Exploring mobile network data for tourism statistics: the collaboration between Istat and Vodafone Business Italia

Lorenzo Cavallo[1], Erika Cerasti[1], Mascia Di Torrice[1], Alessandra Righi[1], Maria Teresa Santoro[1], Tiziana Tuoto[1], Luca Valentino[1], Dario Di Sorte[2], Mauro Rossi[2], Andrea Zaramella[2], Dario Bertocchi[3], Glauco Mantegari[3], Bruno Zamengo[3]

Abstract

The paper describes the collaboration between Istat and Vodafone Business Italia to innovate and enhance tourism statistics. The common goal is to evaluate the potential uses of mobile phone data in current surveys and to investigate new outputs for official statistics, such as visiting routes and means of transport. The analysis concerned inbound tourism (foreigners in Italy), domestic tourism (Italians in Italy), and outbound tourism (Italians abroad). The work presents analyses and results for the Province of Rimini and the Municipality of Roma, referred to August 2019/2020 and April 2020, and a trial of the use of the "Welcome SMS" for the estimate of the residents in Italy who travel to foreign countries. Phone data required specific treatments to meet the definitions of official statistics. Some aspects related to the location and definition of overnight stays will require further investigation.

Keywords: Data for good, privately-held data, mobile phone data, data analytics, tourism statistics, inbound tourism, domestic tourism, outbound tourism.

DOI: 10.1481/ISTATRIVISTASTATISTICAUFFICIALE_3.2022.02

1 Lorenzo Cavallo (cavallo@istat.it); Erika Cerasti (cerasti@istat.it); Mascia Di Torrice (maditorr@istat.it); Alessandra Righi (righi@istat.it); Maria Teresa Santoro (masantor@istat.it); Tiziana Tuoto (tuoto@istat.it); Luca Valentino (luvalent@istat.it), Italian National Institute of Statistics – Istat.

2 Dario Di Sorte (Dario.Disorte@Vodafone.com); Mauro Rossi (Mauro.Rossi2@Vodafone.com); Andrea Zaramella (Andrea.Zaramella@Vodafone.com), Vodafone Business Italia.

3 Dario Bertocchi (dario.bertocchi@motionanalytica.com); Glauco Mantegari (glauco.mantegari@motionanalytica.com); Bruno Zamengo (bruno.zamengo@motionanalytica.com), Motion Analytica.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction

In the framework of a collaboration protocol for the identification of innovative statistical methodologies, the Italian National Institute of Statistics - Istat and Vodafone Business Italia carried out experimentation on themes of domestic, inbound, and outbound tourism. The aim was to evaluate the potential of using mobile phone data in the production of high-quality official statistics. The trial allowed us to assess the value of mobile phone data in complementing and enriching current tourism surveys.

In this work, we describe the statistical information useful for the qualification and quantification of touristic flows that Istat could extract from mobile phone data while highlighting some aspects that need further investigation.

The trial involved the Province of Rimini and the Municipality of Roma, two areas with a strong tourist vocation. We had the opportunity to compare the pre-pandemic tourism flows (August 2019) with flows during the lockdown period (April 2020) and after the tourism restart (August 2020). We developed three case studies:

1. Arrivals and nights spent estimates for domestic tourism and inbound tourism;
2. Same-day visits and visit routes for domestic tourism and inbound tourism;
3. Outbound tourism estimates.

Data come from Vodafone Analytics technology, which continuously collects data on the activity of the mobile phone network, thus ensuring a very granular territorial and temporal detail of information.

We preprocessed mobile phone data with the aim of fitting the international definitions of official tourism statistics in an intense preliminary design phase. Also the developed algorithms aspire to meet the requirements of official statistics for estimate calculation.

Colleagues from Vodafone and Motion Analytica implemented the co-designed algorithms and provided the corresponding results in an aggregated form, suitable to be compared with official Istat estimates.

We believe this work is an interesting example of cooperation between a private company and a National Statistical Institute on the investigation of possible usages of mobile phone data for official statistics on tourism statistics and can inspire further cooperation in this field, facilitated by the multinational dimension of Vodafone company and the active role of Istat in the official statistics community. In addition, parts of this work on definitions and data processing are not restricted to tourism statistics and may be beneficial also for the usage of mobile phone data in other fields of official statistics.

This work is organised as follows: in Section 2 the international context of mobile phone network data usage in official statistics is outlined, with a focus on their applications for tourism statistics; Section 3 describes the main feature of the collaboration between Istat and Vodafone Business Italia; in Section 4 the data from Vodafone mobile network are illustrated. Section 5 introduces methods, concepts, and definitions of tourism applied to the mobile phone network data for the identified case studies. Sections 6, 7, and 8 present results on arrivals and nights spent estimates for domestic tourism and inbound tourism; same-day visits and visit routes for domestic tourism and inbound tourism; outbound tourism estimates, carried out using the "Welcome SMS" received by Italian residents arriving in foreign countries. Concluding remarks and future perspectives are provided in Section 9.

2. The international and national context of mobile phone network data for tourism statistics

Mobile phone network data have been more and more attractive for the production of statistics in the last decades. They have been used for several statistical goals, among others for estimating the present population, for disaster management, already before the COVID-19 pandemic outbreak, for human mobility, and for designing smart cities and special events. Several countries in Europe and all around the world have been experimenting with these data.

The usage of data coming from telecommunication companies for tourism statistics production is part of the agenda of principal international official statistical producers (Ahas *et al.*, 2014; De Meersman *et al*., 2016). As an example, the Department of Economic and Social Affairs - Statistics Division of the United Nations Organisation established an expert task team on this subject (UN-CEBD, 2021). Moreover, since 2016 the European Commission funds collaboration projects among National Statistical Institutes of Member States to use data coming from new sources, including data from telecommunication companies (Eurostat, 2021). In addition, Eurostat established a dedicated task force on mobile network operator data used for official statistics, in order to support member states (European Commission, 2021).

In many countries, the usage of mobile phone data still occurs in an experimental stage, and, as far as concerns tourism statistics, they are used for focussing on particular areas by exploiting their spatial granularity (Saluveer *et al.*, 2020). When mobile phone data are used for the official production of tourism statistics, Statistical Institutes are often supported by companies' experts in data analytics (Kuusika *et al.*, 2014). Results of several analyses confirm the high potential of those sources: they show a high correlation with statistics based on traditional data, especially for some target variables (like occupancy in accommodation establishments) and they allow to complete information on tourism categories not included in collective accommodation establishments (*e.g*. relatives and friends' houses, second houses, *etc*.). They can give also real-time information on the attendance to big events with a strong tourist attraction, and can be further analysed in terms of places of origin, length of stay, and other places visited nearby (Guidotti *et al.*, 2017; Nurmi and Piela, 2019).

When phone data are used for official production on tourism statistics, they have to be "mapped" to the internationally harmonised concepts, definitions, and standards of the official statistics on tourism (UNWTO, 2010). In Europe, NSIs are additionally subject to formal legislation (as will be described in detail in Sections 6 and 7). Sometimes, the difficulties in adapting mobile phone data to derive these specific tourism definitions may create limitations to the usage of mobile phone data in the current production of official tourism statistics. Currently, only a few countries (*e.g.* Indonesia and Estonia) use phone data to complete official statistical production.

The COVID-19 pandemic and the consequent lack of traditional statistics, particularly direct face-to-face surveys, sharply increased the use of mobile phone data for public and official statistics. For instance, the Bank of Italy used mostly phone data in March 2020 to get information on touristic flow for the compilation of the balance of payments at the requested monthly planning, due to the abrupt interruption of the survey because of the COVID-19 pandemic diffusion. The promptness in the activation of this source came as a result of experimental projects realised in previous years (Carboni *et al.*, 2020).

In addition, to enrich the understanding of people's movements in Italy during the outbreak of Coronavirus in 2020, Vodafone proposed a solution to anonymously monitor the daily movements of Vodafone SIMs in Italy (at an aggregated level and different spatial and temporal granularities) to provide insights about the movements of Italians aimed at supporting the decisions taken by local authorities in Italy (Calabrese *et al.*, 2021).

The collaborations among national statistical offices and telecommunication companies raise new challenges for the official statistics, relative to the usage of data collected by private parties (Ricciato *et al.*, 2018).

The indication provided by Eurostat on the outputs of the so-called *Trusted Smart Statistics*[4] produced with privately-held data explicitly recalls the need to leave that the data are processed by the holders and to apply new technologies and methodologies to ensure privacy, as well as transparency, verifiability and public control on the whole elaboration process (Ricciato *et al.*, 2021).

National statistical offices guarantee the quality assessment and transparency commitment of the processes and data produced, consistently with the

4 Statistical outputs derived from services provided by smart systems, embedding auditable and transparent data life-cycles, ensuring the validity and accuracy of the outputs, respecting data subjects' privacy and protecting confidentiality (see https://ec.europa.eu/eurostat/cros/content/trusted-smart-statistics-nutshell_en)

European quality framework, *e.g.* the European Statistics Code of Practice (Eurostat, 2017) and the Fundamental Principles of Official Statistics of the General Assembly of the United Nations (UNSD, 2014). The applications of these principles are well-established, at least in European countries, for statistics derived from surveys and administrative data. When statistics are produced from data collected by private parties, there is still the need for further insights on many quality aspects, and the requirement for unified standardised approaches and solutions to grant quality and transparency still animates the debate at the national and international levels.

Privacy guarantees are critical points. To this extent, it is important to remind that mobile phone data need to be used in full compliance with the privacy legislation, and the General Data Protection Regulation (EU GDPR, European Parliament, 2016), for instance following a Privacy by Design methodology. These requirements are added to the crucial attention to data privacy and confidentiality protection by the official statistics community strongly focussed on the output privacy preserving methodologies.

3. Istat – Vodafone Business Italia cooperation mode

In 2018, with the aim of developing studies on the opportunities for the use of Mobile Network Operators (MNO) data, Istat decided to set up a collaboration with Vodafone Italia, which was going to develop the sector of analytics. After a long phase of internal evaluation and bilateral negotiation, Istat decided to sign a first biennial Collaboration agreement in 2019. Istat has opted for an agreement, free of charge, for the exchange of methodologies and not of data, not yet feeling ready to carry out a tender on the market for the provision of MNO analytics services, since the real information scope of MNO data was not yet sufficiently known by Istat at that time.

The agreement provided that the data was elaborated by the Operator, for evaluating the possible use of mobile phone data in the official statistics production in sectors such as sub-populations, mobility and tourism flows, transport and traffic flow, smart cities, and lifestyles. From the Vodafone Business Italia point of view, an additional goal of the agreement was the comparison of its analytics with the official statistical standards.

The agreement stated that all data, documents, contents, and information exchanged were subject to non-disclosure rules (*i.e.* types of data available on their own databases; information on the services developed, algorithms and techniques to process the information; information derived from market analysis; data descriptors used; results of analyses conducted on aggregated data; information on the platforms, methodologies, and solutions adopted). Only a subsequent agreement, in force since 2021, set a relaxation of these rules, and any disclosure of results, if previously communicated by Istat, would be subject to authorisation from Vodafone.

However, the sole methodological exchange soon proved insufficient to investigate some key aspects of the resulting estimates. It was decided to carry out two projects which would have allowed Istat, against the compensation of the costs of using the machine time: 1) having “almost” direct access to the databases of network traffic information; 2) verifying the operating mode with which Vodafone realises the Analytics; 3) making prototypes, also verifying their accuracy.

This work is the product of one of these projects, which was carried out by means of an innovative kind of collaboration, called “Sprint methodology”.

A group of Istat methodologists and thematic experts worked online together with the experts of Vodafone Business Italia and Motion Analytica by means of a dashboard (Mirò, on software license provided by Vodafone) providing their contributions for building the output prototype, in terms of definitions of aggregates to estimate, algorithms to implement, *etc*.

Vodafone colleagues ran the programmes on their machines and returned the results for validation at each round to arrive at a shared output.

In the Sprint, we adopted careful but limited data exploitation, given the budget available for the machine time. It was necessary to arrive at the common Planning phase with clear objectives regarding the extension (territorial vs temporal) of the elaborations, to find a trade-off between the exploitation of spatial data at a given date and the opportunity to make temporal comparisons (*e.g.* before, during and after COVID-19 pandemic period) for a smaller territorial set.

We discussed concepts, definitions, and methodologies applied to adhere as much as possible to the international standard and definitions. In the presentation of the main results, we stress the points in which the adherence to the international definitions may be problematic and where there is a need for further investigation and elaboration on the conceptual mapping between the official statistics concepts and the mobile phone-derived concepts. We have fruitful insights into the data and methods used by Vodafone analytics, all the specific methods have been discussed and approved by the group, although the non-disclosure agreement limits the description of some methodological choices in this paper.

4. Big Data Analytics in Vodafone's mobile network

The continuous growth of mobile phone services and usage, together with the pervasive deployment of network coverage, have made wide-area mobile networks a valuable source of an extreme amount of data. The analytics of this time-space big data can indeed be an innovative way to explore several insights into subscribers' behaviour, presence, and movements.

Vodafone Analytics is the service offered by Vodafone that allows performing statistical analysis on non-personal data generated by the Vodafone 4G and 4.5G Networks. The essential component at the very base of Vodafone Analytics services is the ability to collect data from a network monitor analyser (probe data) which allows collecting space-time information on which cellphone tower and on which cell of that tower a SIM is connected to at a specific point in time (Calabrese *et al.*, 2014). The maximum spatial resolution is then at the single cell level, which usually covers an area with a radius from a few hundred meters (in urban areas) to a few kilometers (in rural areas). The frequency of observation over time depends on how often the mobile device connects to the network infrastructure, and on the network technology used. For instance, for 4G connections, we can identify events in the order of hundred times during a day (Pinelli *et al.*, 2015). The Vodafone Network consists of more than 200,000 telephone cells located throughout the country, covers almost 99% of the population, and collects up to tens of billions of positions daily, coming from interactions with approximately 20 million human SIMs (see AGCOM 02/22). Furthermore, the availability of a time series up to 18 months allows us to compare the phenomenon considering a complete cycle of seasonality and offers the possibility to measure trends and differences year over year.

The data are anonymised and irreversibly aggregated, in compliance with the privacy legislation, and the provisions of the EU GDPR, according to the Privacy by Design methodology. Customers are informed by Vodafone through the privacy policy that the data generated by the mobile network are used for these purposes in an anonymous and aggregate form, pursuant to and for the purposes of the legislation on data protection and as established in this regard by the Opinion 05/2014 on the anonymisation techniques of the Working Group pursuant to Article 29 for data protection. Vodafone, in compliance with this legislation, guarantees that it is technically impossible

for anyone to trace the personal data that would allow the identification of the interested party from the anonymised data stored in the cloud.

To calculate the stay location of a SIM (SIM position) we need to aggregate network data at the cell network towers level. A dwelling time algorithm to estimate stops in a specific area is used on the probe data. The algorithm receives in input a series of parameters, one of these is the minimum time threshold to be spent in a certain location (*i.e.* under the coverage of a cell if we are interested in a small area, or under the coverage of a group of cells if we are studying the presence, for instance, in a municipality). As a minimum temporal threshold, we used 30 minutes (minimum stop duration); this allows us to filter out noise from the data and also to catch only significant events. This ability to collect big data about SIM positions from a cellular mobile network is a function of the parameters described here below.

Space granularity – The density of mobile radio cells is paramount in guaranteeing the supply of reliable data. Vodafone can rely on more than 200,000 phone cells located across the national territory, which guarantees an advanced granularity of service because the higher the density of cells and the smaller their coverage and thus more precise the information about the SIM positions from the network big data. Within densely populated areas, the radius of a cell can be reduced to a very few hundred meters; on the other side, within rural areas where cellphone towers are definitely less dense, the dimension of a cell coverage may raise up to a few kilometers. An Additional increase of the space granularity can be achieved with the new coverage cells, and the DAS (Distributed Antenna System) technology, which Vodafone regularly installs indoors.

Additional information to improve the precision – The precision with which spatial data is acquired is decisive in obtaining high-quality geo-referenced information. A further improvement to spatial precision measurements, in respect to phone cell coverage, is obtained by accessing other data sources and complex interrelated methods. An example of sources are existing device applications; roughly 10 million devices provide geo-localisation with an accuracy of just a few meters (AGPS). This allows us to better associate the radio coverage of a cell with a geographical area since the mapping algorithm considers not only the output from simulators but also real information from the field (*i.e.* from the devices). It is possible, therefore, to build probabilistic

maps that allow for statistically distribute the users, resulting in an effective increase of spatial detail with respect to analysis based entirely on phone cells.

Time granularity – The frequency of the SIM positions sampling is of the utmost importance to enable the profiling process, along with accurate analysis, especially when it is necessary to know the actual presence or passage number in a limited geographical area (*i.e.* train stations, motorway tolls, border crossings, *etc*.). A sample every 30 minutes or more would not guarantee the effective presence of customers in limited geographic zones and, therefore, it would drastically diminish the reliability of the sample and most of the insights. Vodafone, however, can count on a high-frequency sample that guarantees presence notifications multiple times per minute, thanks to both the Circuit Switched (CS) network and the Packet Switched (PS) network. The ability to access PS data across the entire national territory enables access to data having a frequency in the order of minutes (roughly up to 1,000 records per day per SIM /device), in contrast to the Circuit Switched data that, on average, produces roughly only tens of records per day per SIM/device). CS networks typically generate Call Detail Record (CDR) information, which is collected based on user-generated events (voice calls and SMS), whereas PS networks also register IP-based service (data connections and app usage) and signalling.

Network coverage extension – In the case of the absence of mobile coverage an operator will not be able to collect data or supply data to the contracting authority. Therefore, it is essential for the network coverage extension to be optimal, as the time and costs of dedicated coverage constructions are not compatible with the expectations of the data users. The percentage of the population covered by 2G is close to 100%, whilst the percentage of the population covered by 4G is close to 99% (Vodafone, 2022).

Privacy by design – The Analytics services fully respect the current privacy laws and are compliant with GDPR rules in that all analysed data and information are always strictly anonymised and aggregated by design methodology. Customers are informed via privacy information that the data generated from the mobile network is used anonymously and in an aggregated form. Vodafone guarantees that it is technically impossible to trace back to any individual via data anonymously given and memorised in the cloud. For privacy reasons, data supplied by Vodafone is anonymous and aggregated, showing only cluster groups of more than 15/30 people.

5. Methods, concepts, and definitions

One of the main challenges we met is how to relate phone users and the resident population, which is in general the reference population for official statistics. This challenge requires solving several issues: how the subscribers of a single mobile network operator are representative of the resident population, how to consider people who do not have or use phones, how to deal with people subscribing to more than one contract, and so on. Additional issues are related to the identification of "usual" places (*e.g.* place of residence or place of tourism/leisure, in our case) for people on the territory on the basis of the phone locations and the time aggregation. All those aspects and many others have been discussed and illustrated during the working sessions of the cooperation. As already mentioned, the cooperation agreement leaves to the parties the ownership of the applied algorithms and they cannot be fully revealed at the time this paper is written. Nevertheless, in this Section, we provide some general references on the methods that have been applied.

Concerning the representativeness of the total resident population, Vodafone Analytics is based on a sample of about 1/3 of the Italian population, and geographic information is collected several times per minute. All the provided information is expanded to the entire reference universe. The process to expand the Vodafone sample to the total reference population (all people with a phone that correspond approximately to those who are at least 12 years old) is based on a machine learning algorithm. Indeed, an owner-calibrated algorithm allows to represent the entire universe of users and not only the Vodafone SIM owners or foreigners connected to the Italian network. The main input of this model are:

- Local market share of the operator, obtained from internal market analysis of the Italian SIMs. This is available by province and age groups bases.
- Market share on a national basis of foreign SIMs.
- Market share by type of SIM (business/consumer) obtained from market studies and official reports, such as the telecommunications observatory or AGCOM.
- Socio-demographic characteristics of users from proprietary data. Two levels of data refinement are applied to these indicators:

- "real user" information: Vodafone, through its customer interface points (contact centre, shop, *etc.*) updates the personal information of the real user of the SIM at each contact with its customers. The real user is the person who actually uses the SIM regardless of who signed the contract.
- Correction of any distortions related to the different market penetration between the different age groups of the population using official statistics such as Istat census data.

The combinations of calibration factors have led to the identification of different classes of inferential algorithms according to the types of analysis, creating very specific areas, such as, for example, (non-exhaustive list): indoor presence, at high speed, territorial presence, the passage through limited places. Continuous effort is put into the verification and calibration of the inferential model. When a reliable benchmark for comparison is available (such as, for example, official data of sports and musical events and of rail and air transport), these situations are exploited for testing the inferential algorithms.

The validation of the algorithms used to produce the total resident population and its quality assessment are quite far from the scope of this cooperation. Nevertheless, the discussion on these aspects have been really active during the common working sessions.

Hence, we decided to move forward with the traditional definition of usual residents and adopt a new definition of 'phone residence', to identify the location of the prevalent cell where a user (SIM) spent a night. A SIM is associated with a night phone cell if it is connected for at least 6 hours within the time window from 8 pm to 8 am. The concept of prevalent cell recalls the place where the SIM records the highest number of activity. We built three different user profiles (residents in the area, daily visitors, and tourists) by ascertaining their daily presence during the reference period, by quantifying the time spent in the area, and by identifying places where they spent nights.

We deeply analysed how to better operationalise concepts and definitions used in tourism official statistics using mobile phone data.

In Schema 5.1 the concepts and definitions of the tourism official statistics are connected with the solutions adopted in using mobile phone data.

Schema 5.1 - Concepts and definitions of the tourism official statistics and the corresponding definition adopted using Vodafone data

Concept	Official tourism statistics definition	Definition adopted
Usual place of residence	Municipality of usual residence of the tourist	*Night Cell:* the most frequent cell the user is attached to between 8 pm and 8 am. *Phone residence:* it corresponds to the municipality of the most frequent night cell, in the last 12 months before the reference period.
Tourist	Traveler taking a trip with an overnight stay to the main destination outside his/her municipality of usual residence, for any main purpose (business, leisure, or other personal purposes) other than to be employed by a resident entity in the country or place visited. *Domestic tourist* is a resident in Italy who make trips in Italy, inbound tourist is a resident abroad who makes trips to Italy; outbound tourist is a resident in Italy who makes trips abroad.	A user with a night cell referring to a municipality that differs from his/her *phone residence*.
Arrival	A person (tourist) who arrives at a tourist accommodation establishment and checks in	A tourist is registered as an "*arrival*" on the first day of his/her trip.
Night spent	Each night a guest/tourist actually spends (nights of sleep or stays) in a tourist accommodation establishment. The sum of the *Nights* in the municipality of the destination spent by all the *Tourists* is the *Total nights spent*.	Each night spent by a user with a night cell referring to a municipality that differs from his/her *phone residence*
Same-day visitor	Traveler who makes a visit outside his/her municipality of usual residence for at least three hours without an overnight stay. Visits made during a trip are excluded.	A user who visits a municipality for at least 3 hours being neither a tourist nor a phone resident.
Resident tourist	Tourist resident in Italy	Tourist with an Italian SIM or Foreigner tourists who spent at least 3 months in Italy.
Non-resident tourist	Tourist non-resident in Italy	*Foreign users*: users owning a foreign SIM.

We also used information on the frequency of the presence of a SIM in a municipality in order to distinguish *Frequent (or regular) visitors (i.e.* SIMs who visit the same municipality at least 4 times within a month) from *Infrequent (or non-habitual) visitors*: sum of *Italian* and *Foreign users* net of *Frequent (or regular) Visitors*.

Based on all these definitions, we estimated inbound flows by approximating the foreign nationality of residence with that of the foreign operator who issued the SIM, net of *foreign users resident in Italy*. For the estimation of the domestic tourism flows we used the determination of the phone residence of SIMs observed in the destination to differentiate tourists from users residing in the observed municipality.

6. Domestic and inbound tourism – Arrivals and nights spent

Istat currently measures arrivals and nights spent relative to domestic and inbound tourism through the "Occupancy in collective accommodation establishments" survey. The survey is a census, carried out on a monthly basis and aimed to estimate the tourist flows on the national territory, in particular flows recorded in official accommodation establishments in the country. It responds to the framework of the EU Regulation no. 692/2011 of the European Parliament and the Council concerning the European statistics on tourism (EU Regulation, 2011), as amended by the EU Delegated Regulation no. 2019/1681 (EU Commission, 2019). The scope of this Regulation concerns all tourist accommodation establishments providing as a paid service (although the price might be partially or fully subsidised) short-term or short-stay accommodation services[5] (Eurostat, 2015).

The survey quantifies, for each month and each municipality, arrivals and nights spent by Italian residents (by region of residence) and non-residents in Italy (by country of residence) at tourist accommodation establishments, disaggregated by category of hotels and similar accommodation and by type of other collective accommodation establishments. The official accommodation establishments - hotels and others - involved in the survey are currently over 218 thousand, resulting from the local registers, held and managed by the regional intermediate bodies[6]. The establishments' owners transmit daily data on occupancy to their local public administration in charge of the survey; data are then summarised monthly, at the municipal level, and submitted to Istat by the intermediate bodies. The quality of traditional tourism statistics is certainly high and it has improved over time. Since the survey is intermediated by Regions, quality may /vary slightly from Region to Region, although, it is definitely kept under control by continuous monitoring.

For the trial with MNO data, we identified two areas with a strong tourist vocation but with quite different characteristics. The province of Rimini is indeed a seasonal and marina destination, characterised by the presence of

5 These services are classified and described in the three following groups of the NACE rev.2 classification: 55.1 (hotels and similar accommodation), 55.2 (holiday and other short-stay accommodation) and 55.3 (camping grounds, recreational vehicle parks and trailer parks). Non-rented accommodation are out of scope

6 The "Occupancy in collective accommodation establishments" survey is intermediate by local public administrations in charge of the survey (Statistics Offices in the Regions or Provinces, local tourism bodies) and data collection is entrusted to them.

many second homes. Rome, on the other hand, is extremely attractive, not only for tourists but also for commuters, and less seasonal. For the Province of Rimini, the Istat tourist flows include also of the values registered in private (rented) accommodations that the Emilia-Romagna Region collects and supplies to Istat. The same information is not available for the Municipality of Rome, despite this phenomenon being relevant also for this destination, according to Airdna's elaboration on Airbnb data (AirDNA, 2022). Hence, tourist flows using private rented accommodations might be currently underreported in official figures for the Municipality of Rome.

The analysis concerned *Arrivals* and *Nights spent*, but focussed more in-depth on *Nights spent,* a variable with a clear defining perimeter, for the province of Rimini and the Municipality of Rome, during August 2019 and August 2020, and April 2020. The latter period can be considered as a benchmark of the presence of the population in their usual place of residence, due to the restriction to the movements connected to the Covid-19 pandemic (the whole Italian territory was in lockdown). Nevertheless, April 2020 is not an interesting reference period for tourism, since in principle it should be completely canceled out.

Table 6.1 shows the arrivals and nights spent derived from Vodafone data compared to the Istat figures.

Table 6.1 - Arrivals and nights spent, comparison between Istat and Vodafone (absolute values and percentage differences)

Variable	Territory	Year	Month	Istat	Vodafone	% diff. Vodafone-Istat
Arrivals	Roma	2019	8	810.639	2.664.117	228,6%
		2020	8	188.358	1.036.414	450,2%
	Rimini	2019	8	797.381	1.203.745	51,0%
		2020	8	747.386	1.179.676	57,8%
Nights spent	Roma	2019	8	2.391.301	6.827.457	185,5%
		2020	8	570.568	2.996.464	425,2%
	Rimini	2019	8	4.482.871	3.645.220	-18,7%
		2020	8	3.749.862	4.135.101	10,3%

Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, years 2019-2020

The comparison between mobile phone data and Istat data shows that:

- Vodafone's estimations of nights spent, both for domestic and inbound, are always higher than Istat data, except for domestic nights spent in the province of Rimini in August 2019.

- The gap between Vodafone and Istat data for the municipality of Roma is wider than the one recorded for the province of Rimini.

The first result is quite expected, since Istat data do not include a part of the private accommodations, particularly in Rome, and all the non-rented accommodations (second homes, homes of friends or relatives, *etc*.), according to the EU Regulation. In addition, it seems that the Vodafone data capture also nights related to other phenomena, like commuters, temporary workers, and other temporary present foreigners. The negative difference in August 2019 the province of Rimini requires further investigations, it could be related to local behaviour during summer nights and occasional data quality issues.

The percentage changes in arrivals and nights spent for the month of August 2020, compared with the same month of 2019, are almost equal between the two sources for the municipality of Roma, but different for the province of Rimini, as shown in Table 6.2.

Table 6.2 – Arrivals and nights spent by origin of the guests, comparison between Istat and Vodafone data (percentage changes)**. August 2020 on August 2019**

Territory	Arrivals			Nights spent		
	Domestic	Inbound	Total	Domestic	Inbound	Total
			Istat			
Roma	-12,68	-89,37	-76,76	-11,61	-88,89	-76,14
Rimini	1,58	-41,81	-6,27	-8,96	-48,81	-16,35
			Vodafone			
Roma	-12,26	-75,71	-61,10	-12,21	-70,57	-56,11
Rimini	17,66	-45,50	-2,00	37,60	-40,22	13,44

Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, August 2020

To further investigate the reasons for the large differences between Vodafone and Istat data, we considered April 2020 as the reference time, since that month was characterised by the Covid-19 pandemic lockdown. In principle, at that time the tourist flows were expected to be nearly zero, hence a better alignment between the two sources should apply. Table 6.3 reports the results on arrivals and nights spent in April 2020, showing opposite evidence, presenting even higher divergences than in August 2019 and 2020. This finding confirms the difficulties of the tourism definition of Vodafone data, as well as a potential under-reporting of Istat data. A first explanation could be connected to some specific subgroups of SIMSs considered in the algorithm, *e.g.* the

ones of foreign residents in Italy (therefore "not tourists"). The nationality of the foreign SIMs registered in April largely refers to nationalities present in Italy for business reasons rather than for leisure.

This is particularly evident in Rome in which the presence of foreign residents in Italy is greater than in the province of Rimini.

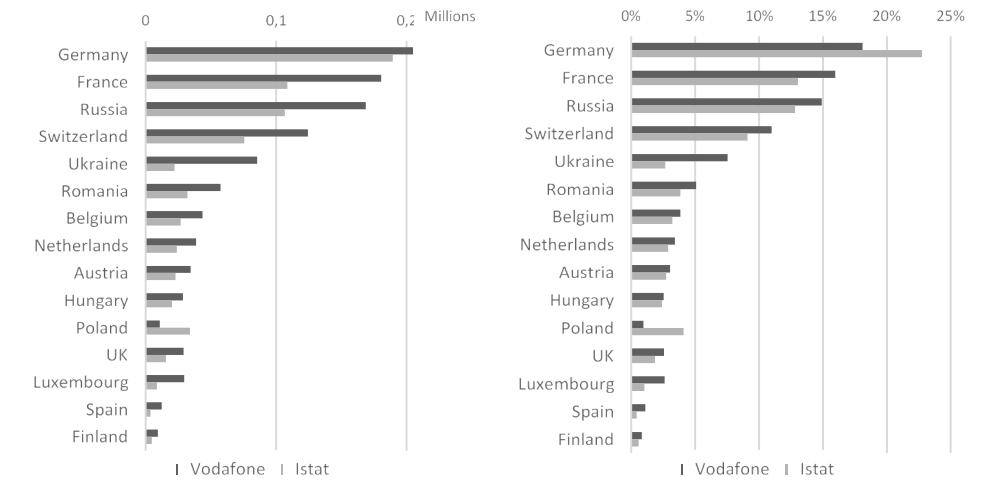
Table 6.3 - Arrivals and nights spent in April 2020, comparison between Istat and Vodafone (absolute values and percentage differences)

Variable	Territory	Year	Month	Istat	Vodafone	% diff. Vodafone-Istat
Arrivals	Roma	2020	4	6.013	154.385	2.467,5%
	Rimini	2020	4	947	18.170	1.818,7%
Nights spent	Roma	2020	4	15.819	1.577.083	9.869,5%
	Rimini	2020	4	47.212	150.032	217,8%

Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, years 2019-2020

Figure 6.1 reports the estimates on the nights spent by inbound tourists by country of origin for the province of Rimini. It shows a good degree of convergence and consistency in terms of percentage shares between the two sources; Germany and Russia are among the most relevant countries.

Figure 6.1 - Nights spent by inbound tourists in the Province of Rimini by country of origin (absolute values and percentage shares)**. August 2020**

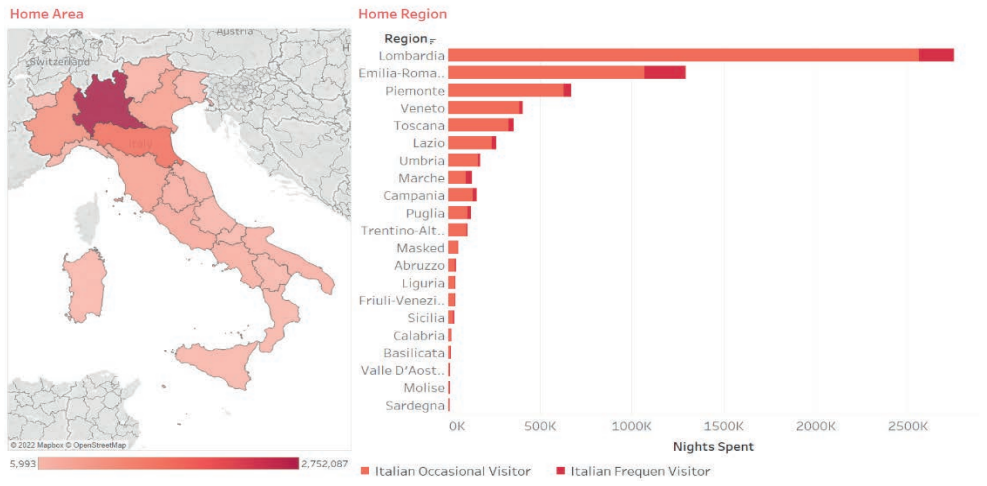


Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, August 2020

Figure 6.2 shows nights spent by domestic and inbound tourists in the municipalities of the Province of Rimini (absolute values) in August 2020.

Both the charts show the home location of the tourists. In the map, the gradation of colour of the Region is proportional to the number of nights spent by people who are from that Region. The bar chart on the right side of the picture reports the same piece of information, distinguishing between occasional and frequent visitors. People from the first category visit the area at most three times in the analysed period, while people from the other one visit the area at least four times in the analysed period. While tourists are mainly from Lombardia, most frequent visitors come from Emilia-Romagna.

Figure 6.2 - Nights spent by inbound and domestic tourists in the municipalities of the Province of Rimini (absolute values)**. August 2020**

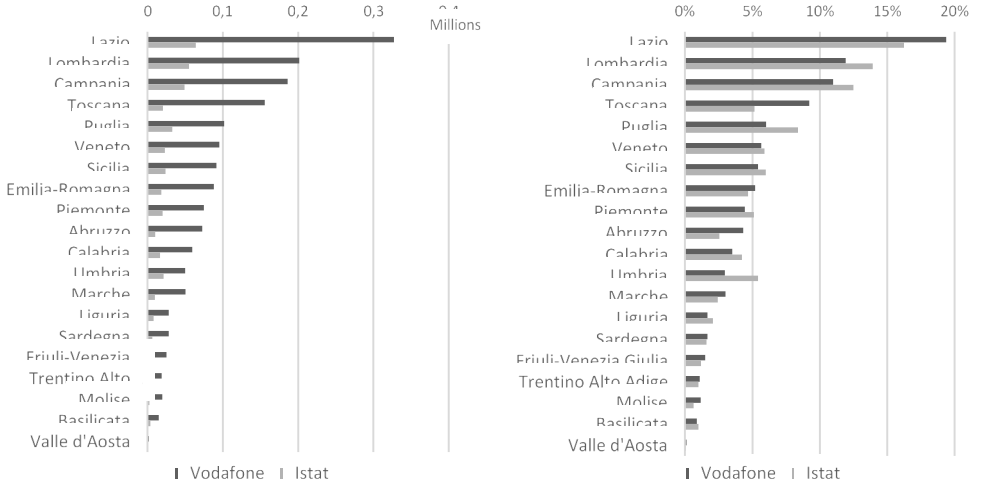


Source: Overnights versus provenience and frequency, Vodafone Analytics, August 2020

Figure 6.3 shows nights spent in Romew in August 2020, by domestic tourists per region of origin. It shows a strong difference in absolute value but a limited one in terms of rank between the two sources. Lazio, Lombardia, and Campania are the territories that originate the highest shares of tourists for both sources.

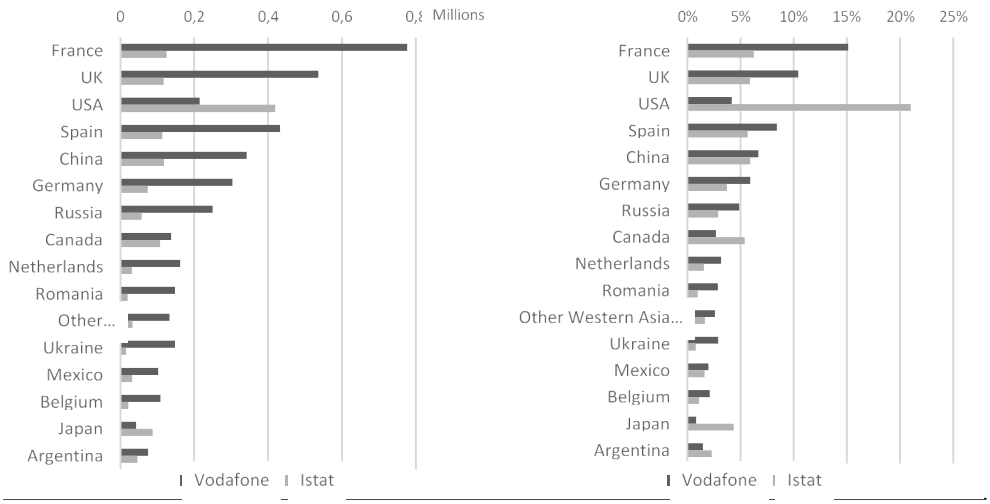
Nights spent by inbound tourists by country of origin are reported in Figure 6.4, showing smaller differences in absolute value but higher differences in the percentage shares for the two main foreign countries of origin (France and the United States) between the two sources. In particular, for the United States, differences are relevant. This is connected to the relevant presence of US citizens temporary living in Rome, working in international organisations, as already underlined above.

Figure 6.3 - Domestic nights spent in the municipality of Roma *(absolute values and percentage shares)***. August 2020**



Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, August 2020

Figure 6.4 - Nights spent by inbound tourists in the municipality of Roma (absolute values and percentage shares)**. August 2020**



Sources: Istat, Occupancy in collective accommodation establishments and Vodafone Analytics, August 2020

7. Domestic and inbound tourism – Same-day visits and visit routes

Same-day visits are visits without overnight stays. They are measured by Istat through the "Trips and holidays" survey, a focus embedded in the sample Household Budget Survey since 2014. It has the goal to obtain information on the tourist movements of the population through direct interviews with households (tourist demand). The resulting estimates are about the number of tourists, trips, overnight stays, expenses (both per trip and per day), and the number of same-day visits in Italy or abroad (Di Torrice, 2018). The Regulation for Tourism Statistics no. 692/2011, which is also the reference for the "Trips and holidays" survey, defines same-day visits as "*visits without overnight stay carried out by residents outside their usual environment starting from their usual place of residence*". The "Trips and holidays" survey currently estimates the domestic aggregate, whereas it does not provide estimates related to the inbound same-day visits, and therefore the Vodafone estimates may represent a potential enrichment for official statistics.

The European guidelines for statistical surveys recommend identifying a same-day visit by surveying the purpose of the visit, the crossing of administrative borders, the duration (at least 3 hours at the place of destination), and the frequency, thus making it possible to exclude visits that do not include a tourist element. These requirements are explicit in the "Trips and holidays" survey questionnaire, but they are difficult to identify using mobile phone data.

In this Section, we present two insights that the granularity of Vodafone data allows to add to the current official statistical production. The new investigations and the increased level of details provided are however subject to the privacy preservation of routes and trips, which Vodafone guarantees by design via the aggregation of the events. Using Vodafone data, we analysed the potential "same-day visits" of visitors (Italian residents and foreigners) who travel to the province of Rimini (domestic and inbound visits, respectively). The analysis is carried out for the week of Mid-August, 2019 (12-18 August 2019), defining the Italian or foreign user as a same-day visitor who spent at least three hours in a municipality in the province of Rimini, without staying overnight and without having the phone residence there or nearby. Therefore, the algorithm applied both the criteria of the duration of the visit and of the municipality of residence border crossing. On the contrary, it was not

possible to exclude the usual behaviours through frequency, given the limited observation period. As the "Trips and holiday" survey does not ask for the exact date of the same-day visit, so it is not possible to narrow the comparison to a specific week but only to the entire month of August.

Figure 7.1 shows the municipalities (first r-w charts) in the province of Rimini visited by Italians by region of the visitors (second-row charts). The most visited municipality is Rimini doubling Riccione and Bellaria-Igea Marina, the-second and third most-visited areas, respectively. Most Italian same-day visitors are from Lombardia and Emilia-Romagna.

In Figure 7.2 are the visited municipalities in the province of Rimini (first row charts) by the foreign country of origin of visitors (second row charts). The most visited municipality is Rimini. Most foreign same-day visitors are French, Germans, and Russians.

These results let us appreciate the level of territorial detail provided by the mobile phone data compared to the sample survey findings, both in terms of the origin and destination of the visit.

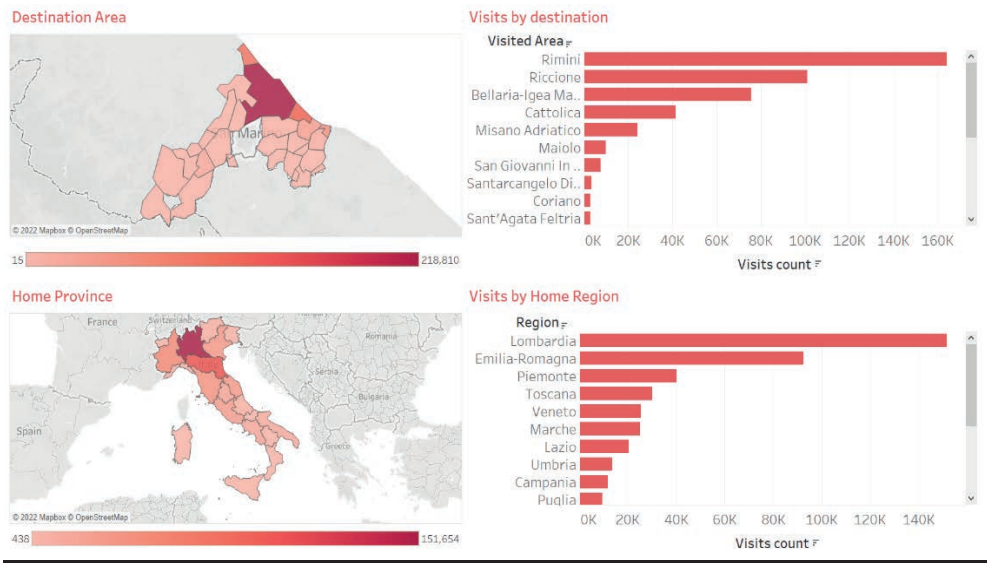
Concerning the origin of the visits of the residents, the results of the sample survey "Trips and holiday" for the same period point out a smaller set of regions, limited to Emilia-Romagna and Marche. However, it must be noted that the presence of regions of origin geographically distant from the province of Rimini provided by Vodafone data indicates that the estimate may be affected by the erroneous inclusion of same-day visits to Rimini as a part of a trip with an overnight stay in an adjacent territory.

This issue is more evident for foreign visitors (not surveyed by the Istat sample survey); even if they come mainly from "neighbouring" countries (France, Germany, and Russia), it is unlikely that they visited the province of Rimini in one day (Figure 7.2). These visits are not strictly defined as tourism, according to the Regulations for Tourism Statistics, as probably they do not originate and end from/in the place of residence.

Also regarding the destination of the visits made by residents, Vodafone data cover a wide set of municipalities, whereas the "Trips and holiday" survey points out only visits to Riccione and Rimini, probably because the tourists may indicate only a municipality in the questionnaire (the main one visited during the day).

All the above considerations can explain the substantial difference between the two sources, as the total amount of visits made by residents estimated by the Istat survey on the whole month is about 22% of the amount provided by Vodafone data, which considers only on the reference week. The most likely reason we identified is the inclusion of same-day visits made during a trip in the Vodafone algorithm, which consequently needs to be refined in order to count only visits starting and ending in the place of usual residence of the tourist (*i.e.* the place of phone residence). In addition, the inclusion of usual same-day visits in the Vodafone estimates can increase discrepancies. Visits made to the same destination on a weekly basis should be excluded as done for frequent trips, but this would be possible only by observing Vodafone data for at least one month.

Figure 7.1 - Same-day visits in the province of Rimini made by residents, by region of origin, and visited municipality (absolute values). Mid-August, 2019 (12-18 August 2019)

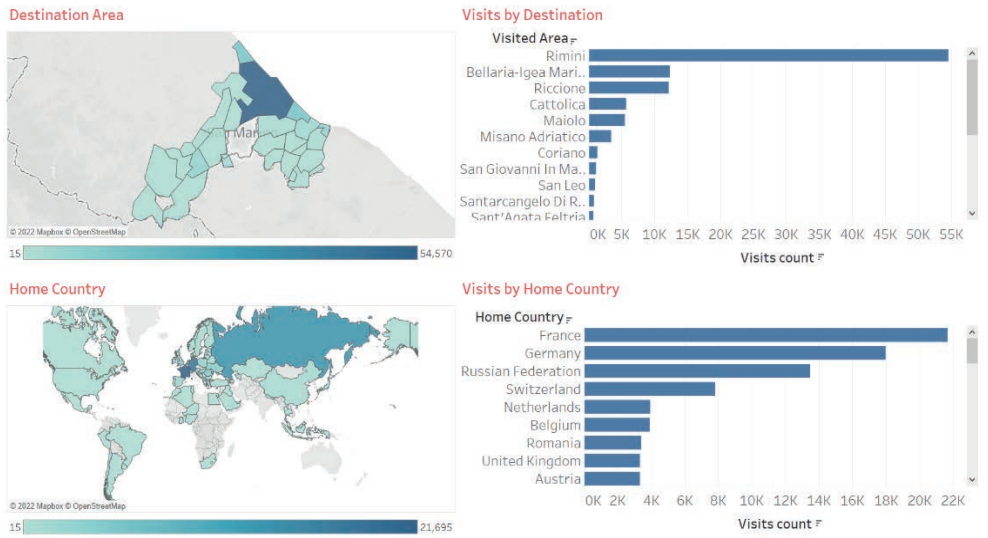


Source: Vodafone Analytics 2019

A further line of analysis that may enrich official statistics is the study of visit routes (or co-visits), namely, different municipalities visited by Italian or foreign tourists during the observed period (August 2019). This analysis provides a significant added value for tourist destinations, as it offers them

the opportunity to know both the most attractive and least frequented "routes" of visit, which, therefore, could be better promoted to decongest the most crowded places. The matrix in Figure 7.3 highlights the shifts between the municipalities (in row and column) co-visited during the observation period, assigning a darker colour to the pairs of those most affected by the phenomenon. The blue map on the left reports the number of foreign tourists per home country, and the red one is the number of Italian tourists per home region.

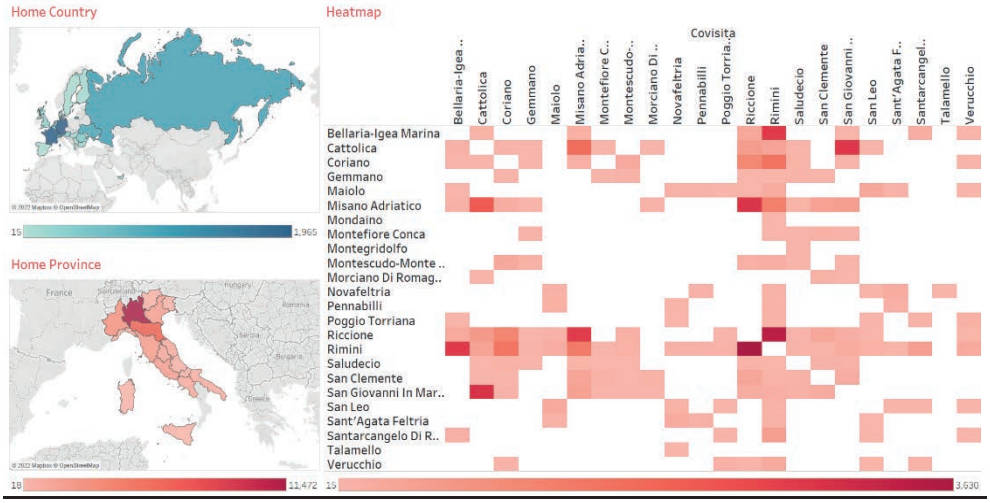
Figure 7.2 - Same-day visits in the province of Rimini made by foreigners, by the foreign country of origin, and visited municipality (absolute values)**. Mid-August, 2019 (12-18 August 2019)**



Source: Vodafone Analytics 2019

Most foreigners who visited at least two municipalities come from France and Germany while Italians who visited at least two municipalities come from Lombardia and Emilia-Romagna. The municipalities most frequently visited by tourists during the same trip in the observed period are Rimini and Riccione. Furthermore, the matrix highlights the close relationship between Riccione and Misano Adriatico, between Rimini and Bellaria-Igea Marina, and between Cattolica and San Giovanni in Marignano and Misano Adriatica.

Figure 7.3 - Visit routes in the province of Rimini by region or country of provenience of the visitor and co-visits among municipalities (August 2019)



Source: Co-visits among municipalities, Vodafone Analytics 2019

8. Outbound Tourism

Istat measures outbound tourism (residents in Italy who travel to foreign locations) through the above-mentioned "Trips and Holidays" survey. Experimental estimates on the use of SIMs during the trips showed that in 80% of trips the travelers used a mobile phone (Dattilo and Sabato, 2017).

It is possible to estimate the size of this type of tourist using the "Welcome SMS" from Vodafone data. When a Vodafone Italy customer connects for the first time to a non-Italian phone network, he/she receives an SMS summarising the contractual conditions and the rates applied. Vodafone Italy logs where (the country) and when (the timestamp) its customers receive these "Welcome SMS". Since these logs contain the SIM identifier, the roaming country, and the first roaming timestamp in that country only, it is not possible to trace any detail of visited location like cities or specific sub-regions (*i.e.* it is not possible to distinguish between "Galicia" and "Catalonia" through these logs since both will be labelled as "Spain"). The same privacy rules previously described in Section 4 about irreversible aggregation and anonymisation of user identification data are applied. To ensure that the Vodafone subscribers are representing all the Italian resident, a similar procedure is adopted, as shortly described in Section 5.

The multi-national dimension of Vodafone Company potentially allows for double checking the outbound tourism flows, for instance an Italian tourist who travel abroad would be addressed via roaming to a foreign Vodafone operational company, if it exists in the visited country. This topic is subject to further developments in the Vodafone international holding.

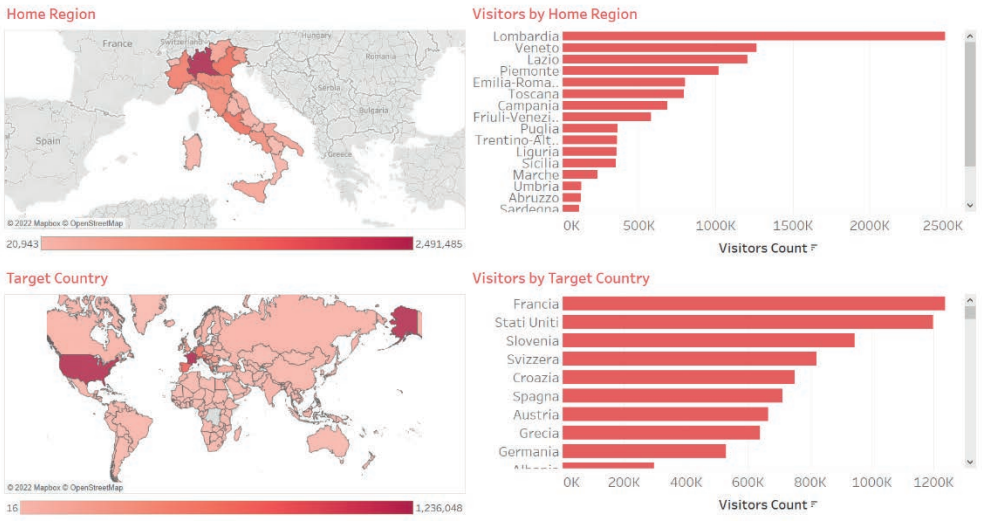
From these data, we developed an analysis aimed to estimating the size of Italian outbound tourism in August 2019 by counting how many Welcome SMS were received per foreign country and then estimating the total outbound Italian population by using the methods described in Section 5. We further enriched the analysis by adding the estimated home location (*i.e.* the phone residence) of SIM at the regional level.

Main results of the trial are in Figure 8.1. The chart at the top shows the distribution of outbound tourists by region of origin. The comparison between Vodafone data, and the "Trips and Holidays" survey estimates, shows a convergence in the top regions of the chart, as both sources indicate

Lombardia and Veneto in the two top positions, followed by Lazio (Vodafone data) or Emilia-Romagna (Istat survey). The chart at the bottom of the Figure shows the distribution of outbound tourists by visited country. France, the USA, and Slovenia appear in the three top positions according to Vodafone data, while the top visited countries are Greece, Germany, and France for the official statistics.

However, results are only partially comparable due to the fact that the official statistics consider only the main destination of the trip, while the Welcome SMS is also delivering information during the intermediate stages of a trip. For example, Slovenia is at the top of the list because is a transit for tourists traveling towards Croatia, and the United States acts as an air stopover for many intercontinental flights. In addition, we need further analysis to differentiate the movements of cross-border workers from tourist flows. This is particularly important concerning the neighboring countries such as France, Slovenia, Switzerland, and Austria, where cross-border workers are common and can cause an overestimation of tourist flows to these countries.

Figure 8.1 - Italian outbound tourists by region of origin and foreign visited country (absolute values)**. August 2019**



Source: Outbound tourism versus region of provenience and country of destination, Vodafone Analytics 2019

9. Conclusions and future developments

The collaboration experience between Istat and Vodafone confirmed the potential of mobile phone data for the production of statistics complementary to the existing official statistics. The big data generated by telecommunications companies may open new perspectives of analysis in the tourism sector, not only for the accurate monitoring on a territorial scale (official statistics are generally at the municipal or regional disaggregation) but also for the greater timeliness and temporal granularity that they can guarantee. Mobile phone data provide the opportunity to monitor the tourism experience in specific areas of interest; while controlling tourism pressure based on the total number of users (people in a given area at a specific time) and time spent in a given destination. Furthermore, it is possible to create groups of users by visit (same-day visit or visit with an overnight stay) and categories (nationality, age classes, types), following their mobility in the area.

The timeliness and the possibility of more detailed temporal breakdowns are growing needs of tourism statistics users, who want information, increasingly close in time to the event, or even estimates in advance. This is especially the case for extraordinarily organised events: short-term events, such as fairs, or longer events, such as the Jubilee or tourism in capitals of culture. This demand emerged clearly during the Covid-19 pandemic, when the need for information became greater while, at the same time, it was difficult to carry on traditional surveys (at least, on households).

A further aspect to underline concerns the coverage, namely the ability of mobile phone data to represent the phenomenon. In terms of overall quantification, the estimate of the total number of tourists (not only those who spend nights in official accommodation) is currently provided by the "Trips and Holiday" sample survey, which also detects tourists in second homes, guests of friends, or those who stay overnight in short-term rentals (for example, Airbnb). However, a sample survey is affected by the limitations of the sample size and the statistical error, which is greater when the event becomes rarefied (as happened to trips during the pandemic). Being able to rely on information derived from MNO data to support and complement traditional surveys can be a significant advantage in maintaining the continuity of tourism statistics and in strengthening their coverage.

However, it should be emphasised that relevant information currently collected by traditional surveys, which is also required by the Regulation on Tourism statistics, cannot be inferred from mobile phone data: for instance, the reason for the trip, the type of accommodation in which the tourist stayed, the quantification of the expenditure incurred. On the other hand, mobile phone data open up the possibility of producing new information and statistics currently not existing, such as co-visits or means of transport upon arrival and departure of the tourist, which would be very useful for users.

In our opinion, the main finding of this collaboration experience is that the mobile phone data and other auxiliary information held by Vodafone represent a valuable contribution to improving the official tourism estimates, forecasts, or eventually for calibrations of sample surveys. The experience described in this work suggests that mobile phone data can supplement and complement the current statistics production, and can provide new insights on emerging topics only partially covered by traditional data sources. It is hard to imagine that mobile phone data can completely replace traditional data sources for the production of tourism statistics, in the short term.

The main challenge remains to refine methodologically and conceptually Vodafone's definitions and data processing algorithms for a greater convergence with the concepts of official statistics (*e.g.* phone residence, resident foreign users, arrivals, outbound tourism), which requires joint testing and verification procedures of implemented algorithms. While in this experience we mainly adopted the Vodafone Analytics perspectives and solutions, discussing and partially adjusting them to the official statistics needs, in future collaborations we envisage the necessity to overcome the current solutions to better identify, for instance, different groups/subpopulations of users, via clustering techniques and AI approaches. In the same spirit, a deeper analysis would be devoted to the assessment of the potential selectivity of the input data, also via dedicated sample surveys, as already used in national statistical offices. An occasional sample survey, with a limited size, would be actually effective for assessing the quality of the results derived from the mobile phone data, as well as from other organically generated data (Zhang, 2019). This way could be useful also to understand how the differences observed in the results are due to misalignment in the reference populations (residents and non-residents) and/or is determined by the ability of the new sources in capturing the tourism flows according to the current official statistics definitions.

Finally, the mandate of official statistics requires the use of transparent algorithms, to this extent, we still require that the international community moves forward with a public-private partnership that guarantees the respect of this principle in a win-win standard agreement for considering data as a public good. Regulated access to statistics derived from mobile phone data is a prerequisite for using this data source in regular tourism statistics production. We hope that the effective collaboration already experienced and described so far will lead to positive results in this regard.

References

Ahas, R., J. Armoogum, S. Esko, M. Ilves, E. Karus, J.-L. Madre, O. Nurmi, F. Potier, D. Schmücker, U. Sonntag, and M. Tiru. 2014. "Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics (Eurostat Contract No. 30501.2012.001-2012.452)". *Consolidated Report*. Luxembourg: Eurostat.

AirDNA. 2022. *Vacation Rental Data to Set You Apart. Insights to Keep You Ahead*.

Autorità per le Garanzie nelle Comunicazioni – AGCOM. 2022. *Osservatorio sulle Comunicazioni*, N. 2/2022. Roma, Italy: AGCOM.

Calabrese, F., E. Cobelli, V. Ferraiuolo, G. Misseri, F. Pinelli, and D. Rodriguez. 2021. "Using Vodafone mobile phone network data to provide insights into citizens mobility in Italy during the Coronavirus outbreak". *Data & Policy*, Volume 3: e22. DOI: 10.1017/dap.2021.18.

Calabrese, F., L. Ferrari, and V.D. Blondel. 2014. "Urban Sensing Using Mobile Phone Network Data: A Survey of Research". *ACM Computing Surveys*, Volume 47, Issue 2, Article N. 25: 1-20.

Calabrese, F., G. Di Lorenzo, L. Liu, and C. Ratti. 2011. "Estimating Origin-Destination Flows Using Mobile Phone Location Data". *IEEE Pervasive Computing*, Volume 10, Issue 4: 36-44.

Carboni, A., C. Doria, e S. Zappa. 2020. "La produzione statistica nell'emergenza Covid19: la stima dei "viaggi" in bilancia dei pagamenti", *Note Covid-19*. Roma, Italy: Banca d'Italia.

Dattilo, B., and M. Sabato. 2017. "Travelling SIM and Trips: An approach to make mobile phone data usable in tourism statistics". Presented at *New Techniques and Technologies for Statistics – NTTS 2017*, Brussels, Belgium, 11-17 March 2017.

De Meersman, F., G. Seynaeve, M. Debusschere, P. Lusyne, P. Dewitte, Y. Baeyens, A. Wirthmann, C. Demunter, F. Reis, and H.I. Reuter. 2016. "Assessing the Quality of Mobile Phone Data as a Source of Statistics". Paper presented at the *European Conference on Quality in Official Statistics - Q2016*, Madrid, Spain, 31 May - 3 June 2016.

Di Torrice, M. (*a cura di*). 2018. "La nuova indagine sulla domanda turistica". *Letture statistiche – Metodi*. Roma, Italy: Istat. https://www.istat.it/it/archivio/222043.

European Commission. 2019. *Commission Delegated Regulation (EU) 2019/1681 of 1 August 2019*, amending Regulation (EU) N. 692/2011 of the European Parliament and of the Council concerning European statistics on tourism, as regards the transmission deadlines and adaptation of Annexes I and II.

European Parliament, and the Council of the European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016* on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

European Parliament, and the Council of the European Union. 2011. *EU Regulation N. 692/2011 of the European Parliament and of the Council of 6 July 2011*, concerning European statistics on tourism and repealing Council Directive 95/57/EC.

Eurostat. 2021. *ESSnet Big Data II, Work package I Mobile Network Data* (all the milestones and deliverables). https://ec.europa.eu/eurostat/cros/content/wpi-milestones-and-deliverables_en.

Eurostat. 2015. "Methodological manual for tourism statistics. Version 3.1 - 2014 edition". *Manuals and Guidelines*. Luxembourg: Publications Office of the European Union.

Eurostat, and the European Statistical System - ESS. 2017. *European Statistics Code of Practice. For the National Statistical Authorities and Eurostat (EU statistical authority)*. Adopted by the European Statistical System Committee 16th November 2017. Luxembourg: Publications Office of the European Union.

Guidotti, R., R. Trasarti, M. Nanni, F. Giannotti, and D. Pedreschi. 2017. "There's a Path for Everyone: A Data-Driven Personal Model Reproducing Mobility Agendas". *Conference Paper:* 303-312. 4th Institute of Electrical and Electronics Engineers - *IEEE International Conference on Data Science and Advanced Analytics (DSAA 2017)*, Tokyo, Japan, 19-21 October 2017.

Kuusik, A., K. Nilbe, Tanel Mehine, and R. Ahas. 2014. "Country as a Free Sample: The Ability of Tourism Events to Generate Repeat Visits. Case Study with Mobile Positioning Data in Estonia". *Procedia - Social and Behavioural Sciences*, Volume 148: 262-270.

Nurmi, O., and P. Piela. 2019. "The Use of Mobile Phone Data in Tourism Statistics". In *Proceedings Special Topic Session*, Volume 4: 135-142. 62nd ISI World Statistics Congress 2019, Kuala Lumpur, Malaysia, 18-23 August 2019. Malaysia: Department of Statistics Malaysia, Central Bank of Malaysia, and Malaysia Institute of Statistics.

Pinelli, F., G. Di Lorenzo, and F. Calabrese. 2015. "Comparing Urban Sensing Applications Using Event and Network-Driven Mobile Phone Location Data". *Conference Paper*: 219-226. 16th Institute of Electrical and Electronics Engineers - *IEEE International Conference on Mobile Data Management*, Pittsburgh, PA, U.S.A., 15-18 June 2015.

Ricciato, F., T. Siil, R. Talviste, B. Kubo, and A. Wirthmann. 2021. "A proof-of-concept solution for the secure private processing of longitudinal Mobile Network Operator data in support of official statistics". *Joint Project* conducted by Eurostat and Cybernetica during 2020-2021 (Project Reference number ESTAT 2019.0232). https://ec.europa.eu/eurostat/cros/sites/default/files/unece2021_estat_cybernetica_v6.pdf.

Ricciato, F., F. De Meersman, A. Wirthmann, G. Seynaeve, and M. Skaliotis. 2018. "Processing of Mobile Network Operator data for Official Statistics: the case for public-private partnership". Paper presented at the DGINS Conference 2018, *The European path towards Trusted Smart Statistics*, Bucharest, Romania, 10-11 October 2018. https://ec.europa.eu/eurostat/cros/system/files/dgins2018_mno-so_ricciato_0.pdf.

Saluveer, E., J. Raun, M. Tiru, L. Altin, J. Kroon, T. Snitsarenko, A. Aasa, and S. Silm. 2020. "Methodological Framework for Producing National Tourism Statistics from Mobile Positioning Data". *Annals of Tourism Research*, Volume 81, Issue C. DOI: 10.1016/j.annals.2020.102895.

United Nations Committee of Experts on Big Data and Data Science for Official Statistics - UN-CEBD. 2021. "Mobile phone data". *Task Team of the UN-CEBD*. New York, NY, U.S.: United Nations. https://unstats.un.org/bigdata/task-teams/mobile-phone/index.cshtml

United Nations, Department of Economic and Social Affairs, Statistics Division, and World Tourism Organization - WTO. 2010. "International Recommendations on Tourism Statistics 2008". *Studies in Methods*, Series M, N. 83/Rev.1. New York, NY, U.S.: United Nations Publication.

United Nations, General Assembly. 2014. Resolution adopted by the General Assembly on 29 January 2014, N. 68/261. *Fundamental Principles of Official Statistics*. New York, NY, U.S.: United Nations. https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

Vodafone. 2022. *Vodafone coverage* 2022. Available at https://www.vodafone.it/portal/Privati/Vantaggi-Vodafone/rete-vodafone-copertura-mobile.

Zhang, L.-C. 2019. "On valid descriptive inference from non-probability sample". *Statistical Theory and Related Fields*, Volume 3, N. 2: 103-113. DOI: 10.1080/24754269.2019.1666241.