

Chapter 19

New Data and Computational Methods Opportunities to Enhance the Knowledge Base of Tourism



Gustavo Romanillos and Borja Moya-Gómez

Abstract Tourism is becoming increasingly relevant at different levels, intensifying its impact on the environmental, the economic and the social spheres. For this reason, the study of this rapidly evolving sector is important for many disciplines and requires to be quickly updated. This chapter provides an overview and general guidelines on the potential use of new data and computational methods to enhance tourism's knowledge base, encourage their institutional adoption and, ultimately, foster a more sustainable tourism.

First, the chapter delivers a brief review of the literature on new data sources and innovative computational methods that can significantly improve our understanding of tourism, addressing the big data revolution and the emergence of new analytic tools, such as artificial intelligence (AI) or machine learning (ML). Then, the chapter provides some guidelines and applications of these new datasets and methods, articulated around three topics: (1) measuring the environmental impacts of tourism, (2) assessing the socio-economic resilience of the tourism sector and (3) uncovering new tourists' preferences, facilitating the digital transition and fostering innovation in the tourism sector.

19.1 Introduction

Tourism is playing an increasingly important role at many levels, and its sector is evolving extraordinarily fast. Thus, the study of tourism, crucial for numerous disciplines, needs to be quickly updated. During the last few years, tourism research is starting to be renovated to keep pace with the ongoing transformations. Nowadays, new data sources and innovative quantitative and qualitative methods offer new possibilities for better analysing and planning tourism (Xu et al., 2020), overcoming many limitations of more conventional approaches.

G. Romanillos (✉) · B. Moya-Gómez
tGIS, Department of Geography, Universidad Complutense de Madrid, Madrid, Spain
e-mail: gustavro@ucm.es; bmoyagomez@ucm.es

Although recent tourism research is exploring and taking advantage of new data sources and methods, there is still a long way to walk on innovation. This chapter aims to provide a general review and some guidelines on the potential use of new data and computational methods to enhance tourism's knowledge base and promote their institutional adoption and, ultimately, more sustainable tourism.

The chapter is articulated around three topics proposed by Barranco et al. and included in a publication of the Joint Research Centre that aimed at collecting the upcoming research needs in terms of policy questions (Bertoni et al., 2022). The first one measures the environmental impacts of tourism. Tourism is, directly and indirectly, consuming an increasing amount of global resources, including fossil fuel consumption with the associated CO₂ emissions, freshwater, land and food use (Gössling & Peeters, 2015). Therefore, assessing the impact of global tourism activity is one of the most relevant potential applications of new data sources and computational methods.

The second topic is assessing the socio-economic resilience of the tourism sector. Tourism economic weight and social impact have become even more evident in the context of the COVID-19 pandemic: the crisis has put between 100 and 120 million direct tourism jobs at risk, many of them in small- and medium-sized enterprises, according to the UNWTO (2021). Hence, it is relevant and urgent to explore how new data sources, analyses and models can contribute to planning a more resilient and balanced tourism sector in socio-economic terms.

Finally, the third topic is uncovering new tourists' preferences, facilitating the digital transition and fostering innovation in the tourism sector. How can we better analyse new tourist patterns? COVID-19 may have accelerated some existing changes in tourism trends, so there is an urgent need for quick analyses and predictions for the very near future, as the emergence of nowcasting techniques evidences it.

19.2 Existing Literature

Over the past years, new data sources and innovative computational methods emerged to significantly improve our understanding of tourism. A summary is provided next.

19.2.1 New Data Sources of Potential Interest for Tourism

Tourism is being transformed at an accelerated pace, and conventional data sources often do not reflect ongoing changes with enough velocity or spatiotemporal resolution to support the urgent studies to be carried out. In this scenario, new data sources emerge as the raw material to open further explorations in tourism.

New datasets can be grouped into different categories according to data sources. Next, a listing of relevant new datasets is provided, classified according to the nature of the data source and its potential interest for tourism studies, offering some specific examples.

First, we must point out big data from specific sources of the tourism sector, such as smart tourism cards or information systems in destinations. These sources provide data directly recorded in tourism points of interest, valid to monitor existing activity and analyse current or past trends. In this category, we can also include other data sources such as booking data from transportation companies (especially flight booking data from airline operators), which can help predict tourism activity quickly, feeding nowcasting models. Additionally, we can highlight online accommodation companies and apps, such as Tripadvisor¹ or Booking,² or new peer-to-peer accommodation online services such as Airbnb³ (Calle Lamelas, 2017). These sources can be helpful not only for anticipating tourism demand but also because of the additional information collected from users, such as opinions, ratings, comments, etc. In addition, data with high spatiotemporal resolution allows us to analyse emergent spatial patterns, for instance, in the location of Airbnb accommodation in heritage cities (Gutiérrez et al., 2017).

Second, it is remarkable the potential use of GPS datasets. GPS data was actually ranked the top of big data in tourism research (accounting for 21%) and the first of device data (58%) according to the classification provided by Li et al. (2018). It is essential to use GPS tracks to study tourists' routes, with an unprecedented level of detail, thanks to the high spatiotemporal resolution of GPS records. In this group, we include the GPS routes recorded by vehicle navigation apps, such as TomTom,⁴ Waze⁵ or Google Maps,⁶ and tracking apps such as Wikiloc⁷ or Strava,⁸ very useful when analysing tourism in natural areas, for instance (Barros et al., 2019), or GPS data collected through the emerging tourist mapping apps (Brilhante et al., 2013; Gupta & Dogra, 2017).

Third, it is also outstanding the interest of user-generated content (UGC), especially datasets obtained from social networks such as Twitter; photo-sharing social networks such as Instagram,⁹ Flickr¹⁰ or Panoramio¹¹; or apps focused on the location of points of interest, such as Foursquare.¹² UGC allows us to explore

¹ Tripadvisor website, for more information: <https://www.tripadvisor.com/>

² Booking website, for more information: <https://www.booking.com/>

³ Airbnb website, for more information: <https://www.airbnb.com/>

⁴ TomTom website, for more information: <https://www.tomtom.com/>

⁵ Waze website, for more information: <https://www.waze.com/>

⁶ Google Maps website, for more information: <https://maps.google.com/>

⁷ Wikiloc website, for more information: <https://www.wikiloc.com/>

⁸ Strava website, for more information: <https://www.strava.com/>

⁹ Instagram website, for more information: <https://www.instagram.com/>

¹⁰ Flickr website, for more information: <https://www.flickr.com/>

¹¹ Panoramio website, for more information: <https://www.panoramio.com/>

¹² Foursquare website, for more information: <https://foursquare.com/>

different tourism dynamics. Semantic analysis of online textual data, such as tweets or travelling blog content, can uncover tourism preferences and trends (Ramanathan & Meyyappan, 2019). Spatial or temporal analyses can also be carried out because most users share data through mobile apps that register GPS coordinates. For instance, Flickr data can be the basis for different temporal analyses, such as estimating tourism demand over a day according to time slots or measuring tourism seasonality in national parks (Barros et al., 2019); also Twitter and Foursquare data can support spatial analyses, such as the identification of multifunction or specialised tourist spaces in cities (Salas-Olmedo et al., 2018) (Fig. 19.1).

Fourth, search engines' data constitute a precious data source, such as Google Trends records. Considering that search engines are a leading tool in planning vacations (Dergiades et al., 2018), these datasets provide information on tourists' interests and plans in advance and can feed models oriented to forecasting tourist arrivals (Havranek & Zeynalov, 2021).

Fifth, we must highlight the interest of datasets obtained from diverse information and communication technologies/devices. The rapid development of the Internet of Things (IoT) provides an increasing amount of Bluetooth data, RFID data and Wi-Fi data (Shoval & Ahas, 2016), which can be helpful to measure, for instance, tourist presence and consumer behaviour over time. Also, in this group, we must emphasise mobile phone data due to its potential use at different scales and for various purposes. The COVID-19 pandemic has accelerated the adoption of mobile phone data to monitor changes in tourism or general mobility trends with a high level of spatiotemporal resolution (Romanillos et al., 2021). This analysis may be extended beyond national borders. Nowadays, roaming services have become crucial for tourists, and roaming data allows us to track tourists globally. Lastly,

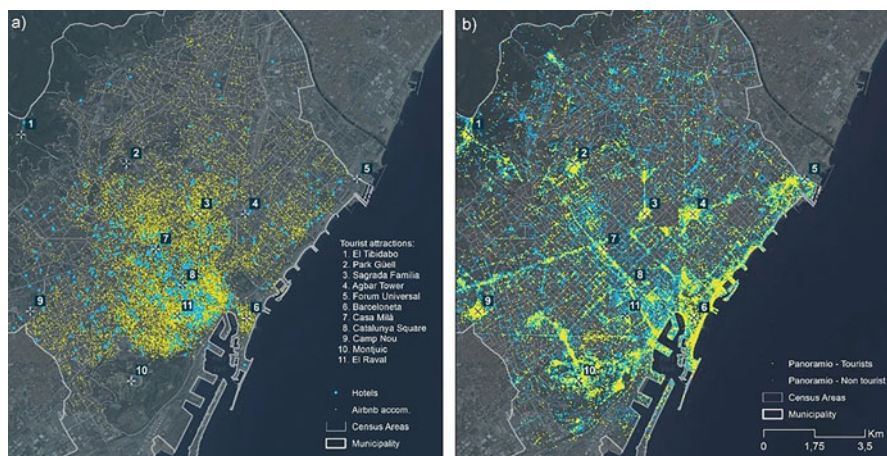


Fig. 19.1 Location of hotel and Airbnb offers (a) and density of photographs taken by tourists and residents (b) in Barcelona. Source: Gutiérrez et al. (2017)

credit card datasets should also be included here, given their potential for tourist consumption and behaviour analyses.

Finally, more conventional data sources can also provide “new” datasets and opportunities, due to improvements in the quality of data or the way data is shared, in real time, through mobile apps and online services. For example, it is the case of meteorological data. Given that weather is an essential factor in tourism demand, incorporating meteorological variables in tourism forecasting models can increase the predictability of tourist arrivals (Álvarez-Díaz & Rosselló-Nadal, 2010).

19.2.2 *New Computational Methods with Application to Tourism Studies*

In recent years, increased computational capacity, part of the *big data* revolution, has allowed for faster and cheaper analysis of massive databases by using new analytic tools, such as *artificial intelligence* (AI) or *machine learning* (ML). Nowadays, tourism analysts may also access an enormous collection of methods for their studies (some are comprehensible; others are like “black boxes”). This section gives a brief and non-exhaustive list of computational methods used in tourism studies, applications and examples.

Unsupervised techniques can identify groups and relationships by analysing explanatory variables themselves: no already known responses exist. Outcomes must be validated – are they logic? – tagged and hypothesised. In tourism, *clustering* techniques were used for detecting the spatial patterns of new touristic accommodations (Carpio-Pinedo & Gutiérrez, 2020) or exploring topics of online tourists’ reviews (Guo et al., 2017), *factor analysis* for uncovering latent motivational and satisfaction variables in tourist (Kau & Lim, 2005) and *association rules mining/learning* for discovering the most frequent and strong sets of visited places with Bluetooth data (Versichele et al., 2014).

Supervised techniques provide models to explain/predict responses. They need complete observations: explained (response) and explanatory variables. Outcomes must be compared to observed datasets. Some models investigate causalities and hypothetical “what-if” scenarios (key results are model’s parameters): *linear regressions* for inferring causes on tourism industry employment and retention (Chen et al., 2021) or *structural equation models* (SEM) for modelling the quality of life in a tourist island (Ridderstaat et al., 2016). Other models, especially AI-based techniques, anticipate responses or classify observations (key results are responses): *autoregressive moving average* (ARMAX) *time series models* for forecasting weekly hotel occupation with online search engine queries and weather data (Pan & Yang, 2017) or *artificial neural networks* (ANN) for predicting tourist expenditures (Palmer et al., 2006).

Some datasets need to be treated before applying the above methods, especially for reusing datasets from other studies or online sources. Observations must be

regrouped into another spatial or temporal unit. While aggregating is a straightforward procedure, disaggregating data needs the use of other techniques; see estimating visitor data from regional to municipality scope (Batista e Silva et al., 2018).

Finally, data and models' outcomes need to be presented and stand out to the target public. They can be shown using innovative designs (word clouds, cartograms, etc.), such as the United Nations World Tourism Organization (UNWTO) tourism data dashboard (UNWTO, n.d.). Part of them should be used on digital social networks or in other analysis processes.

19.3 Guidelines

This section proposes some guidelines and potential applications of the described new data sources and computational methods to the three main topics mentioned in the introduction.

19.3.1 *Assessing the Environmental Impacts of Tourism*

To facilitate the green transition in the tourism sector, we need a concrete EU roadmap with a solid framework and measurable objectives. Working with key performance indicators (KPIs) can help guide and commit the tourism industry and destinations. This section aims to propose a set of KPIs related to central topics regarding the environmental impact of tourism, focusing on new data sources and computational methods.

The first topic concerns tourism mobility. Sustainable tourism should be linked to a concept of sustainable mobility, so we propose a set of KPIs that can reveal to what extent we are advancing in the transition to a more sustainable model (Table 19.1).

The second topic is tourism land consumption. As a consequence of the growth of tourism activity, land in tourist destinations is progressively occupied and degraded. Essential variables in this degradation process are land occupation, land fragmentation and changes in land-use patterns. We propose a set of KPIs that can improve the monitorisation of these variables, with the help of new data sources and methods (Table 19.2).

Finally, the third topic is tourism resources consumption and management. The increasing number of tourists leads to dramatic growth in the consumption of local resources, often leading to unsustainable scenarios. Next, a set of KPIs is proposed to help evaluate tourism resources consumption with the support of new data sources and methods (Table 19.3).

Table 19.1 KP Is for tourism mobility

Topic	KPIs' description [<i>units</i>] (from <i>low</i> (+) to <i>high feasibility</i> (+++))
Tourism mobility	<p><i>National/International tourist trips made by plane, according to distance ranges [%] (+++)</i></p> <ul style="list-style-type: none"> • The lowest the percentage, the lower the environmental impact of the tourism mobility model, since planes are generally the most pollutant transportation mode. Around 40% of tourism emissions are attributed to air travel (Higham et al., 2019) • Mobile phone data can be used to differentiate transportation modes by geolocating call detail records (CDR) and by map-matching estimated routes with transportation networks. Especially effective at the national level but increasingly feasible at the international one with the help of roaming databases (Ahas et al., 2018) • Data from transportation operators, especially airlines • Feasibility: high. Although monitorisation of different mobility and transport dynamics (including tourism) based on mobile phone data is still being adopted in other countries (at a very different pace), data from transportation operators are easily accessible
	<p><i>National/international tourist trips made by collective transport modes [%] (+++)</i></p> <ul style="list-style-type: none"> • Mobile phone data can be used to differentiate some collective transportation modes from private ones (train, planes or boats) but not correctly others (bus) • Data from transportation operators, car rental companies, etc. • Feasibility: high, for the same reasons explained above
	<p><i>Proximity tourism, below a threshold to be defined [%] (++)</i></p> <ul style="list-style-type: none"> • Mobile phone data can be effectively used to differentiate tourist trips, analyse trip distances and estimate tourist flow • Trends in the present and the near future can be estimated through data from online accommodation companies and apps, photo-sharing social networks, online accommodation companies and apps and search engines, such as Google Trends records • Feasibility: moderate–low. Monitorisation tourism based on mobile phone data is not yet adopted by all EU countries, or it does not have the necessary spatiotemporal resolution
	<p><i>Tourism active mobility [%] (++)</i></p> <ul style="list-style-type: none"> • Trekking and cycling trends can be estimated through data from tracking apps, such as Strava, and photo-sharing social networks (Barros et al., 2019) • Regarding urban tourism, cycling trends can be explored through bike share systems records and other tracking and navigation apps such as Google Maps. In addition, walking is starting to be explored through mobile phone data (Hunter et al., 2021) • Mobile phone data can be effectively used to differentiate trekking and cycling activity • Feasibility: moderate. Accessibility to the mentioned data sources depends on partnerships with private companies

(continued)

Table 19.1 (continued)

Topic	KPIs' description [<i>units</i>] (from <i>low</i> (+) to <i>high feasibility</i> (+++))
	<p><i>CO₂ emissions associated with tourist trips [CO₂tonnes]</i> (++)</p> <ul style="list-style-type: none"> • Mobile phone data can be used to differentiate transportation modes and estimate routes and trip distances, before estimating CO₂ emissions • Recently, specific methodologies to estimate the emissions from main touristic activities have been developed (Russo et al., 2020) • Emission inventory is the most reliable and comprehensive data source since it derives from many studies' compilation and analysis • Feasibility: moderate, depending on the accuracy of the estimation desired. Accurate estimations would rely on precise monitorisation of all tourist trips. Some useful data for this purpose (such as mobile phone data) have not been adopted by many EU Member States yet

19.3.2 Socio-Economic Resilience in the Tourism Sector

Tourism is an important sector in the EU economy. EU's tourists spent about \$400 billion on trips across Europe before COVID-19 (Eurostat, 2021b). In 2016, tourism was 10% of the EU's GDP, and it employed 10% of workers in 3.2 million tourism-related enterprises (Eurostat, 2018). However, the tourism sector has high levels of temporal contracts and low retention rates (25%), women employment (~60%), younger workers 15–24 years old (~20%), lower educated workers (~20%) or foreign workers (~1/6) compared to other sectors.

The following KPIs can help key stakeholders assess their tourist offers and benchmark with competitors. These indicators could identify socio-economical relationships, vulnerabilities and weaknesses, undeveloped attractions and upcoming opportunities to make a more resilient sector. KPIs' spatiotemporal dimensions are essential, especially for regions characterised by stationarity. These KPIs should be calculated for several periods, for the whole touristic population in a location (descriptive) or the whole/specific touristic population in competitors (comparison).

This first group of KPIs points out socio-economic impacts of tourism in a region that can be used for comparing them with other industries and competitors (Table 19.4). Some of these KPIs measure tourism impacts directly, but others estimate effects through related activities.

The second set of KPIs concerns assessing tourist models' diversity for detecting excess dependencies on a few attractions and tourist profiles and their stationarity (Table 19.5). Less diverse territories might be very vulnerable to changes in the tourism demand, wildly unexpected events or incompatible weather, among other cases.

Table 19.2 KPIs for tourism land consumption

Topic	KPIs' description [<i>units</i>] (from <i>low</i> (+) to <i>high feasibility</i> (+++))
Tourism land consumption	<p><i>Land occupation related to tourism [%]</i> (+++)</p> <ul style="list-style-type: none"> • In addition to conventional approaches, based on the analysis of satellite/aerial images, new methodologies based on new data sources can help differentiate tourist developments from residential ones and better assess tourism's particular impact • Mobile phone data and diverse online social networks can be analysed over time to discriminate tourism activity across the territory (Salas-Olmedo et al., 2018) • Feasibility: high. Satellite images are widely available, and characterisation of land occupation can be done based on several – more or less – accessible data sources
	<p><i>Coast occupation related to tourism [%]</i> (+++)</p> <ul style="list-style-type: none"> • Defined as the land occupation in a distance range from the coast, generally 500 metres (Riitano et al., 2020) • Similar data sources and methods as in the percentage of land occupation KPI • Feasibility: high, for the same reasons mentioned above
	<p><i>Land fragmentation indexes related to tourism</i> (+++)</p> <ul style="list-style-type: none"> • The European Environment Agency assesses the general land fragmentation pressure through the Effective Mesh Density (seff) Fragmentation Index • Land fragmentation indexes in tourist areas or the contribution of tourism to general land fragmentation could be estimated by differentiating tourist developments from residential ones using mobile phone and online social networks data • Feasibility: high. The datasets for calculating the most common fragmentation indexes are publicly accessible
	<p><i>Tourism share in real estate market [%]</i> (++)</p> <ul style="list-style-type: none"> • New analyses can be based on data from companies monitoring its evolution daily – new opportunities for nowcasting • Tourist developments can be distinguished from residential ones using mobile phone and online social networks data • Feasibility: moderate. Accessibility to the mentioned datasets depends on partnerships with private companies
	<p><i>Presence of short-term rentals platforms (e.g. Airbnb)</i> (++)</p> <ul style="list-style-type: none"> • At EU27, the share of international guest nights at short-stay accommodation offered via collaborative economy in 2019 was 67% (Eurostat, 2021a) • Analysis based on data from accommodation platforms or associated websites (e.g. Inside Airbnb) • Short-term tourism properties can be discriminated through estimations based on mobile phone data and diverse online social networks data • Feasibility: moderate. Accessibility to the mentioned datasets depends on partnerships with private companies

Table 19.3 KPIs for tourism resources consumption and management

Topic	KPIs' description [<i>units</i>] (from <i>low (+)</i> to <i>high feasibility (++++)</i>)
Tourism resources consumption and management	<p><i>Local product consumption in tourist destinations [%] (++)</i></p> <ul style="list-style-type: none"> • A high percentage is related to the less environmental cost of product transportation and warehousing • Credit card datasets could be the basis for estimating the percentage of consumption in local businesses and markets related to local products • Google points of interest (POI) datasets allow estimating/monitoring people's presence in local businesses and markets, is an approximation of consumption • Feasibility: moderate. Accessibility to the mentioned datasets depends on partnerships with private companies
	<p><i>Water consumption in tourist destinations [litres/person-year] (++++)</i></p> <ul style="list-style-type: none"> • New water consumption datasets registered by water management companies, with high granularity • Innovative analyses and short-term forecasting of water consumption patterns can be done based on these new datasets (Candelieri et al., 2015) • New indicators are proposed considering direct and indirect water consumption, estimated from food consumption data, for instance (Gössling, 2015). In addition, food consumption patterns could be analysed based on credit card datasets • Feasibility: high, due to the increasing accessibility to the mentioned datasets
	<p><i>Electric energy consumption per capita (kWh/person-year) (++++)</i></p> <ul style="list-style-type: none"> • New datasets from electricity companies, which are recording consumption with unprecedented granularity • New electric energy consumer characterisation methodologies are proposed, based on data mining techniques (Pérez-Chacón et al., 2018) • Feasibility: high, for the same reasons explained above
	<p><i>Solid waste generation [kg/person-year] and recycled solid waste generation [%] (++++)</i></p> <ul style="list-style-type: none"> • New smart waste management systems are being proposed. For example, bins equipped with sensors can register geolocated data on total/recyclable waste (Folianto et al., 2015) • Feasibility: high, for the same reasons explained above

19.3.3 *Uncovering New Tourists' Preferences, Digital Transition and Innovation in the Tourism Sector*

New information technologies have revolutionised the tourism sector too. This section introduces how new technologies can be used to detect tourists' preferences and better manage touristic businesses and locations.

Table 19.4 KPIs for socio-economic impact of tourism in a region

Topic	KPIs' description [<i>units</i>] (from <i>low (+)</i> to <i>high feasibility (++++)</i>)
Tourism employment	<p><i>Employment in tourism [%] (++++)</i></p> <ul style="list-style-type: none"> • Job portals include a large sample of available positions. The number of tourism positions can be compared with the whole set • Project Skills-OVATE is an example of getting data for this KPI (CEDEFOP, n.d.) • Feasibility: high; job portals usually have a good sample of positions everywhere
	<p><i>Training/educational activities for tourism activities [%] (++)</i></p> <ul style="list-style-type: none"> • Tourism-related post-compulsory educational offers within a location can point out the tourism industry employment demands • It is necessary to set some affinity between activities and tourism employment • Feasibility: moderate; educational offers usually are spread into many websites
Tourism economies	<p><i>New employee base salaries [€] (++++)</i></p> <ul style="list-style-type: none"> • Job positions on employment portals often include base salaries • This data can point out foreseeable income for employees and their relationship to any other sector's new position offer • Feasibility: high; position offers should include it
	<p><i>Expenditure in destination [€] (+)</i></p> <ul style="list-style-type: none"> • Aggregated credit card data can provide insight also regarding their consumptions • Additional data sources (accommodation websites, etc.) must be scanned to include them in the total expenditures • Feasibility: low; it highly depends on available aggregated bank data. Some payments were paid at home

19.3.4 Analysis of Preference Changes in the Tourism Sector

Businesses may use tourist demand data (accommodation booking, car renting) and users' responses (comments or reviews on products or services) to comprehend the needs of (new) customers to develop and/or to update their products and services and to improve their customer care. While the former may reveal tourist preferences based on their choices, the latter may also highlight some declared unsatisfied ones. Analyses of preference changes need benchmarking approaches; competitor performances provide insight into the strengths and weaknesses of the study location/business. Nevertheless, how may new data and methods aid in the detection of preferences and their changes? Some guidelines are provided next.

19.3.4.1 Searching for Holidays and Activities

Many trips or touristic activities begin with an online search. Potential tourists use either general online search engines or specific touristic planner services. Consequently, data on preferences may be extracted by using autocompletion to

Table 19.5 KPIs for assessing tourism diversity

Topic	KPIs' description [<i>units</i>] (from <i>low</i> (+) to <i>high feasibility</i> (+++))
Tourism offers diversity	<p><i>Type/category of tourism attractions [%]</i> (++)</p> <ul style="list-style-type: none"> • Online travel agencies, trip advisor services and local tourist websites gather helpful information about tourist destinations (what to do, when, where, etc.). Businesses and other users feed them too • Data of those services for specific locations and nearby places and categorising them helps approximate the supplied tourist diversity • Feasibility: moderate; some attractions might not be online
	<p><i>Affinity to be recommended [%]</i> (++)</p> <ul style="list-style-type: none"> • Previous services have also provided users with AI-based recommendations for choosing their destinations • These recommendations can show substitutive and complementarity options • These services can be used to know if any location is recommended for any specific tourism model or demand (search ranking, etc.) • Feasibility: moderate; it needs to create comprehensive groups before undertaking data collection
	<p><i>Open attractions by weather [%]</i> (+)</p> <ul style="list-style-type: none"> • Some tourism attractions need particular short-/mid-term weather conditions • Attractions might post online messages regarding disruptions because of weather conditions • Feasibility: low; small attractions do not post it
Tourism profiles diversity	<p><i>Prominent group(s) of tourists per age/trip type [%]</i> (++)</p> <ul style="list-style-type: none"> • Google Trends or similar can show the most frequent additional queries to some locations (main query) • The additional queries can include target words (children, family, party or elderly) that may describe a type of tourist • Longer queries might not retrieve results for having very few searches with that query • Feasibility: moderate; it needs to create comprehensive groups before undertaking data collection
	<p><i>Prominent group(s) of tourists per activity [%]</i> (++)</p> <ul style="list-style-type: none"> • Some tourists post their experiences, opinions and revision on digital platforms. This usually has spatial data or metadata • Aggregating all this data can be used to get an approximation to identify which type of tourist usually visits any place • Feasibility: moderate; some groups might not be well represented (digital divide)
	<p><i>Prominent group(s) of tourists per origin [%]</i> (+)</p> <ul style="list-style-type: none"> • Aggregated mobile phone data can be linked with the country of the user's operator (international and other territorial prefixes) and local antennas • It is possible to approximate which places are attractive to tourists, by origin, by counting their connections • This procedure can also determine national stationarities (Raun et al., 2016) • This KPI can also be obtained using credit card data • Feasibility: low–moderate; it highly depends on the availability of phone mobile/bank data

suggest current trending complete search queries, or using some services like Google Trends for a similar end to observe variations over time. These tools can use queries from specific countries to help segment tourist preferences per origin while planning their holidays. Search query data has been used in many academic studies; Dinis et al. (2019) gathered and summarised some of them into the following topic categories: forecasting, nowcasting, identifying interests and preferences, understanding relationships with official data and others.

19.3.4.2 Text Is a Mine

People use words to communicate, and they can publicly share their opinions, recommendations, suggestions and complaints towards touristic attractions in interactive platforms. An analyst can use *text mining* techniques, such as *natural language processing (NLP)*, to extract the sense of messages (including emojis) and undertake sentiment analyses (converting text into Likert scale values). However, this data may contain brief messages, with abbreviations, because of character restrictions. They must be translated into expanded statements. Also, fake/compulsive users should be dropped to avoid biases. Finally, text mining techniques have difficulties detecting ironic tones.

19.3.4.3 What a Beautiful Picture!

Some tourists also upload their pictures and videos on digital social networks. Unlike texts, images need to be described before automating processes to extract comprehensible data. Simple methods can summarise colours in pictures (they can explain weather conditions or infer day periods). More advanced ones, available in cloud computing services, can also identify locations, buildings and objects. Thus, pictures transformed into texts and previously mentioned *text mining* techniques can help determine preferences. In addition, images can include description text and comments that can be used to uncover revealed preferences. Finally, images' metadata include when and, sometimes, where they were taken. This data can be used for determining spatial preferences of what to take a picture of and from where (viewpoints).

19.3.4.4 Life Is Change

Tourists' preferences can evolve for many reasons (getting older, having children or new job positions or contextual reasons, among others). To detect these changes, it is required to have previous preferences to compare with the new ones and see significant changes. The above-mentioned methods can continuously process data, get further insights or update continuous datasets.

Notice that many tips and ideas use similar strategies for calculating the KPIs introduced in the previous section. Thus, they may also be reinterpreted to help analyse preference changes or warn regarding new successful tourist strategies in competitive locations by KPIs' variations.

19.3.5 Digital Transition and Innovation

We have seen that using new data sources and computational methods can improve our understanding of tourism dynamics and help plan and develop better tourism policies. However, institutions and companies still have a long way to go to use all these new resources. To accelerate what's been called the digital transition and foster innovation, we address several relevant questions in this section.

19.3.5.1 What Are the Main Challenges for Increasing Digitalisation and Innovation in the Tourism Sector? How Can Existing Difficulties Be Overcome?

Small and medium enterprises (SMEs) constitute the majority (around 90%) of Europe's tourism enterprises (UNWTO, 2020). These kinds of enterprises often do not keep pace regarding technological advances, and are behind large companies regarding the digital transition. Furthermore, it has been estimated that up to 25% of jobs in tourism need upskilling.

To maintain the competitiveness of the European travel destinations and satisfy the emerging interests of the travellers towards sustainable travelling options, we need to support the digital transition. Therefore, it is urgent to digitalise services and close the existing skills gap.

The private sector essentially provides this support, with most SMEs relying on a few private tech companies. Public institutions should provide similar platforms or foster new public-private partnerships (PPPs) to increase the accessibility to new technologies and facilitate the upskilling process.

19.3.5.2 What Are the Main Difficulties in Collecting New Data? What Strategies Towards Effective Data Collection Should Be Put in Place?

New datasets essentially come from digital data sources. Fostering digitalisation is, therefore, the first step in the way of increasing the collection of data. However, as previously mentioned, digitalisation is mainly led by a few private big tech companies. Consequently, most of the new datasets come from these companies. Two actions could be necessary, then: first, to foster new or better deals and partnerships with them as data providers and, second, to avoid an excessive

dependence on big tech companies by developing public digital/online platforms and services for SMEs, where the whole ecosystem (companies, institutions and users/tourists) shares data.

19.3.5.3 How to Measure Innovation, Digital Transition and Digital Skills Needs in the Tourism Ecosystem?

Some indicators can reflect the advance in the digital transition or tourism. For example, quantifying the (1) number of public-private partnerships and the (2) budget allocated to these PPPs could be necessary, given the importance of big tech companies in the digital sphere. In addition, when providing license to new digital services, some authorities are pushing agreements in terms of data sharing, so that companies (in the fields of mobility, waste management, energy, etc.) have to make datasets public, which could be helpful for the mentioned analyses and models. Quantifying the (3) number of agreements on data sharing would then be another essential indicator.

19.3.5.4 How to Motivate and Monitor High-Quality Data Collection by the EU Member States?

The Member States must be aware of the usefulness of new data sources and computational methods. All campaigns and initiatives launched to incentivise/facilitate data collection should be supported by services provided in exchange. We need to strengthen the link between sharing data and getting benefits in better analyses and services. It could be a good strategy for incentivising bottom-up data collection initiatives, from users to companies, institutions and, eventually, Member States.

Monitoring the advances in data collection by the EU Member States is crucial and should be coordinated. Initiatives such as the Tourism Satellite Account (TSA)¹³ are essential. As previously mentioned, this reflects the almost absence of indicators calculated based on new data sources, in the reports provided by the Member States. However, annual reports should be replaced by constantly open and updated online platforms that could also inform not only about results but also about Member States' progress, strategies, initiatives or agreements, regarding the digital transition.

Although recent tourism research is exploring and taking advantage of new data sources and methods, there is still a long way to walk on innovation in institutions at

¹³ The Tourism Satellite Account (TSA) is a standard statistical framework and the main tool for the economic measurement of tourism. It has been developed by the World Tourism Organization (UNWTO), the Organisation for Economic Co-operation and Development (OECD), the Statistical Office of the European Commission (Eurostat) and the United Nations Statistics Division (UNSD). More information: <https://www.oecd.org/cfe/tourism/tourismsatelliteaccountrecommendedmethodologicalframework.htm>

the level of the European Union and national, regional or municipal levels. This fact is evidenced in the Tourism Satellite Account (TSA) 2019 (Eurostat, 2019) Annex II. All countries indicate the most relevant data sources used to calculate the related indicators for each TSA table. Annex II shows the almost absence of nontraditional or new data sources, such as “mobile positioning data” or “other Big Data sources”.

19.4 The Way Forward

This chapter briefly discusses the potential of new data and computational methods to help stakeholders better understand and plan tourism.

The above KPIs might be measured almost everywhere in Europe and other regions of the world, in a wide range of periods and spatial scales, since they can be fed with similar data. If data sources are different, data must be reformatted to a common structure in comparative studies. Therefore, due to data’s total/partial interoperability, KPIs can be measured for several locations or industries, including competitors, and undertake comparative studies.

Data, methods and KPIs proposed in this chapter have some limitations. They do not cover all the analyses needed regarding the complex tourism sector. Therefore, other traditional measurement techniques and data sources (surveys) are still required and used complementarily. Moreover, new techniques can create new problems. Some potential issues are:

- *Dependency on the digital footprint.* A significant number of tourists or tourist attractions may leave no or only a few digital footprints. The digital divide and digital infrastructure supply in the analysed locations must be considered while measuring and understanding KPIs’ results. Traditional studies must help determine which tourism segments might be well represented by digital data and help fix biases.
- *User’s privacy.* It must be guaranteed while estimating tourism KPIs and developing new products and strategies (Hall & Ram, 2020). KPIs should reveal the necessary information on a given topic without interfering with people’s personal life and without explicitly fostering changes in their preferences. People can feel insecure if they feel constantly and unconsciously watched or worried that their data may be used against them. However, people may also agree to donate data to feed purely anonymous databases for good by an explicit agreement and fair and transparent methods.
- *Data availability/ownership.* Private companies own valuable data for analysing tourism. They usually restrict access to data since it might be a big part of their business. However, some have already released datasets worldwide or in events like hackathons and datathons. These actions have created new businesses/products and brought new insights into the tourism processes. It is necessary to explore win-win strategies among businesses and the public sector to access relevant data while keeping privacy and industrial know-how permanently (Robin et al., 2016).

Finally, the above KPIs are just values. Although some of those values seem to be easily interpretable (higher values are better than lower ones in some KPIs), they usually need some comparative or normative framework. These ranges must also be defined.

References

- Ahas, R., Silm, S., & Tiru, M. (2018). Measuring transnational migration with roaming datasets. In *Adjunct Proceedings of the 14th International Conference on Location Based Services* (pp. 105–108).
- Álvarez-Díaz, M., & Rosselló-Nadal, J. (2010). Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics*, *16*(1), 153–168.
- Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2019). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*, *0*(0), 1–20. <https://doi.org/10.1080/13683500.2019.1619674>
- Batista e Silva, F., Marín Herrera, M. A., Rosina, K., Ribeiro Barranco, R., Freire, S., & Schiavina, M. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, *68*, 101–115. <https://doi.org/10.1016/j.tourman.2018.02.020>
- Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., & Vespe, M. (Eds.). (2022). *Mapping the demand side of computational social science for policy*. EUR 31017 EN, Luxembourg, Publication Office of the European Union. ISBN 978-92-76-49358-7, <https://doi.org/10.2760/901622>
- Brilhante, I., Macedo, J. A., Nardini, F. M., Perego, R., & Renso, C. (2013). Where shall we go today? Planning touristic tours with TripBuilder. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (pp. 757–762).
- Calle Lamelas, J. V. (2017). Revolución Big Data en el turismo: Análisis de las nuevas fuentes de datos para la creación de conocimiento en los Destinos Patrimonio de la Humanidad de España. *International Journal of Information Systems and Tourism (IJIST)*, *2*(2), 23–39.
- Candelieri, A., Soldi, D., & Archetti, F. (2015). Short-term forecasting of hourly water consumption by using automatic metering readers data. *Procedia Engineering*, *119*, 844–853.
- Carpio-Pinedo, J., & Gutiérrez, J. (2020). Consumption and symbolic capital in the metropolitan space: Integrating ‘old’ retail data sources with social big data. *Cities*, *106*(April), 102859. <https://doi.org/10.1016/j.cities.2020.102859>
- CEDEFOP. (n.d.). *Skills-OVATE: Skills online vacancy analysis tool for Europe*.
- Chen, T.-L., Shen, C. C., & Gosling, M. (2021). To stay or not to stay? The causal effect of interns’ career intention on enhanced employability and retention in the hospitality and tourism industry. *Journal of Hospitality, Leisure, Sport and Tourism Education*, *28*(1), 100305. <https://doi.org/10.1016/j.jhlste.2021.100305>
- Dergiades, T., Mavragani, E., & Pan, B. (2018). Google trends and tourists’ arrivals: Emerging biases and proposed corrections. *Tourism Management*, *66*, 108–120. <https://doi.org/10.1016/j.tourman.2017.10.014>
- Dinis, G., Breda, Z., Costa, C., & Pacheco, O. (2019). Google trends in tourism and hospitality research: A systematic literature review. *Journal of Hospitality and Tourism Technology*, *10*(4), 747–763. <https://doi.org/10.1108/JHTT-08-2018-0086>
- Eurostat. (2018). *Tourism industries—Employment*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=267583#Characteristics_of_jobs_in_tourism_industries
- Eurostat. (2019). *Tourism satellite accounts in Europe—2019 edition*. Retrieved from <https://doi.org/10.2785/78529>

- Eurostat. (2021a). *Share of international guest nights in total annual number of guest nights at short-stay accommodation offered via collaborative economy*. Retrieved from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Share_of_international_guest_nights_in_total_annual_number_of_guest_nights_at_short-stay_accommodation_offered_via_collaborative_economy_2019_\(%25\).png](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Share_of_international_guest_nights_in_total_annual_number_of_guest_nights_at_short-stay_accommodation_offered_via_collaborative_economy_2019_(%25).png)
- Eurostat. (2021b). *Tourism statistics—Expenditure*.
- Folianto, F., Low, Y. S., & Yeow, W. L. (2015). Smartbin: Smart waste management system. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (pp. 1–2).
- Gössling, S. (2015). New performance indicators for water management in tourism. *Tourism Management*, *46*, 233–244.
- Gössling, S., & Peeters, P. (2015). Assessing tourism's global environmental impact 1900–2050. *Journal of Sustainable Tourism*, *23*(5), 639–659. <https://doi.org/10.1080/09669582.2015.1008500>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management*, *59*, 467–483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Gupta, A., & Dogra, N. (2017). Tourist adoption of mapping apps: A UTAUT2 perspective of smart travellers. *Tourism and Hospitality Management*, *23*(2), 145–161.
- Gutiérrez, J., García-Palomares, J. C., Romanillos, G., & Salas-Olmedo, M. H. (2017). The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*, *62*, 278–291. <https://doi.org/10.1016/j.tourman.2017.05.003>
- Hall, C. M., & Ram, Y. (2020). Protecting privacy in tourism – A perspective article. *Tourism Review*, *75*(1), 76–80. <https://doi.org/10.1108/TR-09-2019-0398>
- Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google trends meets mixed-frequency data. *Tourism Economics*, *27*(1), 129–148.
- Higham, J., Ellis, E., & Maclaurin, J. (2019). *Tourist aviation emissions: A problem of collective action*. Retrieved from <https://doi.org/10.1177/0047287518769764>.
- Hunter, R. F., Garcia, L., de Sa, T. H., Zapata-Diomedí, B., Millett, C., Woodcock, J., & Moro, E. (2021). Effect of COVID-19 response policies on walking behavior in US cities. *Nature Communications*, *12*(1), 1–9.
- Kau, A. K., & Lim, P. S. (2005). Clustering of Chinese tourists to Singapore: An analysis of their motivations, values and satisfaction. *International Journal of Tourism Research*, *7*(4–5), 231–248. <https://doi.org/10.1002/jtr.537>
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- Palmer, A., José Montaña, J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, *27*(5), 781–790. <https://doi.org/10.1016/j.tourman.2005.05.006>
- Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, *56*(7), 957–970. <https://doi.org/10.1177/0047287516669050>
- Pérez-Chacón, R., Luna-Romera, J. M., Troncoso, A., Martínez-Álvarez, F., & Riquelme, J. C. (2018). Big data analytics for discovering electricity consumption patterns in smart cities. *Energies*, *11*(3), 683.
- Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. In *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1–5).
- Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, *57*, 202–212.
- Ridderstaat, J., Croes, R., & Nijkamp, P. (2016). A two-way causal chain between tourism development and quality of life in a small island destination: An empirical analysis. *Journal of Sustainable Tourism*, *24*(10), 1461–1479. <https://doi.org/10.1080/09669582.2015.1122016>

- Riitano, N., Dichicco, P., De Fioravante, P., Cavalli, A., Falanga, V., Giuliani, C., Mariani, L., Strollo, A., & Munafo, M. (2020). Land consumption in Italian coastal area. *Environmental Engineering & Management Journal (EEMJ)*, 19(10), 1857.
- Robin, N., Klein, T., & Jutting, J. (2016, February). Public-private partnerships for statistics: Lessons learned, future steps: A focus on the use of non-official data sources for national statistics. In *OECD development co-operation*.
- Romanillos, G., García-Palomares, J. C., Moya-Gómez, B., Gutiérrez, J., Torres, J., López, M., Cantú-Ros, O. G., & Herranz, R. (2021). The city turned off: Urban dynamics during the COVID-19 pandemic based on mobile phone data. *Applied Geography*, 134, 102524. <https://doi.org/10.1016/j.apgeog.2021.102524>
- Russo, M. A., Relvas, H., Gama, C., Lopes, M., Borrego, C., Rodrigues, V., Robaina, M., Madaleno, M., Carneiro, M. J., Eusébio, C., & Monteiro, A. (2020). Estimating emissions from tourism activities. *Atmospheric Environment*, 220(May 2019), 117048. <https://doi.org/10.1016/j.atmosenv.2019.117048>
- Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13–25. <https://doi.org/10.1016/j.tourman.2017.11.001>
- Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies*, 18(5), 587–606. <https://doi.org/10.1080/14616688.2016.1214977>
- UNWTO. (2020). *European Union tourism trends*. World Tourism Organisation - UNWTO.
- UNWTO. (2021). *2020: Worst year in tourism history with 1 billion fewer international arrivals*. Retrieved from <https://www.unwto.org/news/2020-worst-year-in-tourism-history-with-1-billion-fewer-international-arrivals>
- UNWTO. (n.d.). *UNWTO tourism data dashboard*. Retrieved from <https://www.unwto.org/unwto-tourism-dashboard>
- Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, 44, 67–81. <https://doi.org/10.1016/j.tourman.2014.02.009>
- Xu, F., Nash, N., & Whitmarsh, L. (2020). Big data or small data? A methodological review of sustainable tourism. *Journal of Sustainable Tourism*, 28(2), 147–166. <https://doi.org/10.1080/09669582.2019.1631318>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

