twitter and Pinterest with 9% each [13].

The use of digital tools to take advantage of the large volume of information that is on the Internet and social networks will give a competitive advantage to medical tourism service providers, especially considering new generations; as can be seen in the results, they prefer to interact with internet sites, social networks, e-commerce portals, whether through computers or mobile applications, than through a travel agency or a telephone advisor. Therefore, it is a significant opportunity to use recommendation systems based on Artificial Intelligence (AI) algorithms that can filter and channel consumers of these medical tourism services to the specialist, hospital, or clinic that can provide quality care, considering their budget, interests, and condition.

Based on the results obtained from the cluster segmentation analysis, we can conclude that consumers of medical tourism services in Mexico can be categorized into 3 different profiles based on their preferences and consumption habits, where the reservation method is a differentiating factor that is associated with the frequency of travel and the age of the participants within the sample analyzed in this research. Thus, results indicate that digital tools, in fact, positively influence consumer decision-making. Further, having clearly identified profiles of medical tourists will allow service providers to offer a better user experience focused on attention and improve contact conditions, which may increase patient's willingness to consume services in Mexico.

Therefore, we can also conclude according to the results obtained from the applied models that digital tools, such as the Internet, mobile applications, and social networks have a direct and positive effect on the behavior of consumers of medical tourism services in Mexico, which allow determining the type of profile of the traveler accurately, in addition to being increasingly used as a reservation method for the selection and contracting of these services, which confirms what is commented by several of the authors reviewed in the theoretical framework such as [13, 23, 25] and [12].

The study's main limitation is the sample, as results cannot be generalized. As a future line of research, it is contemplated to integrate the segmentation of the clusters into a platform for the data processing of medical tourists, as well as the generation of automatic reports that could be consulted automatically through the cloud.

References

1. Taufik, N.A., Sulistiadi, W.: The impact of medical tourism industry for the hospital services and marketing activities: a systematic review. J. Adm. Rumah Sakit **5**(1), 42–48 (2018)
2. Lam, C.C.C., Du Cros, H., Vong, T.N.L.: Macao's potential for developing regional Chinese medical tourism. Tour. Rev. **66**, 68–82 (2011)
3. Moslehifar, M.A., Ibrahim, N.A., Sandaran, S.C.: Assessing the quality of trust features on website content of top hospitals for medical tourism consumers. J. Komun. Malays. J. Commun. **32**(1), 469–489 (2016)
4. Kim, S., Arcodia, C., Kim, I.: Critical success factors of medical tourism: the case of South Korea. Int. J. Environ. Res. Public Health **16**(24), 4964 (2019)
5. Secretaria de Turismo: El gran motor de la economía nacional (2018)
6. ProMéxico: Turismo de Salud. Tur. Salud, pp. 1–2 (2013)
7. Orozco Núñez, E., et al.: An overview of Mexico's medical tourism industry – the cases of Mexico City and Monterrey (2014)
8. Euromonitor International: Medical Tourism Report (2020)
9. Euromonitor International: Global medical tourism briefing: a fast growing niche market (2011)
10. Becker, S.: Opportunities and caveats of medical tourism in Canada (2014)
11. INEGI: Encuesta de Turismo de Internación (2018)
12. Garcia-Muñoz, C., Pérez Sánchez, B., Navarrete Torres, M.: Las empresas ante el COVID-19. Rev. Investig. Gest. Ind. Ambient. Segur. salud en el Trab. GISST **2934**, 85–101 (2020)
13. Montiel, J.: Estrategias publicitarias en línea utilizadas por la oferta de turismo médico plástico y estético: el caso de Tijuana, Baja California, México. Online advertising strategies used by the supply of plastic and aesthetic medical tourism: Case study of T. Int. J. Mark. Commun. New Media **3**(June), 106–117 (2015)

14. Hardy, T.: (IA: inteligencia artificial). Polis Rev. la Univ. Boliv. **1**(2), 23 (2001)
15. Dwivedi, Y.K., et al.: Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int. J. Inf. Manage. (2019)
16. Marchand, D.A., Peppard, J.: Why IT fumbles analytics: tech projects should focus less on technology and more on information. Harv. Bus. Rev. (January-February 2013), 104–112 (2013)
17. Davenport, T.H., Patil, D.J.: Data scientist: the sexiest job of the 21st century. Harv. Bus. Rev. (October 2012), 70–76 (2012)
18. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. Harv. Bus. Rev. (October 2012), 60–68 (2012)
19. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Heal. Inf. Sci. Syst. **2**, 1–10 (2014)
20. Koh, H.C., Tan, G.: Data mining applications in healthcare. J. Healthc. Inf. Manag. **19**(2), 64–72 (2011)
21. Gandomi, A., Haider, M.: Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manage. **35**(2), 137–144 (2015)
22. Aggarwal, C.C.: An introduction to social network data analytics. In: Aggarwal, C. (ed.) Social Network Data Analytics, pp. 1–15. Springer, Boston (2011). https://doi.org/10.1007/978-1-4419-8462-3_1
23. Angulo Toro, L.S.: Big data y neuromarketing como herramientas útiles para medir el comportamiento del consumidor en la industria de telecomunicaciones con el fin de crear con mayor precisión. Universidad Católica de Santiago de Guayaquil (2019)
24. Naveed, M.: Online learning based contextual model for mobility prediction BT - evolving ambient intelligence. In: O'Grady, M.J., et al. (eds.) Evolving Ambient Intelligence. AmI 2013. Communications in Computer and Information Science, vol. 413, pp. 313–319. Springer, Cham (2013). https://doi.org/10.1007/978-3-319-04406-4_32
25. Christensen, I., Schiaffino, S., Armentano, M.: Social group recommendation in the tourism domain. J. Intell. Inform. Syst. **47**(2), 209–231 (2016). https://doi.org/10.1007/s10844-016-0400-0
26. Hahn, I., Scherer, F., Basso, K., Santos, M.: Consumer trust in and emotional response to advertisements on social media and their influence on brand evaluation. Braz. Bus. Rev. **13**(4), 49–71 (2016)
27. Medhekar, A.: The role of social media for knowledge dissemination in medical tourism. In: Medical Tourism, pp. 132–161 (2017)
28. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. Knowl.-Based Syst. **46**, 109–132 (2013)
29. Burke, R.: Hybrid recommender systems: Survey and experiments. User Model. User-Adapted Interact. **12**(4), 331–370 (2002)
30. Schiaffino, S., Amandi, A.: Building an expert travel agent as a software agent. Expert Syst. Appl. **36**(2 PART 1), 1291–1299 (2009)
31. Martínez, J., Majó, J., Casadesús, M.: El uso de las tecnologías de la información en el sector hotelero. In: Turitec 2006: VI Congreso nacional turismo y tecnologías de la información y las comunicaciones, pp. 1–13 (2006)
32. Ward, J.H.: Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58**(301), 236–244 (1963)
33. Loy-García, G., Rodríguez-Aguilar, R., Marmolejo-Saucedo, J.-A.: An analytical intelligence model to discontinue products in a transnational company. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) ICO 2020. AISC, vol. 1324, pp. 812–822. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68154-8_70

34. Park, H.S., Jun, C.H.: A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. **36**(2 PART 2), 3336–3341 (2009)
35. Banerjee, A., Davé, R.N.: Validating clusters using the Hopkins statistic. In: IEEE International Conference on Fuzzy Systems, vol. 1, pp. 149–153 (2004)
36. Hopkins, B., Skellam, J.G.: A new method for determining the type of distribution of plant individuals. Ann. Bot. **18**(2), 213–227 (1954)
37. Tuerhong, G., Kim, S.B.: Gower distance-based multivariate control charts for a mixture of continuous and categorical variables. Expert Syst. Appl. **41**(4 PART 2), 1701–1707 (2014)
38. Gower, J.C.: A general coefficient of similarity and some of its properties. Biometrics **27**(4), 857 (1971)
39. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. **3**(3), 32–57 (1973)
40. Gentle, J.E., Kaufman, L., Rousseuw, P.J.: Finding groups in data: an introduction to cluster analysis. Biometrics **47**(2), 788 (1991)
41. Musa, G., Thirumoorthi, T., Doshi, D.: Travel behaviour among inbound medical tourists in Kuala Lumpur. Curr. Issues Tour. **15**(6), 525–543 (2012)
42. Hopkins, L., Labonté, R., Runnels, V., Packer, C.: Medical tourism today: What is the state of existing knowledge. J. Public Health Policy **31**(2), 185–198 (2010)
43. Lew, A.A., Hall, C.M., Williams, A.M.: The Wiley Blackwell Companion to Tourism. Wiley, Oxford (2014)
44. Sajjad, R.: Edical tourism in India: an empirical analysis of the demographic profile and perception of medical tourists. MAGNT Res. Rep. **3**(8), 150–168 (2015)
45. Chou, S.Y., Kiser, A.I.T., Rodriguez, E.L.: An expectation confirmation perspective of medical tourism. J. Serv. Sci. Res. **4**(2), 299–318 (2012)