Athens Journal of Tourism - Volume 8, Issue 2, June 2021 – Pages 131-146

How to Measure the Touristic Competitiveness: A Mixed Mode Model Proposal

By Antonella Bianchino[*]*, Daniela Fusco*[±] *& Daniele Pisciottano*[+]

In 2018, according to the World Tourism Organization (WTO), the number of arrivals of international tourists worldwide reached 1.4 billion, which represents enormous potential for the sector and global economies. According to WTO, Italy is in the top ten of the countries with the greatest tourism competitiveness thanks above all to its natural and cultural resources. Today, in the globalized world, tourists are pressed by the opinion of travelers, the number of times that a location is mentioned and in which way influencers marketing consider it. The aim of this work is to create a composite indicator that allows us to evaluate the tourist competitiveness of Italian cities by evaluating both the data on the receptivity and the opinions of travelers. To do this, the official data of Istat have been taken together with Big Data, in particular information from the main holiday home platform and the opinions of travelers expressed on Twitter. Subjective and objective indicators have been produced. The results allow us to build a rating list of Italian touristic cities.

Keywords: *tourism, composite indicator, Big Data, Open Data, ranking*

Introduction

The year 2018 marked a new record for Italian tourism: 428.8 million customers in accommodations (Istat 2019a, b), up by two percentage points compared to the previous year.

Tourism revenues in Italy mainly come from nearby countries, whose residents have the advantage of less distance to travel and, in the case of states, which are part of the European Monetary Union, a common currency (Alivernini et al. 2014).

Traditionally, Germany is the main market for Italy's international tourism. The United States is the second country of origin in terms of tourism revenues for Italy and France is the third most important origin of Italy's incoming tourists.

Holidays are the purpose most frequently cited by international travelers. Nearly 60 percent of the overnight foreign travelers in Italy and half of the Italian overnight travelers abroad were on holiday trips.

Rome is the most popular city with 29 million visitors, followed by Venice and Milan (both with 12.1 million). The first city from the south of the country sits at the eleventh place in the ranking, and it is Naples with 3.7 million presences.

In the past, the choice of the destination of a trip depended by word of mouth, the offers proposed by the agencies, the photos in the catalogs. With the advent of

[*] Office Manager, Italian Institute of Statistics (Istat), Italy.
[±] Statistician, Italian Institute of Statistics (Istat), Italy.
[+] Sociologist, Federico II University, Italy.

https://doi.org/10.30958/ajt.8-2-4 doi=10.30958/ajt.8-2-4

the internet and social networks, travelers find more and more inspiration from influencers, who, by sharing their experiences, through a bottom-up participatory promotion approach, suggest destinations to the tourists.

For this reason, knowing people's opinions on a touristic destination provides qualitative information that drives travelers' choices. From this point of view, social networks have opened a new frontier: from the analysis of the posts, it is possible to understand, beyond the numbers, which are the destinations people like best and for what reasons.

One of the most popular social networks is Twitter. It contains some information as the number of comments, the number of retweets, the number of likes, that are useful to build indicators to evaluate which emotions could have stimulated a holiday destination, how many users have talked about it, where they came from and what influence they have on internal social network.

The aim of this research is the construction of a composite index using official data (objective information) and Big Data from social network (subjective data), in order to evaluate the competitiveness of touristic cities also from the point of view of the traveler. After the extraction of the information from Twitter, the Sentiment Analysis of the tweets was carried out to evaluate the opinions of users on cities. The other information was used to build indicators capable of determining the tweet's weight, its visibility and, where possible, its geographical origin.

The results make it possible to build a ranking of favorite touristic cities. To test the index, it was decided to build it on the main Italian touristic cities: Rome, Venice, Milan and Florence, as well as the city of Naples, the first city in the south of Italy for presences. The tweets of July, the month in which is recorded the peak of tourist presences in Italy, were analyzed.

Literature Review

In an increasingly demanding and competitive tourism market, it is necessary to apply new techniques of investigation in order to improve processes of governance and contribute to the sustainability of tourism destinations (Zerba 2018).

Using data from official sources, it is possible to investigate a phenomenon exclusively from a quantitative point of view, but for the purpose of a more accurate investigation, it is also necessary to take into account qualitative information.

In this context, it is important to introduce the use of Big Data for statistical purposes. These data provide potentially relevant complementary information compared to the standards, because they are based on quite diverse sets of information, they are also available in real time (or almost).

By Big Data we mean a collection of data so extensive in terms of volume, speed and variety that it requires specific analytical technologies and methods for the extraction of value (Curry 2016).

The transition from Big Data to data is possible thanks to the growth in power of today's systems and their computational capacity. More popular, easier, and faster connections, is key to this step. The three main characteristics of this data source are traditionally called the "3 V":

- Volume, a huge amount of data.
- Velocity, the speed with which the data becomes available is so high that it is necessary to use new tools.
- Variety, diversity of sources and formats due to their lack of structuring.

Taking into account the "reviews" of people help us to understand what can attract to a specific city and what are the related problems, giving an immediate feedback at a very low cost, especially taking into account how much this kind of survey, would cost if carried out with traditional methods.

Big Data seem very useful, but they bring with them new challenges starting from data access. Since the quality of the analysis depends on the quality of the information analyzed, one of the main challenges in the use of Big Data is to cope with the enhancement of the "care" of such data in order to simplify its usability (Freitas and Curry 2016). If in statistical surveys it is possible to control each phase of the collection, the use of big data makes it necessary metadata availability (Gozzo et al. 2020).

In this contest, artificial intelligence (AI) will beget social and economic changes far more profound than any other technological revolution in human history (CINI National Lab 2020). Italian community of researchers in AI, are ready to cooperate with the priorities defined by Italian institutions in terms of industrial needs and societal challenge. These aspects must take into account ethical values and ensure respect for human rights and democratic values, following OECD Principles on AI[1].

The AI experiences are various and different. The availability of a platform with data automatically retrieved from the Web sites of TripAdvisor and Google Maps, for example, could be useful for the integrated tourism, defined as the kind of tourism which is explicitly linked to the localities in which it takes place and, in practical terms, has clear connections with local resources, activities, products, production and service industries, and a participatory local community (Lisi and Esposito 2015). The availability of all this information in associated form would be very important for tourism studies. Unfortunately, these integrated platforms are still in an experimental phase and not available globally.

The analysis of the source from which the data comes is therefore essential. Unfortunately, the ownership of many Big Data is of private suppliers or connected to private aspects, so to evaluate the potentially useful sources of Big Data, it is necessary to evaluate the cost of the information, the presence of metadata that can provide additional information and at the same time avoid privacy violations.

Definitely, we cannot think of conducting a survey based on the mutual exclusivity of traditional source or the Big Data. It is important to find a

[1] http://www.oecd.org/going-digital/ai/principles/.

compromise that can lead to increasingly precise estimates by compensating for the failures of one source through another.

Numerous studies examined several ways of collecting twitter text messages, classified the topics discussed and looked at the usability of the information from an official statistics point of view, especially for social statistics studying the opinions, attitudes, and sentiments shared in social media could be interesting (Daas et al. 2012).

Twitter messages are publicly available, meaning that people that are not a member of the sender's network are able to read it, making it a very attractive source of information (Laniado and Mika 2010).

In this paper, it is shown how is possible to summarize, through a set of objective (via official and open source) and subjective indicators (via Twitter), the preferences towards the main touristic destinations.

In this experimental phase, it was decided to focus the attention on main Italian tourist destinations, according to the results of the Italian Institute of Statistic (Istat) survey: Rome, Venice, Milan, Florence and Naples.

For the evaluation of the cities, it was decided to use a composite index, called Competitive City Index for Travelers (CCIT).

The use of composite indices to analyze touristic phenomena is well established in the literature. The Word Trade Organization publishes every two years the Travel & Tourism (T&T) Competitiveness Index (WTO 2019).

It compares the T&T competitiveness of 140 economies and measures the set of factors and policies that push up the sustainable development of the T&T sector, which in turn contributes to a country's development and competitiveness. It consists of 4 sub-indices (Development Capacity, Policy, Infrastructure, Natural and cultural resources), 14 pillars and 90 indicators distributed in the different pillars.

Methodology

The aim of this work is to synthesize, which, among the main Italian tourist destinations, are the most chosen ones according to a set of objective and subjective indicators.

For what concerns the primary experimental phase, we just focused the attention on the most visited among Italian destinations, that is to say Rome, Venice, Milan, Florence and Naples in accordance with the results of the Italian Institute of statistics survey "Occupancy of tourist accommodation establishment".

For the cities' ranking, we decided to use, in a methodological point of view, a synthetic index, the Competitive City Index for Travelers (CCIT).

The sources used, the summary methodology chosen for the construction of the CCIT and the choice of indicators are specified and analyzed below.

The Official European Data

From a quantitative point of view, the National Statistical European Institutes, through the "Trips and holidays survey" and the "Occupancy of tourist accommodation establishment survey" (Regulation for Tourism Statistics 692/ 2011), provides numerous information with a city detail.

The "Occupancy of tourist accommodation establishment survey" survey is carried out every month. The monthly statistics are regularly processed by Istat since 1956 and represent the main source of information on domestic tourism available in Italy.

The survey quantifies, for each month and for each city, the arrivals and presences of customers (resident and non-residents) according to the category of establishment and the type of structure and according to the foreign country or the Italian region of residence.

This is the unique official survey allows the availability of touristic data at the municipal level for all Italian cities. It allows both to estimate the tourist movement and the relevance of the sector at the local level. Unfortunately, direct surveys that collect important information such as traveler spending or the weight of international tourism on GDP do not disseminate data at the local level (Bank of Italy 2020). For this reason, it was decided to use only this statistical source at this stage.

The most interesting aspect of the data lies in the possibility of articulating the tourist movement of customers according to all the possible combinations of the variables considered, in order to allow an in-depth analysis of the relationships between them. Istat also calculates the utilization rates of beds and rooms in hotel-type accommodation facilities.

The survey units are the accommodation establishments present on the national territory, divided between hotel facilities and extra-hotel facilities:

- Hotel establishments: hotels classified into five categories divided by number of stars and tourist-hotel residences.
- Non-hotel establishments: campsites, tourist villages, mixed forms of campsites and tourist villages, rental accommodation managed in an entrepreneurial form, farmhouses, youth hostels, holiday homes, mountain refuges, bed and breakfasts and other accommodation establishments.

Data collection is entrusted to the intermediate bodies. The survey is conducted according to the rules contained in the Istat annual circulars.

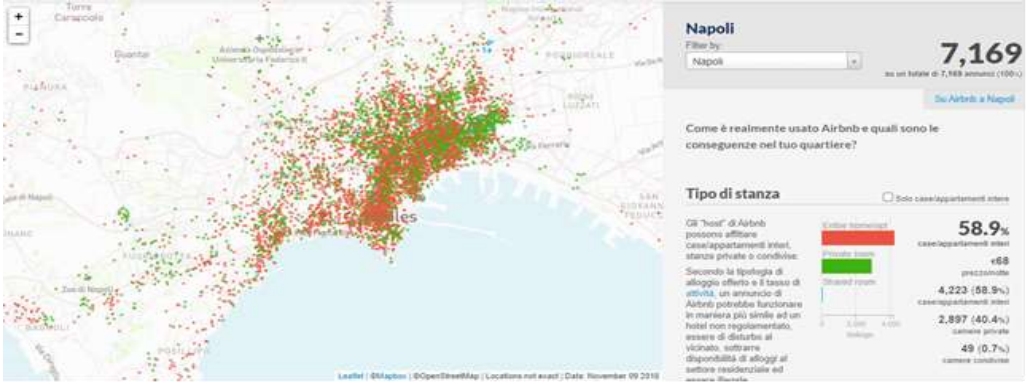
The Open Data

Istat data allow us to have a general view of tourism trends in Italy. The AirBnb data (AirBnb is widely adopted among the under 50s) make it possible to obtain further information on the choices of tourists, considering a type of accommodation not included among the observation units of the Istat survey.

Through the "Inside AirBnB" platform it is possible to obtain a lot of information about the accommodations in the most important Italian tourist destinations, including Milan, Rome, Venice, Florence and Naples, which are the subject of the survey.

Furthermore, it is easy to represent these BnBs graphically on a geographic map of the reference city, if there is a need to consider the location, perhaps to relate it to the price (it is reasonable to think that accommodation located in the city center are less economical than those in the suburbs). Figure 1 shows the dynamic graphic representation of the Naples map of AirBnB accommodations classified by color according to type (whole house, single room or shared room). There are also data on the number of accommodations and average prices per night.

Figure 1. *Dynamic Localization of AirBnb Structures*



Napoli

Filter by:

Napoli

7,169

su un totale di 7,169 annunci (100%)

Su Airbnb a Napoli

Come è realmente usato Airbnb e quali sono le conseguenze nel tuo quartiere?

Tipo di stanza

□ Solo case/appartamenti intere

Gli "host" di Airbnb possono affittare case/appartamenti interi, stanze private o condivise.

Secondo la tipologia di alloggio offerto e il tasso di attività, un annuncio di Airbnb potrebbe funzionare in maniera più simile ad un hotel non regolamentato, essere di disturbo al vicinato, sottrarre disponibilità di alloggi al settore residenziale ed

58.9% case/appartamenti interi

€68 prezzo/notte

4,223 (58.9%) case/appartamenti interi

2,897 (40.4%) camere private

49 (0.7%) camere condivise

Source: AirBnB data.

Twitter as Informative Source

To obtain representative data of popular opinion, it is useful to retrace the scraping procedures of social networks.

By scraping we mean the extraction of data from a website through the use of software or algorithms designed for this use.

Several social networks are very restrictive regarding the release of data, especially after the 2016 Cambridge Analytica scandal involving Facebook. Since then, it has been almost impossible to obtain data from platforms belonging to Mark Zuckerberg such as Facebook or Instagram, however there is still a rather available social network regarding scraping: Twitter.

In fact, Twitter lends itself perfectly to this procedure thanks to the very limited character threshold and its purely textual posts. Furthermore, it is possible to request access to the API through a form that can be filled in directly on the site, without having to resort to third parties. The form requires a very specific compilation and numerous data to be entered, but within a reasonable time, and with the right reasons, you can get access to the developer account, which allows you to develop your own scraping app. However, it is necessary to specify that access to free APIs suffers from various limitations, including, particularly as an obstacle to the investigation, that of the possibility of accessing only tweets dating back to a maximum of 7 days prior to the search.

In this case, all the IT procedures such as scraping, sentiment analysis, etc. were carried out through algorithms written in Python (Mitchell 2018). Python is an object-oriented programming language created with the aim of being easier to understand and use than its competitors such as Java.

Its strength lies precisely in its versatility and ease of learning; it is also equipped with a rather large community of users that offers open-source libraries capable of performing numerous functions.

There are several variables that can be observed in a tweet, not all are fundamental in the data collection phase, each variable can be more or less useful depending on the purpose pursued. However, it is useful to analyze them:

- The length of the tweet (currently the maximum number of characters is 280).
- The number of comments.
- The number of retweets.
- The number of likes.
- The hashtags.
- Whether or not there is a multimedia element (photo/video/gif).
- The language in which it is written.
- If activated by the user, the geolocation of the tweet.
- Date and time of publication.
- Number of followers and following of the profile that tweeted.
- Number of tweets of the profile he tweeted.

Through an algorithm written in Python it was possible to collect all this information from a large number of tweets, selecting them based on the city thanks to two factors:

- The geolocation of the tweet.
- Hashtags.

It was useful for the research also to take in consideration the use of hashtags regarding the reference city as several tweets did not always come from the geolocated city (for example, some geolocated tweets in Naples actually concerned Sorrento).

Through two Python libraries, TextBlob and Pandas, it was possible to analyze the sentiment coming from each tweet.

Sentiment Analysis, also known as Opinion Mining, is a field within Natural Language Processing (NLP), the purpose of which is the analysis of a text with the aim of identifying and classifying the information present in it. Usually, in addition to identifying the opinion, these systems extract the attributes of the expression such as:

- Polarity: positive or negative opinion.

- Subject: what we talk about.
- Opinion holder: the person or entity who expresses the opinion.

In other words, sentiment analysis is used to learn about brand perception (where brand means any object about which you want to express an opinion) through user interaction exchanges on social networks or more generally on the web.

It was necessary to select the most useful information among those extracted, in order to construct the subjective synthetic indicators that will be subsequently analyzed, this information concerns:

- The text of the tweet.
- The number of followers and following of the account author of the tweet.
- The number of posts published by the account author of the tweet.
- The number of retweets.
- The number of likes.
- The sentiment that comes from the tweet.

The texts were subjected to an important pre-processing, necessary for the processing of the same for statistical purposes. In addition to empty spaces, links and symbols, the so-called stop words have been removed, i.e., the set of words commonly used in any language, such as conjunctions and adverbs, which create "noise" in the analysis.

The Composite Indicator

A composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model. The composite indicator should ideally measure multidimensional concepts which cannot be captured by a single indicator, e.g., competitiveness, industrialization, sustainability, single market integration, knowledge-based society, etc. (OECD 2008).

A composite indicator is easier to interpret than a battery of many separate indicators, even if it may invite simplistic policy conclusions.

The literature on composite indicators is vast and almost every month new proposals are published on specific methodological aspects potentially relevant for the development of composite indicators.

It’s interesting to mention in this context a recent work, the Semantic Brand Score (SBS) (Fronzetti Colladon 2018). This synthetic index measures the importance of a brand when it is possible to analyze textual data sources (particularly geared towards big data).

Taking advantage of graph theory, text mining and social network analysis, this measure combines three fundamental indicators (by making a standardized sum between them):

- Prevalence (measures how much a brand is mentioned in a speech).
- Diversity (heterogeneity of the brand’s textual associations).

- Connectivity (the connecting power of the brand, what is at the heart of the speech).

All the indicators considered have the same importance and only combined are useful for measuring the importance of a brand. Many concepts can be considered as brand (politicians, areas, etc.) and be analyzed in order to assign a value to the concept.

As it is known, the applicable methodology for the construction of any synthetic index involves the following phases:

a) Defining the concept. The definition should give the reader a clear sense of what is being measured by the composite indicator.
b) Selecting variables.
c) Normalization of data. Normalization is required prior to any data aggregation as the indicators in a data set often have different measurement units.
d) Weighting and aggregation. In any case, equal weighting does not mean "no weights", but implicitly implies that the weights are equal.

The aggregation methods are usually based on additive, geometric models, or non-compensatory multi-criteria approach (MCA).

Relevant for the study analyzed is the Mazziotta-Pareto Index (MPI), based on additive approach. It is based on a non-linear function which, starting from the arithmetic mean of the normalized indicators, introduces a penalty for the units with unbalanced values of the indicators (De Muro 2011).

Two version of the index have been proposed: (a) MPI, and (b) adjusted MPI (AMPI). The first version is the best solution for a 'static' analysis (e.g., a single-year analysis), whereas the second one is the best solution for a 'dynamic' analysis (e.g., a multi-year analysis).

The composite index is given by:

$$MPI_i{+/-} = M_{zi} +/- S_{zi}\,CV_i$$

Where:
M is the mean of normalized matrix values for unit i.
S is the standard deviation.
CV is the coefficient of variation.

The sign +/- depends on the kind of phenomenon to be measured. If the composite index is 'increasing' or 'positive', i.e., increasing values of the index correspond to positive variations of the phenomenon (e.g., socio-economic development), then MPI- is used. On the contrary, if the composite index is 'decreasing' or 'negative', i.e., increasing values of the index correspond to negative variations of the phenomenon (e.g., poverty), then MPI+ is used. In any cases, an unbalance among indicators will have a negative effect on the value of the index.

The synthesis of the indicators using MPI_i allows to realize, in a simple and immediate way, descriptive analyses aimed at temporal and space comparisons beyond the state of complex phenomena.

The Indicators Chosen

Once the data from the institutional sources and the Big Data from Twitter were collected, simple indicators were built, divided into two categories: subjective and objective.

The subjective indicators were built through the data collected by Twitter, in order to obtain direct feedback from tourists on the cities they visited:

- Popularity of the source: followers/account followed (indicates how popular the author's account of the tweet is and can consequently affect a large number of people).
- Profile regularity: number of posts published (a more active profile is a more authoritative profile; his tweets are more visible to the public and are more credible).
- Diffusion: (retweet + like)/total number of tweets (a tweet with a high number of likes and shares appears in different profiles and can gather consensus and influence the thinking of others).
- Profile engagement: (likes + retweet)/number of followers (the ability of a profile to generate interest in its followers through a tweet).

The objective indicators are the result of the data collected through Istat surveys and the information available on the "Inside AirBnB" site:

- Accommodation density: number of beds/km^2 (the possibility of a tourist city to accommodate more or less tourists).
- Economic results: number of presences/number of beds (per year) (how much the available beds in the city yield during the year).
- Social sustainability: number of presences/populations.
- Average accessibility to BnB: average opening days out of 365/average price of Air BnB.

These indicators have contributed to the construction of the CCIT.

Results

The results of the objective indicators show that the cities with the greatest tourist flow have the highest accommodation density, i.e., Milan and Rome, while the city with the fewest beds per square kilometer is Naples (Table 1).

Related to the social sustainability indicator, the city of Venice has the largest tourist flow compared to the resident population, a situation that obviously drags the ordinary activities of residents. The analysis of the economic results achieved is interesting: the city of Naples, despite having a smaller number of tourists, counted, in 2018, 1,021 guests for each bed, a result that confirms the tourist growth trend that the city is living in recent times.

Table 1. *Objective Indicators for Cities, Year 2018*

Cities	Accommodation density	Economic results	Social sustainability	Accessibility average BnB
Milan	39.6	251.9	4.1	1.5
Venice	20.4	725.1	18.6	1.9
Florence	12.9	337.3	9.6	2.2
Rome	34.3	175.5	5.3	2.4
Naples	11.8	1,020.8	4	3.9

Source: Our processing on Istat and AirBnb data.

The average accessibility to BnB shows that the city of Naples is the most competitive related to the average number of days of occupied structures compared to the price of the single structure (3.8).

For the sentiment analysis, the tweets published in the first week of July 2020 were taken into account. It was decided to study this month since July is the favorite month for Italian tourism. Obviously, in the year under consideration, the advent of the pandemic led to a reduction in both local and, above all, foreign tourists. For this reason, it was decided to focus the attention on the tweets in Italian language.

Altogether, about 1,000 tweets were analyzed, mainly relating to the city of Milan (36%), followed by Rome (24%), Florence (17%) and Naples and Venice, both with 11%.

Considering the polarity of the tweets, which is how the cities visited in the period considered were expressed, it tends mainly to neutrality. In fact, if we consider the total number of tweets, in 60% of cases users have expressed themselves with words that express neither enthusiasm nor disapproval.

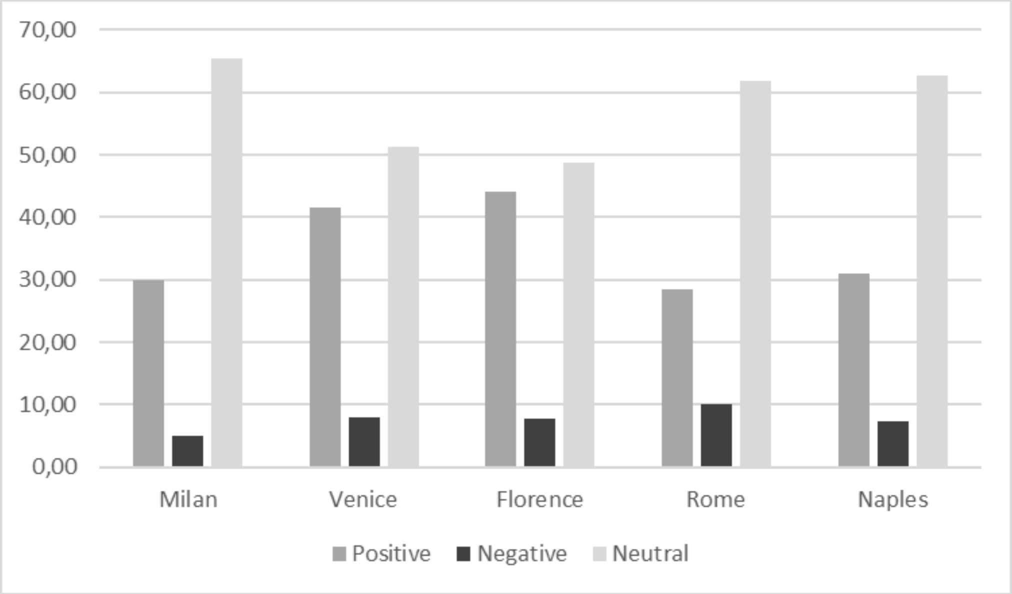
The cities for which users have shown the most positive feelings are Florence (44%) and Venice (42%), while the one for which a negative sentiment prevails compared to the total number of tweets is Rome (10%) (Figure 2).

Table 2 shows the results of the subjective indicators considered. The polarity has been summarized in a single indicator that expresses the number of positive tweets on the total tweets for each city.

Starting from the first indicator, popularity of the source, it is clear that the most popular users, in the period considered, have mentioned the city of Naples. They are also the most authoritative users, because the profile Regularity indicator shows that the users considered also have a greater presence on the social network in terms of published tweets.

The tweets most retweeted, and with the greater number of users, are related to the city of Venice, followed by Rome and Naples. In any case, the tweets considered did not raise great interest (Engagement indicator); slightly higher values for the cities of Milan and Rome.

Figure 2. *Polarity of Tweets by City, 1st Week of July 2020*



Source: Our processing on Twitter data.

Table 2. *Subjective Indicators by City, 1st Week of July 2020*

City	Source popularity	Users regularity	Polarity	Diffusion	Engageme nt
Milan	7.27	22,164.34	0.29	1.53	0.02
Venice	8.71	11,141.61	0.42	9.96	0.01
Florence	4.39	25,316.52	0.44	1.37	0.01
Rome	4.6	15,930.01	0.29	5.32	0.02
Naples	11.14	26,766.46	0.31	5.18	0.01

Source: Our processing on Twitter data.

The summary of the indicators obtained with the MPI index with negative penalty, allows us to obtain the final ranking of Italian cities based on the CCIT.

All indicators have a positive sign, because they contribute positively to the composite indicator. The only negative sign is for social sustainability since high values of the indicator show the presence of a tourist flow that may not be bearable for the resident population.

In the period analyzed, the city of Naples was the favorite destination for travelers, followed in order by Milan, Rome, Venice and Florence.

The cities are compared with eight different methods: the Mazziotta-Pareto Index (MPI) in the two variants (positive and negative); the taxonomic method of Wroclaw (Wroclaw); the mean of the mean of the standardized values (M1Z); the ranking method (Grad.RNK); the method of relative indices (IR); the method of the arithmetic mean of the basic average index numbers (ANIM); the method of the geometric mean of the basic mean index numbers (GNIM); the method of the square mean of the basic mean index numbers (QNIM). The results are similar for each method used.

Conclusion

The work shows how subjective aspects, thanks to the high informative and analytical value, can contribute to analyzing phenomena such as tourism.

Subjective indicators are complementary to strictly objective indicators, as they allow us to assess any divergences between what people report and what by objective indicators captured. The observation of these indicators allows us to have a more articulated and complete vision of the phenomena.

In this specific case, according to official data, Rome is the main Italian tourist destination, followed by Milan and Venice. Considering the opinions of travelers, the first city is Naples, which would otherwise be the eleventh if we considered only the tourist flow.

Unfortunately, the results were influenced by the pandemic that drastically reduced world tourism; however, the results underline the objective tourist growth of the city of Naples greater than the main Italian tourist destinations.

The aim of the work was the construction, as an explorative exercise, of a tourism competitiveness index at the municipal level. This limited the choice of statistical sources by forbidding the use of strictly economic data. In the future, other sources could be used by evaluating the practicability of small areas estimation techniques, based on the use of linear models with mixed effects referred to the unit or small area level (Rao 2013). Another possibility could be to integrate the survey micro data with administrative data available locally or with the statistical register of active companies in order to integrate the index with economic indicators.

The social network used for the sentiment analysis is Twitter. It has 330 million monthly active users in the world. In Italy, according to the telecommunications guarantees authority (Agcom), it has about 13 million of members, a little number compared to Facebook’s 38 million, but as previously specified, it has not restrictive policy regarding the release of data and the limited size of the number of tweets characters makes Twitter better than other social networks for the sentiment analysis application. Obviously, the opinions of twitter users are spontaneous and concern a reduced number of tourists, but this is the limit of any social network. However, in the analysis described, the number of indicators chosen is small, due to the limitations posed by the use of free APIs in data scraping.

A paid version would allow to use a greater number of variables and therefore to build additional subjective indicators. It would also allow downloading information not limited to a week, but relating to an entire year, making the analysis of user opinions more reliable. In this manner, it would be possible to use the subjective indicators referring to the same reference period as the objective ones, which inevitably suffer from diffusion delays compared to the survey period.

Ultimately, for the future development of the CCIT it will be necessary to limit scraping to real tourists and non-residents commenting on the city in which they live. However, the information on the city of residence sometimes is not expressed by users and could represent a limitation for the study.

In any case, the results of the research have shown that Twitter represents the ideal source for this type of analysis: first of all, for the privacy policy of the site which leaves the public nature of the information as a default setting; secondly for the great potential of hashtags that tag each tweet based on the topic to which it refers, giving the possibility to gather discussions related to the same topic, even if started by users who have no connection with each other.

Acknowledgments

Our thanks to Armando Spigno, Cà Foscari Challenge School, and Manuel Spigno, Rotterdam Business School, for English language consulting.

References

Alivernini A, Breda E, Iannario E (2014) *International tourism in Italy (1997–2012).* Questioni di Economia e Finanza (Occasional Papers). Bank of Italy.

Bank of Italy (2020) *Indagine sul turismo internazionale*. (Survey on international tourism). Statistiche, Bank of Italy.

Cini National Lab (2020) *AI for future Italy*: *The CINI vision and recommendations for Italian AI*. Lab CINI-AIIS.

Curry E (2016) The Big Data value chain: definitions, concepts, and theoretical approaches. In JM Cavanillas, E Curry, W Wahlster (eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe.* Ginevra: Springer Open.

Daas PJH, Roos M, van de Ven M, Neroni J (2012) *Twitter as a potential data source for statistics.* Discussion Paper (201221). Den Haag/Heerlen: Centraal Bureau voor de Statistiek.

De Muro P, Mazziotta M, Pareto A (2011) Composite indices of development and poverty: an application to MDGs. *Social Indicators Research* 104(1): 1–18.

Freitas A, Curry E (2016) Big Data curation. In JM Cavanillas, E Curry, W Wahlster (eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe.* Ginevra: Springer Open.

Fronzetti Colladon A (2018) The semantic band score. *Journal of Business Research Volume* 88(Jul): 150–160.

Gozzo S, Pennisi C, Arsero V, Sampugnaro R (2020) *Big Data e processi decisionali: strumenti per l'analisi delle decisioni giuridiche, politiche, economiche e sociali.* (Big Data and decision-making processes: tools for the analysis of legal, political, economic and social decisions). Egea.

Istat (2019a) *Movimento turistico in Italia, anno 2018*. (Tourist movement in Italy, year 2018). Statistica Report – 7 novembre 2019. Roma: Istat.

Istat (2019b) *BES 2019 – Il benessere equo sostenibile in Italia*. (BES 2019 - Sustainable fair well-being in Italy). Roma: Istat.

Laniado D, Mika P (2010) Making sense of Twitter. Conference Paper. In *9th International Semantic Web Conference, November 7–11, Shanghai, China*.

Lisi FA, Esposito F (2015) *An AI application to integrated tourism planning.* Conference Paper. In *XIV International Conference of the Italian Association for Artificial Intelligence, 2015, Ferrara, Italy*. Volume: LNAI 9336.

Mitchell R (2018) *Web scraping with python: collection more data from the modern web*. Sebastopoli, USA: O'Reilly Media, Inc.

Organisation for Economic Co-operation and Development – OECD (2008) *Handbook on constructing composite indicators: methodology and user guide*. OECD Publications.

Rao JNK (2013) Small area estimation: methods, applications and new developments. Conference Paper. In *NTTS 2013 Conference, March 2013, Brussels, Belgium.*

World Trade Organization – WTO (2019) *The travel & tourism competitiveness report 2019*. Ginevra: World Economic Forum's Platform for Shaping the Future of Mobility.

Zerba F (2018) Big Data tools and tourism market intelligence. Conference Paper. In *15th Global Forum on Tourism Statistics, 28–30 November 2018, Cusco, Peru*.