ATU PRESS

Analysis of the Behavior of Tourists in Iran Based on Data Mining and Search Rate on Google

Mahrokh Makhsoos (iD)
PhD Candidate, Department of Business Management, Faculty of Humanities, Hamedan Branch, Islamic Azad University, Hamedan, Iran

Naderehosadat Najafizadeh (iD) *
Assistant professor , Department of Business Management, Arak branch , Islamic Azad University,Arak, Ira

Habibollah Javanmard (iD)
Associate Professor, Department of Management, Arak Branch, Islamic Azad University ,Arak , Iran

Abstract

Understanding tourist behavior is a requirement for marketing planning for the supply of goods and services to satisfy tourists. Nowadays, many tourists decide to travel to any place by searching through internet explorers. The present study was conducted with the aim of analyzing the behavior of tourists in Iran based on data mining (search rate on Google Trend). The research is applied and has been done with a causal descriptive method. The method of collecting data is by searching keywords on Google. . To collect the keywords needed for the research, we had to turn to the experts and authorities on the field as well as the relevant scientific articles. The process of data collection is that a series of key words in tourism were selected and accordingly, it was determined how much people used these words in different places in Iran. Correlation matrix (covariance) model has been used for data analysis. In this research, the structural information of these keywords was obtained and edited based on the keywords related to tourism extracted from Google Trend by time series and

with the help of Pearson correlation matrix. Nine keywords were selected for search, including hotel, entertainment, pilgrimage, tourism, nature, places of interest, archeology, travel and travel tour. The keywords are not searched at random, but they are related and correlated and stem from a structural thinking. The results of the data analysis have shown the type and intensity of connections between the words that had a communication structure. It was also found that the words pilgrimage, recreation, and archeology have less connection with other words.

Introduction

Due to its positive social, cultural and economic effects, tourism recieves serious attention in various countries today. The tourism industry is becoming the first industry in the world in terms of income generation and impact on economic growth, so that many countries are trying to exploit the benefits of this industry by creating suitable fields and infrastructures. For most of these countries, tourism is an important source for business activities, income generation, job creation and foreign exchanges. Many of these countries earn a lot of income from this industry every year. (Rahm Del, 1385).

The tourism industry is known as one of the largest industries in terms of income and employment generation which tends to grow really fast. In fact, tourism is a key tool and the number of international tourists is used to measure its growth. The United Nations states that the tourism rate is increasing and it is predicted that by 2020, America will account for 4% of the total tourist population, East Asia will account for 7.2%, the Middle East will account for 6.2%, and South Asia will account for 5.8%. It accounts for all international tourists (Dmytro, 2017).

Henderson et al. (2018), Tang and Tan (2015) and Li and Chang (2008) stated in their research that tourism is an important item in the country's economic growth and that tourists are vital for the continued growth of the tourism industry and its economic effects. . Tourism creates many benefits for the society and affects the quality of life of the people of that society and brings many benefits to the people of the society. In fact, it improves people's social life, and this obliges the government to pay more attention to supporting the tourism industry and respect the increase in the quality of people's lives and the increase in the quality of people's needs (Liu et al., 2017). Capra (2018) stated in his research that the government should create a growing horizon for the development of tourism through the tourism organization and thereby highly influencing the sustainable development of tourism.

Iran, having thousands of historical monuments and places of interest and enjoying features of cultural legacy such as Persepolis, Pasargad, Bistun, Falak ol-Aflak, etc., which have been registered in the list of universal monuments, must have an acceptable presence in the target markets of worldwide advertisements, but in effect a very small share of the tourism worldwide income belongs to Iran. By

reviewing the published statistics, it turns out that most of the tourists have traveled to Iran due to family , work or otherwise problems. Another issue is the low amount of domestic tourism in the country, one of the reasons for this is the lack of information of foreign and domestic tourists and of course the lack of plans and measures to attract tourists and meet their needs. In order to attract tourists and keep them attracted, it is necessary to first identify the needs, destinations and goals of the tourists and then make a comprehensive plan to meet the needs of the tourists.

One of the methods of determining the need of tourists as to travel is to understand and analyze the behavior of tourists. What the tourist searches for is to a large extent representative of the tourist's goals and intentions. Tourists use internet search tools to find the necessary information.

The purpose of collecting information for tourists is to plan based on the amount of tourists' searches to meet their needs and goals. For this, it is necessary to obtain information from tourists. It is necessary to know what the interests and needs of people's travel are, so that suggestions can be made to the managers and decision makers of tourism affairs. For example, you can tell them what the tourists are looking for and what strategy is needed to attract tourists

Nowadays, by using Google Trend, it is possible to get economic and generally comprehensive information from all over the world, different provinces and cities and at different times. Conventional methods and tools such as questionnaires in collecting data from tourists include weak points such as small statistical population, unavailability of all people, incorrect and incomplete answers, waste of time and lack of variety, and using Google Trends shows who , where from , when and to which places they intend to travel or have traveled. This information is obtained through the search of tourists on the Internet. The advantages of this method are access to a large number of samples around the world, accuracy and correctness of data, high speed of data preparation and availability of data.

Therefore, to identify which of the variables or keywords in the field of tourism has a greater contribution to the choice of tourists and whether it is possible to categorize the searched variables or keywords and reach a specific pattern in these searches, so that we can get the tourists to help us choose the time and place of travel, the present research will provide a model helping tourists by establishing a

relationship among the variables of the research . Therefore, the current research with the aim of analyzing the behavior of tourists in Iran based on data mining (search rate in Google Trend), shows how by accessing the data of tourists, we will achieve statistical results , whereby we can determine the purchasing behavior of tourists and predict or determine where tourists travel to, how they make their traveling decisions, what they expect of a trip, what their needs are ,and finally, what results the analysis of the behavior of tourists in Iran yields based on data mining (search rate on Google Trend).

Research literature
Tourism market and marketing
Tourism market refers to a set of countries from which tourists come . In other words, the countries that send tourists are considered as destination countries for tourism markets. Accordingly, whenever tourists travel in their own country, it is considered a starting point for the destination of domestic tourism markets.

Marketing in the tourism industry is a managerial (administrative) process that includes forecasting the needs and satisfaction of current and future tourists, so that a travel company and suppliers are based on   competition with each other. Exchange management is guided by personal benefit, social benefit, or both, and the success of a long-term method depends on the mutual satisfaction of the supplier and the consumer. In other words, it can be said that fulfilling the tourist's environmental and social demands is as important as getting his satisfaction and they can no longer be indifferent to each other's characteristics (Delavar et al., 2013).

Tourism marketing is all the activities and processes carried out in the way of bringing buyers and sellers together. These actions include creating, distributing, promoting, expanding, pricing and presenting new ideas and opinions in order to facilitate the exchange that takes place in a dynamic environment (Lundberg, 2013).The tourism target market is a part of the potential market in which the tourist attractions will sell the most. Since the tourism target market has the highest competitive advantage, organized marketing efforts are necessarily focused on their needs and desires. (Islam, 2003).

Investigating consumer behavior in tourism

Consumer behavior in successful marketing tourism begins with understanding   why   and   how   the   consumer   behaves.   Consumer

behavior includes various psychological and social processes that exist before and after purchasing activities. Behind every purchasing action lies an important decision-making process that needs to be explored. The steps a buyer goes through to decide what kind of product or service to buy is called the buying decision process. This process consists of five steps as shown in the figure below. Each buyer first feels a need, then searches for information about obtaining that product or service, and according to the purchase criteria and the information obtained, a selection is made from amidst the products and services, and the decision to purchase is made. In the end, the behavior after the purchase, i.e. the agreement of the goods or services with the expected demands, causes the satisfaction or dissatisfaction of the consumer, which also causes the repetition or non-repetition of the purchase.
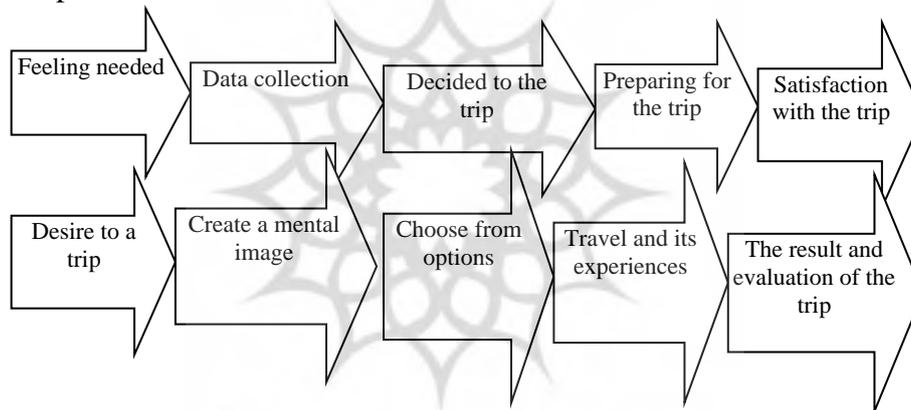
Diagram 1. Related to the purchase decision process

Deciding to buy tourism products and services is associated with a lot of risk. This risk involves a large amount of money being spent on things that cannot be seen or evaluated before purchase. If a holiday is spoiled, it will not be compensated for another year. Most people don't have vacations or extra money to make up for a bad vacation.

The internal and external factors that affect the decision to spend time on vacations are evident in the above diagram. Marketing professionals cannot plan without considering these factors. But the classification of these factors and its impact on consumer decision-making clarifies a more specific framework. Because deciding on tourism products and services depends on choosing a series of different products and services, which makes it difficult to make a

purchase decision (Bahramian and Shams al-Dini, 2017). The complex commodity or product the tourist seeks, consists of the 4 following components:

1- The attractiveness of the destination that is its image in the tourist's mind or his impression from that place

2- Accessibility of the destination (ease of access to the destination)

3- The spirit of hospitality and cultural resources of the destination

4- Facilities at the destination (infrastructure and superstructures)

On which we elaborate below:

- The attraction or attractions of the destination may be natural or artificial or temporary or permanent. Natural attractions are mostly permanent and artificial attractions are mostly temporary and changeable. By natural structure we mean weather and historical monuments, waterfalls, sunny beaches and religious cities and artificial structure of tourism products can include souvenirs and handicrafts, which according to each region have their own special characteristics.

- The facilities normally do not motivate or attract tourists and few people travel to use the facilities . But the important point about this element is that the existence of these facilities has a direct relationship with evocation of a desire in the tourist to travel and use the attractions of the destination, and it will increase or decrease his subsequent trips to the destination. These facilities include tourism infrastructures and superstructures. Infrastructure facilities include road communication networks, airports, sources and facilities for supplying drinking water, electricity, etc.

- The ability or ease of access to the destination as well as the ease of return is one of the main components of the tourist complex product. This accessibility is directly related to the type of transportation that the tourist chooses according to the distance dimension and according to his taste and income.

In a sense here the economic distance is more important than the physical distance. Economic distance, in turn, takes into account air, road, rail and sea facilities. These facilities are different for tourists with different incomes and they are chosen according to the advantages of each transportation vehicle, for example, airplane is one of the expensive means of transportation where comfort and speed of

getting to the destination is higher. Train is a means on which the tourist can see the scenery better, but the speed of reaching the destination is lower than the plane. Another component of the complex tourist product is the spirit of hospitality and the atmosphere that dominates the environment. The desire to provide acceptable, warm and friendly services on the part of the host is very important (Marut, 2017).
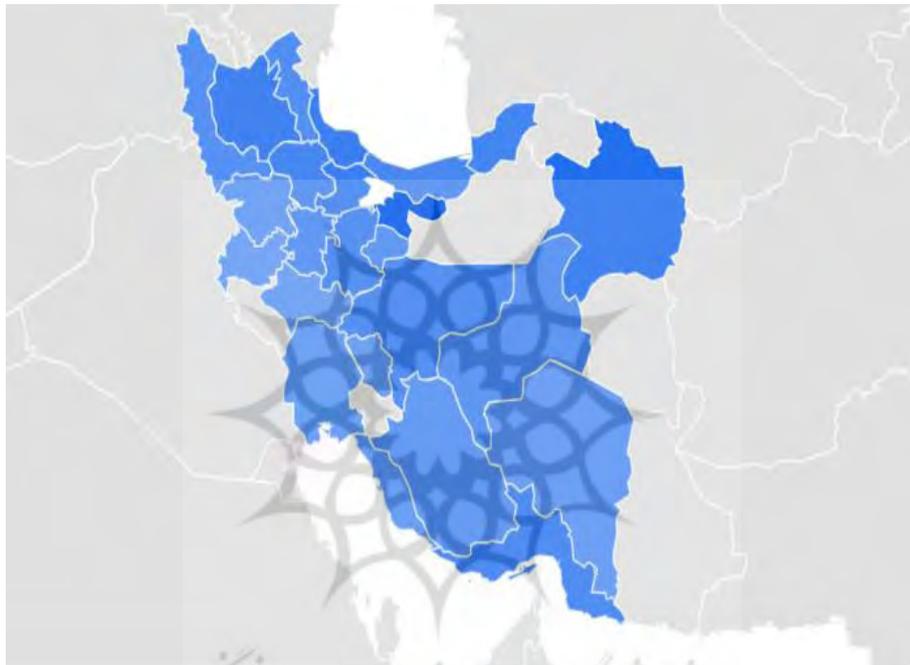

Figure (1) tourism map

Method
The research is applied and has been done with a causal descriptive method. The method of collecting data is searching keywords in Google. Experts in this field and relevant scientific articles were consulted to collect the key words needed for the research. The process of data collection is that a series of key words in tourism were selected and accordingly, it was determined to what extent people used these words in different places in Iran. In this research, we intend to survey the circumstances and particulars involving people`s traveling and how the Internet(Google Trend) facilitates this process for them. First, the tourism key words in the table below were entered

in Google Trend as singular and compound words, and the correlations and connections between the words were examined. The data of this research is determined through Google Trend, which shows the search history of these words in the form of time, location and amount, and Pearson's correlation matrix statistics method was used to analyze the data. The source of the research is Google Trend from 01/06/2013 to 12/24/2017.

Table 1.The searched keywords

keywords	keywords
Hotel	Pilgrimage
Entertainment	Tourism
Nature tour	Places of interest
Archaeological sites	A trip
Travel tour	

If we show each key word with x, the time series will be displayed as x(t), indicating that this word has been searched in Google many times at time t. The correlation matrix between these keywords is as follows:

$$= \{x(1), \ldots x(i) \ldots x(t)\} \qquad x(t)$$

$x(t)$ It shows the number of searches for keywords in Google (it shows the time series of the number of searches in Google). In this article, these keywords are listed as follows:

Hotel - Entertainment - Pilgrimage - Tourism - Nature - Places of interest - Archaeological Sites - Travel - Travel tour-

Table 2. Matrix Components

$x_1 =$ Hotel	$x_6 =$ places of interest
$x_2 =$ Entertainment	$x_7 =$ Archaeological Sites
$x_3 =$ pilgrimage	$x_8 = ATrip$
$x_4 = Tourism$	$x_9 = Travel\ tour$
$x_5 =$ Nature tour	

The components of the Pearson correlation matrix follow the following equation.

$$C_{ij} = \frac{1}{T} \sum_{t=0}^{T} \left(x_i(t) - \bar{x_i} \right) \left(x_j(t) - \bar{x_j} \right)$$

C_{ij} The components of the correlation matrix indicate the degree of correlation of the keyword $x_j(t)$ $x_i(t)$, is in the time interval T.

Findings

Research question - How is the analysis of tourists' behavior in Iran based on data mining (search rate on Google Trend)?

Data analysis has been done with the help of graph theory and based on threshold limit and inverse contribution. To analyze the data, first the normality of the distribution of the data searched in Google and the linearity of their relationships were checked. The correlation between the nodes of the keyword network in the tourism phenomenon was investigated with the help of MATLAB software. In this regard, the N×N correlation matrix, where N is the number of keywords searched in Google, was obtained with the help of Pearson's correlation coefficient. The results of this matrix are presented in Table 2. Data analysis according to the statistical method, the average and standard deviation of each keyword were placed in the diameter of the obtained matrices. In the matrix of each keyword, it plays the role of a vertex.

According to Table 2, the correlation of all nodes for the entire research sample with the value (P<0.01) is significant . positive correlation means aligned correlation; That is, the increase of the value of one node with a certain probability is accompanied by the increase of the value of another node. Negative correlation means anti-correlation. That is, the increase of the value of one node with a certain probability is accompanied by the decrease of the value of another node.

Table 3. The correlation matrix of the search rate of keywords in the tourism industry in Google

	Tourism	Hotel	Pilgrimage	a trip	Entertainm ent	nature tour	Places of interest	Archaeologi cal sites	Travel Tour
Tourism	M***= 7/55 SD****= 1/66	0/44**	0/56**	0/31**	0/35**	0/26**	0/29**	0/23**	0/31**
Hotel	0/44**	5/19 3/76	0/28**	0/71**	0/38**	0/30**	0/39**	0/35**	0/40**
Pilgrimage	0/56**	0/28**	6/95 2/04	0/45**	-0/30**	-0/14**	-0/15**	-0/20**	0/24**
a trip	0/31**	0/71**	0/45**	5/94 4/01	0/38**	0/22**	0/32**	0/30**	0/35**
Entertainm ent	0/35**	0/38**	-0/30**	0/38**	20/59 5/05	0/24**	0/33**	0/33**	0/46**
nature tour	0/26**	0/30**	-0/14**	0/22**	0/24**	60/41 11/37	0/56**	0/39**	0/48**
Places of interest	0/29**	0/39**	-0/15**	0/32**	0/33**	0/56**	78/05 19/24	0/52**	0/63**
Archaeolog ical sites	0/23**	0/35**	-0/20**	0/30**	0/33**	0/39**	0/52**	38/37 12/90	0/49**
Travel Tour	0/31**	0/40**	0/24**	0/35**	0/46**	0/48**	0/63**	0/49**	4/62 3/47

*** Mean
**** Standard deviation
** P<0.01
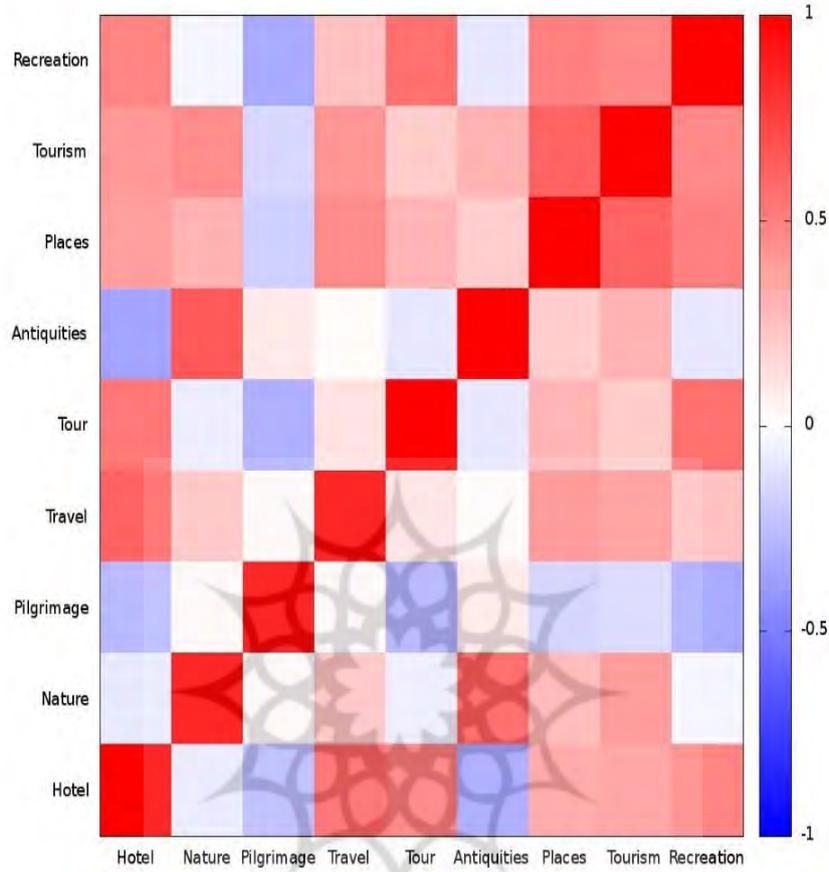
Figure 2. Visual representation of the correlation matrix of the search rate of keywords in the tourism industry in Google
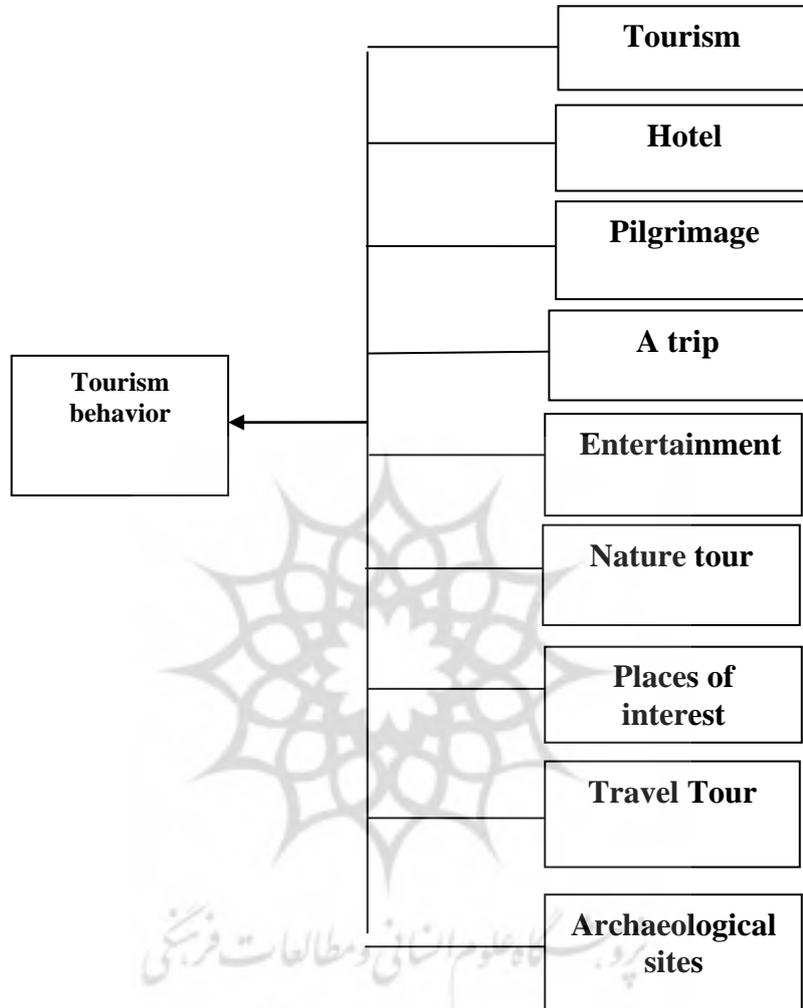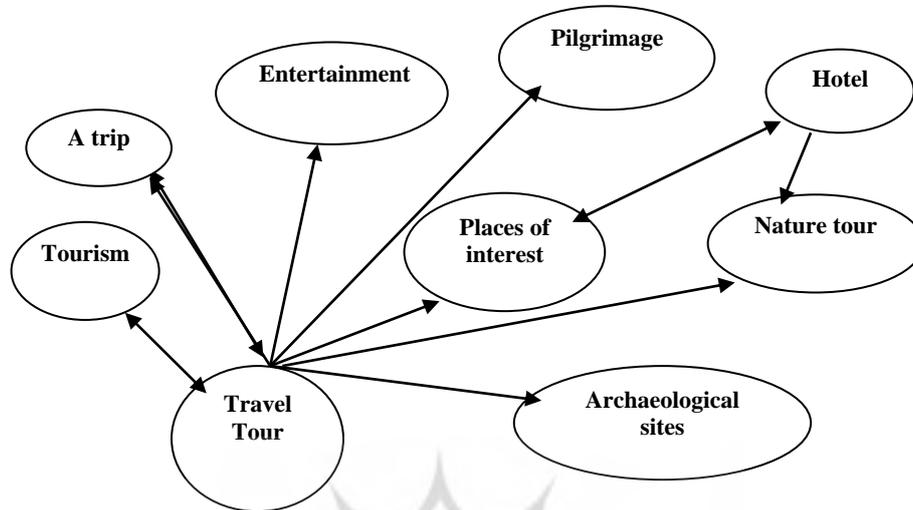
Figure (3) conceptual model of research derived from Google search from 2013 to 2017

Figure (4) relationship between keywords based on correlation matrix

According to data mining and Pearson's correlation test, Figure (3) was presented as a proposed model for analyzing the behavior of tourists in Iran based on the amount of search in Google Trend. In this model, the tourism and travel behavior of people who intend to travel increases with the searches of the keywords : hotel, pilgrimage, travel, nature tourism, places of interest, ancient monuments and tourism. The correlation matrix showed that there is a positive correlation between the tourism variable and the key word variables of hotel, pilgrimage, travel, nature tourism, sightseeing, ancient monuments, and tourism. This correlation is bilateral. That is to say, people tend to travel more frequently to destinations where good hotels and further amenities join beautiful scenery, natural resorts and possibly sanctuaries to visit as the results of data mining analysis show. The highest correlation of the variable of tourism is with the variable of pilgrimage, because the visit of tourists and tourism in an area that is a place of pilgrimage is more than tourism of places with pristine and beautiful nature, ancient monuments and places of interest.

Discussion and conclusion
Tourism is a clean (green) and complex industry, generating income and starting the engine of all-round growth and development, and

after the oil industry and automotive industry, it occupies the third place in the world. According to the UNESCO ranking, Iran ranks fifth in terms of biodiversity and natural attractions and is among the top 10 countries in the world in terms of tourist attractions. Unfortunately, the rate of tourist acceptance and tourism income of Iran is not even in the 40th place in the world, and a very small share of the country's income is allocated to tourism (Mousavi, 2015). Considering Iran's rich civilization and culture, it can be an attractive country for researchers and seekers of history and human science, and many tourists choose this country for travel. Therefore, the current research aims to analyze the behavior of tourists in choosing Iran as a tourist destination through data mining of tourist search results.

The results of this research showed that tourism increases with hotel, pilgrimage, excursions, nature tourism, sightseeing, ancient monuments . The results of the correlation matrix of this variable showed that there is a positive correlation between hotel tourism, pilgrimage, travel, nature tourism, sightseeing, ancient monuments and tourism.

The next variable is pilgrimage, in fact, the place of pilgrimage will attract tourists and travel to that area will increase. Pilgrimage is most correlated with tourism, and the more pilgrimage sites there are in the region, the more tourism will be in that region. Pilgrimage has anti-correlation with nature tourism, sightseeing places and ancient monuments and sites. Considering the high correlation between pilgrimage and tourism, it is suggested that tourists do not limit themselves to pilgrimage in that area and if there is a place with ancient monuments, they should not miss that either. The tourism organization of these places should cooperate with tourists and look at their travel as an opportunity for the region.

The next variable is travel and the highest degree of correlation exists between travel and hotel, the more the amount of travel to a region, the more investors try to make investments there, establish accommodation hotels, and thus the number of hotels will increase. After the hotel, the highest correlation is with travel tours. When there is more travel to an area, travel agencies of that area will register more and become more prosperous. It is suggested that the tourism organization provide more affordable hotels to tourists and meet the needs of tourists to enter a safe and clean environment.

The next variable is entertainment and the highest correlation of entertainment is with travel tours and then travel. When there are more recreational places in a region, there are more travel tours for that region and more trips to those regions. It is suggested that the travel tours will inform the travelers about the place they are going to stay or visit by giving detailed information about the recreational places and providing the travelers with more services.

Nature tourism was the next variable that influenced the behavior of tourists. Nature tourism had the highest correlation with sightseeing places. In other words, the more beautiful and pristine the natural scenery of an area is, the more places there are to visit in that area. For example, investors in these areas can use this potentiality by establishing canteens, parks and more scenic spaces. It is suggested that tourists pay attention to their preservation and cleanliness when they go camping. The tourism organization also prevents prospective transgressors from harming nature by devising and enforcing the necessary rules and regulations to preserve these places.

The next variable is sightseeing places, the highest correlation exists between sightseeing places and nature tourism and this indicates that the more sightseeing places there are in a region, the more nature tourism there will be. It is suggested that the tourism organization spend every effort to preserve the places of interest, for the next generations , so that they will also be able to enjoy the pristine and beautiful natural scenery of the region.

Archaeological sites and travel tours were variables that also influenced the behavior of tourists. The highest correlation exists between archaeological sites and travel agencies. In fact, it can be said that the more ancient monuments there are in a region, the more travel tours for that region will be. Travel tours also have the highest correlation with sightseeing places. When the travel tour to a region increases, visiting sightseeing places increases and tourists would rather visit the sightseeing places of that region than the ancient monuments. It is suggested that the officials of travel tours provide tourists with detailed information about ancient monuments and places of interest by providing detailed descriptions contained in informative relevant booklets and catalogues.

References

Rahm Del (1385), "Khorramabad Underdevelopment Factors Seminar", Cultural Heritage Organization, Volume 1.

Bahramian, Mohsen; Shams al-Dini, Ali (2017), "Quantitative analysis of the role of urban management in promoting the tourism industry in cities (case study: Maroodasht city)", *Journal of Urban Planning and Research*, 9th year, serial number 32, 131-140

Delavar, Ali; Kadri, Ismail; Madi, Nima (2013), "The place of social media in the development of e-tourism marketing in Iran", *scientific-research quarterly of tourism management studies*, 9th year, number 28, 1-16

ii motoooo ihii stina 000077) ,''The uuest oor a rr actccac appooach to morality and the tourism inuustry'', *Journal of Hospitality and Tourism Management* 31 , 45-51

Edvardson, Bo,(2005), "*Service Quality: Beyond cognitive Assessment*" managing service Quality, 15/2 , pp 127-132

dend eosono Isaac Leii ; Aii s kakk ; oon g Kan ,s,,, ,a, 22l1)) ,'' *Testing discontinuous innovations in the tourism industry: The case of scenic airship services''*, Tourism Management 66 ,167-179

Higham, J., Cohen, S.A., Cavaliere, C. T., Reis, A., & Finkler, W(2016), ''iiim ate chan,e, tiiiist aii tlalel anl raii lal emissions ueuu't''n'' *'Journal of Cleaner Production*, 111 (Part B), 336-347

Islam, Ali Akbar (1382), "*Domestic Tourism Market*", Tehran, first edition, Zahad Publications

rarera rrr 000,, ''uustabnalbe torr ism vevepppment errrrt s by ooæal oorern ments in Po'an''',*Sustainable Cities and Society,* https://doi.org/10.1016/j.scs.2018.05.001

Lee, C.-C., & Chang, C.-P2200,,, ''Tuuiism vevelppment and eoonomic rr wwtth A rrrser rrkk at pane's''' *Tourism Management*, 29(1), 180-192 ##

Liu, J., Nijkamp, P., Huang, X., & Lin, D(2017), ''bbban iiiaiiiity and tourism development in China: Analysis of sustainable development by means oo slatlal lan ed data'',*Habitant International*, 68, 99-107 ##

Lohmann, M. and E. kaim, (1999), "*weather and holiday destination preferences, image attitude and experience*" The Tourist riview. P: 12, 72 ##

dad ii son, ,, 22)))) , "*in search of warmer climates? The impact of climate change of flower of British tourists. In the Amenity valae of Global climate*MMavkdoson ,, (e)) E lär thsean, oonoon5 555 P5̌5 ##

Marut, Habib; Salem, Ali Asghar; servant, blessing; Elahi, Mahbubeh (2017), "Identification of factors affecting the demand of foreign tourism", Economic Research Journal, 18th volume, No. 69, 275-306

oo rrrrt hr M,, & uu nt, ((0066,,''*Tourism and sustainability Development, globalisation and new tourism in the third world*'', uuu tgeggeg wew York (NY)

Taleghani,Gholam;Ghafary, Reza Ali; Asgharpour, Seyed Esmail, Bekiai, rary am22l1)) ,''An Investggatoon ff the Barrrers r edatedtd oo iiism Industry Development in Iran'', *Procedia - Social and Behavioral Sciences*, Volume 120, 19 March 2014, Pages 772-778

gang , .. ,,, & ,an, E. 0000,,,, ''*Does tourism effectively stimulate Malaysia's economic growth*''' iiiii sm Management, 4,, 558-163 ##

Wang, Q., Li, X.-Y., Zhang, Z.-T., Tao, Z.-L., & Tang, B.- 000088,,''oaooon emossoos ueuuotoon in tccacc c rr imary rr ggessgng line: *A case study in China. Journal of Cleaner Production*gg ,,,, ,, - 28