 doi:10.20944/preprints202205.0182.v1

Article

Tourism Sustainability Index: Measuring Tourism Sustainability Based on the ETIS Toolkit, by Exploiting Tourist Satisfaction via Sentiment Analysis

Damiano De Marchi [1,*], Rudy Becarelli [1] and Leonardo Di Sarli[1]

[1] The Data Appeal Company, Via del Tiratoio, 1 50124 Firenze, Italy
* Correspondence: damiano.demarchi@datappeal.io Tel.: (+39 345 1563011)

Abstract: The importance of measuring sustainability in tourism has been significantly advancing in recent years, following the need to manage the impact of tourism on territories and hosting communities. It was further boosted by the pandemic, where sustainability has been defined as one of the central elements to restart global tourism. The ETIS model, developed by the European Commission, is a point of reference based on self-assessment, data collection and analysis by the destinations themselves. The application of ETIS toolkit has faced many challenges, especially at subnational level, mostly related to the lack of available and updated data to feed the model. The hypothesis explored by the authors is to solve the implementation issues, developing an indicator based on the use of the Sentiment Analysis to frame e-reputation and tourism satisfaction, and further combining it with other open data sources. The Tourism Sustainability Index (TSI) can provide a scalable and geo-referenced evaluation of tourism sustainability, measuring the four pillars and sub-components referenced to ETIS criteria, applicable to any tourism destination. Results show that the TSI can be seen as a consistent and valid tool for destinations to analyze sustainability, monitor its evolution through time periods and sub-areas, and compare it to other benchmark or competitive areas.

Keywords: Tourism; Measuring sustainability; Tourist satisfaction; E-reputation; Sustainable development; Sentiment analysis; ETIS; Open data; Geospatial Index

1. Introduction

1.1. Measuring Tourism Sustainability

Sustainable tourism has been advancing since the early 1990s, with the UNWTO promoting the use of sustainable tourism indicators as essential instruments for policy making, planning and management processes at destinations [1]. With the new Millennium, the sustainable tourism assessment became increasingly relevant in national and international agendas, culminating in the UN declaration of 2017 as the International Year of Sustainable Tourism for Development, following the adoption of the United Nations 2030 Development Agenda and the Sustainable Development Goals. Recognizing the potential that tourism has in contributing, directly or indirectly, to all the goals, boosted the need of an initiative to develop a common framework to measure the impacts of tourism on the environmental, economic, and socio-cultural aspects at any level. The initiative towards a Statistical Framework for Measuring the Sustainability of Tourism launched by UNWTO with the support of the United Nations Statistics Division, aimed to develop an international statistical framework for measuring tourism's role in sustainable development, including economic, environmental, and social dimensions and support universal, cross-sectorial, and sustainable tourism policies and practices that work from an integrated, coherent, and robust information base.

 doi:10.20944/preprints202205.0182.v1

Although scientific literature and practices have pointed out the difficulties of transferring the principles of sustainable development to a specific sector, the need for a valid measurement for tourism sustainability is evident across all stakeholder groups, cutting across global, national, and subnational (including local) levels, and a key tool for the adoption of the principles of sustainability by the tourism industry [2,3]. It has been acknowledged that defining a set of indicators - internationally comparable - enabling an assessment of the level and of the improvements towards sustainability at the destination level will allow tourism managers to monitor the situation and eventually identify the actions required to improve the level of sustainability across the different elements of tourism products [4]. In this regard, the European Commission in 2016 developed an 'European Tourism Indicators System' (ETIS) to help tourist destinations measure their performance with respect to sustainability, which is essential, by using a common comparable approach. The ETIS model is composed of 4 pillars and 43 core indicators plus an indicative set of supplementary indicators. Its monitoring results are based on self-assessment, observations, data collection and analysis by the destinations themselves [5]. The flexibility provided by the system - destinations can choose for themselves the most relevant indicators they wish to adopt - was initially designed to improve the potential of feasibility and success, but then the task of tailoring the indicators to their destination became sometimes an obstacle to the actual implementation [6]. Moreover, another problem for tourism destinations was the lack of data needed to feed the system accompanied by a tepid interest in destination stakeholders in a long-term investment for its adoption, to cover for example the costs of conducting surveys on visitors, residents, and tourism businesses [7-9]. Despite the challenges, ETIS is still the most known, cited, and valuable methodology to implement a quantitative and effective measurement of tourism sustainability at destinations' level.

1.2. Tourism sustainability and reputation

The concept of sustainability has been widely accepted to mitigate the impacts of tourism, this is why it has been intensively linked to the concept of carrying capacity of tourism destinations since the early 90s , showing the critical range of elements of capacity leads to a level of unsustainability, harming the destinations elements of attraction [10] and generating additional negative effects (externalities) on third parties not directly involved in tourism activity [11-12]. In most recent times, satisfaction has been added as another key element to be considered when discussing the tourism carrying capacity [13], which is possible to be directly related to the different elements of sustainability as perceived by the different users of a tourism destination: residents (both working in the tourism industry or not), commuters, visitors, tourists. Perception became an important element of study since the importance of the concept of risk started rising. In the history of the study of consumer purchasing behavior a lot of space has been given to the concept of risk, since the increasing of welfare and the consequent increase in the average life span led individuals to reflect more and more on this topic, creating a conscience risk, which feeds back on society leading it to demand ever higher levels of safety [14].

The search for timely and precise information is a determining factor for an individual or collective user in the process of mitigating the perception of risk, whether referring to the statistically and objectively demonstrable one or the perceived affective risk, which is more related to the cognitive, emotional, and social dimensions [15]. Over the years, most of the literature on risk perception has demonstrated the importance of emotion in risk perception and risk-taking behaviors, and the fundamental role of the emotional reactions playing in risk-judgement and decision making [16,17]. Reducing uncertainty therefore plays a key role in the process that leads to the decision to make a purchase or not, especially when we refer to a virtual transaction, where the customer adds the financial risk of the transaction with the potential technical complexity. The development of

 doi:10.20944/preprints202205.0182.v1

the Web 2.0 in the early 2000s has encouraged people to express their opinions about products and services through Social Media Companies [18] and many studies focused on the correlation between user-generated content (UGC) and sales: online product and service reviews and the related features have a significant impact on consumers' purchase decisions and sales [19-22].

This is the reason why, as e-commerce started soaring, marketplaces such as Amazon, Ebay and Booking.com included ways for registered users to share opinions about their buying experiences. By reducing the risk for actual customers by providing information before making their actual purchase decisions [23] and attracting new prospects by providing an online platform that enables customers to exchange their consumption experiences [24], online companies developed more sophisticated methods to enrich the review experience, e.g., including the rating, picture, and popularity of the reviewer. In tourism, risk mitigation becomes even more important, for a variety of factors: the very characteristics of tourism products - such as intangibility, inseparability from provider and customer, absence of ownership and heterogeneity -; the distribution which is mainly online; the cost which is money, time, and the emotional potential, since the consequences of a bad tourism experience may be perceived as significant. By investigating the perceived risk in the field of tourism, many conceptualizations of risk perception have been traced [25]. It is also amply demonstrated how reputation has a direct influence not only on the purchasing process of tourism services, but also how visitors approach and experience a place and consequently developing a process of involvement and trust which is a leverage of success for businesses at any size and destinations [26-28]. Hence, to understand reputation by visitors' opinions, a new research field and consequent activity emerged during the last decade: the Sentiment Analysis, which is generally defined as the computational and computerized process of recognizing, detecting, and determining the orientation of human opinion or emotion and its polarity [29,30], using as bases mostly big data and Tourists' UGC [31]. The use of sentiment analysis in tourism can positively contribute to the decision-making process of governments and businesses to foster sustainability and to understand consumer behavior patterns [32,33].

1.3. Towards a Tourism Sustainability Index (TSI)

"Sustainability must no longer be a niche part of tourism but must be the new norm for every part of our sector. This is one of the central elements of our Global Guidelines to Restart Tourism. It is in our hands to transform tourism and that emerging from COVID-19 becomes a turning point for sustainability."

With these words in 2020, UNWTO Secretary-General Zurab Pololikashvili addressed The One Planet Vision for the Responsible Recovery of the Tourism Sector: a repository of inspiring initiatives, tools, and strategic thinking, representing a common vision for better tourism for people, planet, and prosperity [34]. Considering the fundamental importance for tourism of sustainability, it is indispensable that its principles should be adopted by any typology of tourism destination to manage and mitigate the impacts of the tourism phenomenon. Moreover, sustainability must be measured, as well as its elements as governance, environment, economy, and the social-cultural sphere, to make it a strategic objective in management plans and actions. The challenges of a widespread implementation of the ETIS methodology haven't been solved by the existing Tourism Sustainability Indexes: the Sustainable Tourism Index created by The Economist Intelligence Unit [35] or the Tourism Sustainable Development Index developed by the European Space Agency in collaboration with Murmuration [36] can be applied only at country-level; the Global Destination Sustainability Index designed by the GDS Movement can be applied at destination level, but it is based on a questionnaire where the quantitative

 doi:10.20944/preprints202205.0182.v1

and qualitative data must be entirely collected then submitted by the destinations themselves [37]; the ISOST Index [38] remained at test-level and hasn't found a real and continuative implementation by tourist destinations. The main contribution of this paper lies in its exploration of the hypothesis that a real and widespread implementation of an actual measurement of tourism sustainability can be achieved only by an Index based on data sources which are self-updating and at the same time maintaining an international comparability and an applicability to any typology of tourism destination, from the sub-urban to the macro-regional level. The proposed Tourism Sustainability Index (TSI) a georeferenced indicator evaluating in a quantitative scale (0-100) the sustainability of tourism activity in a destination. The TSI is scalable, suitable for any location worldwide, with a monthly time frame, based on various sources of data as described in the next section and including the sentiment analysis as one of the axes to any of the four pillars defined in ETIS methodology. The authors present the TSI to the scientific community with visualizations of test applications to real territories and possible future implications.

2. Methodology

The methodology presented in this paper, describes the authors' approach to design the Tourism Sustainability Index within the procedures and proprietary algorithms of The Data Appeal Company (TDAC), based on a machine learning model which analyzes online conversations and provides a multi-class score of any text [39].

2.1. The Process

In order to compute the TSI, the process starts from the definition of Points Of Interest (POIs), defined as economic, cultural and naturalistic geographical points for which a set of textual contents can be retrieved and univocally related to. TDAC collects and monitors the digital presence of POIs through the analysis of over 130 online platforms, websites, OTAs and social media channels (e.g., Google, Booking.com, Airbnb, TripAdvisor, Facebook, Twitter, and many others) and explores all the present characteristics of the POIs, the UGCs and their makers. A complex automatic process, aims to maximize the probability that the channels explored are related to the same POI, comparing, and completing the information through the various channels to build the digital identity of the POI. Considering the amount of data involved, a crucial factor for the correct use of this data is the activation of data quality procedures, to identify and possibly correct outliers or anomalies and potential inconsistencies. Also, the data quality process must be automatized, consisting of fast and precise algorithms, designed for the specific problem to be monitored [40]. Once the ingestion and data quality processes have been completed, the data will be stored in The Data Appeal Company's private Data Lake. At this time, the Semantic Analysis takes place, processing the textual contents and aiming not only to provide a polarity score (sentiment) for each content, but to also identify the main topics and subjects and opinions connected with these topics. The algorithm is basically composed of 3 models:

- Name Entity Recognition (NER) model, which aims to classify words or phrases within predefined categories starting from unstructured text such as that of reviews;
- Dependency Parser model, analyzing the grammatical structure of texts and identifying the connections between the various words;
- Sentiment Analysis model: it is a classic machine learning model in NLP (Natural Language Processing) that looks for non-linear dependencies between the various words to "understand" at a computational level the logics that represent satisfaction and, more in general polarity, of a generic text.

The Sentiment Analysis is the key model for the development of the TSI. Further detailed in the next subsection, the TSI is computed by two kinds of sources: public data (here referenced as Open Data) and proprietary data. The first one can assess the objective

 doi:10.20944/preprints202205.0182.v1

component of the index, while proprietary data is mostly used to assess the subjective one. Proprietary data, as result of the Sentiment Analysis, provides the customers' perception of the experience related to the tourism industry. Since the sentiment analysis is applied to social media contents related to each point of interest, and these contents vary over time, it's possible to average the sentiment grouping of POIs by type of sector, time and on a geographic basis (e.g., the average/median sentiment computed for the hospitality-related POIs in a certain county in a certain time period). The same approach can be used to deduce the sentiment related to a certain territory or to a particular event if the sentiment analysis is applied to social media contents filtered by hashtag. The implementation of an objective and a subjective point of view provides unprecedented insight into the tourism sustainability topic since, not only the typical environmental parameters are considered, but even the perception of the policy efficiencies aimed to preserve the socio-cultural heritage and local productions. Finally, to ingest and aggregate publicly available data, three different activities had been carried out:

- Investigating data sources that could be useful to characterize all the aspects of Sustainability as defined in the ETIS methodology;
- Defining and implementing processes that must be put in place for data cleaning, data quality and data understanding. This activity is necessary just because heterogeneous data, coming from the ingestion of the previously selected data sources, must be synchronized on the same timeline and missing and spurious data must be imputed and/or geographically resampled in order to profitably aggregate a georeferenced indicator;
- Designing, creating, and aggregating each component of the final index. Each component is implemented as a specific model that tries to describe, at its best, each aspect of the final index, following an additive / multiplicative aggregation approach where a compensation between the base indicators can be admitted [41]. The list of the components used in the final index computation, their aggregation in main pillars and their rationales will be detailed in the last subsection of the methodology.

2.2. The data sources

A heterogeneous data set has been selected to estimate how much a tourist destination is compliant with the objective of the TSI. These data must have the following characteristics: being scalable (i.e., covering the entire planet even at the cost of changing accuracy and resolution, depending on the country they are representing), georeferenced (i.e., they can be localized on a map with predicted precision) and storified (i.e., they can be ingested according to a certain schedule as they change over time. Frequency can be variable depending on the nature of the content). To cover a large area and to deal with data outliers, the index is computed with respect to the tile coverage provided by the Bing Maps Tile System [42]. A tile size of 16 has been selected so that the index could estimate the sustainability compliance of the destination with sufficient detail while not losing its statistical significance. Furthermore, Data Sources can be divided in two different categories: Open Data Sources and Proprietary Data Sources. The following paragraphs detail their characteristics and their ingestion technique at some degree.

2.2.1. Open data

Open data refers to any data that is openly accessible, exploitable, editable, and shared by anyone for any purpose. In the remainder of this analysis, the term open data is used for those datasets provided by two different sources:

- Copernicus [43] is the European Union's Earth observation program, looking at our planet and its environment to benefit all European citizens. It offers information services that make raw and preprocessed data available, gathered by satellites continuously observing and measuring our planet's environmental parameters. Another set of relevant data comes from the direct (in situ) measurements accomplished by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0182.v1

means of ground based, airborne and seaborne platforms. This additional data comes with a higher degree of accuracy and precision at the cost of a less scalable coverage. Where available, however, this data provides a significant boost to the data quality and allows for a higher degree of spatial and temporal resolution. The European Commission manages the Program It is implemented in partnership with the Member States, the European Space Agency (ESA) [44], the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) [45], and the European Centre for Medium-Range Weather Forecasts (ECMWF) [46]. This vast amount of global and local data enables policy makers, public authorities in general, and privates to undertake informed analysis and make advised decisions to improve European citizens' quality of life and beyond. The information services are free and openly accessible to any registered user;

- The World Bank Data Portal [47] provides access to global economic and development statistics including World Development Indicators, International Debt Statistics, Millennium Development Indicators and data on poverty, education and gender related issues (i.e., disparity on economical revenues and social/human rights related to the gender). This kind of data is useful in the definition of the social and economic context for some components of the final index, namely for the ones trying to measure the impact of tourism on the local culture and economy.

2.2.2. Proprietary Data

Proprietary data refers to data that is stored and managed in The Data Appeal Company's data lake. The Data Appeal Company combines 3 key, real-time intelligence elements at once: Location, Sentiment and Market Intelligence, providing access to both historical and forecasting insights for any point of interest or territory. Across the qualitative value of that data, several studies conducted during the past 2 years in collaboration with the actual customers (i.e., Italian Government, Veneto Region, Tuscany region and others at sub-regional level as provinces or municipalities) show that the correlation between the changes in online contents and tourism arrivals is, with a Pearson correlation coefficient always > 0,90 [48]. The proprietary data forms the backbone of all the components of the index, while datasets coming from Open Data Sources act as collateral, but still relevant, information useful to orchestrate and harmonize all the index components within and among the tiles covering a certain destination.

2.3. The components of the Tourism Sustainability Index

According to the ETIS methodology, the TSI has four main components (pillars), each of them composed of several sub-indicators to map the ETIS toolkit for sustainable destinations criteria. All indicators and all pillars are computed in a range between 0 and 100 and combined on a weighted mean. This specific range allows the user to compare different behaviors in an easy way as well as facilitate calculation of the TSI.

Table 1. Linkages between ETIS criteria and TSI sub-indicators, with detail of data sources used

ETIS Criteria Reference	TSI sub-indicators	Sentiment Analysis	Data Source
A.2 Customer satisfaction	Customer Satisfaction Index	Yes	Proprietary data
	Short Term Confidence	Yes	Proprietary data
B.1 Tourism flow at destination	Seasonality Balance	Yes	Proprietary data
	Volume of Visitors	No	Proprietary data
	Percentage of Business Tourism	No	Proprietary data
B.2 Tourism enterprise(s) performance	OTA Penetration	No	Proprietary data
C.1 Community / social impact	Number of Short-Term Rentals	Yes	Proprietary data
	Tourism Pressure	No	Proprietary and Open data

 doi:10.20944/preprints202205.0182.v1

	Tourism Supply Pressure	No	Proprietary and Open data
C.2 Health and safety	Good Health Index	Yes	Proprietary data
C.4 Inclusion / accessibility	Accessibility Index	No	Proprietary data
C.5 Protecting and enhancing cultural heritage, local identity, and assets	Cultural Index	Yes	Proprietary data
	Urban Green Index	Yes	Proprietary data
D.1 Reducing transport impact	Public Transportation Accessibility	Yes	Proprietary data
D.2 Climate change	CO2 Index	No	Open data
	Air Quality Index	No	Open data
	Sea Quality Index	No	Open data
D.7 Landscape and biodiversity protection	Natural Coverage	No	Open data

2.3.1. Destination Management Pillar

Since no data source with the above characteristics can provide the governance of a destination, this pillar emphasizes the perception, confidence, and success of a tourist destination as a consequence of its destination management. The main data source used to compute the component indicators is proprietary data, in particular the sentiment and the contents time series analysis.

2.3.2. Overtourism Pillar

At this pilot stage, it was impossible to reach an economic indicator based on data sources with the above characteristics. This is the reason why the authors propose a pillar emphasizing the over tourism grade in a destination, mixing different indicators regarding the tourist flows and the tourism pressure on the supply and on the hosting community also in terms of satisfaction, and the percentage of accommodation which we estimate has been booked via online travel agencies (OTAs). These indicators are computed with property data such as sentiment, the time series analysis of online content, and the point of interest composition of a destination combined with World Bank Data, such as population density.

2.3.3. Social and Cultural Pillar

The Social and Cultural pillar focuses on the effects of tourism on both social and cultural sides in the destination, such as the development and efficiency of the healthcare system, the presence and quality of the urban green, the cultural vitality and reputation, the presence of events as well as business / MICE tourism. A special focus has been given also to accessibility, seen as how many POIs are accessible to disabled, but also identifying how the whole destination is accessible in terms of public transportation. The main data source used to compute the following indicators is the proprietary data, in particular the sentiment and the contents time series analysis.

2.3.4. Environment Pillar

The Environment pillar focuses on those elements that are critical to the sustainability of the environment of the destination, highlighting the importance of investing in its protection and how tourism supports or affects it. High quality of air, water and great biodiversity helps ensure the sustainability of natural areas, benefits the destination's image, and attracts visitors. This highlights the importance of investing in landscape and biodiversity protection and the tourism sector's role in supporting this process. This pillar and its indicators are computed on Copernicus Open Data.

3. Results

In order to test, explore, compare and inspect the TSI and its components, a front-end data visualization tool has been deployed by the authors: in this way it is possible to apply the TSI to a custom area and test the results, and selection of snapshots has been presented analyzing the province of Milan as target destination.

 doi:10.20944/preprints202205.0182.v1

There are several reasons for this area to be designated: it is a wide area containing a metropolitan city (Milan), other towns and small villages. The area presents different characteristics: the presence of high-quality hospitals and health centers, public services, and facilities, yet it also has a severe concentration of air pollutants and CO2 as well as an intensive land use. Milan is a major tourism destination for domestic and international visitors, combining several different purposes such as leisure and recreation, business and professional, visiting friends and relatives, health, study, events.

The first test results in the following spider plot, with the four pillars computed in the province of Milan for the whole year 2021, the Tourism Sustainable Index has been computed as a weighted average of the four pillars. As the TSI, each pillar or sub-component score has a value in a range between 0 and 100: a higher score means a higher level of sustainability.

Destination Pillar	Overtourism Pillar	Social Cultural Pillar	Environment Pillar	TSI
76.29	41.42	62.35	27.69	47.07

Figure 1. TSI and its pillars 2021 - Province of Milan

A selection of snapshots from the front-end tool has been presented to deep dive into the target area and explore one component per TSI pillar: Tourism Pressure (Overtourism Pillar), the Good Health Index (Social and Cultural Impact Pillar) and Air Quality Index (Environment Pillar) for the county of Milan in the year 2021. As well as the TSI, higher the score or a darker the tile, a higher level of sustainability would correspond.

In Figure 2 the Tourism Pressure indicator has been applied: it measures the ratio of tourism online contents on residents. The component shows how the center of Milan has a higher tourism pressure compared to its peripheral area, which is exactly corresponding with the official tourism data flows registered by the statistics authority.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0182.v1

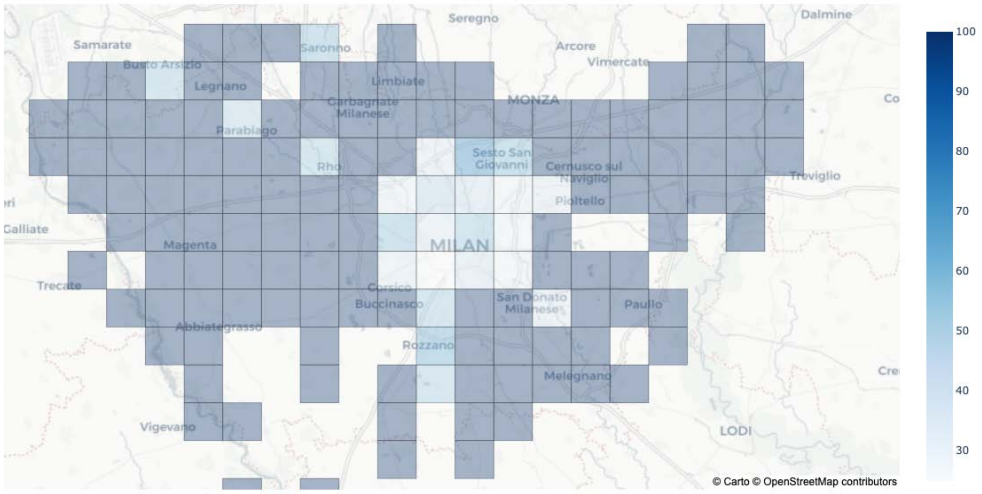

Figure 2. Tourism Pressure (Overtourism Pillar) 2021 - Province of Milan

In Figure 3 the Good Health Index has been applied: it measures the concentration, quality, and e-reputation of the healthcare centers. The best medical centers of this area are located in the two darkest tiles, which correspond with the area of 5 hospitals, which are listed in the World Best Hospitals 2022 Report [49].

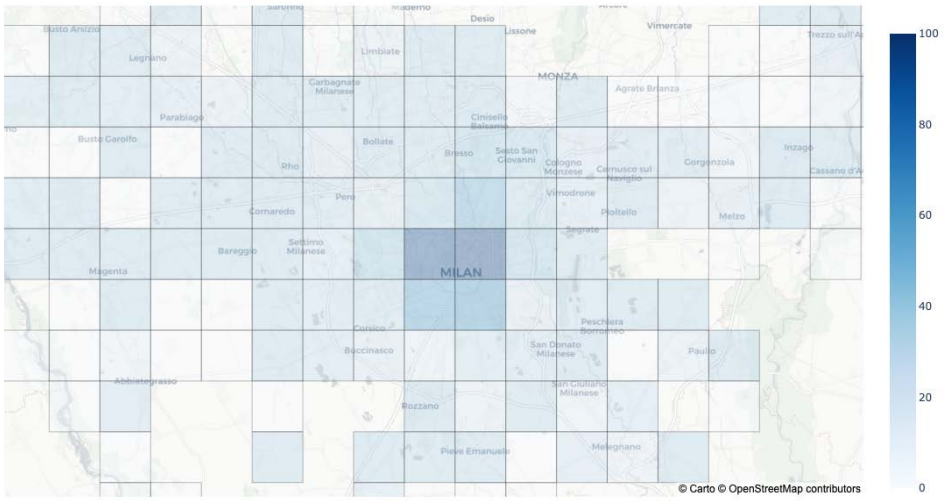

Figure 3. Good Heatlh Index (Social Cultural Pillar) 2021 - Province of Milan

In Figure 4 the Air Quality Index has been applied: it measures the concentration in terms of several air pollutants: Pm2.5, Pm10, NO2, SO2, O3. It is evident how the town of Milan, as well as some other industrial towns, suffered from bad quality air, corresponding with the results of a recent analysis of Legambiente, reporting that Milan was one of the most polluted cities of the entire Italy [50].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0182.v1

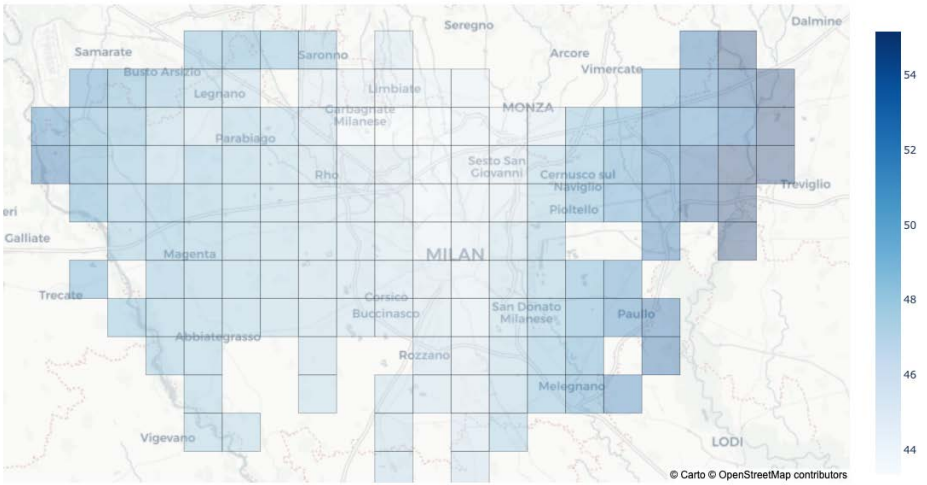

Figure 4. Air Quality Index (Environment Pillar) 2021 - Province of Milan

4. Discussion

The first tests applying TSI to a heterogenous area as the province of Milan, lead to results which find a direct correspondence with other data sources, confuting what has been reported, or evaluated by other entities and research centers. Assessing actual the level of sustainability for any tourism destination worldwide would be easily reached without providing any data by destinations themselves, which was one of the main obstacles deriving from ETIS approach. The managerial implications the authors outline for DMOs, Tourism Authorities and private business managers, have been exemplified by data visualization of the application of the TSI to other tourism-related areas.

4.1. Destination Analysis

Computing the TSI and its sub-components on a Bing tile system, determines the possibility to assess sustainability to any area, even for destinations whose boundaries don't necessarily follow administrative borders. In fact, the literature has pointed out that administrative boundaries, or other political boundaries, may potentially divide destination, in particular regarding natural areas, and potentially hindering tourism development [51]. A most related consumers' perspective should be taken into consideration, even if instances of actual attempts of structuring tourism geographies on the basis of visitors' consumption patterns are scant [52].

Moreover, within a destination, tourism activity may be distributed unequally: being highly concentrated in some sub-areas and some others almost not being affected. The level of detail chosen by the authors, Bing Tile size of 16, permits to deepen the sustainability analysis at sub-destination level, which is a fundamental plus for any strategic and operative planning. In figure 5 an example of the application on the areas of the municipalities of Milan and Florence of one of the TSI sub-component, Tourism Pressure, i.e. the ratio between visitors and residents: lighter the tile, higher the pressure. The data visualization immediately shows how much different neighborhoods can diverge when assessed and the consequent significance of a sub-destination level measurement in terms of managerial implications.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0182.v1

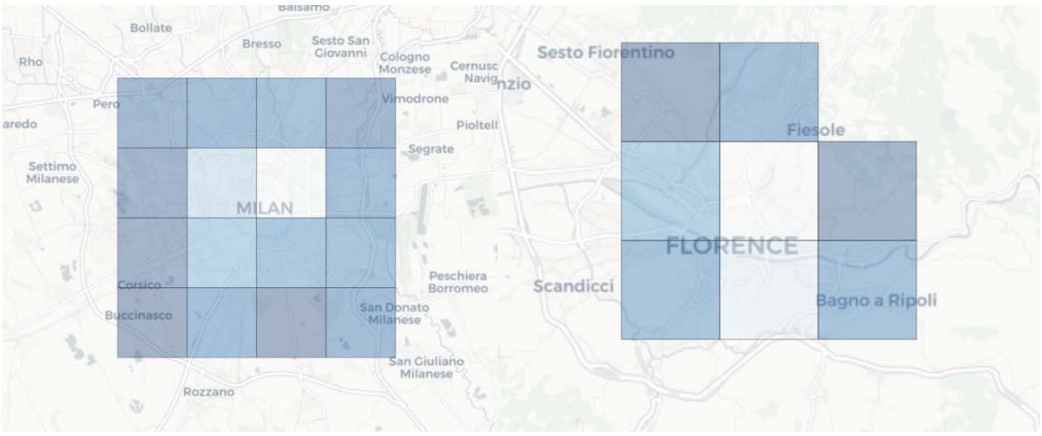

Figure 5. Tourism Pressure - Municipalities of Milan and Florence

Furthermore, the monthly update permits a real time analysis of the evolution of the sustainability assessment, both for destination as a whole or at sub-destination level. Figure 6 shows the application of the Overtourism pillar on the municipality of Florence in the same month over the last 3 years: the lighter the tile, the higher the overtourism level.

Figure 6. Overtourism Pillar (2019-2021) - Municipality Florence

4.2. Destinations Comparison

Comparing sustainability assessment between any type of tourism destinations with a monthly update is still something unseen. The methodology proposed allows not only to assess a holistic indicator as the TSI to be compared, but also its sub-components, to benchmark the different dimensions of tourism sustainability. Figure 7 reports the application of the TSI four pillars to a well known comparison for the ones studying Italian tourism: Florence vs Venice. Both the cities reveal a similar quadrilateral shape in the spider plot: the historical city of Venice shows a higher grade of sustainability on the environmental and social-cultural aspects, while Florence registers a higher level of sustainability on the over tourism and destination pillars.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0182.v1

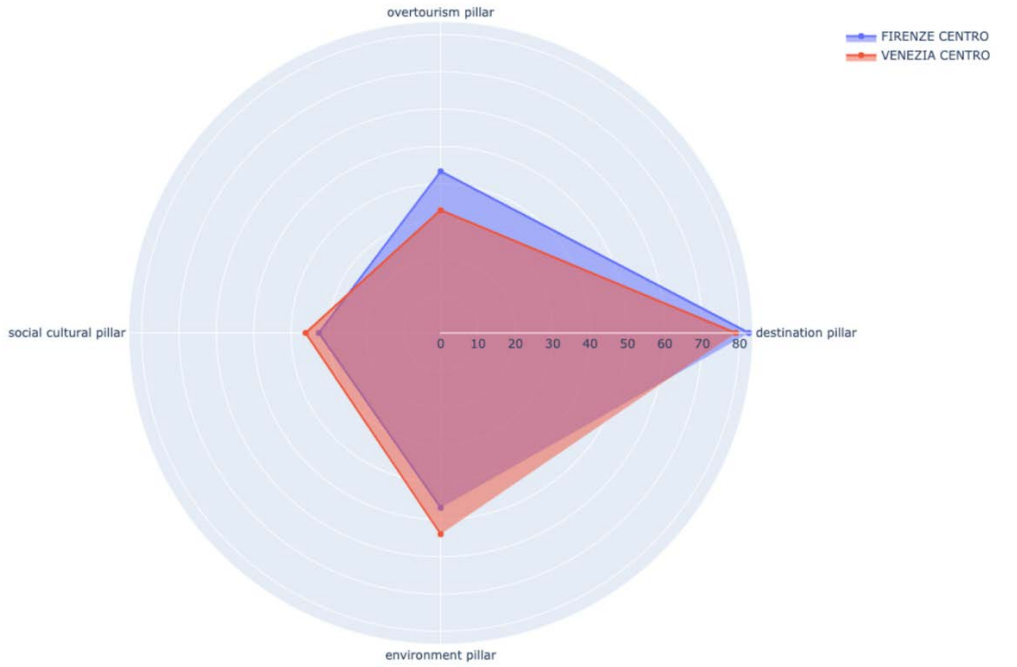

Figure 7. TSI pillars - Venice vs Florence

A second comparison was applied to the two places holding the next Olympic Winter Games in 2026: Milan and Cortina. In this case the quadrilateral shape in the spider plot is quite different between the two destinations. On a managerial perspective, Cortina and Milan seem like they almost compensate each other in terms of sustainability, and this could be interpreted as a new way also to assess the model of joint candidatures or tourism projects involving different destinations.

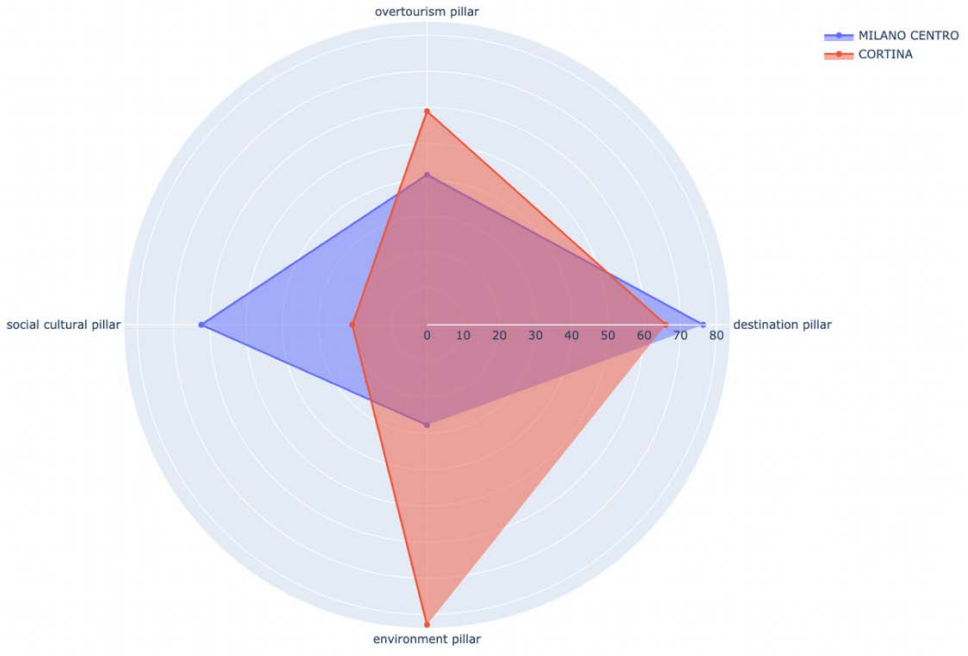

Figure 8. TSI pillars - Milan vs Cortina

 doi:10.20944/preprints202205.0182.v1

5. Conclusions

In this paper a pragmatic and original approach to the definition of a Tourism Sustainability Index (TSI) is presented. Having the ETIS model as reference and bearing in mind the challenges of the ETIS applicability, the authors proposed to use a combination of Open Data Sources and Proprietary Data based on Sentiment Analysis to achieve a scalable and monthly updated index. The core of the proposed index is to estimate both the objective and subjective parameters that can affect the sustainability of a destination. While the objective parameters can be deduced by the analysis of a set of Open Data Sources, the subjective parameters are mostly based on the Sentiment Analysis that can be carried on the social media contents related to a certain number of POIs within any typology of tourism destination. Spatially, the index can be computed both at destination level (i.e., with respect to the portion of territory of tourist interest) and at tile level, using the Bing Maps Tile System. This type of analysis can be useful for the monitoring, comparing, and planning sustainability policies that the entitled DMOs and private business can implement.

The ongoing validation process, consisting in reviews by domain experts for the methodology and the possible managerial implication, will be followed by a first implementation within 2022. At the same time, authors are already following paths to extend the effectiveness and context of the sustainability analysis by integrating new data sources and refining the semantic analysis tools used to create the TSI. With reference to the ability to perform benchmark analysis between different destinations or tiles, a possible evolution could consider the clustering of tourism destinations to identify any common patterns, beyond the actual product-oriented classification.

6. Patents

Author Contributions: Conceptualization, D.DM. and R.B..; methodology, D.DM. and R.B.; validation, D.DM., R.B. and L.DS.; formal analysis, D.DM., R.B. and L.DS.; data visualization, L.DS.; resources, D.DM., R.B. and L.DS.; writing—original draft preparation, D.DM., R.B. and L.DS.; writing—review and editing, D.DM. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: "Not applicable"

Acknowledgments: This study was supported by the Data Appeal Company.

Conflicts of Interest: "The authors declare no conflict of interest."

Abbreviations

DMO: Destination Management Organization
ECMWF: European Centre for Medium-Range Weather Forecasts
ESA: European Space Agency
ETIS: European Tourism Indicators System
EUMETSAT: European Organization for the Exploitation of Meteorological Satellites
MICE: Meetings, Incentives, Conferences and Exibitions
NER: Name Entity Recognition
OTA: Online Travel Agency
POI: Point Of Interest
TDAC: The Data Appeal Company
TSI: Tourism Sustainability Index
UGC: User-Generated Content
UN: United Nations
UNWTO: World Tourism Organization

References

1. UNWTO. Indicators for Sustainable Development of Tourism Destinations: A Guidebook; World Tourism Organization: Madrid, Spain, 2004

2. Torres-Delgado, A.; Palomeque, F. The growth and spread of the concept of sustainable tourism: The contribution of institutional initiatives to tourism policy, Tourism Management Perspectives, 2012
3. Alfaro Navarro, J.; Martínez, M.A.; Mondéjar Jiménez, J. An approach to measuring sustainable tourism at the local level in Europe, Current Issues in Tourism, 23:4, 2020; pp. 423-437
4. Lozano-Oyola, M.; Blancas, F.J.; González, M.; Caballero, R. Sustainable tourism indicators as planning tools in cultural destinations, Ecological Indicators, Volume 18, 2012; pp. 659-675
5. European Commission. The European Tourism Indicator System—ETIS Toolkit for Sustainable Destination Management, Publications Office of the European Union: Luxembourg, 2016
6. Zabetta, M.C.; Sacerdotti, S.L.; Mauro, S. Community-based monitoring in tourism sector: An application of the European Tourism Indicators System, in the "ATL del Cuneese". Leadership and Governance for Sustainable Tourism: Proceeding of Summer School, 2014
7. Tudorache, D.M.; Tamara, S.; Frenț, C.; Mărioara M. Difficulties and Challenges in Applying the European Tourism Indicators System (ETIS) for Sustainable Tourist Destinations: The Case of Braşov County in the Romanian Carpathians" Sustainability 9, no. 10: 1879, 2017
8. Modica, P.; Capocchi, A.; Foroni, I.; Zenga, M. An assessment of the implementation of the European tourism indicator system for sustainable destinations in Italy. Sustainability, 10(9), 2018
9. Font X., Torres-Delgado A., Crabolu G., Martinez J., Kantenbacher J., Miller G., The impact of sustainable tourism indicators on destination competitiveness: the European Tourism Indicator System, Journal of Sustainable Tourism, 2021
10. Butler, R. Tourism, Environment, and Sustainable Development. Environmental Conservation, 18(3), 1991; pp. 201-209
11. Canestrelli, E.; Costa, P. Tourist carrying capacity: A fuzzy approach. Annals of Tourism Research 18, 1991; pp. 295-311
12. De Marchi, D.; Manente, M. Key themes for tourism development management, in Overtourism: Tourism Management and Solutions (1st ed.). Routledge, 2019 https://doi.org/10.4324/9780429197987
13. Tokarchuk, O.; Barr, J.; Cozzio, C.How much is too much? Estimating tourism carrying capacity in urban context using sentiment analysis, Tourism Management, 2022
14. Slovic, P. Trust, Emotion, Sex, Politics, and Science: Surveying the Risk Assessment Battlefield, 1997, University of Chicago Legal Forum: Vol. 1997, Article 4
15. Slovic, P.; Peters, E. Risk perception and affect. Curr. Direct. Psychol. Sci. 15, 2006; pp. 322–325
16. Loewenstein, G. F.; Weber, E. U.; Hsee, C. K.; Welch, N. Risk as feelings. Psychological Bulletin, 127(2), 2001; pp. 267–286
17. Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2007). The affect heuristic. Eur. J. Operat. Res. 177, 1333–1352
18. Tuten, Tracy L. Advertising 2.0: Social Media Marketing in a Web 2.0 World. Westport, Conn: Praeger, 2008
19. Duan, W., Gu, B., Whinston, A.B. The dynamics of online word-of-mouth and product sales-An empirical investigation of the movie industry, Journal of Retailing, vol. 84, no. 2, 2008; pp. 233-242
20. Forman C., Ghose A., Wiesenfeld, B. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets, Information Systems Research 19, 2008; pp 291-313
21. Elwalda, A.; Lü, K.; Ali, M. Perceived derived attributes of online customer reviews, in Computers in Human Behavior 56, 2016; pp. 306-319
22. Chen, R.; He, F. Examination of brand knowledge, perceived risk and consumers' intention to adopt an online retailer, Total Quality Management & Business Excellence, 14:6, 2003; pp. 677-693
23. Zhu, F.; Zhang X. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics, J. Mark. 74, 2010; pp. 133–148
24. Mudambi, S.M.; Schuff D. What makes a helpful online review? A study of customer reviews on amazon.com, MIS Q. 34, 2010; pp. 185–200
25. Wolff, K.; Larsen, S.; Øgaard, T. How to define and measure risk perceptions, 2019, Ann. Tourism Res. 79
26. Gretzel U., Yoo KH. Use and impact of online travel reviews. Information and communication technologies in tourism 2008, 2008; pp. 35-46
27. Almeida-Santana, A.; Moreno-Gil, S. New trends in information search and their influence on destination loyalty: Digital destinations and relationship marketing, Journal of Destination Marketing & Management, Volume 6, Issue 2, 2017; pp. 150-161
28. Yachin, J.M. The 'customer journey': Learning from customers in tourism experience encounters, Tourism Management Perspectives, 2018
29. Kang, D.; Park, Y. Review based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 2014; pp. 1041–1050
30. Prabowo, R.; Thelwall, M. Sentiment analysis: A combined approach. Journal of Informetrics, 2013; pp. 143-157
31. Alaei, A.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research. 58, 2017
32. De Marchi, D.; Lalli, M.; Mancini, A. Monitoring online perception of environmental issues on coasts of Sicily, in Proceedings e report, Firenze University Press, 2020; pp. 214-218 https://doi.org/10.36253%2F978-88-5518-147-1.21
33. Franciele, C.; Domareski Ruiz, T.C. Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review. Journal of Tourism, Heritage & Services Marketing, Tourlab, the International Hellenic University,7 (2), 2021; pp.16-27
34. Sustainability as the new normal. Available online: https://www.unwto.org/covid-19-oneplanet-responsible-recovery (accessed on 23/04/2022)

35. The Sustainable Tourism Index. Available online: https://impact.economist.com/perspectives/sites/default/files/Sustainable_Tourism_Index.pdf (accessed on 23/04/2022)
36. Tourism Sustainable Development Index. Available online: https://business.esa.int/projects/sustainable-tourism-indicators (accessed on 23/04/2022)
37. Global Destination Sustainability Index. Available online: https://www.gds.earth/wp-content/uploads/Methodology-2022.pdf (accessed on 23/04/2022)
38. Torres-Delgado, A.; Palomeque, F. The ISOST index: A tool for studying sustainable tourism. Journal of Destination Marketing & Management, 8, 2017
39. The Data Appeal Company. Available online: https://www.datappeal.io/about-us/ (accessed on 23/04/2022)
40. Doaa Mohey El-Din, M.H. A survey on sentiment analysis challenges, Journal of King Saud University - Engineering Sciences, Volume 30, Issue 4, 2018; pp. 330-338
41. Blancas Peral, F.; Oyola, M. Sustainable tourism evaluation using a composite indicator with different compensatory levels. Environmental Impact Assessment Review. 93, 2022
42. Bing Maps Tile System. Available online: https://docs.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system (accessed on 23/04/2022)
43. Copernicus - Europe's eyes on Earth. Available online: https://www.copernicus.eu/en (accessed on 23/04/2022)
44. The European Space Agency. Available online: https://www.esa.int/ (accessed on 23/04/2022)
45. EUMETSAT. Available online: https://www.eumetsat.int/ (accessed on 23/04/2022)
46. ECMWF. Available online: https://www.ecmwf.int/ (accessed on 23/04/2022)
47. The World Bank - IBRD-IDA - Data. Available online: https://data.worldbank.org/ (accessed on 23/04/2022)
48. Appeal is real. Available online: https://www.datappeal.io/appeal-is-real/ (accessed on 23/04/2022)
49. World Best Hospitals 2022 report. Available online: https://www.newsweek.com/worlds-best-hospitals-2022 (accessed on 23/04/2022)
50. Legambiente: malaria di città. Available online: https://www.legambiente.it/wp-content/uploads/2021/11/malaria-di-citta-2022.pdf (accessed on 23/04/2022)
51. Paulino Valldepérez, I, Redefining Tourism Destination Boundaries from a Consumption-based perspective, 2020
52. Blasco, D.; Guia, J.; Prats L. Tourism destination zoning in mountain regions: a consumer-based approach, Tourism Geographies, 16:3, 2014; pp. - 512-528, DOI: 10.1080/14616688.2013.851267

Authors' Biographies

Damiano De Marchi is Senior Analyst and Tourism Domain Expert at The Data Appeal Company. He received a master's degree in Tourism Economics and Management at Ca' Foscari University of Venice in 2005. Since then, he has been working in the tourism industry in Italy and abroad in private companies, institutions and research centers mainly involved in strategic and operative analysis, planning and project management. Lecturer and professional trainer since 2010, in 2019 has been hired by UNWTO as external expert on statistics to strengthen the system of tourism statistics in Myanmar and Tajikistan. His current main interests of applied research are statistics, tourism sustainability and destination management.

Rudy Becarelli, PhD - Leading Data Scientist at The Data Appeal Company. He received an MA degree at Università degli Studi di Firenze - School of Engineering (2001) in Video Motion Estimation techniques and a PhD degree at Università degli studi di Firenze - School of Engineering (2016) in Image and Video Authentication. He spent 15 years in scientific research on image and video processing. Its activities included also Big Data (Images and Video) storage and server-side provisioning. His current main interest are Big Data Analysis and Key Feature Engineering for Location Intelligence and Geospatial Analysis.

Leonardo Di Sarli is Data scientist at The Data Appeal Company since February 2021. He received a master's degree in data science and business informatics at the University of Pisa (2019), then he worked in consulting companies as data scientist and business intelligence consultant into finance, luxury, and public administration. He is interested in Data Science, Machine Learning and Location Intelligence.