

145 Экономика впечатлений:  
менеджмент в индустрии гостеприимства и туризме

## 1. Задача 1

Полный балл 100

Прочитайте статью и сделайте её критический анализ на русском языке.

Статья подготовлена на основе: Giglio S., Bertacchini F., Bilotta E., Pantano P. Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*. 2019. Vol. 72.

### 1. Introduction

Nowadays the technological revolution, through the increased use of digital devices, has allowed people to generate tremendous amounts of much heterogeneous data called Big Data, a genuine revolution where by, through increasingly innovative technologies, the stakeholders' attention is focused on consumer needs (*Che et al, 2013*). Big Data refers to voluminous amount of data characterized by large volume, variety and velocity requiring new processes to enable enhanced decision making, insight discovery and process optimization (*Siddiqa et al, 2016*). In the tourist context, the focus is on management of tourist data and marketing strategy and the ability to directly reach users through social media allows creating new opportunities for service providers (*Pantano et al, 2017*). The importance of social networks and of the information they generate has helped to change the user's approach to tourism services and products. Users post and share various kinds of data on the Web, for instance the images posted on Flickr provide information on space and time (i.e. when and where tourist took a certain picture). Such data contributes to a common knowledge about places visited and people's behaviours, information that could be very important for local administrations, travel agencies and other organizations.

The images shared online provide evidence of travels or vacations and data produced and registered have helped researchers create and define new methods of observation and analysis of a place to understand tourism dynamics. In particular, sharing data have allowed creating a bridge between real and digital worlds and have offered possibilities to know users' preferences and behaviour (*Bertacchini et al, 2018*). Images posted on the web by users, especially by tourists, reflect people interests and preferred activities while on vacation. In fact, geo-tagged images, available online, reveal where the user has been and, if the user has uploaded more than one image, suggest the sequence of locations visited (*Mirkovic et al, 2011*). They indicate landmark identification, in particular the trajectories of users since incorporate information on latitudes and longitude coordinates (*Sun et al, 2015*).

Social platforms like Flickr, Facebook, Twitter and other services of geo-location are prominent providers of geographic information. Images and metadata can be stored on these platforms representing a rich source of users-generated content, used by data

mining scientists. Images provide information on different types of user activities, places, landmarks, and the ways people are attracted to them. To uncover the relationship that users have with tourist destinations, we have analysed specific data to identify certain area of interest and its cultural context. First, we consider geo-referenced images of six Italian cities available on Flickr, an online platform where users share and embed personal photographs with more than 87 million registered members and more than 3.5 million new images uploaded daily ([www.flickr.com](http://www.flickr.com)).

Flickr is an online photo library, through which a user provides spatial (latitudes and longitude coordinates) and temporal (date and time of day) information in the form of coordinates. Users of this social platform connect with each other by becoming friends and followers and interact by favouriting and commenting on photos (*Hollenstein, Purves, 2010; Gopalakrishnan et al, 2016*). Data is downloaded through the Flickr Application Programming interfaces (*Flickr APIs – <http://www.Flickr.com/services/api/>*), a set of procedures that allow modelling operations, useful for a variety of computer applications (*Banta et al, 2006*). Finally, according to the number of downloaded images, we determined the relationship between images' location and content in order to study the attractiveness of various places and recognise the places visited and points of interests (POIs) that are identified by spatiotemporal analysis. This study confirms the importance of social media data as a benefit in analysing users' behaviour. Findings contribute to strengthen a general understanding of behaviours, interests and experiences of tourists and offer a methodology for the analysis and conversion of Big Data into meaningful information to support decision-making.

## 2. Methodology

### 2.1. Procedure

We analysed the data using a specific software, Mathematica<sup>®</sup> (*Wolfram, 2016*), a tool for computing, simulation and mathematical modelling. The Mathematica language employs several methods and tools to reduce the complexity of data and provides a number of state-of-the-art methods that allows to carry out a series of operations automatically such as the analysis of tourists' reviews (*Pantano et al, 2017*), the identification of customer-robot interaction (*Bertacchini et al, 2017*) and the classification of cellular automaton (*Diaz, 2017*). We based our analysis on Machine Learning (ML) models approach (*Zhou et al, 2017*). ML consist of the study and modelling of algorithms that can learn from the data, extrapolating from them new knowledge. Thanks to the technologies progress and computer engineering it is possible to bring the analytical capacity of the machine closer to the human capabilities. Through ML algorithms, we can extract patterns, build hidden predictive models to traditional algorithms and really discover unknown values in Big Data.

In data analysis, we have adopted a specific type of ML for image recognition process. Given a set of data composed by a series of thumbnails downloaded from Flickr, the machine can distinguish the subject of images, with the module *ImageIdentify*. The proposed approach is used to identifying what is the object of the image. As we have

already said, the identification method refers to artificial neural networks. The machine is been trained using millions of images supplied as examples from which it progressively learns to make distinctions. This function operates in a probabilistic way, generating multiple possible results with different estimated probabilities. In addition to automatic image recognition, another ML technique was adopted in the data analysis process: clustering techniques. Cluster analysis is a multivariate data analysis technique that allows the selection and grouping of homogeneous elements in a set of data (*Zhou et al, 2017*). Clustering applications include object recognition, gene sequence analysis and market research. Given a set of data, in this case geographic coordinates, the machine can generate clusters automatically, with the module *FindCluster*. This function classifies similar items with unsupervised machine learning, where the machine compares data without receiving any information about it.

## **2.2. Data collection**

Using public Flickr APIs, we have collected 26.392 images shared by 4.205 users. Dataset include three years of data, from January 2014 to December 2016, in six Italian cities: Milan, Venice, Florence, Rome, Naples and Palermo, popular and highly photographed tourist destinations. We choose these cities due to their cultural impact and according to the classification published by the Italian National Institute of Statistics (ISTAT). ISTAT in the 2016 annual official statistics shows on tourist flows and their features in Italy. In 2016 arrivals in Italy were 117 million (+ 3.1%) of tourist compared to the previous year and these six cities are been tourist destinations with the main the activity of travelling made by visitors. The six cities chosen have full in historic building, churches, castles, museums and are the ideal tourist destinations. All of them are art cities and preserve traces of the past. They marked by the activity of great artists and themselves are works of art. Our goal is to verify the attractiveness of some places based on images taken by foreign tourists and tourists on site.

This information will allow us to (i) identify the places they visited, (ii) determine the city and country of origin of users, (iii) discover how many foreign tourists and how many domestic tourists arrive at an area, and (iv) show places and POIs through space-time analysis. Our research group implemented a platform able to run a query using Flickr APIs to collect data. Users during their travels use GPS devices that allow to automatically record geographic information of photos. During the first step, we downloaded photos with metadata from Flickr website using an open APIs. For each of these publicly available images, we retrieved the following metadata: Location, Location Id, Latitude and Longitude, Username, User ID, Date-time, Link of image. The same process was conducted for each city and the research was made on a 5 Km radius from the city center. Once we've defined the six Italian cities, we started the analysis ranking the total number of photographs taken by tourists from all over the world. The number of photographs taken in each city differs among the three years period.

Some cities have witnessed an increase in photographic activity in terms of tourist activity (Milan 5%, Naples 11%, Florence 43% and Palermo 25% respectively), while a decrease in photographic activity was recorded for Rome and Venice (5.3% and

16.3% respectively). Results show that the most photographed city is Rome with 6.988 photographs, followed by Milan with 5.596 photos. Using Mathematica<sup>®</sup>, we have identified users' home town. In order to identify the dynamics of tourism, we examined data in depth. We have included image recognition (image classification) and cluster analysis. In the first analysis to have an image classifier, we use a recognition algorithm takes an image and offers as a result the contents of image. This is a non-supervised machine learning technique where you need to train an image recognition algorithm with thousands of images.

The result is a class label. In our analysis, the algorithm can distinguish the subject of the photo with the module *ImageIdentify*. The classification runs through a software-integrated classifier and characterized by a large set of images tagged with the concepts of Mathematica language. We run a subsequent cluster analysis to identify landmark. Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in data. Through module *FindCluster* geo-located data were clustered and different clusters were generated (Bertacchini et al, 2017). Clustering method is applied for the geo-tagged images of tourists to identify the POI of cities.

### 3. Results

#### 3.1. Users incoming tourists

Our dataset includes user profile, through Mathematica software, we created a connection to Flickr API with a query, ServiceConnect ["Flickr"], and we identified users' home town and country of origin. The aim was to discover which places are attractive for Italian tourists and which places are for international ones. After determining the source of users, we have clustered them by their nationality. We have labelled users based on their home country, for example, a Madrid user and a Barcelona one are both under the Spain label. Findings show that for more than 75% the users' home town is identifiable. The rest of users are defined as unavailable. Some users have indicated as their home a fake city or imaginative place of origin ("Fantasy", "On a quest to discover the world"), that have been assembled to the unavailable data. The results show that Italian visitors are more than foreigners in all cities. However, a meaningful number is represented by visitors coming from United Kingdom, United States and Germany, especially in Rome and Venice.

#### 3.2. Analysis of pictures

After calculating the number of images taken by users for each city and year, we have evaluated what is being photographed in a certain place. First, we have selected the images taken in each city. Secondly, Mathematica Machine Learning allowed identifying images' contents through the command *ImageIdentify* (Figure 1). Finally, all the images have been classified into categories based on the content. The data process was repeated automatically for each city. Results show the extent to which in each city the most photographed elements are people, followed by houses (including building) and cultural assets such as towers and arches (Figure 2).



Figure 1. Parts of codes and output for *ImageIdentify* module.

| roma            |     |  |
|-----------------|-----|--|
| person          | 783 |  |
| tower           | 56  |  |
| arch            | 50  |  |
| house           | 42  |  |
| building        | 33  |  |
| ballroom        | 32  |  |
| church          | 30  |  |
| road            | 25  |  |
| store           | 21  |  |
| Tesla coil      | 17  |  |
| person          | 265 |  |
| arch            | 29  |  |
| house           | 27  |  |
| road            | 24  |  |
| altarpiece      | 22  |  |
| coast           | 15  |  |
| store           | 12  |  |
| hotel           | 12  |  |
| building        | 12  |  |
| tower           | 9   |  |
| person          | 81  |  |
| house           | 47  |  |
| church          | 30  |  |
| altarpiece      | 30  |  |
| arch            | 28  |  |
| building        | 17  |  |
| road            | 10  |  |
| woody plant     | 10  |  |
| tower           | 8   |  |
| coniferous tree | 8   |  |

Figure 2. The ten most photographed elements between 2014 and 2016 in Rome.

### 3.3. Analysis of visited location and point of interest (POI)

The places that users prefer photographing are of key importance to understand what drives people to photograph an item in a specific location. We started the work analysing Location ID, metadata included in our initial dataset. With Mathematica software we identified all places visited by users considering the number of times they appear in our workbook. For example, in Naples, the fifth most photographed place is Vasto, a neighbourhood where there is a large number of churches erected in the early

1900s. In Palermo, the second most photographed place is Guadagna, where there is a shopping center. Of cultural nature are the places photographed by users in the cities of Florence, Rome and Venice. For example, the Giardino di Boboli in Florence is an historic city park considered as an outdoor museum. Campo Marzio in Rome is a district of the capital where many historical buildings can be found. In Venice, Piazza San Marco is one of the most important Italian squares, known all over the world for its architectural beauty (*Figure 3*).

|   |   |
|---|---|
| <b>Firenze</b><br>Firenze<br>Giardino di Boboli - Firenze<br>San Frediano - Firenze<br>San Niccolo - Firenze<br>Santo Spirito - Firenze             | <b>Palermo</b><br>Fiera Dei Mediterraneo - Palermo<br>Guadagna - Palermo<br>Leoni - Palermo<br>Palermo<br>Perpignano - Palermo<br>Sampolo - Palermo<br>Santicelli - Palermo |
| <b>Milano</b><br>Boffalora - Milano<br>Milano<br>Moncucco - Milano<br>Oggiono - Milano<br>Segnanino - Milano<br>Vigentino - Milano                  | <b>Roma</b><br>Campo Marzio - Roma<br>Flamino - Roma<br>Gianicolo - Roma<br>Rione di Trastevere - Roma<br>Roma<br>San Paolo - Roma  |
| <b>Napoli</b><br>Napoli<br>Porto di Santa Lucia - Napoli<br>Rione I.n.c.i.s. - Napoli<br>S.ferdinando - Napoli<br>Vasto - Napoli<br>Vomero - Napoli | <b>Venezia</b><br>Castello - Venezia<br>San Marco - Venezia<br>San Polo - Venezia<br>Venezia  |

*Figure 3.* Analysis of places photographed by users 2014-2016 in six Italian cities.

In order to investigate attractiveness of cities, we also identified the distribution of users in attractive areas. The attractions within cities are not homogeneous but there are areas considered “hotspots” and then less attractive areas. These places called “hotspots” represent the key places of the city's identity. In our work, we consider “hotspot” the regions with a large number of clusters. Thanks to the geographic coordinates of a sequence of photos downloaded by Flickr, we have identified spatial distribution of pictures taken by users to investigate attractiveness of each city. Geographic data has been clustered through the use of Mathematica Machine Learning and run the



command *FindCluster*. We have created maps for all six cities and we plotted users' location points to define POIs (*Figure 4*).

|   |                    |
|---|--------------------|
| Museo Galileo. Istituto e museo di storia della scienza | Museo              |
| Piazza della Signoria                                   | Punto di Interesse |
| Loggia dei Lanzi  | Monumento          |
| Loggia del Porcellino - Mercato Nuovo                   | Monumento          |
| Museo di Palazzo Vecchio                                | Edificio storico   |
| Corridoio Vasariano                                     | Museo              |
| Gli Uffizi  | Museo              |
| Gucci museo   | Museo              |



*Figure 4.* List of cultural POIs, cluster & POIs in an area of Florence.

Below *Table 1* that summarizes the different tools used in this work, the interactions among the latter and the analysis and results obtained.

Table 1. Tools used, the interactions between them and the analysis and results obtained

| Tools used and their description  | Interactions between them   | Analyses  | Results and recommendations  |
|---|---|---|--|
| <i>ServiceConnect</i> ["Flickr"] creates a connection to the Flickr API. This function allows to collect information from Flickr            | Through User ID the software searches user's home country write on his Flickr account | The function identifies users nationality. The users are labelled based on their home country in order to create a classification   | Findings show which places are attractive for Italian tourists and which places are for international ones. The recommendation is to identify the tourist cultural background to devise a strategic plan that improves services and quality of hospitality   |
| <i>ImageIdentify</i> identifies the objects included in the pictures. It is a Machine Learning algorithm implemented on Wolfram Mathematica | Cities pictures are import into software and processed by function                    | The analysis consists on the image identification of the objects included in the pictures. The function classifies the objects in a certain category  | Results show the extent to which in each city some objects are photographed. The most photographed elements are people. Finding shows the importance of user-generated content in terms of pictures useful to investigate on tourists attention and places influence. Pictures shared online could be related to positive or negative experiences and can provide insight for decisions making |
| <i>FindCluster</i> classifies data into clusters of similar elements. It is a Machine Learning algorithm implemented on Wolfram Mathematica | Geographic coordinates are import into software                                       | During cluster analysis geographic coordinates are processed and grouped in different cluster. These cluster are plotted on city maps. Through the map the places and their attractiveness are identified | Findings show which are POIs in the Italian cities. This analysis generates benefits to the territory. Indeed using these technologies stakeholders have the opportunity to find information on tourists cluster and can to provide them knowledges about history and culture of place visited   |

#### 4. Discussion and conclusion

The present work analysing Big Data extracted from social media and highlight the importance that digital data represented to understand the dynamics of tourism. Our system of analysis reveals clear benefits in being able to quickly identify the tourist dynamics linked to the economic and cultural environment of a city, examining on real-time data extracted from the social media. Furthermore, allows to provide functional tools to move inside the city. Gathering images posted on Flickr, we created maps of



the behaviour and of the attractiveness for every city and POIs visited by tourists. Through the analysis of users, we have identified the annual trend of photographic activity in the cities. We have identified city and country users to find out which places are attractive to Italians and which places for foreigners.

Furthermore, recognise objects of photo was essential to understanding what the user likes to photograph and investigate which are POIs in the cities generate benefits for the territory. In this work, attention was focused on six Italian cities: Florence, Milan, Naples, Palermo, Rome and Venice with important cultural and tourist relevance, highlighting the importance of online photo sharing to investigate about tourist behaviour. Several different researchers have proven that every aspect of the tourist experience is full of psychological mechanisms (*Larsen, 2007*). During the trip, users seek feelings and emotions, to appreciate and experience new experiences. They create social relationships, which are the product of the identity and attitude of the tourist immersed within an organization. Resuming a concept expressed by *Pearce (1982)*, a tourist turns into a natural ambassador of human-lived experiences that can be amplified and made known through social media. Based on this knowledge, our study of a user's behaviour acquires, in contemporary research, a privileged position to propose suitable marketing strategies to promote a site.

Therefore, interested parties could take this analysis of social data freely and rapidly accessible online, to know the current state of a particular tourist destination and to improve it. Social media data can be a significant component for interactions between users and service providers (*Uşaklı et al, 2017*). Data sharing platforms represent socio-cultural and economic systems that fit business models to adapt to technological diversity and different types of users (*Jeacle, Carter, 2011; Scott, Orlikowski, 2012*). Through appropriate business models, the tourism industry can influence the choices of potential tourists while searching for information on a trip. Some tourists want to relax during their travels while others want to have fun and enjoy an unprecedented experience. Significant and attractive images, persuasive language and fascinating content, can help attract the attention of the users and get visitors (*Chen, Hsu, 2000*).

The use of technologies and the internet can allow tourism destinations to improve their competitiveness (*Buhalis, 2000*). Providing tools such as our can to help improve the efficiency of a destination and its development. An important contribution of our research is that through this analysis can to identify the dynamics associated with a tourist site and potentially influence the future destination choice. Using the methodology that is adapted in this work for data analysis our future purpose is to realize predictive models in order to formulate new tourism-cultural scenarios and provide customized services for individuals (both international and national tourists, and inhabitants) and public administration.

**Вопросы для размышления**

1) При подготовке данного задания была использована публикация типа Research note, отличие которой от Research Article заключается, в том числе, и в краткости представления материала. Это касается и литературного обзора – раздела, традиционно присутствующего в Research Article. Представьте, что на основании проведённого исследования Вы сразу готовите публикацию для подачи в разные журналы (последовательно – в случае, если в одном будет отказ) в двух возможных форматах – полнотекстовой статьи и Research note. Изложите своё видение структуры раздела Literature review публикации типа Research Article для данного исследования.

2) Обозначьте ограничения проведённого исследования.

3) Какие из приведенных в статье выводов по результатам исследования представляются Вам спорными, недостаточно обоснованными? Аргументируйте свой ответ.

4) Выдвинуты ли авторами гипотезы? Нашли ли они подтверждение?

5) Предложите своё исследование, которое можно провести, продолжая основную идею статьи. Сформулируйте цель и задачи соответствующего исследования. Предложите методологию, методы сбора и анализа данных своего исследования.