December 2023
Paper 1

Working Papers Collection

Spatial Analysis of Citizens' Travel Data: The Pollicino Project

Fondazione Transform Transport ETS
Research and Innovation in
Mobility and Transport Planning

Via Lovanio, 8
20121 Milan, Italy
T + 39 02 6231191

www.transformtransport.org
info@transformtransport.org

transform
transport

Spatial Analysis of Citizens' Travel Data: The Pollicino Project

Giulia Ceccarelli[✉1], Abubakr Albashir[1], Andrea Gorrini[1], Massimo Ciuffini[2], Luca Refrigeri[2], & Sofia Asperti[2]

[1] Fondazione Transform Transport ETS. Milan, Italy
[2] Fondazione per lo Sviluppo Sostenibile, Rome, Italy

✉ Corresponding author: Giulia Ceccarelli – *g.ceccarelli@transformtransport.org*

Abstract

Shifting mobility paradigms are leading to fragmented and flexible travel patterns in cities, no longer captured by traditional survey methods. Sensor data provided by multiple devices, including smartphones, enables ever more sophisticated analyses, that meet the necessity in understanding our environment. In this context, this research investigated citizens' travel behavior in Bologna, using GPS data collected through the Pollicino project, an app-based mobility survey carried out between May and June 2022. The study is structured in a data cleaning, processing, and mapping methodology, with the objective to identify spatial behaviors of citizens. The proposed analysis uncovered soft modes as predominant and concentrated in central zones of the city. Temporal assessments unveiled peak travel times, while aggregation reveals unique mobility patterns. Leisure emerges as the leading trip purpose. Notably, results highlighted that GPS data sourced through a mobility survey produces important insights to understand granular mobility behaviors, shedding light on key information for city planning and sustainable transportation.

Citation: Ceccarelli, Albashir, Gorrini, Ciuffini, Refrigeri, & Asperti (2023). Spatial Analysis of Citizens' Travel Data: The Pollicino Project. *Transform Transport Working Papers Collection, 1.* doi: https://doi.org/10.5281/zenodo.10361489

Made in Milan by Transform Transport

Keywords

Spatial analysis, GPS data, mobility patterns, GIS mapping, sustainable mobility.

1. Introduction

With the growth of sensor data provided by multiple devices, including smartphones, and due to the rapid development of analysis tools and platforms, the necessity of using the data provided by sensors is increasing and becoming crucial to understand our cities. In a similar manner, the GPS trajectories data recorded through smartphones and generated by moving objects provides the research community with a valuable resource for revealing patterns of human activities, transportation science, and urban studies (Yang et al., 2018).
This research focused on understanding the spatial mobility patterns of the GPS data collected through the project Pollicino[1] carried out by Osservatorio Nazionale Sharing Mobility[2] e Fondazione per lo Sviluppo Sostenibile[3] in the city of Bologna (Italy), between May and June 2022 (see *Figure 1*). In this context, this study was carried out by Transform Transport, engaged in the project Pollicino working group in the data analysis

[1] See: https://osservatoriosharingmobility.it/pollicino/
[2] See: https://osservatoriosharingmobility.it/
[3] See: https://www.fondazionesvilupposostenibile.org/

phase. Within the scope of the Pollicino project, a smartphone app-based mobility survey was developed and used to collect citizens' travel data in compliance with the GDPR policy. Following the Data Altruism concept, participants voluntarily participated in the study, using the mobile app to track their daily movements. This interactive collaboration of the participants in the data collection improves the awareness about mobility behaviors and in addition, represents a tool for involving the public in the framework of political choices to improve the cities (Ciuffini et al., 2023).

Figure 1 *– The Pollicino project, carried out by Osservatorio Nazionale Sharing Mobility and Fondazione per lo Sviluppo Sostenibile (freely taken from: https://osservatoriosharingmobility.it/pollicino)*

The main objective of this research was to unlock the potential of the Big Data collected by the mobile sensors by analyzing its spatial components in order to extract the mobility patterns, citizens' behaviors related to modal choice, and to understand its relation with both the urban characteristics and city infrastructure, as follows:

- Produce thematic maps (e.g., multimodal, socio-demographic, etc.) to describe the spatial distribution of trips;
- Visualize spatial clustering analyses, with the aim of identifying similarities in modal choice of participants, in relation to the origin, destination and route of each trip.

The paper is structured as follows, Section 2 outlines enabling data and methodology used to map the spatial component of citizens' travel behaviors; Section 3 identifies key results of the analysis, allowing for a comparative study of mobility behaviors, based on modal choice, trip purpose and mobility profiles of participants; Section 4 provides a reflection on how Big Data enables a deep understanding of mobility patterns, creating new possibilities for optimized and tailored city policies.

1.1 Related Works

This research focused on understanding the spatial mobility patterns of the GPS data collected through the project Pollicino carried out by Osservatorio Nazionale Sharing Mobility e Fondazione per lo Sviluppo Sostenibile in the city of Bologna (Italy), between May and June 2022. The project Pollicino was based on the voluntary participation of a sample of 600 citizens, which tracked their daily movements for seven consecutive days, annotating relevant information on modal choice and travel purpose on each trajectory and sharing their socio-demographic characteristics.

Related works to this study include research investigating trajectory data mining and methods to preprocess and analyze raw GPS pings, aiming at improving data accuracy in urban environments and creating best practices for integrating this data source in urban planning frameworks. Wang et al. (2020) structured a systematic literature review on trajectory data mining, providing definitions of trajectory data based on acquisition methods and purpose of application. Following Wang et al. (2020) definition, the Pollicino dataset is composed of Explicit Trajectory Data, defined as a type of well-structured data which directly provide time and location information and have strong spatiotemporal continuity. Explicit Trajectory Data composed by raw GPS pings is susceptible to errors due to many factors as well as many of the sensor's observations. These include the ionospheric delay of the satellite signal, multipath effects in urban areas, and signal blockage (Adhinugraha et al., 2020). For this reason, multiple data cleaning and outliers' removals methodologies have been implemented. Prabha and Kabadi (2020) provide a definition on the methods for GPS data acquisition and list a comprehensive set of

algorithms for data preprocessing. Meng et al. (2019) provide a systematic literature review on trajectories outlier detection, identifying classification based methods, clustering based methods, density based methods, statistical based methods for outliers' removal. Then, the authors identify the core steps in trajectories preprocessing methodologies, defined as noise filtering, segmentation, and map-matching. The latter corresponds to the process of finding the most probable corresponding points of the recorded waypoints of a trajectory on a road network (Saki & Hagen, 2022).
Furthermore, related works include mobility studies enabled by GPS data collection through smartphone apps. These can be undertaken for understanding multiple aspects of mobility dynamics across different sections of the population. Barbosa et al. (2018) listed a comprehensive overview of mobility data gathering methodologies, modelling approaches, metrics, and applications, stating that GPS based data provides the most granular and accurate source for movements trajectories. Moro et al. (2019) gathered the Breadcrumbs mobility dataset, collected from multiple sensors (e.g., GPS, GSM, WiFi, Bluetooth) on the smartphones of 81 participants and containing ground truth information on visited POIs and demographics of participants. Molloy et al. (2022) describe the MOBIS mobility dataset, collected in the context of a transport pricing survey in Switzerland. The dataset was gathered through a GPS tracking app, Catch-my-Day, and contains daily travel trajectories segmented by transport modes of 3680 participants nationwide. Larroya et al. (2023) structured a GPS records dataset from single-day home-to-school pedestrian mobility from multiple schools in the Barcelona metropolitan area, exploring citizens science practices involving youngsters.

2. Enabling Data and Methodology

The project Pollicino, carried out by Osservatorio Nazionale Sharing Mobility e Fondazione per lo Sviluppo Sostenibile in the city of Bologna (Italy), enabled the gathering of mobility data relative to a sample of citizens, voluntarily participating in the survey. The project took place over six weeks, between May and June 2022. During this time frame:

- 1827 citizens activated the IoPollicino app;
- 955 citizens completed the mobility survey, recording their activities for seven days;

Among them, 600 participants were selected based on socio-demographic characteristics and place of residency to be included in the analyses. These constitute a representative sample of the population, based on two demographic indicators: age group and gender. The daily movements recorded through the tracking app IoPollicino are structured in three datasets.
The first set of data contains the anonymous profile of each participant.

- Anonymized Participant ID;
- Socio-demographic (*e.g., gender, age group, salary group, car ownership*).

Furthermore, the dataset includes information on the mobility profiles of participants, these are identified based on their modal choices and behavioral patterns:

- *Metabolici*: participants always moving on foot, by bike and with sharing mobility modes;
- *Sostenibili*: participants moving by shared vehicles, by bike or on foot; they never use motorized vehicles;
- *Megamixer*: participants mixing all transport modes;
- *Autonomi*: participants moving by private means of transport (car, motorbike, bike + feet); they never use shared vehicles;
- *Auto/ moto Dipendenti*: participants using private motorized transport modes only.

The second dataset includes descriptive information about trips:

- Trip ID;
- Anonymized Participant ID;
- OD matrix (latitude, longitude).

The third dataset consists of the GPS point records between the origin and the destination of each trip to describe the full trajectory of the trip:

- Point ID;

- Trip ID;
- Latitude, longitude;
- Location accuracy [m];
- Speed [m/s];
- Time [dd/mm/yyyy];
- Travel mode (*e.g., car, bus, walking*).

2.1 Methodology

The objective of this research was to derive spatial insights, describing mobility patterns of participants of the Pollicino study. To this end, six methodological steps were performed (see *Figure 2*).

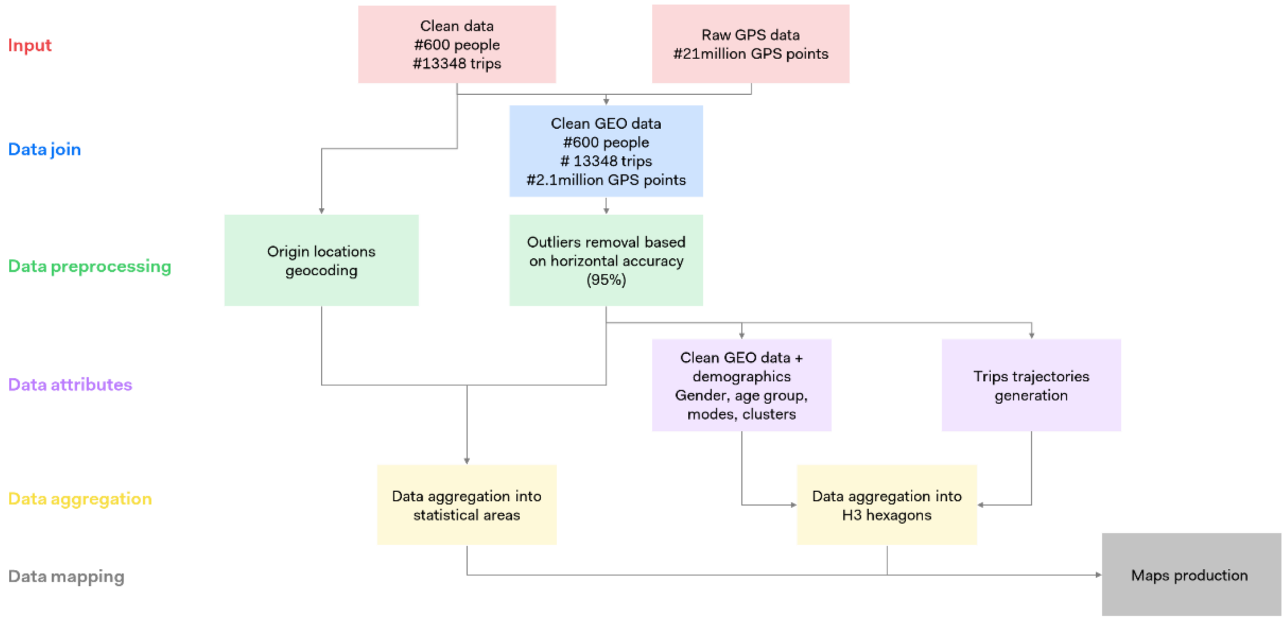

Figure 2 *– Methodological steps*

The first methodological step had the objective to gain an understanding of the data structure of the mobility dataset, containing roughly 21 million GPS points, of the clean participants dataset and of trips dataset described above. In the second step, mobility data, including origin and destination locations of trips and raw GPS points were joined based on their Trip ID, to create a complete mobility dataset. In the preprocessing step, geocoding was performed on the home locations declared by participants to analyze their aggregated distribution in the study area, in an aggregated an anonymized manner. Geocoding refers to the process of retrieving geographical coordinates from a text address. Parallelly, data cleaning was performed on raw GPS points, with the objective to reduce positioning errors in the dataset. The 95th percentile of the GPS data accuracy detected by the app sensor was used as a threshold to filter out the outliers from the data, following Kubo et al. (2020). Then, a final data validation step was performed to remove the trips that had lost most of their in-between points while performing the previous filtering steps. After the data cleaning phase, the trips were reduced from 13348 to 10819.

The fourth methodological step consisted in the generation of trajectories, which are composed of consecutive points having the same Trip ID being connected by a line segment (Wang et al., 2020). Trajectories were then associated to the Open Street Map (OSM) network of Bologna through a map-matching methodology, using the Valhalla algorithm (Saki and Hagen, 2022). Furthermore, the fourth step consisted in merging the generated trajectories with the socio-demographic information contained in the participants dataset, to derive granular spatial insights. Then, the fifth step consisted in aggregating data into homogeneous areas. The mobility data was aggregated into the hexagonal H3 grid, a hierarchical geospatial index based on hexagons developed by

Uber[4], with the goal develop a quantitative analysis on the distribution and volumes of trips passing by each homogeneous spatial unit. Parallelly, the geocoded origins were aggregated into ninety statistical areas, with the aim to obtain a comparison between the number of residents and the sample of participants of the Pollicino project. Last, aggregated data was mapped with the aim to unveil mobility patterns of related to different modal choices, temporal segments, and mobility clusters.

3. Results

This research investigated citizens' travel behavior in Bologna, using GPS data collected through the Pollicino project, the methodological process aimed at structuring participants trajectories from the raw GPS coordinates, minimizing positioning errors while maintaining consistency. Furthermore, the aim was to unveil spatial patterns in the mobility survey data, highlighting the relevance of continuous monitoring and GPS data sources for understanding granular travel behavior and target city policies.

3.1 Zoning

In this research, two zoning systems were used to analyze the datasets. The municipality of Bologna is divided into 90 statistical areas, these are used as a base for the preliminary data analysis phase (see *Figure 3*).

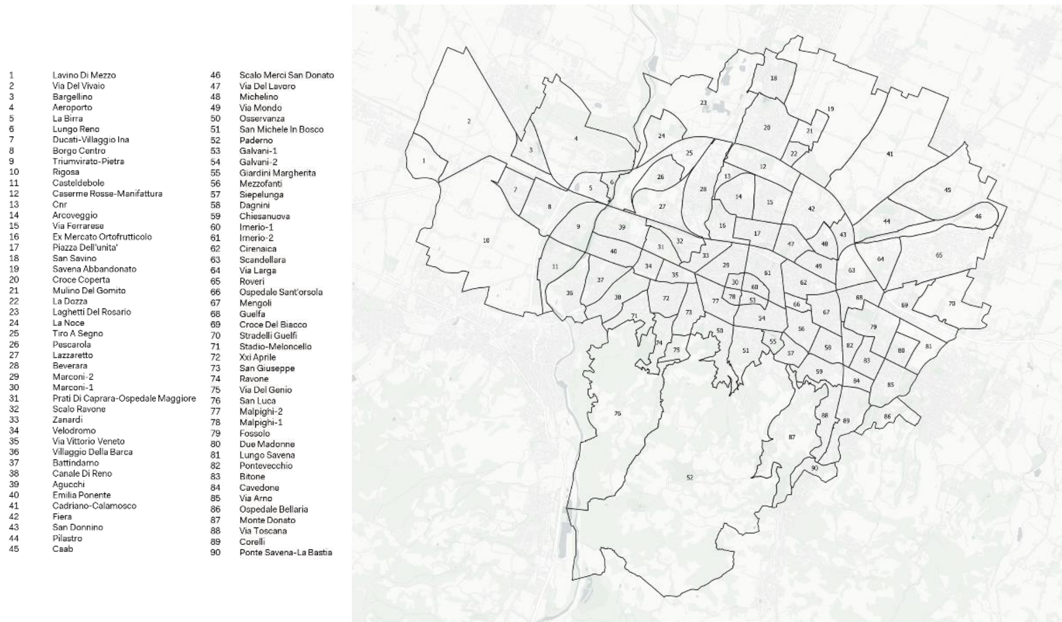

Figure 3 *– Zoning – Statistical areas (90)*

The hexagonal H3 grid, a hierarchical geospatial index based on hexagons developed by Uber, was used as a base for the mapping phase, to quantify the number of trips passing by each area. The municipality of Bologna is structured into 1305 hexagons, with resolution 9, with an edge length of 0.20km[5] (see *Figure 4*).

[4] See: https://h3geo.org/

[5] See: https://h3geo.org/docs/core-library/restable/

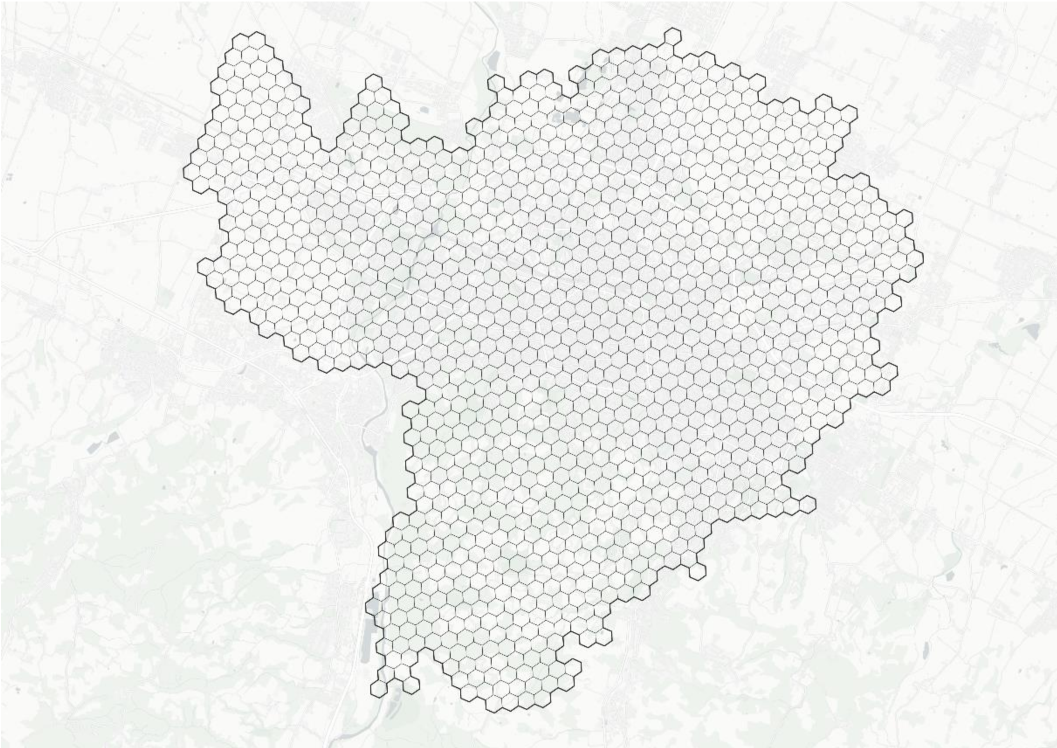

Figure 4 *– Zoning – H3 hexagons (1305)*

3.2 Preliminary data analysis

In the preliminary data analysis phase, the first aim was to gain an understanding of the distribution of participants and their representativity from a spatial perspective. Among participants in the study, 566 were living in the municipality of Bologna, while 34 lived in neighboring municipalities. In particular San Lazzaro di Savena and Casalecchio di Reno had 17 and 11 participants respectively (see *Figure 5*, *Figure 6*).

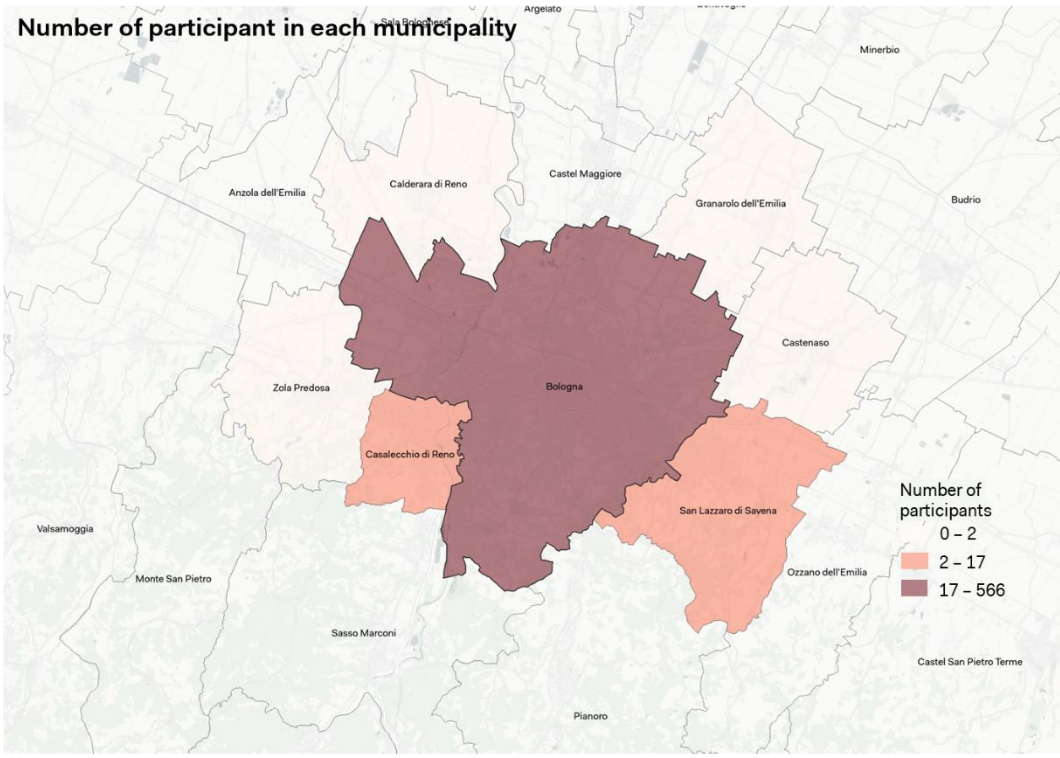

Figure 5 *– Bologna Metropolitan Area – number of participants by municipality*

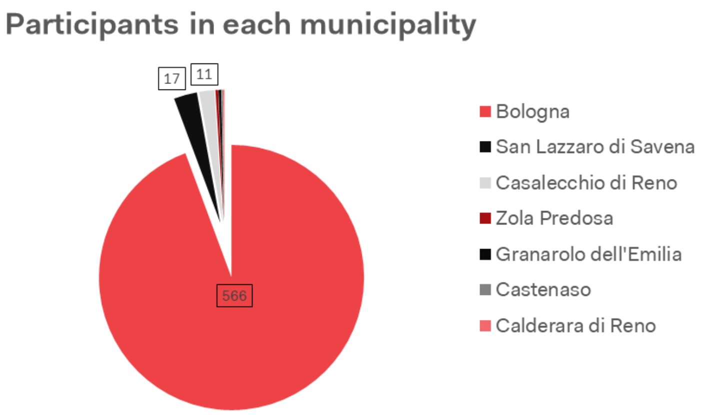

Figure 6 *– Bologna Metropolitan Area – number of participants by municipality*

Within the municipality of Bologna, participants were distributed mainly among central areas, zones 72 (XXV Aprile), 14 (Arcoveggio), 67 (Mengoli) and 61 (Irnerio-2) had the largest number of participants, respectively 26, 24, 21, 20 (see *Figure 7*). Overall, the spatial distribution of participants corresponded to a population's coverage 0.15% on average, with several areas that are not represented in the dataset. A peak of 2.67% in population coverage is seen in zone 66 (Ospedale Sant'Orsola) (see *Figure 8*).

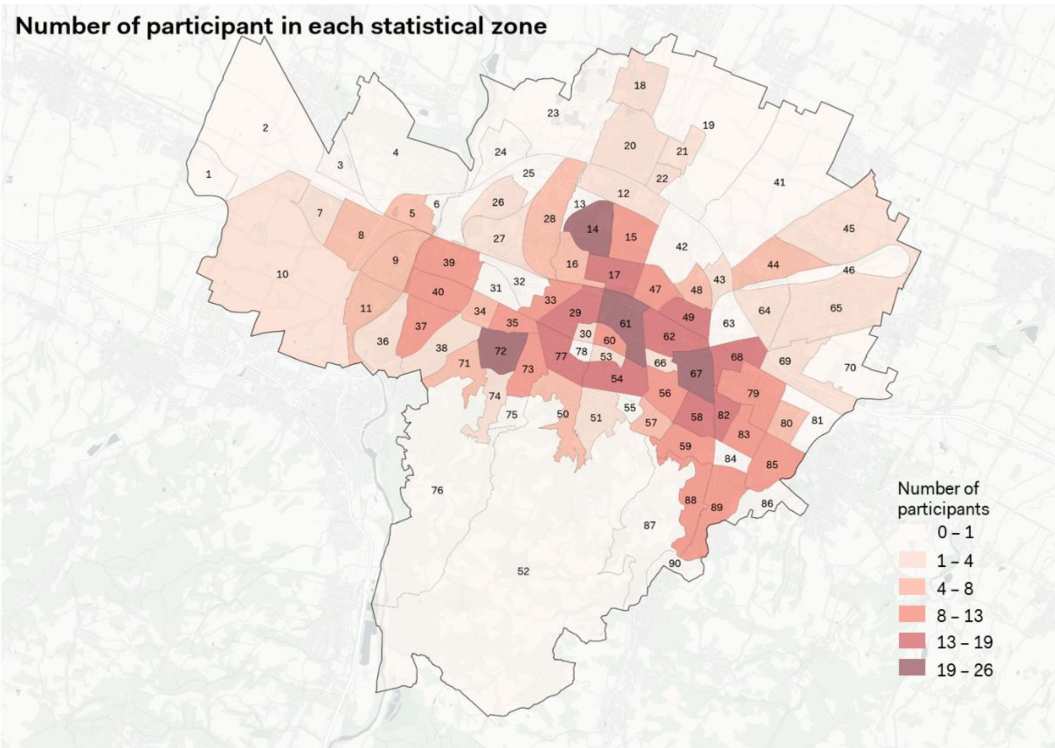

Figure 7 *– Bologna Municipality – number of participants by statistical area*

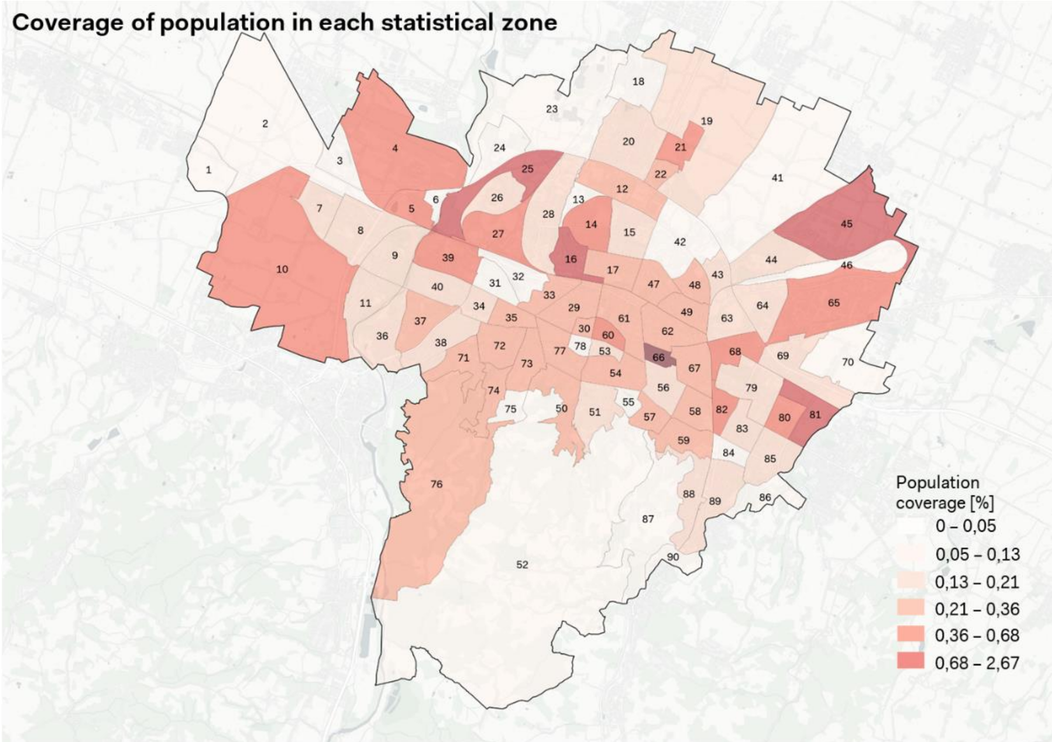

Figure 8 *– Bologna Municipality – population coverage by statistical area*

The second aim of the preliminary data analysis phase was to remove outliers from the GPS data. To this end, the 95th percentile of the GPS data accuracy detected by the smartphone app was used as a threshold to filter out the outliers. Results reveal significant improvements in the quality of the dataset, with the mean GPS accuracy shifting from 16m to 10m (see *Figure 9*, *Figure 10*). The greater improvements can be seen in Zone 4 (Aeroporto).

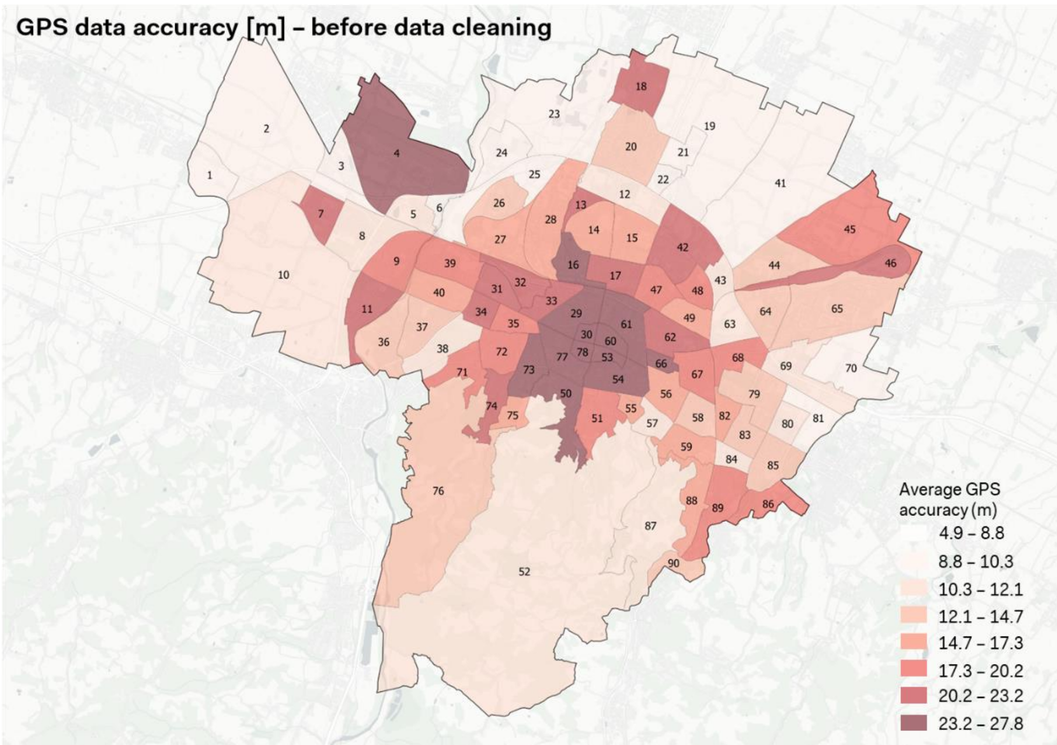

Figure 9 *– Preliminary data analysis – GPS data accuracy before data cleaning [m]*

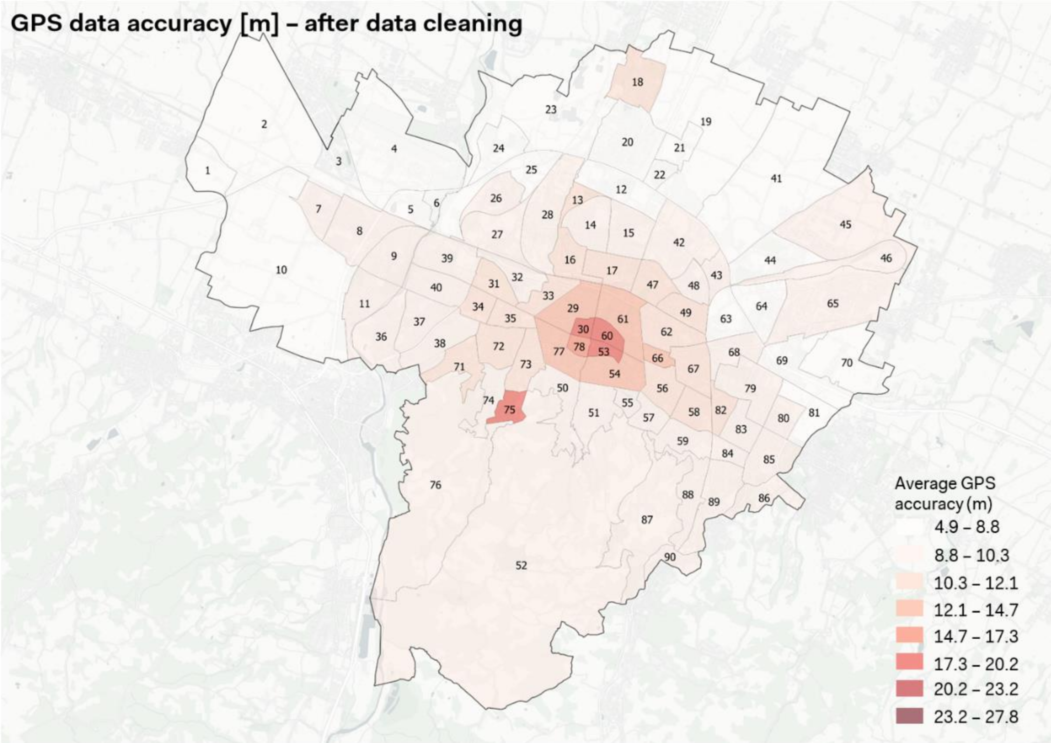

Figure 10 *– Preliminary data analysis – GPS data accuracy after data cleaning [m]*

3.3 Data mapping

In the data mapping phase, trips data was quantified and structured according to three domains:

- Modal share: data was mapped in three groups, soft modes (i.e., walking, cycling, bike sharing, scooter), public transport modes (i.e., bus, train, people mover), and motorized modes (i.e., car, passenger car, car sharing, motorcycle, passenger motorcycle, taxi);
- Trip purpose: data was mapped in three groups, home trips, work trips, leisure trips;
- Mobility profiles: data was mapped in five groups, following the classification of mobility profiles defined in the first analysis phase of the Pollicino project, these are: *Metabolici, Sostenibili, Megamixer, Autonomi, Auto/ moto Dipendenti*.

In Bologna, soft modes represent 48.4% of the modal share, while motorized modes and public transport represent 37.4% and 14.2% each (see *Figure 11*).

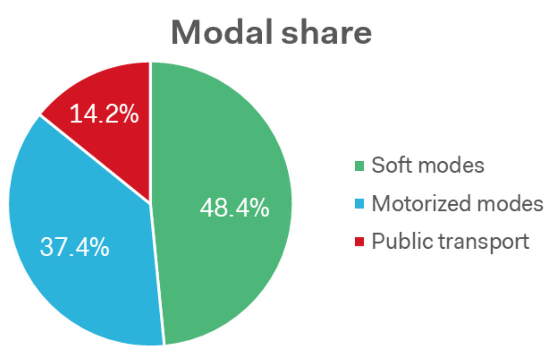

Figure 11 *– Modal share, Pollicino*

The three modal groups present different a spatial distribution, these are outlined in *Figure 12*. Here, trajectories have been structured as a sequence of connected and cleaned GPS points. This step allowed for the spatial understanding and classification of movements, into three main modal groups: soft modes, motorized modes, and public transport. Trajectories were overlapped spatially, outlining areas where higher number of trips pass by over the street network of Bologna. These differ for the three modal groups, in particular, soft modes trips are mainly concentrated in the city center, while motorized modes trips are mainly distributed over the highway and outskirts of the city.

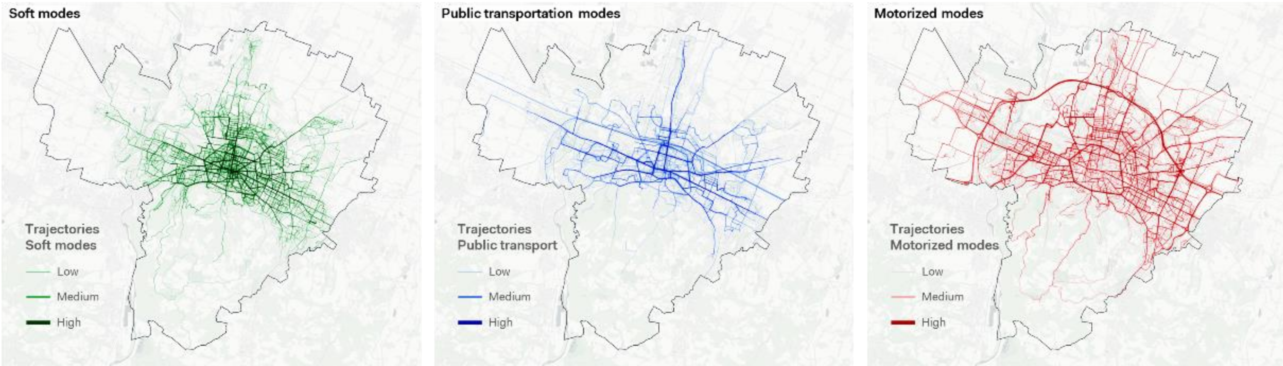

Figure 12 *– Trajectories by modal groups, soft modes, public transport, motorized modes*

Following data cleaning processes and the structuring of trajectories, data was aggregated into the hexagonal H3 grid, a hierarchical geospatial index based on hexagons developed by Uber. The goal of this step was to quantify the intensity of trips volumes in Bologna, using a homogeneous spatial unit. *Figure 13* represents passing-by trips volumes, quantified as total number of trips passing by each hexagonal area in the H3 grid during the study period, and the spatial distribution of three modal groups in Bologna, motorized modes trajectories, public transport trajectories and soft modes trajectories. Passing-by trips volumes are overlapped spatially on the map, revealing where most of the trips occur by modal group the spatial changes of modal co-presence around the municipality area.

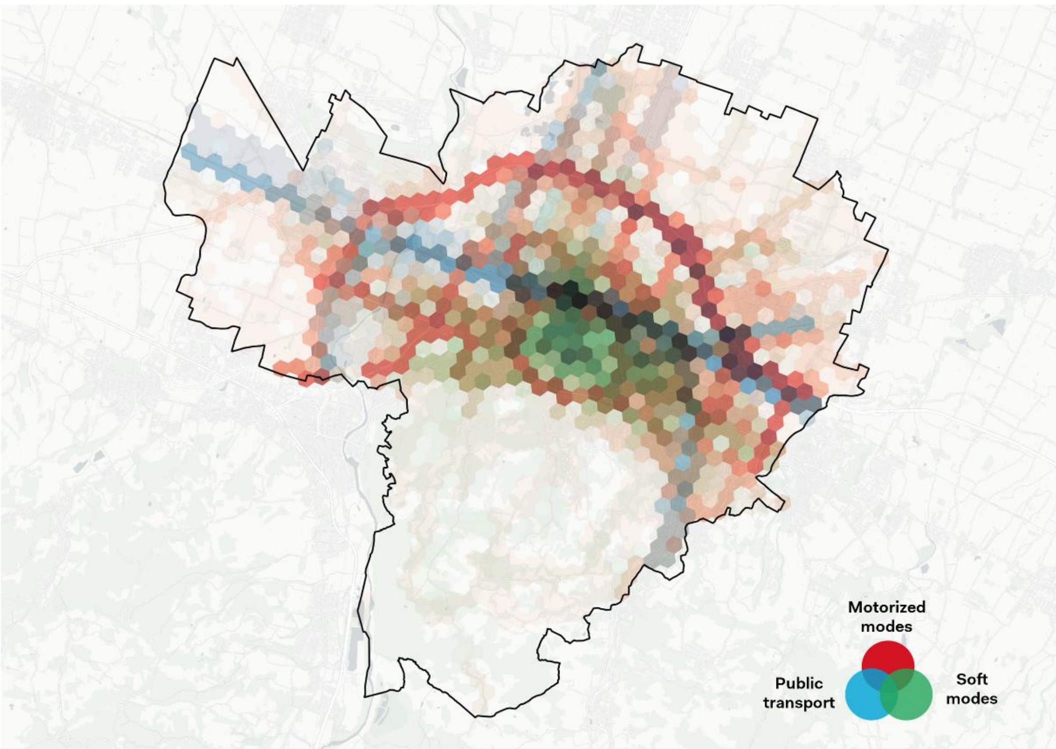

Figure 13 *– Modal spatial distribution, trip volumes*

These results are disaggregated into *Figure 14*, *Figure 15* and *Figure 16*. The analysis showed that most of the trips were traveled far from the southern part of the city (hills zones) to the city center. However, motorized trips were the most widely distributed, covering around 81% of the study area, and were mainly concentrated on the highway and the inner circular road. The public transportation modes served about 60% of the study area and were distributed along the eastern-western axis of the city, following the spatial distribution of public transport infrastructures, with a higher concentration in the city center. The soft mode trips, on the other hand, were concentrated in the central areas of the city characterized by pedestrian zones and covered roughly 56% of the study area.

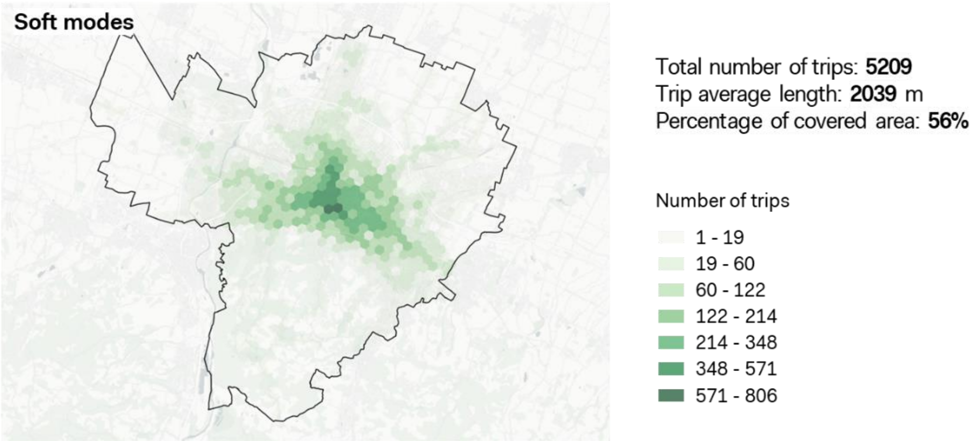

Figure 14 *– Soft modes, trip volumes*

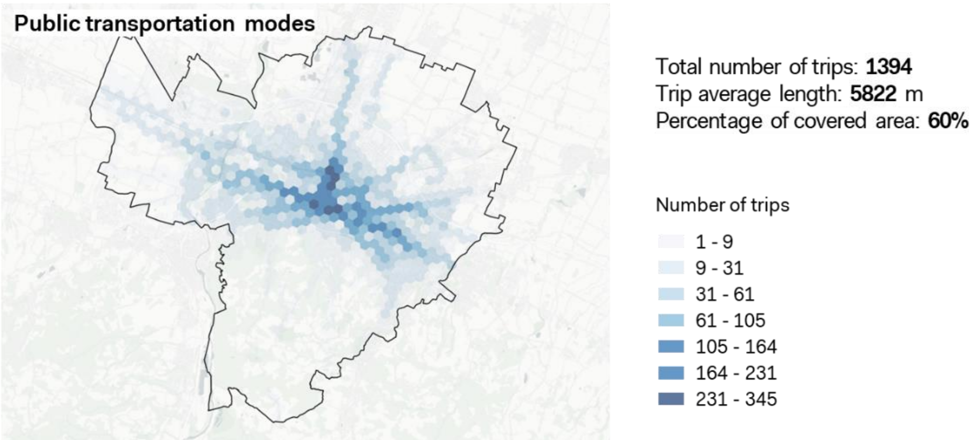

Figure 15 *– Public transport, trip volumes*

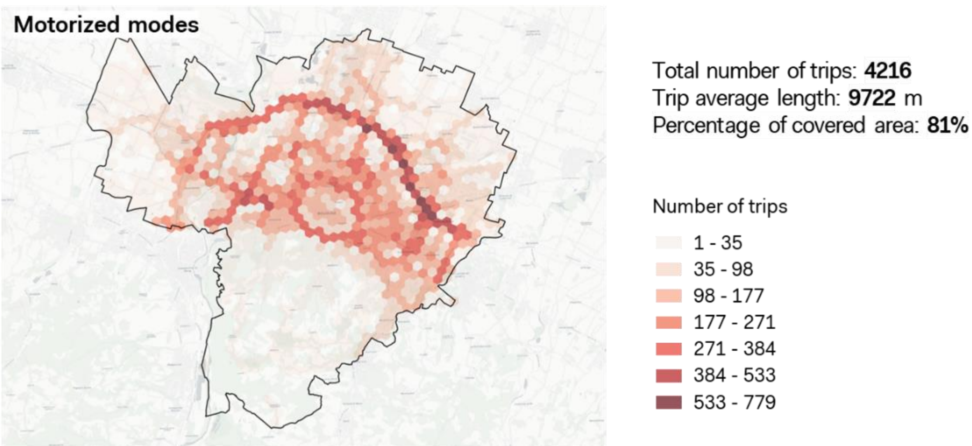

Figure 16 *– Motorized modes, trip volumes*

Then, trip purposes were grouped based to the following groups: *i*) Home trips; *ii*) Work trips, and *iii*) Leisure trips (see *Figure 17*, *Figure 18* and *Figure 19*). Results showed that leisure was the most common purpose for the considered trips, with 5354 trips representing 53% of all trips and covering 75% of the study area. These were followed by home-based trips, which represent 32% of all trips and covered 66% of the study area. Work trips, instead, represented the least portion of the total trips (15%). The leisure and work trips presented an overall similar spatial distribution, not depicting any mobility profile, while the home trips revealed a low density around the highways.

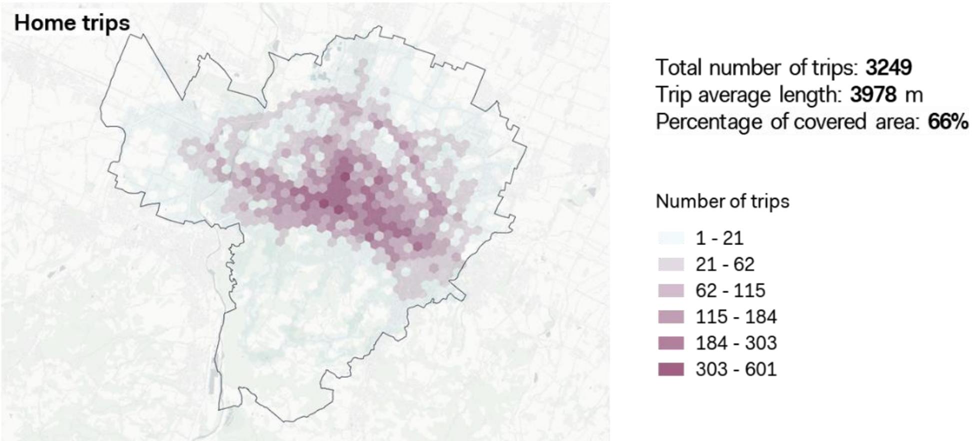

Figure 17 *– Home trips, trip volumes*

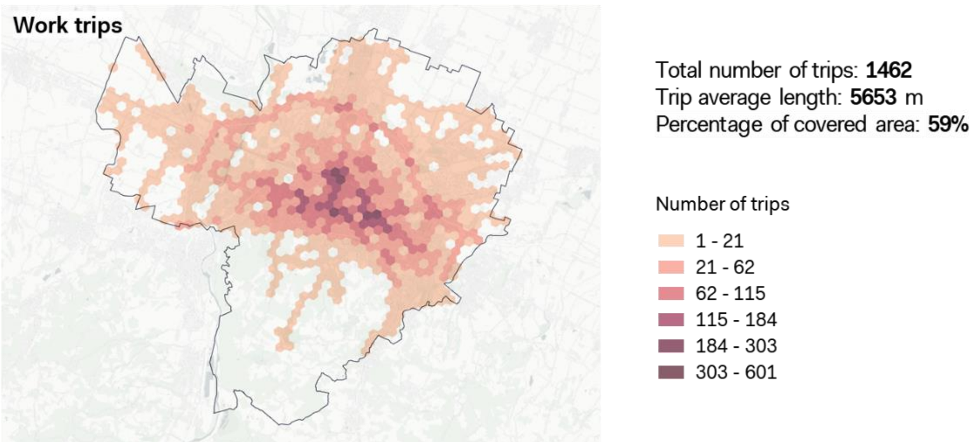

Figure 18 *– Work trips, trip volumes*

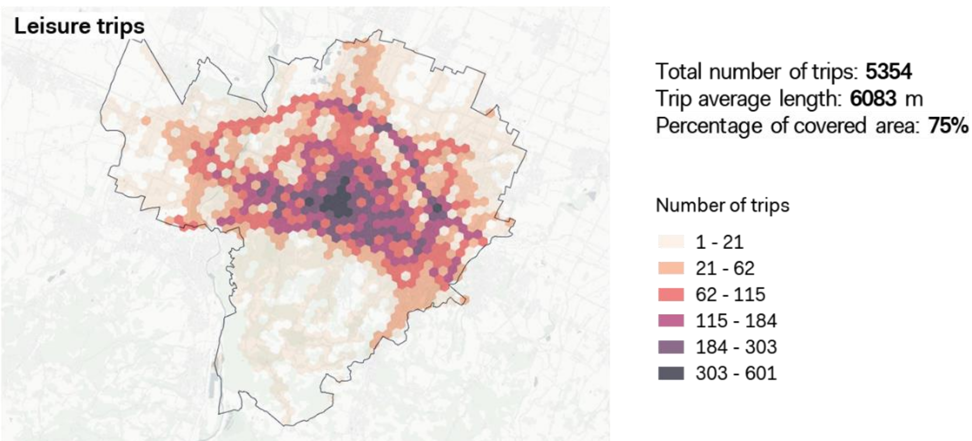

Figure 19 *– Leisure trips, trip volumes*

Last, the movements were analyzed based on the mobility profiles of participants to highlight their mobility patterns (*see Figure 20 to Figure 24*). Results showed that the *Metabolici* cluster trips were characterized by noticeably short journeys within the center of the city, covering 23% of the study area and representing 5% of all trips. The *Sostenibili* trips mainly occurred nearby public transportation infrastructures and represented 9% of all trips. The *Megamixer* and *Autonomi* trips were highly concentrated both on the highways and in the central areas of the city and they were characterized by longer distances traveled. Instead, the *Auto/moto Dipendenti* were mainly distributed on the highways and in the outskirts of the city.

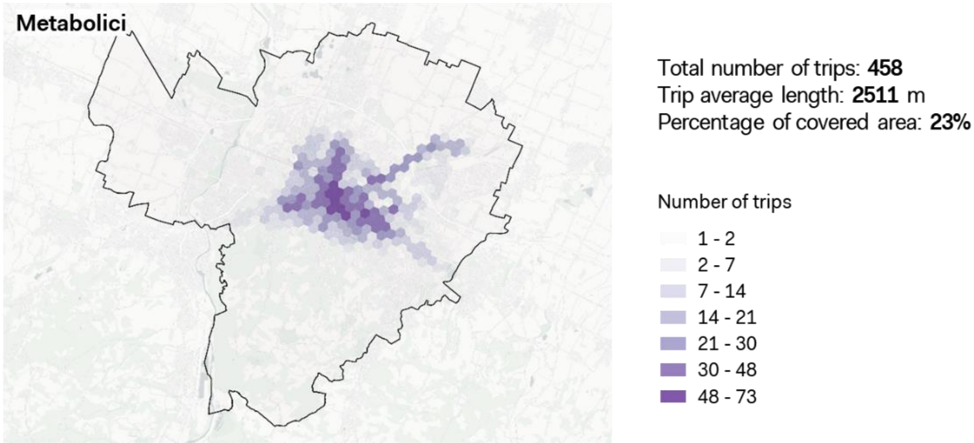

Figure 20 *– Metabolici cluster, trip volumes*

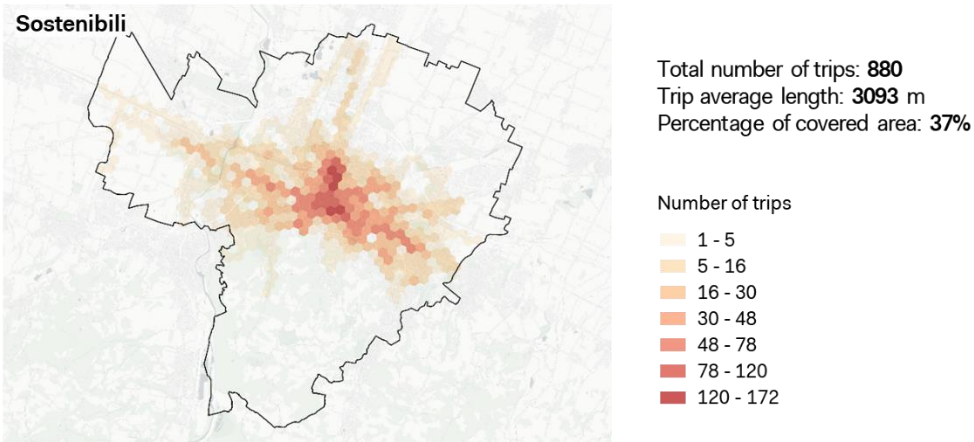

Figure 21 *– Sostenibili cluster, trip volumes*

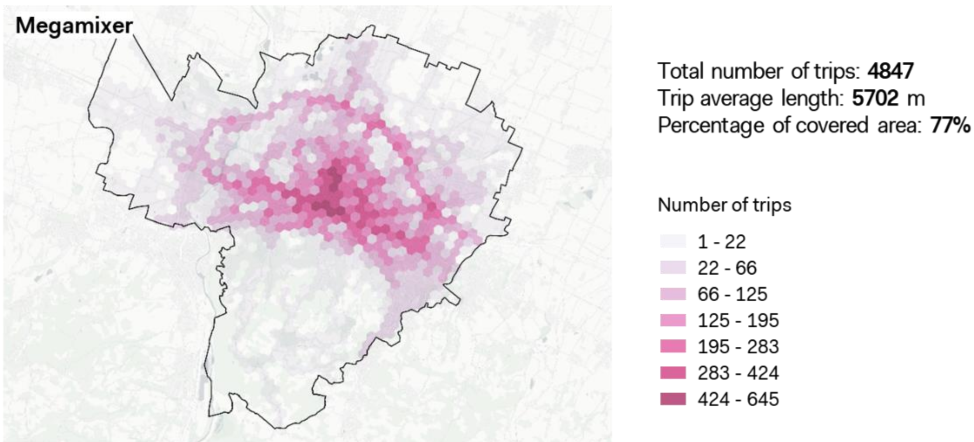

Figure 22 *– Megamixer cluster, trip volumes*

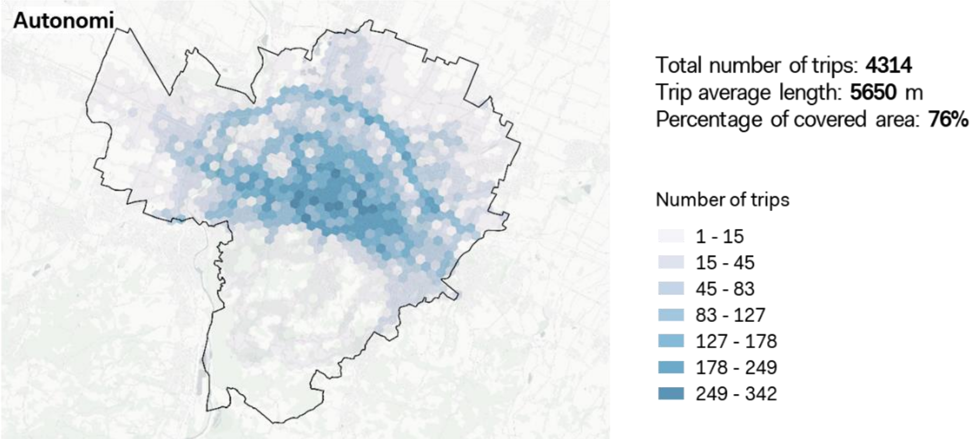

Figure 23 *– Autonomi cluster, trip volumes*

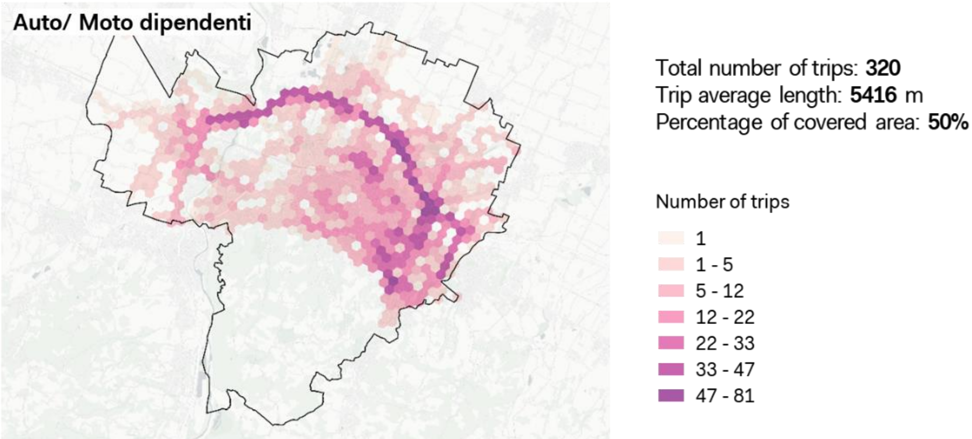

Figure 24 *– Auto/ Moto dipendenti cluster, trip volumes*

4. Conclusions and Future Works

The main goal of this study was to extract the spatial mobility patterns in the city of Bologna, with the aim support the decision-making process of local administrations in the field of sustainable mobility. The research was based on the GPS data collected through the app IoPollicino, a mobile-based survey developed by Osservatorio Nazionale Sharing Mobility and Fondazione per lo Sviluppo Sostenibile. To this end, we built a methodological framework to clean, process and aggregate GPS pings into trajectories, which connect consecutive points with a line. Then, we aggregated trajectories into the H3 hexagonal spatial grid, with the aim to perform a quantitative analysis on the trips, structured into multiple classes (i.e., *modal share, trip purpose, mobility profile of participants*). Results showed that soft modes represent the dominant modal means (48%), with walking being the most frequently used transport mode. Furthermore, the majority of the public transport trips were made by bus, while most of the motorized trips were made by cars. The three modal groups present a different spatial distribution, central areas of the city were characterized by the highest density of overall trips, where most of the soft trips were concentrated. Motorized trips, instead, were often spread out on the highway and the inner circular road. The dominant trip purpose was leisure (53% of trips), followed by home trips (32% of trips), while work trips represent the least portion of the total trips (15% of trips). Last, mobility profiles of participants, which were identified based on their modal choices and behavioral patterns, are well reflected by their spatial patterns. Future works include refining the methodological process developed to clean and preprocess raw GPS points. Furthermore, future works could cover the development of spatial clustering analyses, with the aim of identifying similarities in modal choice of participants, in relation to the origin and destination of each trip, trip purposes and mobility profile. Then, future developments could include a comparative analysis of cycling trajectories and the cycle lanes infrastructure in Bologna, to identify the differences between the shortest path and the route used by participants, in relation to the existing transport infrastructures and services. Last, they could involve the analysis of the annotated trajectories to feed a modal classification model based on parameters as speed and distances between the GPS points.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Giulia Ceccarelli: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization.
Abubakr Albashir: Methodology, Software, Formal analysis, Investigation, Writing – review & editing, Visualization.
Andrea Gorrini: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration.
Massimo Ciuffini: Conceptualization, Resources, Writing – review & editing, Supervision.
Luca Refrigeri: Conceptualization, Resources, Writing – review & editing, Supervision.
Sofia Asperti: Conceptualization, Resources, Writing – review & editing, Supervision.
All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The analyzed data were treated according to the GDPR-General Data Protection Regulation (EU, 2016/679). This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The results of this research were presented at the 7th edition of the Conferenza Nazionale della Sharing Mobility, 5 October 2023, Rome (Italy).

References

Adhinugraha, K., Rahayu, W., Hara, T., & Taniar, D. (2021). Dealing with noise in crowdsourced GPS human trajectory logging data. *Concurrency and Computation: Practice and Experience*, *33*(19), e6139. https://doi.org/10.1002/cpe.6139

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, *734*, 1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Ciuffini, M., Asperti, S., Ciuffini, F., Gentili, V., Orsini, R. and Refrigeri, L. (2023). *Rapporto di ricerca: Pollicino: i cittadini raccontano come si muove la città*. Fondazione per lo Sviluppo Sostenibile, Rome, Italy. Available at: https://osservatoriosharingmobility.it/wp-content/uploads/2023/03/Refrigeri-e-Asperti_OSM.pdf

Kubo, N., Kobayashi, K. and Furukawa, R. (2020). GNSS Multipath Detection Using Continuous Time-Series C/N0. *Sensors*, *20*(14), 4059. https://doi.org/10.3390/s20144059

Larroya, F., Díaz, O., Sagarra, O., Colomer Simón, P., Ferré, S., Moro, E., & Perelló, J. (2023). Home-to-school pedestrian mobility GPS data from a citizen science experiment in the Barcelona area. *Scientific Data*, *10*(1), Articolo 1. https://doi.org/10.1038/s41597-023-02328-3

McCarty, D. A., & Kim, H. W. (2023). A Standardized European Hexagon Gridded Dataset Based on OpenStreetMap POIs. *Data in Brief*, *49*, 109315. https://doi.org/10.1016/j.dib.2023.109315

Meng, F., Yuan, G., Lv, S., Wang, Z., & Xia, S. (2019). An overview on trajectory outlier detection. *Artificial Intelligence Review*, *52*(4), 2437–2456. https://doi.org/10.1007/s10462-018-9619-1

Molloy, J., Castro, A., Götschi, T., Schoeman, B., Tchervenkov, C., Tomic, U., Hintermann, B., & Axhausen, K. W. (2022). The MOBIS dataset: A large GPS dataset of mobility behaviour in Switzerland. *Transportation*, 1–25. https://doi.org/10.1007/s11116-022-10299-4

Moro, A., Kulkarni, V., Ghiringhelli, P.-A., Chapuis, B., Huguenin, K., & Garbinato, B. (2019). Breadcrumbs: A Rich Mobility Dataset with Point-of-Interest Annotations. *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 508–511. https://doi.org/10.1145/3347146.3359341

Prabha, R., & Kabadi, M. G. (2020). A comprehensive insight towards Pre-processing Methodologies applied on GPS data. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(3). https://doi.org/10.11591/ijece.v10i3.pp2742-2754

Saki, S., & Hagen, T. (2022). A practical guide to an open-source map-matching approach for big GPS data. *SN Computer Science*, *3*(5), 415. https://doi.org/10.1007/s42979-022-01340-5

Wang, D., Miwa, T., & Morikawa, T. (2020). Big trajectory data mining: a survey of methods, applications, and services. *Sensors*, *20*(16), 4571. https://doi.org/10.3390/s20164571

Yang, X., Stewart, K., Tang, L., Xie, Z., & Li, Q. (2018). A review of GPS trajectories classification based on transportation mode. *Sensors*, *18*(11), 3741. https://doi.org/10.3390/s18113741