



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Collection and Sources

22-24 May 2024, Geneva, Switzerland



Mobile Phone Data for Enhanced Tourism Statistics in Italy: Insights from Vodafone-Istat Project Foundation

Maria Teresa Santoro, Lorenzo Cavallo and Silvia Di Sante (Istat, Italy)

LORENZO CAVALLO

Istat | Directorate for Environmental and Territorial Statistics (DCAT) - Rome

Introduction

Tourism information is fundamental and increasingly in demand by various data users, particularly policymakers at both national and local levels, sector operators, researchers, and analysts. The shift towards utilizing alternative data sources for producing official statistics has emerged as a prominent theme

This work presents an overview of the activities and findings from **Project Foundation**, a collaborative initiative between Istat, Vodafone (a Mobile Network Operator), and Motion Analytica



Its aim is to identify innovative methodologies for statistics and ensure their quality while testing **the potential use of mobile phone data in producing official statistics for the tourism sector**

Use of Mobile Network Operator data for tourism statistics



The **Project Foundation** began with a preliminary phase, the **Sprint**, conducted from September to November 2020

Its primary goal was to initially explore the potential of mobile data in providing estimations and trends regarding **arrivals and nights** spent in Italy, encompassing both *domestic and inbound tourism*

Subsequently, the **PF** delved into various aspects, notably focusing on enhancing **the convergence of conceptual definitions and classifications**

Some of the themes explored included:

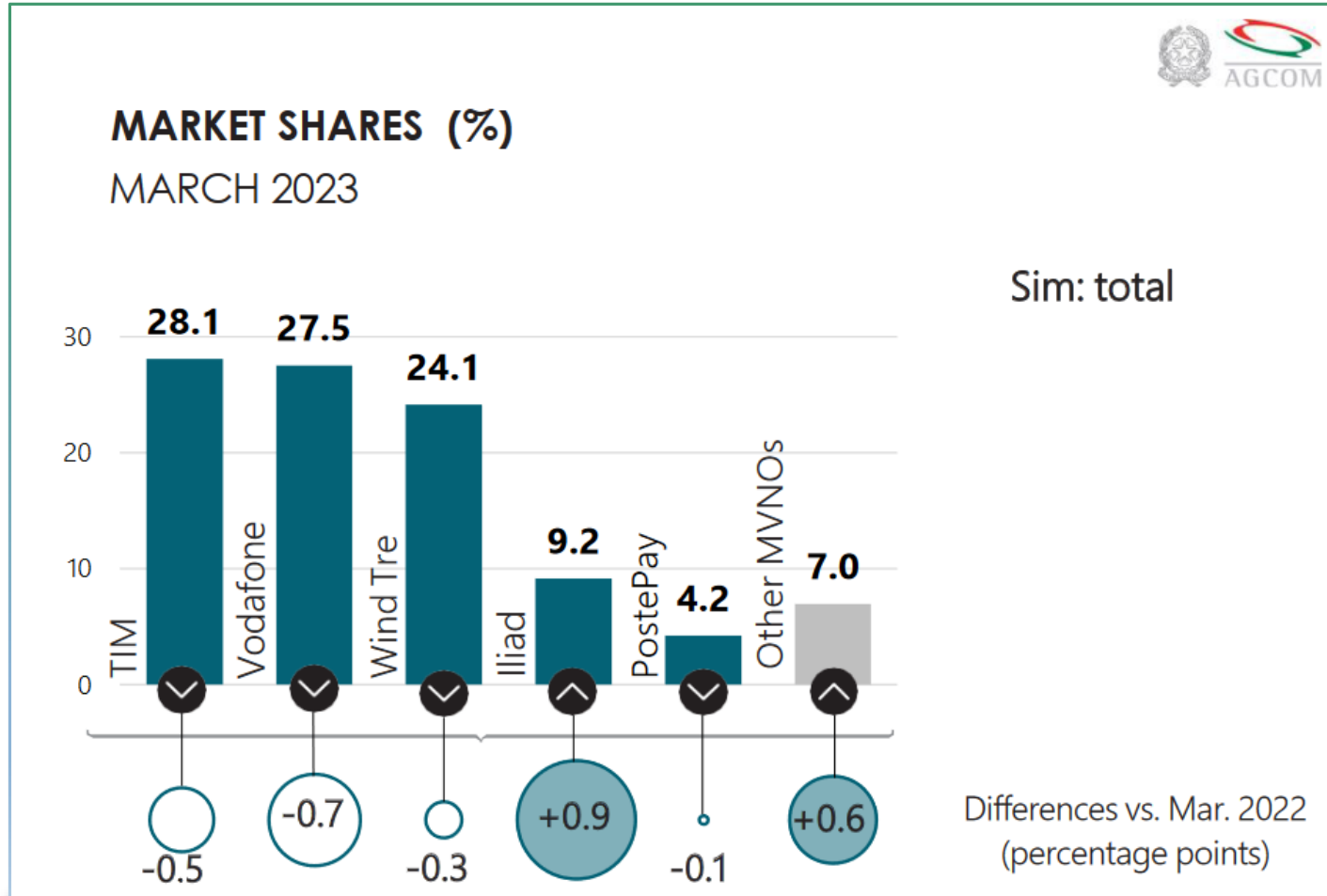
- The potential use of mobile data to further disaggregate tourist flows at territorial and/or temporal levels
- The feasibility of leveraging its timeliness
- The exploration of using mobile phone data for estimating, forecasting, or calibrating surveys
- Identifying possible new areas of interest

At first glance, it appears very promising for complementing and enriching current investigations on tourism, especially due to its ability to provide **highly timely information with detailed territorial coverage**

Mobile Network Operator



Features of Vodafone



- ❖ Beyond 200,000 cell phone calls
- ❖ Covers nearly 99% of the Italian territory
- ❖ Data collection occurs every minute
- ❖ Spatial data resolution is higher in urban areas and lower in rural areas

Characteristics of data



We can say that the MNO data provide information on:

- **Who:** Code SIM (Subscriber Identity Module)
- **Where:** Area covered by antenna
- **When:** Day/Time

*What **DON'T** they tell us?*

WHY the SIM is located in a specific area
WHY it moved there

We accept some forcing on the **WHY** (one SIM \neq one person) and **WHERE** (territorial assignment)

The technology employed by Vodafone Analytics intercepts and collects real-time data flow between smartphones and the mobile phone network (**signaling data**). This process ensures granular temporal and spatial information through proprietary estimation algorithms, although some limitations exist. For instance, the territorial assignment based on cells may lack precision

Cell tower records enable the generation of a table that minutely identifies the connecting cell for each user. Each cell corresponds to a specific municipality based on its coverage area. Consequently, at any given minute, each user is assigned to a particular municipality

From *Sprint* to *Project Foundation*

Sprint

Perimeter of the analysis: Province of Rimini and the municipality of Rome (2 areas with a strong tourist vocation)

Time periods: 3 periods - August 2019 and 2020, and April 2020 (period of *lockdown* with tourist flows equal to 0)

Case study (from **supply side**): arrivals and nights spent of tourists resident in Italy (*domestic*) and coming from abroad (*inbound*)



Project Foundation

Perimeter of the analysis: Emilia Romagna and Lazio Regions (*detailed data at municipal, provincial and regional level*)

Time periods: From June 2022 to October 2023

Case study (from **supply side**): arrivals and nights spent **by residence of the guests** (Regions for domestic tourists, **Countries of residence** for international tourists)



Methodological Issues

One of the primary challenges we encounter is how to establish a relationship between the resident population and **Telco** users, encompassing all individuals with smartphones or SIM cards.

It's essential to note that all information provided extends to the entire reference universe

To overcome these challenges, a **proprietary algorithm** developed by Vodafone Analytics is employed. This algorithm is specifically designed to represent the entire user universe, extending beyond solely Vodafone SIM owners or foreigners connected to the Italian network.

It is informed by:

1. Local market share analysis of the operator, considering Italian SIMs by province and age groups
2. National market share of foreign SIMs
3. Market share based on SIM type from market studies and official reports (AGCOM)
4. Socio-demographic characteristics of users derived from proprietary data

Other Issues addressed

A second set of issues was addressed during the activities prior to the Sprint and subsequently during the Project Foundation:

- difficulties in precisely assigning a SIM user to a specific municipality, both in terms of **destination** (due to antenna coverage not tied to municipal administrative boundaries) and in terms of **origin** (difficulty in determining the **Phone Residence**)
- challenges in accurately identifying **residents in Italy (domestic users)** from those coming from abroad (**inbound users**) regardless of the nationality of the SIM card in possession
- association of specific concepts from tourism surveys (clearly defined by international classifications) with information from mobile phone data (definition of arrival, night spent, residence, etc.)

Definitions adopted

Concepts	Official Definitions	Definitions adopted
Usual residence	Municipality of usual residence of the tourist	<p>Prevalent Night Cell: the most frequent cell the user is attached to between 8 PM and 8 AM</p> <p>Prevalent Municipality: This is the municipality where the user spends the most hours. To define a valid overnight stay, the user must spend at least six hours within the municipality</p> <p>Phone Residence: the municipality of the Prevalent Night Cell in the last 12 months before the reference period</p>
Tourist	Traveler making an overnight stay in a destination outside the municipality of usual residence, for any purpose	A user in a Prevalent Night Cell in a municipality other than the Phone Residence (defined, therefore, as the Municipality of Destination)
Arrivals	No. of guests who carry out the <i>check-in</i> in accommodation establishments (of a given municipality) in the reference period	A Tourist is registered in the Municipality of Destination as an Arrival on the first day of his trip
Nights spent	No. of nights spent by guests in accommodation establishments (in a given municipality) in the reference period	Sum of nights spent by Tourist in the reference period, where Nights = daily sum of Tourist in the Municipality of Destination
Resident tourist (domestic)	Tourist resident in Italy	Tourist with an Italian SIM or Foreign tourists (Foreign SIMs) who spent at least 3 months in Italy
Non resident tourist (inbound)	Tourist Non resident in Italy	Users who do not use an Italian SIM and who have spent less than 90 days in Italy in the 12 reference months

Initial data analysis

The initial tests on the adopted definitions revealed a significant discrepancy between the tourist data estimated from mobile phone sources and those from **Istat supply side** (only accommodation establishments, **no second homes, private accommodations, guests of relatives and friends, etc.**)

To further investigate the quality of the definition adopted and the reasons for the large discrepancies, we considered April 2020 as the reference time (a month characterized 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

The April 2020 data, however, showed an even greater disparity between the two sources

However, these initial analyses highlighted some problems to be addressed:

- considering only **one** municipality in the **Usual residence** overestimated **domestic flows** (due to the presence of workplaces, second homes, frequently visited destinations, etc., considered among tourist destination municipalities)
- abnormal origins of **foreign tourists** from certain nations (i.e., the nationality of the foreign SIMs registered in April largely refers to nationalities present in Italy for business reasons rather than for leisure)

Improvements introduced with the *Project Foundation*

The *Project Foundation* allowed us to make some steps forward.

- Improvement in the definition of **Usual residence**: from **Prevalent Municipality** to **Prevalent NIGHT Municipality** (i.e., the municipality where the user has spent the most nights in the twelve reference months)
- Improvement in the definition of **Non resident tourist**: users who do not use an Italian SIM **and who have spent less than 90 days in Italy in the 12 reference months** (to exclude resident foreign workers in Italy from the count of tourists)
- Introduction of the concept of **Usual Environment** (for better estimate tourists **Resident in Italy**) also considering a “buffer” of municipality of origin

Usual Environment

The **Usual Environment (UE)** of a user comprises municipalities where they habitually spend time; their presence in these municipalities doesn't signify tourist behavior.

A municipality is considered part of their UE if the user exceeds a specific hourly threshold for a given number of months

Expanding the UE definition reduces the tourist phenomenon by including more municipalities where the user isn't deemed a tourist

Different UE definitions were tested, each with a 'buffer' version, adding neighboring municipalities to the set obtained

USUAL ENVIRONMENT

- **"onlyhome"**: only the municipality of usual residence
- **"6m-50h"**: municipalities with at least 6 months of presence and a no. of hours exceeding 50 times the no. of months, plus the municipality of usual residence
- **"4m-50h"**: municipalities with at least 4 months of presence and a no. of hours exceeding 50 times the no. of months, plus the municipality of usual residence;
- **"4m-25h"**: municipalities with at least 4 months of presence and a no. of hours exceeding 25 times the no. of months, plus the municipality of usual residence;
- **"6m-0h"**: municipalities with at least 6 months of presence or municipalities with 3 months of presence and at least 125 hours, 4 months and at least 100 hours, or 5 months and at least 75 hours, plus the municipality of usual residence

Main Results

- ✓ The **Project Foundation** undoubtedly allowed us to make great strides forward compared to the Sprint
- ✓ The study on **Usual Environment** enabled us to enhance the estimates of **domestic tourism**
- ✓ A significant improvement has been the change in the definition of **non-resident tourists**, allowing for the almost complete elimination of the anomalous presence of tourists from country typically present for work-related reasons rather than tourism
- ✓ Nights spent calculated using Telco data are consistently higher than those detected by Istat (NB: Telco data do not allow us to distinguish where the tourists stayed!)
- ✓ The estimation of arrivals appears to have improved compared to the Sprint, but further insights and refinement of definitional convergence are still necessary
- ✓ It should be noted that these findings pertain only to the two territories analyzed; a national analysis could yield different results, hence caution must be exercised

Evaluations and prospects

The experience has confirmed the potential of MNO data, but...

- At the moment, we are still far from being able to use mobile phone data outright to produce official statistics on tourism. However, the analyzed trends have highlighted that it might be possible to use these data to estimate territorial trends
- It is also necessary to further refine definitional convergence
- MNO data can open new analytical perspectives in the tourism sector, exploiting territorial granularity and timeliness
- It is difficult to imagine that mobile data can replace traditional data sources for the production of tourism statistics
- In the short term, MNO data are not ready for use: constant interaction with mobile operators is necessary to test definitions, algorithms, and evaluate the quality of the data

Thank you

LORENZO CAVALLO | cavallo@istat.it