

Ministry of Tourism

---

Undersecretariat of Planning and Tourism Policy

---

Research on Statistics & Economics of Tourism

No. 2014-1

---

# **PRODUCTIVE USE OF BIG DATA AND SOCIAL NETWORKS IN TOURISM**

---

---

Research on Statistics & Economics of Tourism provide preliminary research results. Said studies were carried out in order to generate an exchange and debate of ideas on the development of the tourism sector. The content, as well as the arguments expressed, are of the sole responsibility of their authors and do not necessarily reflect the point of view of the Ministry of Tourism.

---

[www.datatur.sectur.gob.mx](http://www.datatur.sectur.gob.mx)

**November 2015**

**DIRECTORY**

**Enrique de la Madrid Cordero**

Ministry of Tourism

**Salvador Sánchez Estrada**

Undersecretary of Tourism Planning and Policy

**Pedro Manuel Lichtle Fragoso**

General Coordinator

**Juan Carlos Sánchez Salinas**

Analysis Coordinator

## PRODUCTIVE USE OF BIG DATA AND SOCIAL NETWORKS IN TOURISM

Pedro Manuel Lichtle Fragoso  
Ministry of Tourism\*

Juan Carlos Sánchez Salinas  
Ministry of Tourism\*\*

Prepared in collaboration with the BIG DATA Work Team of the National Institute of Statistics and Geography (INEGI), consisting of Gerardo Leyva, Alfredo Bustos, Juan Muñoz, Silvia Fraustro and Abel Coronado.

---

### Summary

Some instruments described in this paper have been recently utilize for the productive use of information from non-conventional sources and also for social networks in the tourism sector. This type of processes speed up the building up of empirical models to explain the individual behavior of tourists.

The objective of the paper is to study the movements of a Twitter user's profile that is representative of the average tourist, with a pilot test in the states of Puebla and Guanajuato. The Twitter information was put together in collaboration with INEGI. This exercise consists of determining the place of origin of tourists during February 1, 2 and 3, 2014; as well as contrasting this place of origin that stems from Twitter with official statistics of Puebla and Guanajuato. This study represents an alternative to the high costs of state surveys carried out to follow-up the domestic flow of tourists.

---

\* Email: plichtlef@sectur.gob.mx

\*\* Email: jsanchezs@sectur.gob.mx

## 1. Introduction.

Some instruments are described in this paper that have been recently utilized for the productive use of Big Data and also for social networks in the tourism sector. Although economic and social activities require performing data analysis, nowadays it is indispensable to simultaneously consider the volume, speed and variety of the data under analysis of social behavior according to some objectives and scopes. For example, there are instruments that extract information automatically on-line and transform it into a database with a defined structure. This type of processes speed up the building up of empirical models to explain individual behavior.

In the context of the development of e-government, open-government and Big Data, the subsequent review of the literature looks for empirical evidence on how governments may benefit from this type of platforms and from the world's most popular social networks. Data extraction does not end with the internet: text messages, cell phones and geo-locations, electronic registries, all of them are part of the new generation of information. To this regard, the objective of this paper is to obtain evidence of the productive use of information and to contribute to the development and use of this type of instruments in Mexico's tourism sector.

Hal Varian (2009) studied the predictive attributions of "Google Trends"<sup>1</sup> and used models adjusted to the trends of variables such as retail sales, car sales, home sales and even movement of travelers. According to the author, how travel is commonly nowadays planned over the internet; the data resulting from the search of travel destinations may be useful for predicting visits to a specific travel destination.

In this rationale, Tilly (2013) considers trust as a major variable for using an electronic tourism service. The content generated by users generates more trust than information from travel agencies, official tourism web sites and media. Travelers perceive the content generated by social networks users as an approximation to the experience at the travel destination, however, the question that is still under discussion is the possibility of approximating to tourist's behavior with the information that comes from both Big Data and social networks.

In the past, tourists would only search and use information in person at travel agencies, using travel books and brochures made by promoters of travel destinations. The travel demand was focused on classic travel packages scheduled by local operators, with fixed itineraries and with low-feedback input. The characteristic of the productive use of information was a restricted access to travel destinations, a static flow of information and tourists with passive attitude to locate, recommend and evaluate travel services. Furthermore, the decision on where to travel would result from recommendations and comments of close persons, without considering information massively generated.

Nowadays, tourists search and use information remotely through digital travel services such as web sites, all kinds of applications and communities in social networks. Travel demand is diverse ranging from accommodation, leisure activities and even gastronomy options, with

---

<sup>1</sup> It provides a consultation volume index of Google by geographic location and by category.

flexible and customizable itineraries, and a high degree of feedback. The information regarding to tourism has more volume, variety and speed. A characteristic considered as essential for productive use of information is to have diversified access to travel destinations, a dynamic flow of information and having an active tourist in emerging communication with travel services. As a result, the decision of where to travel arises from recommendations and comments that may be reflected in the information contained in Big Data and social networks of tourists.

Considering this change of search and use of information model, tourists' trends, both domestic and international, focus on using pictures, maps, and iterative searches of restaurants, activities and attractions. To this regard, the so-called "experiential tourism" is based on the emotions and on an adaptation to the client's needs, as well as on exceeding travel expectations, with the complement of communicating messages in such an easier way and with a simple language because the social networks demand so.

To sum up, the trust that tourists have in the information included in Big Data and social networks on destination sites and travel services, and the change in searching and using the information model are the platform to generate the possibility of approximating the tourist's behavior with information from non-structured data and social networks.

This document contains the following sections: 2. Non-structured data and social networks; 3. Productive use of the information; 4. Big Data as source of information; 5. Twitter as source of information; and 6. Conclusions.

## 2. Non-structured data and social networks.

Firstly, non-structured data are defined as open data of great volume, accessible to on-line users; such data are regularly free and have an open or indefinite format. In the tourism sector, you may encounter non-structured data on the geo-localization of users, pictures, registries in all kinds of hotels and searches of leisure activities and tourism attractions.

Secondly, social networks are an on-line service among people who, for example, share interests and activities. A social network service consists of representing each user (profile), their social ties, and a variety of additional services.<sup>2</sup> In the tourism sector, the social networks are utilized to transmitting the experiences and sensations lived by the tourist, and to recognize the destinations based on the activities and services offered.

In this case, the following limitations on the productive use of both non-structured data and social networks information should be considered: data are by no means clear and final and there is no systematic registry stable through time. However, these problems ordinarily referred are part of the struggle faced by social behavior researchers.

---

<sup>2</sup> Most social network services (such as Facebook, Twitter, Google+) are based on using the network in real time and they provide interfaces so that users interact through means such as email and instant messenger.

Furthermore, the biggest challenge is not to collect data from the internet, but in the next stage: processing. This process goes through three stages, the extraction of the information sought, its deuration and conversion into useful information to make statistics.<sup>3</sup> For example, a hotel will frequently publish information about the services it offers in more than one web site, being also published in a number of aggregated sites. This means that after collecting the data, a process against duplication of information has to be implemented.

### 3. Productive use of the information.

Recently, social sciences research is focused on improving mathematical models to predict social and economic trends. However, the availability of digital data opens new options, such as searching technology over the internet, mobile phones, email and social networks. The new instruments offer detailed registries of behaviors and compensate for the classical problems of generating statistics such as the high cost of the national census and surveys, as well as the lack of precise administrative registries. It is worth highlighting that the surveys have biases in the outcomes due to the possibility obtaining subjective information from the interviewed, as well as presenting opinion trends but not reflecting the actual behavior the interviewed.

Although the World Tourism Organization (WTO) has made recommendations to manage domestic tourism, which are accepted worldwide, and today the debate is focused on the generation of regional and sub-domestic statistics. For example, it is recognized that administrative registries are missing in some geographic areas to measure the domestic flows in the countries, or in areas where there is free flow of people, as well as how costly is to do surveys to obtain information; this makes difficult measuring domestic tourism.

As a result, the research on the productive uses of information from non-structured data and social networks, as well as of new estimating methods, acquire great relevance to generate statistics in the tourism sector. According to Heerschap (2014), internet data may be collected in real time and such data are more complete than the data collected through traditional surveys. However, in order to evaluate the quality of the internet data, it is always advisable to compare them with data from other available sources such as official statistics.

A research case in tourism is the one carried out by Wu (2013), where internet consultation and searching engine technology is used. The authors consider that this type of technology has made possible to obtain information at a minimum cost and in a practically instantaneous way. Furthermore, it is pointed out that the collected information is valuable because it reveals the individual characteristics of the consumer and the possibility of making an economic transaction, and the knowledge of this type of information may be used to predict the future demand.

The information used in Wu's work comes from the volume of searching consultations on the internet related to real property of "Google Trends" with weekly and monthly reports on consultations statistics for a number of industries. In this case, the consultations allow users to find information that correspond to a specific phrase, for example: "home prices".

---

<sup>3</sup> Heerschap (2014).

The experience of the referred work may be applied to specific searches within the tourism sector. For example, consult information of users with phrases that contain the words with more use among tourists: “hotel”, “motel”, “restaurant”, “attraction”, “fun”, etc. In addition, as a result of the productive use of the information knowledge on the profile of visitors may be obtained and then being able to generate public policies to attract this type of segment for a travel destination.

Mobile phones and social networks are closely related because a “smartphone” has the most advanced computing capacity as well as base connectivity. For example, there is a study of *Big Data and Tourism*<sup>4</sup> that allows us to know better the Spanish tourism sector. The purpose of the study was to explore the possibilities that the information of Big Data may offer to the sector and especially to the hotel industry, to incorporate macrodata of marketing studies of the electronic activity of anonymous foreign tourist.

This study managed to put together and cross data of two different companies: Telefónica Móviles España and BBVA. The former is an operator integrated to telecommunications, leader in Spain and the latter is a private financial institution with international presence. From the data obtained the following information is presented: i) principal origin of visitors; ii) duration of the stay by country; iii) moving between two travel destinations; iv) days and zones where foreign visitors prefer to stay; and v) average daily expense and accumulated expense throughout the stay.

The recommendations stated by the study focus on increasing the obtaining of clients and determine in what countries it is recommended to focus on commercial action, to determine the areas of the cities where the commercial transactions are made and to guarantee an attractive product adapted the true client’s needs. As a result, the productive use of the information is the generation of business models that increase the attraction of the travel destination.

Additionally, Spencer (2013) studied the relationship between the location of the pictures of the “Flickr” page and the visit rates at the different places of the world. This is the first study in the field of using data coming from social networks to predict visit rates in 836 cities of 31 countries. The authors make a comparison between the density of the pictures with geo-localization (latitude and longitude coordinates) and the estimated average occupation rate with local data of travel attractions.

Accordingly, the density of the pictures generates an approximate value of the average occupation rate based on power.<sup>5</sup> It is pointed out that the study central hypothesis is the following: the images may indicate the attractions for visitors and they may be recorded for the future, learning what the people options as visitors are and providing useful information about the place.

The authors conclude that the information from social networks (in this case Flickr), not only may break out the problem of the cost of the empirical data, but it may also generate effective

<sup>4</sup> Oliver (2014).

<sup>5</sup>  $Y = Y_0 X^\beta$ ; donde  $\beta = 0.698$ ,  $Y_0 = 13.7$

predictions. However, social networks and their complements may also provide revealing information to answer questions about why and where the people look for recreation. Here, the productive use of the information gets to an approximation of occupation rates in a number of cities around the world.

Thus, it is seen that there is no consensus on the productive use of information from non-structured data and social networks, but even looking for a consensus on its use in the tourism sector is a more difficult task. According to the objective of this paper, a brief review of the instruments used in Big Data and Twitter is carried out as follows.

In Mexico, there is a pioneering research on the productive use of information of social networks, carried out by the National Digital Strategy.<sup>6</sup> The objective was to respond to the following research questions: i) Is there any relationship between the social dynamics during a natural disaster and the records of telephone calls?; ii) May the information from telephone antennas be used in the future to improve the response to natural disasters such as the Tabasco flooding in 2009?; iii) Do the findings obtained from the antennas records contrast with the information that provided by official entities?; and iv) How do floods affect the usual patterns of communication in the region?

The information obtained in this type of studies may be useful as a basis to determine zones that need immediate help or the sending of additional resources for shelters near the supply zone and to generate an efficient location of distribution centers. In addition, to analyze the places to where the population is moving and to determine where temporary shelters are located helps to evaluate whether they are at the right place, as well as to better calculate the input and resources that would be necessary until people return to their homes.

#### 4. Big Data as source of information.

According to Ronald Jansen,<sup>7</sup> using of the new era of information through large “Big Data” data bases represent the merging of three large international trends: i) on-line transactions (purchase, sale, transportation); ii) on-line interactions (social groups, remote specials); and iii) information processing (processing techniques, storing, visualization). Consequently, in order to analyze the great variety and increasing speed of new databases the characteristics of the information should be considered first, then the tangible benefits for the analyst, and lastly, the sources of information.

---

<sup>6</sup> *Movement of the population and natural disasters: data analysis, cellular antennas and Big Data during the Tabasco floods in 2009*, presented on June 16, 2014.

<sup>7</sup> Chief, International Trade Statistics, United Nations Statistics Division.

Table 1. Considerations for Big Data analysis.

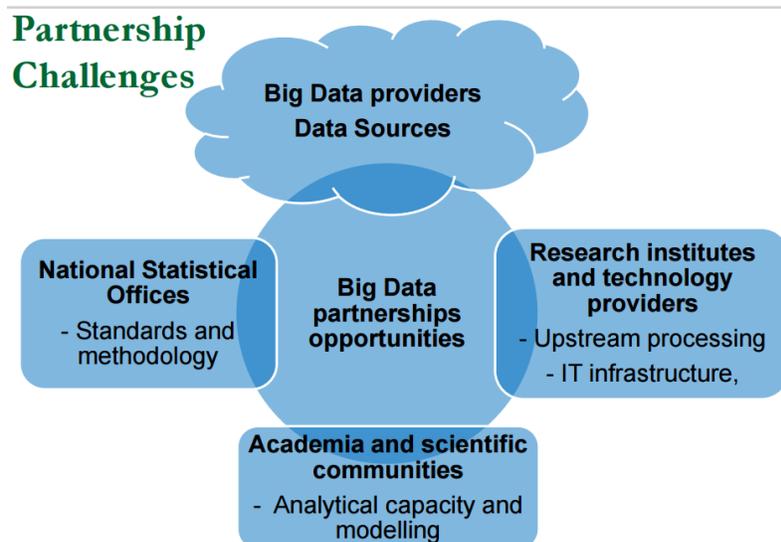
Characteristics of the information	Benefits for the analyst	Sources of information
Highly distributed	Data generated automatically	Mobile devices
Little structured	Timely data	Digital transactions
Great volume	Potentially relevant	Social networks

Source: Ronald Jansen, *Big Data and Official Statistics*, presented on June 16, 2014.

The specialist considers that this type of information is useful in a number of economic sectors, including the tourism industry. However, he also considers that there are major challenges to continue Big Data analysis. The challenges may be group together in the development of calculation methodologies, privacy and legality issues, human capital and associations' development for the analysis. Taking that as reference of the latter, Ronald Jansen considers that the highest booming has been present in suppliers and data sources, because they are the first agents that intervene in the generation of information chain process.

It is important to highlight that the internet searching engines major suppliers (Google, Bing, Yahoo) and the intermediaries of users such as social networks (Facebook, Twitter, Flickr) may be the departing point to generate groups that benefit Big Data analysis. Connections with the three agents that also interact within and outward the associations' chain from the generation of information may be generated. An associations' interaction scheme is presented as follows as Ronald Jansen mention.

Diagram 1. Development of associations for Big Data analysis.



Source: Ronald Jansen, *Big Data and Official Statistics*, presented on June 16, 2014.

- National Statistical Offices: They have knowledge on data validity standards and inference methodologies to be used with data with statistical significance.
- Research institutes and technology providers: They know processing for transmitting and spreading of information, as well as for using state-of-the-art infrastructure in information technologies.
- Academia and scientific communities: They offer the human capital necessary for generating models and a source of innovation.

Other sources of information. Other experiences in different countries are quoted in Spencer (2013):

- Israel. Registry of motion patterns between a group of tourists that voluntarily used a GPS localization system.
- Sweden. Use of a system that complements the making of surveys with GPS localization systems and air images to estimate the number of people who attend a specific event.

Lovelace (2014) states that geographic information from social networks sites have a great potential to improve the geographic models of behavior, especially if the volume of social networks geo-tagged messages continues to increase. However, he points out that the information should be used only as a complement of most consistent data sources or when a set of official data is not available.

It may be claimed that on-line information is not statistically representative, that refers to a small minority of the population who use unusual accessories or digital devices. The geo-referenced tweets used by Lovelace (2014), for example, constitute around 2% of the production of information in Twitter but it may not be assumed that it represents all the tweets.

Therefore, scholars have been cautious in this type of matters: This is used in models with voluntary geographic information and especially when data use to explain irregular and complex travel patterns.

This author uses a space interaction model of visitors to 15 museums of 179 administrative zones in the United Kingdom, to explain the behavior of the flows to the museum. The behavior prediction models, according to Lovelace (2014) are:

- i) multi-purpose travel: travel chain, for example, the home-school-work-store-work-gym-bar-house travel chain, it also may be represented as a simple travel from home to work;
- ii) seasonal travel, the weekly and day variation at the moment and travel frequency; and
- iii) information updated by user, adoption of smartphones, for example, traffic information in real time of Google or directly to navigation systems in the car by satellite.

## 5. Twitter as source of information.

One of the social networks with more penetration in the world is Twitter, which allows sending short text messages (with a maximum of 140 characters), that are shown in the main page of the user. Users may also group messages in the same theme by using words or phrases that begin with “#” (hash) known as “hashtag”. Additionally, Twitter have a number of applications that provide entertainment and information.

Using this tool provides a number of benefits for the analyst: low costs for more volume of information; quick and easy access to a considerable number of users; and contextual information, that is, the geographic region from where data were extracted is considered.

### 5.1 User’s activity in Twitter.

If social networks are considered as a system where users share opinions and experiences, the network effect engages in managing and conveying messages towards a great mass of information and a possible behavioral response is especially produced. In this case, users participate in the social spaces on-line in order to mainly obtain three benefits: i) acquiring information; ii) identifying or belonging to a community; and iii) fun and entertainment.

As a result, the interaction of tourists in their social networks may be a complement of an analysis in order to learn about the average behavior in some physical space such as a travel destination where different types of fun and entertainment may be offered. Therefore, the objective of the Twitter analysis may be focused on learning the movements of a profile of users of social networks representative of the average tourist having a twitter active account in use with access through a mobile device.

### 5.2 Twitter User’s activity in Mexico.

In Mexico, 95% of Twitter users’ age range between 18 and 49.<sup>8</sup> This percentage displays a tourist’ profile at city and beach destinations, where 87% of tourists range between 18 and 55 years of age.<sup>9</sup> Therefore, a higher percentage of tourists are potential users of this social network. When analyzing in more detail the profile of social networks users with respect to the tourist profile in Mexico, both may be linked and chained and therefore, information that is more reliable may be generated when comparing their behavior.

Additionally, the socioeconomic status of Twitter users is mainly medium and high. Consequently, the deciles of homes with higher income tend to allot a higher proportion of their income on leisure. To this regard, the making of a trip to go on vacation and making purchases takes the higher expense made by the tourist, followed by personal expenses.<sup>10</sup>

Finally, it is worth mentioning that in Mexico the use of the internet and social networks is on the rise. The number of internet users accessing to social networks increased to 15

---

<sup>8</sup> Consulta Mitofsky (2011), p.6.

<sup>9</sup> SECTUR (2010), p.7.

<sup>10</sup> SECTUR, Expense in Travel per Home Survey, México, 2011, p.30.

million in 2013. The penetration of the internet is approximately 46 million of nationwide. Consequently, the growth at using these tools in Mexico is extraordinary, as well as its potential growth in the years to come.

### 5.3 Productive use of Twitter information in the Tourism Sector.

Today, the using of Information and Communication Technologies (ICT's), as well as new estimating methods acquires the greatest relevance for generating statistics in the tourism sector. To this regard, spreading the use of social networks in the tourism sector opens a window of opportunities to make up for the lack of administrative registries and to provide information for the decision-making process in the sector.

### 5.4 Study Case: Twitter and domestic flows in Puebla and Guanajuato.

In response to the benefits and advantages described about the possibility of productively using information from Twitter, the Ministry of Tourism in collaboration with INEGI<sup>11</sup> performed the following exercise to obtain a view of the origin of state domestic tourists in Puebla and Guanajuato.

Out of a collection of 60 million of twits recorded from the last week of January, 2014 through the first week of July, 2014,<sup>12</sup> geo-referenced twits were selected especially those coming from within Mexico. Then, a geographic selection was made in the polygons corresponding to the states of Puebla and Guanajuato. After that, a temporary selection was made for days 1, 2 and 3 of February 2014, considered as a long weekend or commonly referred in Mexico as a "holiday period".<sup>13</sup>

Twits were ordered consecutively by Twitter user, eliminating duplicates in order to obtain a list of Twitter users with activity in Puebla and Guanajuato during that long weekend. In total, 7,955 Twitter users were obtained who twittered in the states of Guanajuato (48%) and Puebla (52%) during the long weekend.<sup>14</sup> Then, from the original database of 60 million of twits, twits were identified especially related to the 7955 Twitter users with activity in Puebla and Guanajuato during that long weekend, in order to identify from what state they were issuing twits during the entire semester, totaling 827,424 twits.

These 827,424 twits were ordered by Twitter user and chronologically to identify each change of state by Twitter user, and to determine in this way the change sequences with destination in Puebla or Guanajuato where the stay at these destinations was not longer than 15 days. From this result, the visits by each state of origin were counted. After that,

---

<sup>11</sup> Results from a preliminary analysis of a sample of tweets within the context of a Big Data concept test.

<sup>12</sup> It is a sample of tweets generated through geo-referencing within Mexico, collected from Twitter from the last week of January and the first week of July 2014.

<sup>13</sup> According to the Decree dated August 23, 2013 through which the Official Calendar in Mexico is established: the first Monday of February is observed to commemorate the 5 of February, anniversary of the Battle of Puebla. Official Federal Gazette.

<sup>14</sup> A longer period may be considered in a second exercise, as summer season.

the results were mapped (Map 1 and Map 2) and a comparison was made with the statistics generated by the corresponding travel observatories of Puebla and Guanajuato.

In Map 1 and Map 2, the color tends to red when there is a lower number of tourists coming from such state and tends to green when there is a higher number of tourists coming from such state.<sup>15</sup> In this case, the states of origin were also order progressively by the number of Twitter users (not staying longer than 15 days).

When making a hierarchical comparison of the first 10 states tourists come from visiting Puebla and Guanajuato (considering the twits exercise) and the first 10 states tourist come from (considering the statistics from travel observatories); the following is found:

- In Puebla, the 10 main states domestic tourists come from total 95% of the twits analyzed during the 1, 2 and 3 of February 2014; corresponding to the trend shown by the Puebla Tourism Observatory in 9 of 10 states.<sup>16</sup>
- In Guanajuato, the 10 main states domestic tourists come from total 96% of the twits analyzed during the 1, 2 and 3 of February of 2014; corresponding to the trend shown by the Puebla Tourism Observatory in 9 of 10 states.<sup>17</sup>

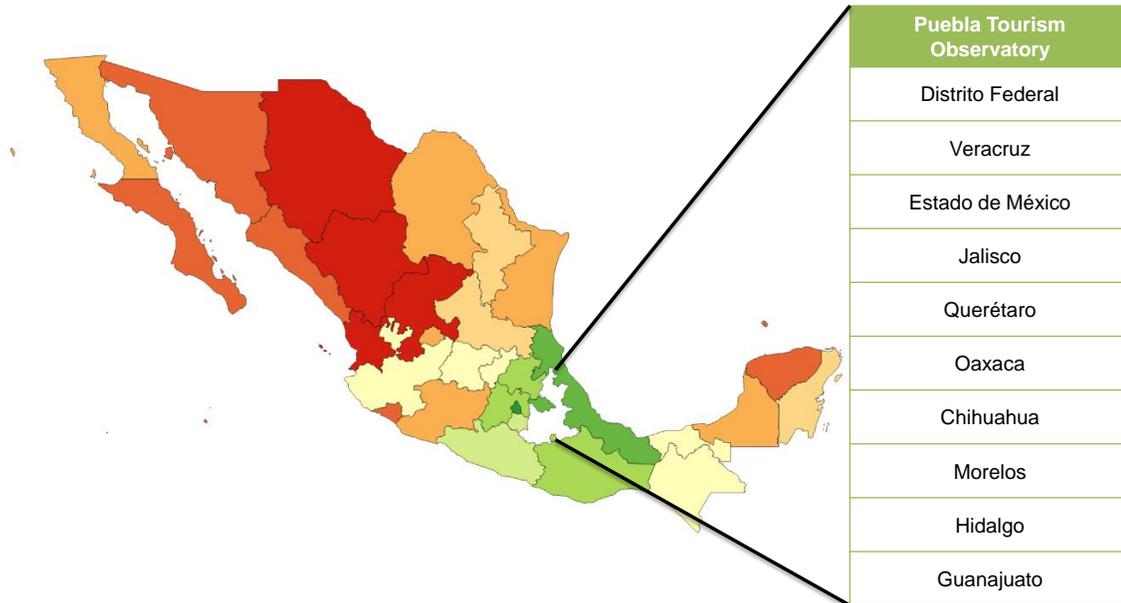
---

<sup>15</sup> The segmentation used for the tweets distribution was by percentiles and the observations were stratified in 9 strata.

<sup>16</sup> The correlation coefficient in Puebla, considering the percentage composition of each state of origin in the tweets exercise and the percentage of each state provided by the Tourism Observatory is 0.71.

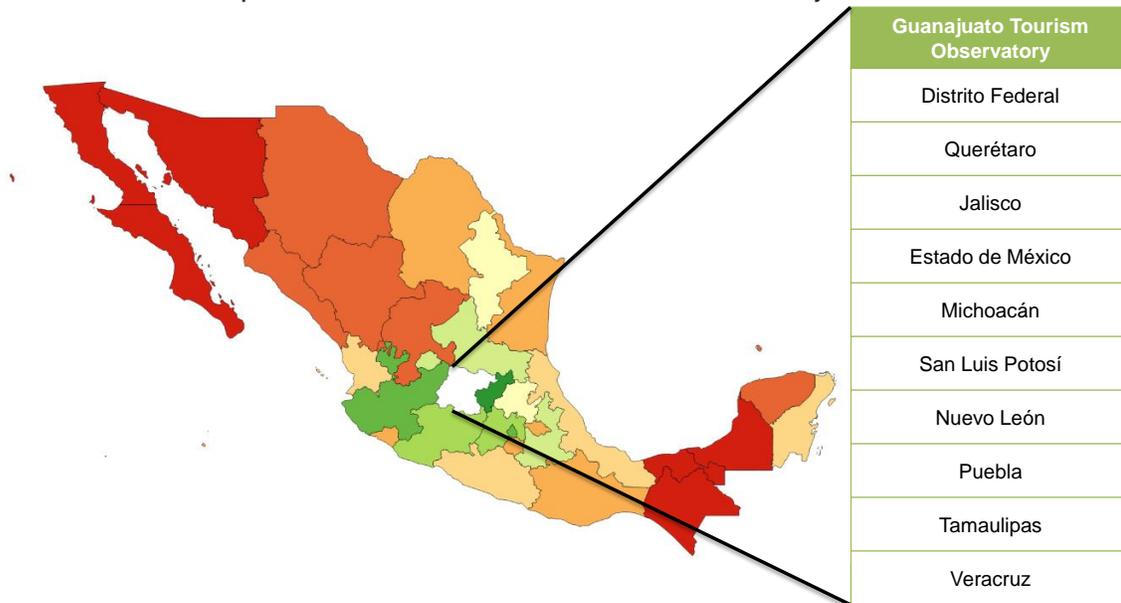
<sup>17</sup> The correlation coefficient in Guanajuato, considering the percentage composition of the tweets exercise for each state of origin and the percentage of each state of the Tourism Observatory is 0.82.

Map 1. Twitter and domestic flows in Puebla.



Source: Made by SECTUR and INEGI based on Twitter information and information from the Puebla Tourism Observatory.

Map 2. Twitter and domestic flows in Guanajuato.



Source: Made by SECTUR e INEGI based on Twitter information and information from the Guanajuato Tourism Observatory.

For example, in Map 1 the states in dark green (Distrito Federal, Veracruz and Tlaxcala) correspond to the states with a higher number of tourists that visited Puebla during the observation period; while the states in dark red (Coahuila, Durango and Zacatecas) correspond to the states with a lower number of tourists that visited Puebla.

The same occurs in Map 2: Querétaro, Jalisco and the Distrito Federal (dark green) correspond to the states with a higher number of visitors in Guanajuato; and Baja California, Baja California Sur and Sonora (dark red) are the states with a lower number of visitors in Guanajuato during the observation period.

The results may be clarified by identifying the access routes to both Puebla and Guanajuato. There may be natural barriers (orography) that prevent tourists from going to another state and vice versa. Furthermore, there may be different infrastructure development for better moving at some states in comparison to others, for example, air connectivity.

First of all, considering the orography of Mexico, both Guanajuato and Puebla are on the neovolcanic axis which may represent a natural barrier for tourists that travel by car, due to the cost to access and possibly because it takes longer to arrive to the destination. Consequently, the states with a higher number of tourists in both cases correspond to the nearby states.

Secondly, considering the air connectivity of Mexico, it is noticed that large cities have centralized routes, and this may represent a difference in infrastructure to get to Puebla or Guanajuato from far states located in the north or southeast of Mexico. However, Tijuana has an airport domestically important connection wise and it is noticed that Baja California is not a state having a lower number of visitors to Puebla. Therefore, the results shall not be generalized and a more precise review is required for each case.

For practical purposes, it is noticed that the travel influence zone in Puebla and Guanajuato is limited to the neighboring states, in green in each map. This may be explained because a three-day period took place, restricting making travel over longer distances. It is possible to increase the timeframe selected in vacation periods of longer duration and this will possibly decrease the prediction error, and therefore, it will increase the correlation coefficient of the hierarchical order between the tweets exercise and the Tourism Observatory of each state.

Additionally, it is important to do a similarity analysis among domestic flows for states located in the same region. On the one hand, if we have similar flows, projects referred in searches of tourist corridors between the states may be carried out. On the other hand, if we have opposite flows, there is an opportunity to promote similar product lines in each corresponding state and to improve the volume of tourists' flow diversifying their offer to other states that are not currently the origin of tourists.

To this regard, the exercise may be replicated in each states of Mexico and see with more detail the access routes for any and all of the states; in order to improve the conditions referred by the hypothesis of natural barriers and infrastructure domestic tourists face. This may even be useful to identify better access routes and to improve the infrastructure conditions in some cases.

The previous exercise may help to improve the understanding and follow-up of the domestic tourists flow and it is also useful as an alternative to the high costs of performing national

surveys. Bear in mind that the information that stems from Twitter has itself no cost, the cost comes from of the extraction, collection and safekeeping of it. In conclusion, the productive use of Twitter in tourism may provide analysts with the following benefits: i) low costs given the great volume of information; ii) quick access and in real time of a considerable number of users; e iii) information contextual, considering the geographic region of the extraction of the data.

## 6. Conclusions.

In a context of a lack of financial resources and seeking to generate information to improve the decision-making process in the public, private and civil society sectors, it is necessary to use sources of information susceptible to contribute with a set of facts in order to design tourism-oriented public policies.

This taking into account the use of appropriate methodologies and algorithms to manage major volumes of information, as well as new graphic and statistical methods to make known the findings and validation of the results obtained. In conclusion, the review of the productive use of information both from Big Data such and social networks constitute a process that demands having effective institutions where there is a close cooperation among governments and private organizations.

In the Twitter study case; there are opportunity windows to find more correlations between the 10 main states domestic tourists come from heading to both Puebla and Guanajuato, with the trend shown by state travel observatories. For example, more follow-up days of Twitter users and take also into account information that is more recent from observatories.

In general, this paper is the first step for Big Data and social networks analysis of tourism in Mexico. Methodology may be improved in the future by integrating of new information access tools and with the possible contribution of observations, opinions and suggestions by participants that wish to engage in this type of analysis. In conclusion, taking advantage of the development of Big Data analysis by the tourism sector may be a prolific.

References.

- [1] Castro, Carlos (2014), "*Datos para el Desarrollo. Desastres Naturales Tabasco 2009*". Presentación en el Seminario Internacional: Big-Data para la Información Oficial y la Toma de Decisiones. INEGI. CIDE. INFOTEC, 16-17 junio 2014.
- [2] Consulta Mitofsky (2011), "*Perfil de usuarios de redes sociales en Internet (Facebook y Twitter)*", Diciembre 2011.
- [3] Elsevier (2014), "*Open Government, Open Data and Digital Government*", Editorial, Government Information Quarterly, No. 31, 4-5, 2014.
- [4] Hal Varian & Hyunyoung (2009), "*Predicting the Present with Google Trends*", Google Inc.
- [5] Heerschap, Nico; Ortega, Shirley; Priem, Alex and Orrermans (2014), "*Innovation of tourism statistics through the use of new big data resources*", The Hague, The Netherlands, 27 March 2014.
- [6] Liran Einav & Jonathan D. Levin (2013), "*The Data Revolution and Economic Analysis*", NBER Working Paper No. 19035.
- [7] Lovelace, Robin; Nick Malleon, Kirk Harland and Mark Birkin (2014), "*Geotagged tweets to inform a spatial interaction model: a case study of museums*", School of Geography, University of Leeds, Leeds, UK.
- [8] Oliver, Víctor; García Enrique y otros (2014), "*Big Data y Turismo: Nuevos Indicadores para la Gestión Turística*", Barcelona, Telefónica y RocaSalvatalla.
- [9] Organisation for Economic Co-operation and Development (2012), "*OECD Workshop on e-government indicators*", Public Governance and Territorial Development Directorate, Paris, December 2012.
- [10] Pablo Barberá (2013), "*NYU Politics Data Lab Workshop: Scraping Twitter and Web Data Using R*", Department of Politics, New York University.
- [11] Ronald Jansen (2014) "*Big Data and Official Statistics*". Presentación en el Seminario Internacional: Big-Data para la Información Oficial y la Toma de Decisiones. INEGI. CIDE. INFOTEC. 16-17 junio 2014.
- [12] Secretaría de Turismo (2010), "*Perfil y Grado de Satisfacción del Turista*", CESTUR.
- [13] Secretaría de Turismo (2011), "*Encuesta del Gasto en Turismo en Hogares*".
- [14] Secretaría de Turismo del Gobierno del Estado de Guanajuato (2013). Perfil del visitante al Estado de Guanajuato 2013. Disponible en: <http://www.observatorioturistico.org/>
- [15] Secretaría de Turismo del Gobierno del Estado de Puebla (2013). Perfil del visitante, Diciembre 2013. Disponible en: <http://www.observatoriopuebla.org>.
- [16] Sencer A. Wood, Anne D. Guerry, Jessica M. Silver & Martin Lacayo (2013), "*Using social media to quantify nature-based tourism and recreation*". Scientific Reports. Stanford, CA, USA.
- [17] Tilly, Roman; Fischbach, Kai; and Schoder, Detlef (2013), "*Assessing The Potential Of Social Media To Reflect Global Tourism*" ECIS, Research in Progress. Paper 17.
- [18] Ubaldi, Barbara (2013), "*Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*", OECD Working Papers on Public Governance, No. 22, OECD, Publishing.
- [19] Wu, Lynn & Erik Brynjolfsson (2013), "*The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales*". The National Bureau of Economic Research. May 2014.

---

Research on Statistics & Economics of Tourism provide preliminary research results. Said studies were carried out in order to generate an exchange and debate of ideas on the development of the tourism sector. The content, as well as the arguments expressed, are of the sole responsibility of their authors and do not necessarily reflect the point of view of the Ministry of Tourism.

---

La calidad de la traducción y su coherencia con el idioma del texto original son exclusiva responsabilidad del traductor. En caso de cualquier discrepancia entre ambos documentos el texto del documento original prevalecerá: 2014-1 Uso Productivo de Big Data y Redes Sociales en el Sector Turismo.