

# **Big Data Analytics in mobile cellular networks as enabler for innovative statistics to evaluate the effects of Recovery and Resilience Plan actions**

## ***L'Analisi dei Big Data delle reti radiomobili cellulari come abilitatore di statistiche innovative per valutare l'effetto delle azioni del Piano di Ripresa e Resilienza***

Andrea Zaramella<sup>1</sup>, Dario Di Sorte<sup>2</sup>, Denis Cappellari<sup>3</sup>, Bruno Zamengo<sup>4</sup>

**Abstract** – The large deployment of packet-switched communication paradigm and 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 on subscriber behaviour, presence and flows, while being strictly compliant with the GDPR requirements. In this paper we specifically focus on the tourism framework and put forward new algorithms and KPIs to investigate dynamics which can be overlooked by the official statistics. This should allow to be more effective/efficient to plan/monitor NRRP projects, the goal of which is to improve the touristic offer.

**Abstract** – *L'ampia diffusione del paradigma di comunicazione a commutazione di pacchetto e la continua crescita dei servizi e dell'utilizzo della telefonia mobile, insieme alla diffusione pervasiva della copertura del segnale, hanno reso le reti mobili geografiche una preziosa e immensa fonte di dati referenziati nel dominio del tempo e dello spazio. L'analisi di questi big-data è uno strumento innovativo per esplorare comportamenti, presenze e flussi degli utenti, in conformità assoluta ai requisiti di privacy del GDPR. In questo articolo ci concentriamo specificamente sul turismo e presentiamo nuovi algoritmi e KPI per indagare dinamiche che possono sfuggire alle statistiche ufficiali, con l'obiettivo di essere più efficaci/efficienti nella pianificazione/monitoraggio dei progetti PNRR in ambito offerta turistica.*

**Key words:** NRRP, mobile phone big-data, business analytics, tourism case-study

---

<sup>1</sup> Andrea Zaramella, Vodafone Business Italia, email: andrea.zaramella@vodafone.com

<sup>2</sup> Dario Di Sorte, Vodafone Business Italia, email: dario.disorte@vodafone.com

<sup>3</sup> Denis Cappellari, Motion Analytica, email: denis.cappellari@motionanalytica.com

<sup>4</sup> Bruno Zamengo, Motion Analytica, email: bruno.zamengo@motionanalytica.com

## 1 Introduction

The big-data generated by cellular mobile networks may open new perspectives of analysis about users' behaviour, presence and flows in several area analysis related to the tourism and transport/mobility sectors. In Vodafone, the platform to collect and elaborate raw data to quickly calculate insights is called Vodafone Analytics.

Mobile network big-data can help to better understand the behavior of visitors, their movement independently of overnight stays, their preferences (coast lovers, country lovers, explorers), co-visits, the trajectory of their movements. These analyses can be carried out also for specific segments of users (gender, age, nationality, residence). Also, it is also possible to perform cluster analysis to identify experience tourist areas (e.g., wine territory, spa locations) outside the main and classic paths, to proactively know and quantify new trends (e.g., bike tourism), to describe over and under tourism phenomenon, to measure the seasonal adjustment of tourism request. The analytics of big data can be useful to be more effective and efficient to both design and monitor projects as a function of the mobility of people; historical data (up to 18 months backwards), the low-latency and high frequency of data refresh (near-real-time) are precious planning and monitoring instruments. In the framework of the Italian NRRP (National Recovery and Resilience Plan), a field of application of Vodafone Analytics is the area of actions related to tourism, especially those related to the attractivity of *villages and small historical centers* ("borghi"). Similar applications can also be found in the NRRP missions relevant to mobility and transport (sustainable mobility) and ecological transition. Also, some analyses can be carried out to study some aspects crossed to several sectors such as, for instance, the measure of *ecological footprint* of tourists in a geographical area, correlated to presence, mobility and means of transport.

In the following Sections, we first introduce of the main concepts related to big data in cellular mobile networks, then we give an overview of the basic definitions of tourism statistics and map them within the mobile data algorithms. Finally, we describe a case study of the municipality of Padova to show how Vodafone Analytics can integrate the official statistics to give additional interesting insights into the mobility of tourists outside the municipality where they spend the night.

## 2 Mobile phone Big-Data

The ability of the Vodafone Analytics big data platform to collect (time, space) data from the field and immediately elaborate them depends on:

*Space granularity*: the density of mobile radio cells is paramount in guaranteeing the supply of reliable data. Vodafone can rely on 200.000 network cells; within densely populated areas the diameter of a cell can be reduced to a few hundred meters, and even to a smaller dimension if a dedicated network coverage to specifically cover certain locations (e.g., malls, stadiums, train stations, airports);

*Time granularity*: the sampling frequency of phone/SIM position (i.e., the cell is connected to) is of the utmost importance to enable the profiling process together with an accurate analysis. Vodafone can rely on a high frequency sample that guarantees presence notifications thanks to the monitoring of raw data from all the packet-switched interactions between phones and the network (calls, messaging,

notifications, data connections, app interactions etc). A phone can be sampled up to more than one thousand time per day, and this represents a proxy of continuity;

*Network coverage extension:* the Vodafone mobile network is widely recognized for its quality and strength throughout the whole national territory with a percentage of population (i) covered by 2G close to 100% and (ii) covered by 4G close to 99%;

*Customer base:* Vodafone counts 23 million of human Italian SIMs which generate a number of raw time-space data in the order of tens of billions;

*Privacy by design:* Vodafone Analytics services are designed according to the principles of privacy-by-design in compliance with GDPR rules (n. 05/2014, Working Group ex Art. 29) and all data are irreversibly anonymized and aggregated. Vodafone Analytics aims at studying the behaviour of homogeneous groups of people and not the behaviour of the single user. Finally, thanks to a proprietary calibration algorithm, the number of people in a cluster is projected to the entire universe of users and not only those connected to the Vodafone mobile network.

### 3 Tourism Case Study

Here below the classic definitions of tourism statistics:

- visitor: traveller taking a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose (e.g., business, leisure) other than to be employed by a resident being and is classified as:
  - *tourist* (or overnight visitor), if the trip includes an overnight stay;
  - *same-day visitor* (or excursionist), if the trip does not include a night.

Through Vodafone Analytics it is possible to qualify travellers, visitors, tourists and same-day visitors by observing the most active area during the night:

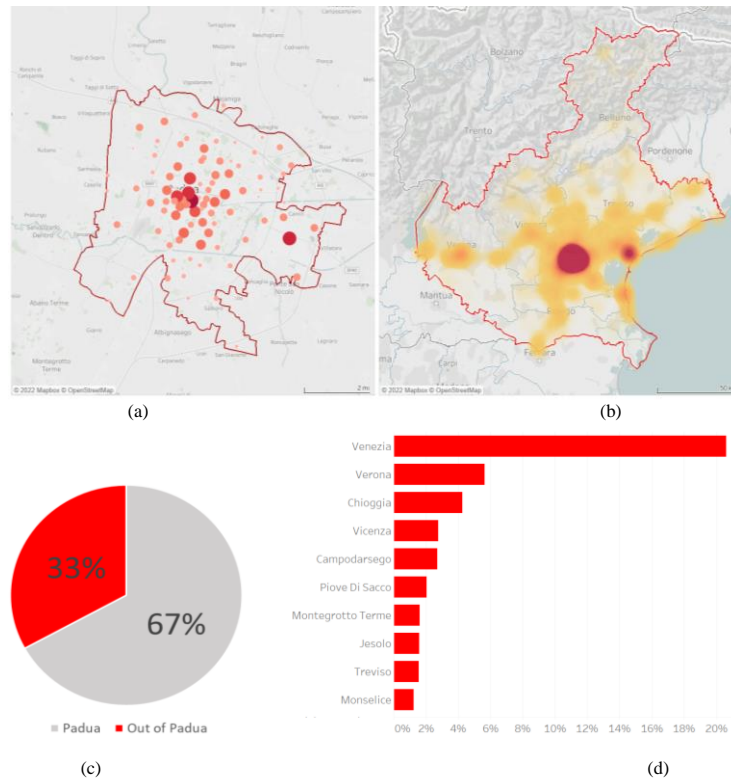
- prevailing night area: this information approximates the location of the overnight stay of the tourist and corresponds to the coverage of the prevailing network cells where the Italian SIM is mainly registered during the night;
- telco home location (or phone residence): corresponds to the municipality where the Italian SIM has been registered more frequently in the past sliding window, typically 6-12 months long. This way Italian users can be disaggregated by region/province/municipality, based on the telephone residence. Also, as a first approximation, the residence of mobile phone users with a foreign SIM can be associated with the nationality of the SIM.

Thus, we can classify as tourists those users whose prevailing night area differs from their telco home location and as same-day visitors those who visit a municipality for at least 3 hours without having the phone residence there or nearby.

We now represent a trial carried out in the city of Padova, which is commonly recognized as a base of overnight stays for tourists visiting the Veneto region. ISTAT traditionally measures arrivals and nights spent in the territory through a census from recorded flows in official accommodation establishment, carried out on a monthly basis. What we want to study is the daily activity of tourists to Padova; where do they move during the holiday? We have considered all people spending a night in Padova in July 2021 and classified them as follows: (i) inhabitants of Padova and nearby, (ii) inhabitants of Veneto, (iii) Italians, (iv) Foreigners.

A subset of main results is shown in Fig.1. Fig.1a presents the geographical distribution of overnight stays of tourists in the area of Padova municipality. Fig.1b gives a visual representation of tourist mobility the day after their overnight stay in

Padova for both Italians and Foreigners. Fig.1c provides a quantitative description of the time they spend within and outside Padova municipality; a significant result is that a large percentage of daytime is spent outside Padova, with a strong presence, as expected, in Venice (Fig.1d). The behaviours of Italians and Foreigners are quite similar; this is the reason why they have been always represented together.



**Figure 1:** (a) Geographical distribution of overnight stays in Padova; (b) Mobility of tourists of Padova in Veneto region; (c) % daily time (8-22) spent within and outside Padova (d) ranking of municipalities visited by the tourists of Padova (% of total time outside Padova).

## 4 Conclusions

In this paper we have presented the capacity of mobile phone big data to deeply and promptly describe the behaviour of people on the territory with a high granularity on both time and space domain. NPRR is indeed a great opportunity to deeply investigate and deploy the analytical skills of big data generated near-real-time by mobile network operator for the accurate territorial monitoring. New perspectives of investigations, algorithms and KPIs can be opened in the framework of tourism, sustainable mobility, and ecological transition.

We have presented an application case in the touristic field to show how Vodafone Analytics metrics can support and complement traditional surveys to identify and quantify the presence of visitors in all the small municipalities and/or touristic destinations where there are no official accommodation establishments and the relevant data from survey/census.