doi.org/10.26398/asaproc.0089

Enhancing mobility and tourism through data analytics and generative AI

Denis Cappellari , Glauco Mantegari , Bruno Zamengo,
Motion Analytica, Udine, Italy.

1. Introduction

Today, vast amounts of data from mobile operators, floating car data (FCD), traffic loops, public transport ticketing, and GPS apps generate valuable insights for decision-makers in urban planning, tourism, and sustainable mobility.

This presentation showcases applications where anonymized big data - telecommunications, FCD, and GPS - complement official statistics to analyze urban presence and mobility. These insights help quantify visitor patterns, transport preferences, traffic congestion, trip lengths, and more. In urban planning, mobility data reveal commuting patterns, peak traffic times, and public transport usage, aiding infrastructure optimization. In tourism, visitor flow analysis improves resource allocation and enhances experiences. In sustainable mobility, data-driven insights help reduce congestion and promote eco-friendly transport.

Finally, advancements in large language models (LLMs), such as OpenAI's GPT, Google's Gemini, and Meta's Llama, enable intuitive data analysis beyond SQL queries and dashboards. We will explore how these technologies make insights accessible to a broader audience, including non-technical users.

The presentation is organized as follows: Section 2 is dedicated to tourism, Section 3 to the analysis of civil aviation, Section 4 to the study of mobility, and Section 5 to our approach to using AI and Generative AI.

2. Tourism

We first explored mobile big data in tourism to provide faster, more detailed statistics than official sources. Telecom data has been widely studied for assessing tourism's impact on residents (Baldin et al., 2024), analyzing overtourism (Bertocchi et al., 2021), and enhancing official statistics, F. Ricciato, et al., (2018), E. Saluveer, et al., (2020), United Nations Committee of Experts on Big Data and Data Science for Official Statistics (2021), Nurmi et al., (2019), Ricciato, et al., (2021).

A study published in 2024, (Scotti et al., 2024), based on data provided by Motion Analytica and Vodafone Business Italia, addresses the drivers of overnight stays and same-day visits in the tourism of the Lombardy region of Italy. Mobile network data proved to be more granular and richer than official statistics and not affected by the biases that are typical of other sources (e.g., geolocated photos), which are more commonly used by specific segments of the population.

Key considerations when using mobile network data include its uneven infrastructure distribution, favoring dense areas and affecting precision by region. To address this, we apply techniques ensuring robust statistics, with greater accuracy in well-covered areas and larger-scale phenomena. Additionally, as the data reflects only smartphone users, we account for this limitation when extrapolating results, integrating other sources and official statistics where possible.

The produced statistics were then used to study which factors might influence the duration of stays, seeking correlations with the presence of specific services (e.g., ski slopes) or specific points of interest (e.g., cultural attractions).

Thanks to the ability to study the behaviour of day-trippers — considered one of the additional strengths of this data source — the authors were able to verify that events such as festivals or specific intermodal nodes influence the level of day-trippers attracted to a destination.

The authors also suggest that this type of data and the derived analyses can be useful for defining specific policies for a territory or a group of nearby destinations. They propose, for example, the identification and creation of common hubs for accommodation, transport, and major attractions, such as cultural sites, using more peripheral areas to offer various leisure and recreational attractions that will draw in day-trippers.

3. Aviation

The world of civil aviation includes two main types of actors: airports and airlines. For an airport, a strategic aspect is its centrality within the network of air connections to attract as many travelers as possible.

The centrality of an airport can be viewed from two perspectives: inbound traffic analysis, which refers to the airport's capacity to serve as an entry point for travelers wishing to reach the airport's region, and outbound traffic analysis, which refers to the airport's capacity to act as a gateway to remote destinations of interest for people living near the airport.

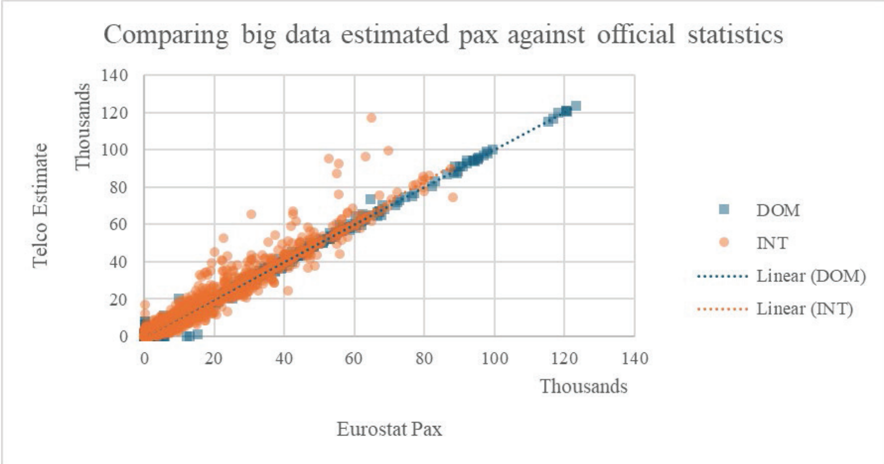
In this context, mobile phone data, cross-referenced and enriched with other data sources such as timetables, has proven to be a valuable resource (Motion Analytica, 2022) compared to other market solutions or official sources, (Associazione Italiana Gestori Aeroporti, 2022), (Eurostat, 2022). This data provides airports with tools to estimate their catchment and distribution areas based on the actual behavior of users.

These basic models were initially supplemented by an unsupervised study using clustering techniques, which gradually evolved into more sophisticated machine-learning solutions that integrated mobile phone data with the actual departure/arrival times of flights at airports. The integration of official statistics from a preceding period allowed for precise expansion of estimates from the sample population to the entire target population including all individuals who flew from any of the over forty major Italian airports, not just Vodafone Italy customers or those roaming on their network. In Figure 1 we report the comparison of our estimates with official data for routes (see Eurostat, (2022), month of August 2022) showing that the overall correlation per route exceeds 99% for domestic flights and 95% for international flights, with linear coefficients of 1.00 and 1.04, respectively. In Figure 2 additionally, as expected, we inspected the MAE index (see Morley *et al.*, 2018), which indicates that the accuracy of the estimates increases with the size of the phenomenon: routes with higher passenger traffic report significantly lower MAE indices. Consequently, where phenomena are smaller and variability is greater, we typically aggregate these data into statistics that describe larger phenomena, without adding additional heuristics or forcing the data.

One of the added values of big data estimates compared to official statistics is the ability to investigate additional details. In this case, for example, it is possible to delve into route data concerning the estimated habitual residence of users based on their behavior.

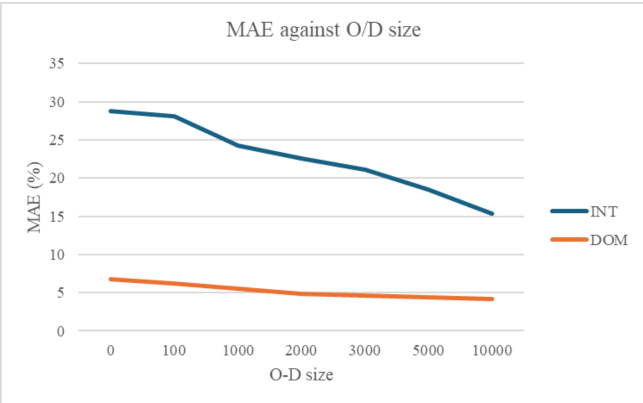
The maps in Figure 3 show the market share among the airports of Ancona, Bologna, Firenze, and Pisa by the municipality of residence of travelers. The first map covers all international routes while the other map shows routes to Germany. For privacy reasons, only municipalities with a certain minimum number of travelers are reported.

The ability to determine the origin of travellers from a particular airport or even a specific route can be used to understand the airport's catchment area based on actual user behavior. It can also help quantify phenomena such as leakage, where travellers choose to use an airport that is farther from their home because they might be interested in a route not available at the nearby airport, or due to the convenience of the airline's offerings, or better connections at the farther airport. For instance, figures extracted from image Figure 3 show that Bologna airport appears to be the strongest, having the broadest catchment area. Florence and Pisa show an interesting pattern: overall, Pisa's catchment area is larger than Florence's but, when analyzing connections with Germany, Florence's catchment area becomes more extensive.



Market	R^2	Slope
DOM	99%	1.00
INT	95%	1.04

Figure 1 The x-axis shows number of passengers recorded by official statistics, while the y-axis our big data estimated value.



O-D size	INT	DOM
0	28.74	6.73
100	28.09	6.22
1000	24.27	5.51
2000	22.61	4.85
3000	21.11	4.67
5000	18.53	4.44
10000	15.31	4.15

Figure 2 We grouped O-D (origin-destination) pairs into bins depending on monthly passenger count. The "O-D size label" represents the bin's lower bound (e.g. "100" indicates O-D pairs with 100-999 passengers). The blue line and the "INT" column in the table show the MAE index for international O-Ds, while the orange line and the "DOM" column refer to domestic O-Ds.

International air travelers share of competitors

Air travelers to Germany share of competitors

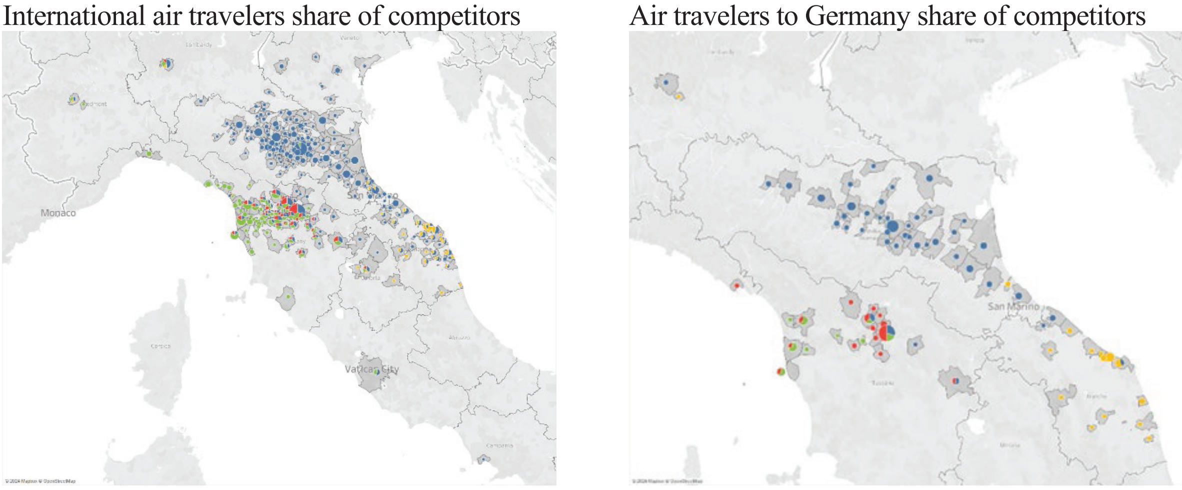


Figure 3 Air traveller market share distribution among Ancona (yellow), Bologna (blue), Firenze (red) and Pisa (green) airport at municipal level. Each pie chart represents the share of travellers in various municipality. The maps provide a visual comparison of competitor dominance and regional preferences in each travel category.

4. Big data for general mobility

The study of mobility through big data and ML/AI experienced significant acceleration during the COVID-19 pandemic, when numerous studies were developed using different data sources to detect changes continuously and extensively in people's movement habits.

Some studies focused on the use of GPS data, as seen in (DePhillipo et. al., (2021), where Google COVID-19 mobility reports were used alongside with social distancing measures to find significant associations between geographic activity and prevalence of COVID-19 infections in the U.S Other studies leveraged mobile network data to monitor mobility behaviors during the pandemic (Jewell et al., 2021), (Xiong et al., 2020), (Calabrese et al., 2021). These studies represent an evolution of earlier research (Calabrese et al., 2011), (Ricciato et al., 2020) which explored the possibility of using mobile network data to measure mobility. For instance, they developed Origin-Destination matrices to study actual mobility demand on a broad sample. Moreover, GPS data tend to be biased since they are generated by a specific population that uses this technology more frequently. Consequently, a strategy to obtain more robust statistics consists of combining insights derived from both sources—MNO and GPS—leveraging their respective strengths.

In addition to these data sources, which typically detect the movement of people, other contexts have seen the development of studies examining the mobility of specific categories of vehicles. For example, the use of floating car data (FCD) from vehicle black boxes was explored in (Mannini *et al.*, 2018) and (Nalubega, 2023) for examining parking demand, estimate parking search time, and measure traffic congestion. Finally, in (Costa & Silvestri, 2021) and (Cavattoni et al., 2024) shared mobility data, such as bicycles and scooters, was used to optimize bike spreading and to measure COVID-19 impact on shared mobility. In the case of FCD and shared mobility data, these are closely tied to the market penetration of the provider and, since they pertain to vehicles, are not directly connectable to other sources focused on people. However, the integration of various sources, while accounting for their limitations and unique characteristics, can provide a more nuanced and complete picture of the phenomenon.

As in other sectors, the use of big data for mobility and the insights derived from it can serve various purposes. Two major Italian institutions, (Ministero delle Infrastrutture e della Mobilità Sostenibili, 2022) and (Tartaglia *et al.*, 2024), identify the following main applications: (i) Measure the real demand for mobility comprehensively, detailing it by means of transportation; (ii) Plan and improve public transport services and infrastructure; (iii) Plan the ecological transition to support decisions that promote the shift towards more sustainable mobility and (iv) Monitor the evolution of mobility behaviors over time and measure the effectiveness of chosen strategies and implemented interventions.

These data have also been used to explore the possibility of supplementing official statistics. For example, in our joint work with ISTAT and Vodafone (Radini et al., 2021), we studied the population that habitually resides in an area and constructed commuting matrices. From these experiments and other experiences by official statistical agencies, several European projects have emerged with the aim of defining a common framework for all member countries. This framework seeks to supplement traditional sources with these new data sources in the production of official statistics for the European Community (Ricciato et al., 2018), (Ricciato, 2018), (Ricciato et al., 2021).

5. AI & GenAI

In November 2022, ChatGPT was announced and released to the public, creating a new wave of interest in AI and ML. However, several works utilizing these types of technologies had already been in use for some time. For example, the solution presented in Section 3 also employs typical ML techniques to correctly associate a passenger with a flight.

A publicly documented case, which is more extensively covered, can be found in (Kim *et al.*

(2022) and further explored in its bibliography), where the authors propose a machine learning model to classify modes of travel based on movement detected by GPS data, using a convolutional neural network.

Thanks to recent technological advancements pioneered by OpenAI, followed by contributions from Google, Microsoft, and Meta, Large Language Models (LLMs) are proving to be an extremely versatile tool. One potential application we explored in (Padoan *et al.*, 2024), involves the use of the reasoning and communication abilities of a LLM to interact with data through natural language. This approach reduces the dependence on dashboards or SQL queries, thereby making data more accessible to a broader audience. To implement this, we designed a cloud-based architecture using (i) Langchain, a Python library designed to streamline the creation of conversational AI applications; (ii) Llamaindex, a data framework designed to support the development of applications based on LLMs; (iii) Azure AI, cloud-based service offering a variety of AI tools for building and deploying AI applications and (iv) Chainlit, an open-source Python package designed for building production-ready conversational AI applications.

From our experiments, we found that by appropriately adapting the general model to specific datasets and providing the LLM with the necessary context, it is possible to create chatbots capable of answering domain questions based on both structured and unstructured data. This required instructing the application with all the contextual elements necessary to interpret the data structure and the semantic value of the contained data. This was achieved through a JSON file containing these metadata and several examples of questions and answers shown to the chatbot engine. The chatbot developed for these experiments was able to deliver basic descriptive statistics as well as perform more complex analyses that required additional computations. Moreover, by equipping the LLM with the right tools, we were able to use it to generate customized charts.

References

Associazione Italiana Gestori Aeroporti (2022). Tabella dati di traffico.

Baldin, A., Bertocchi, D., Camatti, N. (2024). Crowding-in and (temporary) crowding-out in Venice. The effect of cultural events on residents. *Tourism Economics*, 0(0).

Bertocchi D., Camatti, N., Van der Borg, J. (2021) Tourism peaks on the three peaks. Using big data to monitor where, when and how many visitors impact the Dolomites UNESCO World Heritage Site, *Rivista Geografica Italiana*, vol. 3.

Calabrese, F., Ferrari, L., Blondel, V. D. (2014). Urban sensing using mobile phone network data: A survey of research. *ACM Computing Surveys*, 47(2), pp. 1–20.

Calabrese, F., *et al.* (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4), pp. 36–44.

Calabrese, F., *et al.* (2021). Using Vodafone mobile phone network data to provide insights into citizens' mobility in Italy during the Coronavirus outbreak. *Data & Policy*, 3, p. e22.

Cavallo, L., *et al.* (2022). Exploring mobile network data for tourism statistics: The collaboration between Istat and Vodafone Business Italia. *Rivista di Statistica Ufficiale*, 3, pp. 43–76.

Cavattoni, M, Comin, M., Silvestri, F. (2024). COVID-19 pandemic's enduring impact on urban mobility: The case of free-floating bike sharing in Padova, Italy, *Transportation Research Record.*

Confente, I., *et al.* (2024). Integrating tourists' walk and talk: a methodological approach for tracking and analysing tourists' real behaviours for more sustainable destinations. *Journal of Sustainable Tourism*, 1, pp. 1–21.

Costa, E., Silvestri, F. (2021). On the bike spreading problem.

DePhillipo, N. N. et al., (2021) Mobile phone GPS data and prevalence of COVID-19 infections: Quantifying parameters of social distancing in the U.S.,*Arch. Bone Jt. Surg.*, 9, (2), pp. 217–223,

Eurostat (2022). Air transport measurement – passengers and freight.

Jewell, S., et al. (2021). It's complicated: Characterizing the time-varying relationship between cell phone mobility and COVID-19 spread in the US. *npj Digital Medicine*, 4(152).

Kim, J., Kim, J. H., Lee, G. (2022). GPS data-based mobility mode inference model using long-term recurrent convolutional networks. *Transportation Research Part C: Emerging Technologies*, 135, p. 103523.

Mannini, L., Cipriani, E., Crisalli, U., Gemma, A. (2018). FCD data for on-street parking search time estimation. *IET Intelligent Transport Systems*, 12, pp. 664–672.

Ministero delle Infrastrutture e della Mobilità Sostenibili (2022). *Mobilità e logistica sostenibili – Analisi e indirizzi strategici per il futuro*.

Motion Analytica (2022). Analysis of the catchment area of Milan airports.

Morley, S. K., Brito, T. V., Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space Weather*, 16, pp. 69–88.

Nalin, A., *et al.* (2024). Assessing veracity of big data: An in-depth evaluation process from the comparison of mobile phone traces and ground truth data in traffic monitoring. *Journal of Transport Geography*, 118,

Nalubega, S. I. (2023). Development of a floating car data (FCD) model to evaluate traffic congestion: A case of Kampala, Uganda. Stellenbosch University.

Nurmi, O., Piela, P. (2019). The use of mobile phone data in tourism statistics. In *Proceedings of the Special Topic Session*, 62nd ISI World Statistics Congress, vol. 4, pp. 135–142, Kuala Lumpur, Malaysia, Aug. 18–23.

Padoan, L. et al. (2024). Mobility ChatBot: Supporting decision making in mobility data with chatbots. In *Proceedings of the 25th IEEE International Conference on Mobile Data Management (MDM)*, Brussels, Belgium, pp. 295–30.

Radini, R. et al. (2021). Come i dati di telefonia mobile possono contribuire alle statistiche sperimentali sulla mobilità.

Ricciato, F. et al. (2018). Processing of mobile network operator data for official statistics: The case for public-private partnership. Presented at the *DGINS Conference*, *Bucharest, Romania, Oct. 10–11.*

Ricciato, F., (2018). Towards a reference methodological framework for processing MNO data for official statistics, presented at the *15th Global Forum on Tourism Statistics*, *Cusco, Peru, Nov. 30, 2018.*

Ricciato, F. et al. (2020). Towards a methodological framework for estimating present population density from mobile network operator data. *Pervasive and Mobile Computing*, 68.

Ricciato, F. *et al.* (2021). A proof-of-concept solution for the secure private processing of longitudinal Mobile Network Operator data in support of official statistics. *Eurostat and Cybernetica*.

Ricciato, F., Coluccia, A. (2023). On the estimation of spatial density from mobile network operator data. *IEEE Transactions on Mobile Computing*, 22(6), pp. 3541–3557.

Scotti, F. et al. (2024). Exploring drivers of overnight stays and same-day visits in the tourism sector. *Scientific Reports*, 14, p. 9840.

Saluveer, E. et al. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research*, 81.

Tartaglia, M. et al. (2024). Methodological aspects for passenger mobility analysis using big data.

United Nations Committee of Experts on Big Data and Data Science for Official Statistics - UN-CEBD. (2021). *Mobile phone data*. United Nations, NY, USA.

Xiong, C. et al. (2020). Mobile device location data reveal human mobility response to state-level stay-at-home orders during the COVID-19 pandemic in the USA.