UNIVERSITY OF TARTU
Institute of Computer Science
Conversion Masters in IT Curriculum

Merlin Raud

The influence of infrastructure to cyclist route choice in Bologna

Master thesis (15 EAP)

Supervisor(s): Rajesh Sharma
Flavio Bertini
Amnir Hadachi

Tartu 2024

The influence of infrastructure to cyclist route choice in Bologna

Abstract:

Due to urbanization and increased environmental awareness, lot of effort has been paid to promoting sustainable mobility modes. Cycling is one option to replace motorized transport in cities. To achieve this shift, the urban environment must be transformed to more suitable for cyclist. The goal of this thesis is to study the spatial and temporal patterns of cyclist and study the influence of cycling infrastructure to the cyclist's route choice.

Data collected from Bologna for 6 months was used to gain insight into cyclist behaviour and preferences in urban environment. The cycling in Bologna is used as a transport mode mostly in short and medium length distance trips. The results revealed the routine of cyclist in different periods – peak hours in workday morning and evening and periods with lower activities during weekends. Generally, cyclist prefer streets with good cycling infrastructure and avoid streets with intense traffic. In the suburban area, cyclist prefer bigger roads which lead towards city centre, while in city centre preference is on roads which avoid narrow streets in old town. The results of this thesis will help to see the reasons behind specific route preferences, and this can be used to find opportunities to improve the cycling infrastructure for a better and safer urban environment.

Keywords:

Cycling, Urban mobility, Mobility analysis

CERCS: P170, Computer science, numerical analysis, systems, control

Rattainfrastruktuuri mõju jalgratturite teekonna valikule Bolognas

Lühikokkuvõte:

Linnastumise ja keskkonnateadlikkuse suurenemise tõttu on säästvate liikumisviiside propageerimiseks kerkinud oluliseks teemaks. Jalgrattasõit on üks võimalus asendada linnades mootortransporti, kuid selleks on vaja muuta linnakeskkond jalgratturile sobivamaks. Lõputöö eesmärk ongi uurida jalgratturi ruumilisi ja ajalisi mustreid ning uurida jalgratta infrastruktuuri mõju jalgratturi marsruudi valikule.

Selleks et koguda täpseid andmeid jalgratturite käitumisest ja eelistustest linnakeskkonnas kasutati Bologna linnas 6 kuu jooksul kogutud andmeid. Jalgrattasõitu kasutatakse transpordiviisina peamiselt lühikeste ja keskmiste pikkusega reisidel. Tulemustest selgus jalgratturite rutiin erinevatel perioodidel – tipptunnid tööpäeva hommikul ja õhtul ning madalama aktiivsusega perioodid nädalavahetustel. Üldjuhul eelistab jalgrattur hea rattainfrastruktuuriga tänavaid ja väldib tiheda liiklusega tänavaid. Äärelinnas eelistab jalgrattur suuremaid teid, mis viivad kesklinna poole, kesklinnas aga teid, mis väldivad vanalinna kitsaid tänavaid. Lõputöö tulemused toovad välja konkreetsete marsruudieelistuste tagamaid, mis aitavad leida võimalusi jalgratta infrastruktuuri parendamiseks parema ja turvalisgfema linnakeskkonna nimel.

Võtmesõnad:

Rattasõit, Liikuvus linnas, Liikuvuse analüüs

CERCS: P170, Arvutiteadus, arvutusmeetodid, süsteemid, kontroll

Table of Contents

1. Introduction

Due to climate change, energy crisis, depletion of fossil fuels, overpopulation and traffic congestion, increased attention has been paid to sustainable ways of life. One important part of this is sustainable transport modes in highly populated urban areas. One possibility to decrease the usage of motorized vehicles and thereby decrease traffic volumes is to increase cycling as a transport mode in cities. Cycling is sustainable, healthy, economical transport mode and a reasonable and fast alternative for short-distance trips, especially, in congested in urban environment.

To promote and support cycling as a mode of transport, better understanding of various aspects travel behaviour is required. Mobility is the result of individual behaviours, habits and describes the movements of individual objects [1]. Mobility studies can reveal daily and routine human mobility patterns [2]. This information is valuable for a wide range of applications in urban areas like urban planning, developing transportation models, promoting healthy lifestyles [3] and can benefit traffic managers and policymakers to make targeted decisions at different levels [4].

In recent years, the increasing availability of mobile-phone records, Global Positioning System (GPS) data and other datasets have been seen as one possible way to collect location data in more detail way for accurate analysis and thereby patterns in human mobility can be revealed. The GPS and other positioning technologies enable to collect vast amounts of mobility data from moving objects [5]. By using this data and by applying modern data science methods it is possible to extract valuable information about movement patterns and behaviours of people – where, how, and when people move around city.

Data and data analysis is central point in cycling research and planning. One important research field is the route choice behaviour of cyclists. It is important to know what factors and at what magnitude influences the route decisions of cyclists. By knowing more about where, why, and how cyclists move it is possible to improve the transport network and thereby promote cycling as a transportation mode in cities. Many papers on cyclist route choice models for different cities are published. Most of them use localized data and applying diverse infrastructure variables to examine their impact on route selection. However, no studies were found where detailed routes between two locations and multiple alternative trajectories between these locations were analysed.

This thesis addresses this gap by using the Bologna Bella Mossa mobility study data to investigate the cycling mobility patterns in Bologna, Italy. This thesis presents a framework of analysing mobility flows in urban areas by origin-destination tiles and studying the people's routes between them.

The primary objective was to study the influence of cycling infrastructure to the route choice made by cyclists. Understanding these patterns is necessary to see the reasons behind specific route preferences and find opportunities to improve the cycling infrastructure for a better and safer urban environment. Using the results of this study it is possible to make recommendations for infrastructure improvements can be developed, aiming to foster cycling as a viable urban transportation method.

Research goals (RG) addressed with this thesis:

- RG1: Study the spatial and temporal cycling mobility patterns in Bologna to understand when cyclist move and where do they start and end their journeys;

Findings: Our analysis presents the daily patterns of cyclist – cycling was higher in the city centre area and in periods when people commuted to school and work in the morning or returned home in the evening;

- RG2: Study the use of infrastructure concentrating on different type such as urban highways, different type of bike lanes, and streets to understand what infrastructure types are used for cycling;
  Findings: Results reveal that cyclists prefer streets and roads with smaller traffic intensities and with dedicated cycling roads
- RG3: Identify most used flows and routes cyclist use in Bologna for transport purposes to understand where cyclist start and end their trips and what routes are preferred;
  Findings: Cyclist flows were determined and 20 most used flows were selected for further analysis
- RG4: Study the influence of road type and its characteristics to trajectory choice made by cyclist to understand their influence on cyclist route choice;
  Findings: The cyclist prefer to use streets with cycling infrastructure and when possible, narrow streets in the old town are avoided;

The thesis is outlined in the following order: in first paragraph, the overview to the cycling mobility and mobility analysis methods are given; in the second part the framework and methods of the analysis is introduced. The results of the analysis are presented in the third paragraph and fourth part of the thesis includes conclusions.

2. Cycling mobility data and its analysis

2.1 Cycling as urban mobility

The rise in urbanisation and increase in motorized vehicles are causing traffic congestion and different types of urban pollution which is becoming a public health problem. To reduce these negative effects, greener means of transport should be promoted and chosen. Motor vehicles have different transport alternatives, primarily public transport and shared mobility systems which reduce the traffic by decreasing the number of private car trips and reduce the greenhouse gas emissions [6]. Cycling is also considered as one of the best options to improve sustainable mobility in urban environments and reduce the negative effect of motorized traffic [6,7]. It is considered to be environmentally friendly, cost-effective, inclusive, and convenient mode of transportation because of its affordability, adaptability, beneficial impacts on physical and mental well-being, and zero greenhouse gas emissions [8,9].

Cycling in urban environment can have many purposes – cycling for as commuting or utilitarian purposes, leisure cycling and sport cycling. Transport cycling is oriented towards practical and necessary journeys and may include daily activities. Therefore, it is considered an "obligatory" form of transport, distinguished by its practical nature, logistical objectives, and lower cost compared to other cycling forms. Leisure cycling, in contrast, occurs outside working hours and areas and serves purposes like social reasons, tourism, and relaxation. Most research attention has been paid on cycling as a transport [8].

Choice of travel mode depends in various variables, but most important factors are built environment and trip characteristics. Most influential trip characteristics are trip distance, travel time and trip cost [9]. The cycling is preferred over motorized modes of transport in case of short and medium distance trips (<20 min.), while car and public transport (bus and tram) is preferred in case of longer commutes (more than 30 minutes) [10,11]. If the travel distance increases, the likelihood of choosing motorised vehicles as a primary travel mode also increases and this can cause traffic congestion. On the other hand, shorter travel times and distances in compact cities encourages the use of active transport modes like cycling [12,13]. The achievement of switch from motorized transport to cycling is a concern for many cities. However, achieving well-designed, efficient, and cost-effective cycling infrastructure network is difficult since urban space is scares and fitting cycling infrastructure into existing urban space is difficult [11].

2.2 Cycling infrastructure

In urban areas, street network consists in different infrastructure types for various user types – cars, bicycles, pedestrians, mixed users etc. [14]. Bicycle network and infrastructure involves various infrastructure elements and takes several forms - from fully separated paths to various types of on-street markings, parking facilities, changing amenities, and signs which devote street space to cyclists [15]. In terms of their design and structure, bikeways can be constructed at ground level, elevated above other traffic forms, and even in tunnels [11]. The design of cycling infrastructure includes multiple elements to consider, for example road and lane configurations, speed control strategy, traffic signals, and connection road networks and broader transportation infrastructure [16]. The availability of dedicated cycling infrastructure, such as the continuous, separate, and protected bikeway network significantly encourages and promotes cycling [11].

According to Wysling et al there are four basic types of cycling paths [17] :

- stand-alone paths which are often found in parks and are sometimes shared with pedestrians,
- cycling lanes that physically separate bicycles from motorized traffic using devices such as bollards, planters, or a parking lane,
- cycling lanes marked on roads but do not physically separate cyclists from motorized traffic,
- roads where cyclists ride in mixed traffic where speed limits and traffic volumes are low.

Bikeways and cycling paths can be also categorized based on their intended purpose, such as greenways designed for leisure and recreation, or bicycle boulevards that give priority to bicycle traffic. Greenways are linear public spaces which can be present in many ways, including parks, reclaimed trails, and riverfronts, frequently designated as open spaces for leisure [11]. Bicycle boulevards are low traffic intensity and speed roads which are designated and designed to prioritize bicycle travel. These boulevards include secure and convenient bicycle crossings by employing speed and traffic volume management practices [18].

Different measures have been proposed to assess the suitability of cycling infrastructure and cycling paths. Most common are cycling suitability and bikeability. Cycling suitability refers to the perceived comfort and safety of individual street segments for cyclists. However, bikeability evaluates the suitability of cycling infrastructure to facilitate cycling, considering factors that influence the choice of the bicycle as the primary mode of transportation, along with environmental factors that promote bike travel [17,19].

Many studies have been made to investigate the effect of different variables to these measures. The most important are presence of dedicated cycling paths along streets and traffic calming measures. This provides safer environments as bikes do not have to share same urban space with motorized vehicles [19]. Cycling lanes separated from motorised vehicles are more costly to construct and demand more street space compared to cycling lanes marked on roads. Therefore, the share lane markings are widely used in United States and European cities due to their cost-effectiveness and space efficiency. The study made in Madrid indicated that only 9% on the cycling infrastructure comprised of bike lanes physically separated from other vehicles while the 50% are cycling lanes marked on the roads [20]

In addition to cycling paths, the continuity and safety of them is also important features [19]. Cyclists prefer well-established and connected cycling infrastructure where different areas of the city, but also different cycling paths are connected. In addition, density of cycling network is important. The establishment if a single cycling path does not increase bikeability but it must be part of a wider cycling infrastructure for a more substantial improvement in the cycling possibilities [19,21].

2.3 Route choice selection

Cyclist route choice is based on a combination of route features, individual characteristics, the purpose of their cycling journey and even on the region. Route choice is a decision-making process in which a person selects a specific route, based on its attributes, between an origin and a destination [22]. Many papers handling bicycle infrastructure, built environment and its effect on cycling route choice have been published. This knowledge is very important to understand how cycling infrastructure could be improved to further promote cycling [23–25].

Most common factors influencing cyclist route choice are length of the route, travel time, steepness, traffic volume, speed limit, type of intersection, and presence of bicycle infrastructure [17,26]. Generally, cyclists will opt for the shortest route, either in terms of distance or travel time, to move from origin to destination. In addition, cyclists avoid steep hills and prefer moderate terrains. Nevertheless, many studies and reviews have shown that there are additional factors that impact these decisions. This results in cyclists frequently selecting routes that are longer than the absolute shortest ones [17,27].

Cyclists avoid unsafe routes with high speeds and traffic density but highly value dedicated infrastructure to minimize interactions with cars [28] therefore they are willing to take longer route to reach a road with a bicycle-friendly infrastructure and avoid roads without bicycle infrastructure [26]. Roads with a stronger separation between cycling and other forms of transportation are preferred and arterial roads with heavy traffic are avoided [24]. The most preferred forms of infrastructure are residential roads featuring clearly marked bicycle lanes, with cycleways (dedicated paths designed exclusively for bicycles) coming in second. Cycle path situated at sidewalk level is considered superior to a cycle lane on the road that is only separated by painted lane markings, or even worse, by dashed lane markings [24,28] .

Additional influencing factors are routes with less stops and intersections [28,29], attractive surrounding and streetscape [24,28], less traffic, less traffic markings and signals, less major cross streets and turns [30]. However, the work of Lu et al showed that cyclists are willing to take longer routes if they include positive features such as bicycle facilities and low traffic volumes even if it increases in number of turns and number of intersections [30]. A study in Bologna also showed that route choice also depends on the cyclist gender as women prefer routes which minimize interferences with other traffic streams and avoid using roads with frequent intersections and unsafe manoeuvres such as left turns [31].

2.4 Cycling data collection

Different technologies are available to record mobility data. Data can be collected using social media or cellular network which save location information in case of social media activity or when call is made using cellular device. Mobile phones can be used also to collect location information using phones GPS device, cellular network, or wireless networking technology. Additionally, using local area networking of wireless network devices enable to save location info. Most common location information data source is technology based on GPS (Global Positioning System) [32] which is now used in nearly two-thirds of studies related to bicycle route choice analysis [33].

Since 1990s, GPS technology has been widely used for many purposes such as analysing travel patterns, research of transport safety and efficiency, and evaluating travel impacts [34]. GPS is a satellite-based worldwide navigation system that enable to determine a precise location of any point on the Earth's surface by triangulating 4 or more signals from different satellites [35]. GPS tracking enables to gather information globally [36] on longitude, latitude, date, time, speed, altitude, and direction of movement [34]. Compared to interview-based approach, this technology is less expensive, and enables [37] save mobility data in real time and more detail scale than with surveys [38].

According to Lee et al., the most perspective for bicycle data collection are various regional mobility tracking apps. If resources are available for developing the app and recruiting users it is possible to gain a higher level of data quality, accuracy, and user details than with the other methods [39]. Bicycle data can be collected also using fitness tracking apps. However, this data might include mostly recreational cycling activities and not for cycling for transport [26].

Passive data collection enables to collect continuously low frequency data which includes also records that are not related to mobility. On the other hand, the active methods require users to record data only during mobility which allows to collect high frequency data containing mainly mobility records [37,39,40].

2.5 Trajectory analysis

A trajectory is a path formed by a moving object in geographical spaces, typically represented by a sequence of ordered points in time. Trajectories are stored as a polyline, can be represented as $p_1 \rightarrow p_2 \rightarrow ... \rightarrow p_n$ and denoted as $Tr_n=(p_1, p_2, \dots p_n)$. The p_n represents single GPS points which are ordered by time to represent the sequence of GPS points [41]. Trajectories can be created by humans, animals, transportation vehicles, animals, or even natural phenomena. The points include geospatial coordinates and a time stamp but may also include some additional descriptive information [42–45]. Trajectories can have varying durations, may include vast number of moving objects with different sampling rates from every a few seconds to hours and even days [46].

Advancements in satellite and tracking technology has enabled collection of substantial amounts of trajectory data concerning the movements of different objects [47]. This data can be used in various applications like intelligent transport systems, urban management, environmental protection, meteorological forecasting and more [46,48]. Despite the substantial amount of data already collected and potential of collecting this, there are still many challenges to effectively use data due to privacy concerns, commercial considerations, the presence of missing values, and the high costs associated with acquiring the data [49].

GPS data is not fully accurate and suffers from major limitations in terms of data quality. Errors in GPS accuracy can be caused by random noise and interferences and therefore recorded locations may deviate from true locations. Also, missing data between recorded points can be caused by blocked GPS signal indoor or underground places or due to satellite, receiver, signal reflections and other disturbances during data collection [50,51]. To overcome these problems and extract relevant information from large amount of location information, GPS data needs cleaning and preprocessing before further analysis to avoid to false results and misleading conclusions [40,48,50].

The first and fundamental step in trajectory analysis is preprocessing to prepare data for further analysis [52]. Most used methods are smoothing, compression, segmentation, and map matching. The influence of random GPS errors which or minor discrepancies from true value can be reduced by smoothing methods, such as Gaussian smoothing and Kalman filtering. Systematic errors – invalid GPS pointes due to external reasons can be corrected using visual inspection and automatic filtering methods. The goal of compression is to reduce the size of the dataset while maintaining the utility of trajectory to enable less complex calculations. The trajectory segmentation, however, enables to split the trajectories into multiple smaller fragments based on movement characteristics or other rules. The determination of stops where to cut the trajectory can consider various spatio-temporal parameters, including position density, velocity, and direction of movement. The map matching is used to refine the trajectory by replacing each location with the network point that represents the most probable location of the moving object on road infrastructure or to convert the trajectory into sequence of road segments [53].

The GPS trajectory data contains basic location information of movements however further analysis of the data is necessary to be able to predict unknown or future values or reveal hidden behavioural patterns [54]. The analysis methods can be divided into primary methods which aim to categorize the trajectories based on their properties - classification and clustering. Trajectory data clustering seeks to characterize trajectories by grouping them into distinct clusters based on movement characteristics of trajectories. Trajectories similar to each other are grouped in the same cluster, and characteristics of these trajectories are distinct from those trajectories in other clusters [47,55,56]. The classification assigns trajectories into pre-defined classes based on predefined set of labels relying on the features exhibited by the trajectories [56].

The secondary trajectory data mining methods are aimed towards uncovering and describing the hidden movement patterns in trajectories [56] and to predict the future behaviour of moving object [36,45]. Pattern mining focuses on finding interesting, significant, or unexpected patterns from data. The patterns can be periodic, frequent, or collective depending on in which manner the trajectories are reoccurring [42]. This periodic pattern can be going from home to work every morning, to church every Sunday and may include various aspects such as recurrent trajectories, sub-trajectories, or areas of frequent co-occurring movements [57]. The outlier identification in trajectory data mining involves identifying trajectories that are not consistent with majority of trajectories. The prediction involves prediction of future locations and prediction of trajectories within a road network [42].

2.6 OSM road network

OpenStreetMap (OSM) is as initiative to create, provide and distribute a free, openly licensed, volunteer-contributed geographic data. The data is free to use, share and modify if credit is given to OSM and its contributors [58]. The project started in 2004 with a focus on streets and roads [59] however, now OSM includes very detailed road network along with a point of interests (POI) and features like building, land use, etc. [60]. The complete mapping process of OSM is done by volunteer contribution and all users can have access to the information and edit view and/or share the whole data at no cost [61]. However, the OSM information contributed by individual users may have problems associated with data quality like inadequate sampling, potential bias, and inaccuracies in tagging [62].

OSM data contains road network information, points of interest and other relevant geographic information from all over the world. The OSM road network is represented as a graph that consists in basic data elements of OSM data model - nodes, ways, and relations. The nodes are single points and are defined by latitude, longitude, and node identifier. The ways are a collection of connected nodes and represent various linear features (road, path etc) and area boundaries (building, area etc). The relations are used to show the connection between different data elements [63].

The OSM data elements can be characterized using tags that describe their meaning or function. A tag consists of two items, a key and a value. The key is used to describe a topic, category, or feature type while the value provides specific details about the key [15,64]. OSM data elements can include a multitude of tags and various combinations of tags and keys. The greater the number of tags present, the richer the available information becomes [65]. In the case of bicycle infrastructure, the ways with highway or cycleway as tag keys are mostly used. Table 1 includes information about different road types for cycling and their OSM tags.

Table 1. List of most used cycling infrastructure types and respective OSM tags (adapted from Ferster et al., 2020) [15]

Category	Description	Generalized OSM tag
Cycle track	A paved facility alongside a city street, separated by a curb or barrier, intended for bicycle-only use.	Highway = cycleway Highway = * AND cycleway = track
On-street bicycle lane	A painted bike lane on the street, with or without parked cars.	Highway = * AND cycleway = lane
Path (bicycle only or multiuse)	An off-street paved path, either bicycle only or shared with pedestrians.	Highway = path AND bicycle = yes/designated Highway = footway AND bicycle = yes Highway = service (or unclassified) AND bicycle = yes/designated AND motor_vehicle = no
Local street bikeway	A designated bicycle route with signs, and possibly cyclist activated traffic signals/traffic calming.	Cycleway = shared bicycle = designated
Ambiguous infrastructure		Highway = cycleway No other tags

*Tags have variations and additional descriptor tags.

2.7 Map matching

Due to measurement error and noise from multiple sources, the recorded GPS point can deviate from true location. Therefore, it is often necessary to align the observed GPS points with the road network on a specific digital map [66,67]. Map matching is the approach to correct the errors and link the recorded GPS points to a road network graph [51]. It serves as a fundamental pre-processing step for many applications which include location analysis [66]. After two decades of research, many map matching algorithms have been proposed which can be classified into geometric, topology, probabilistic and advanced map matching algorithms based on used approach [68].

One of the most widely used approaches for map matching is based on HMM (Hidden Markov Model) [41]. The HMM is based on Markov Chain, defined as “A stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event” [69]. The map matching algorithm uses HMM to find the most probable path out of all possible paths available. In the map matching process, each GPS point in the trajectory is treated as an observed measurement and the actual position of the object is unobserved (hidden) state. The road segments near the observed measurement are potential locations (states) [41] where the GPS point actually might have been. The probabilities of different paths are computed using probability of GPS measurement being observed in road segment (emission probability) and probability of moving from one segment to another (transition probability). The most probable path is the order of road segments (states) with most likely transitions from one segment to another [70].

The example of transitions between states and optimal path calculation is shown in Figure 1. The possible road segment candidates are found for each measured GPS point - Figure 1, Point P1 with possible candidate edges c_1^1 and c_1^2. The possible road segments are usually limited within a certain radius from the GPS point to reduce computation effort. The measurement probability represents the likelihood of a GPS point P1 being located on that road segment. The road segments further from the measured point have a smaller probability. The transition probabilities are the probability of an object moving from one road segment to another [41,51]. The optimal path is calculated based on both probabilities using different algorithms such as Viterbi [51].

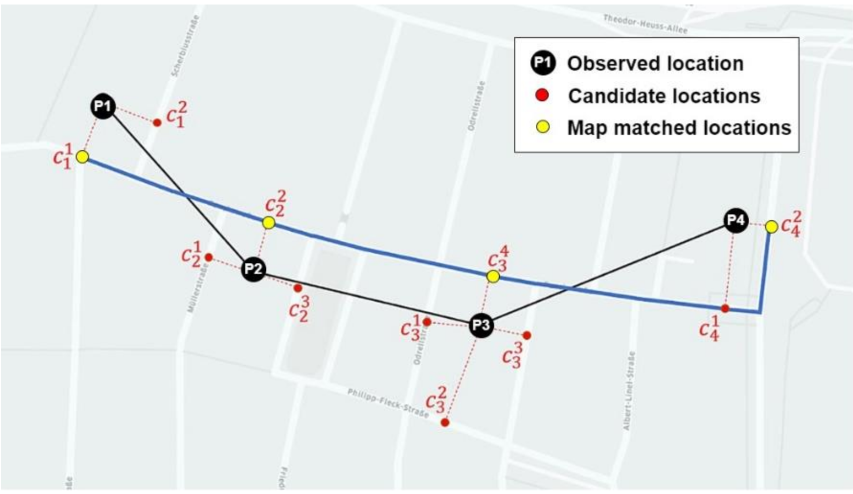

Figure 1. Possible transitions between states and optimal path. Figure adapted from [51].

3. Data exploration and insights discovery

3.1 Data

The data was collected from municipal area of Bologna (44.498955°N; 11.327591°E), which is a capital of Emilia-Romagna Region (Italy). Bologna is characterized by a dense and compact historical centre with industrial areas and fundamental transportation hub for road, rapid and air transport in its periphery [71]. Location of study area is shown in Figure 2.

Data was collected within the Bella Mossa program. This was a program of the City of Bologna to promote a healthy lifestyle and sustainable mobility. During the period from April 1, 2017, to September 30, 2017, around 15 000 people participated in the program, 895 000 trips were recorded which was a total of 3,7 million km. The program was designed to encourage travelling using sustainable transportation modes. During the program, points were awarded for cycling, walking, and using public transportation which could be redeemed for discounts or payments for various services. Participants downloaded a free smartphone app that used GPS for automatic recording of their trips. The data was sent to a database and used anonymously without reference to user [72].

Figure 2. Location of Bologna in Italy. Study area, City of Bologna, is bordered with red line.

The basic information describing the collected data is summarized in Table 2. Each row of the dataset includes information about one GPS point. Each activity has a unique ActivityId. ActivityType is selected by user and indicates the mode of transportation. Data also includes information about the time, location, and other variables as listed in Table 2. In the scope of this program, one activity was considered as a single trip made by one random person. The dataset is anonymized. Therefore, the ActivityId-s can not be linked back to users, nor they can be grouped by users. In this work ActivityId, ActivityType, time, latitude, longitude and Accuracy information were used. The data contained several mobility types but only datapoints labelled with ActivityType as “cycling” were included in this work.

The overview of the data analysis framework for this thesis is outlined in Figure 3. Two types of data were used – cycling mobility data from Bella Mossa program and freely available OSM street network data. The main procedures are data preprocessing, map-matching tessellation, flow analysis which includes trajectory clustering and road type analysis.

Table 2. Description of variables for the dataset.

Variable	Description of variable
ActivityId	a unique number that identifies an activity (not a user), that is a trip
ActivityType	user-declared activity type: Car_Share, Car_Share_Passenger, Cycle, Walk, Bus, Train, Car_Club, Survey, In_Vehicle, Run, Shuttle_Bus
Time	when the activity was done
Latitude	GPS coordinates
Longitude	GPS coordinates
Accuracy	GPS accuracy
Speed	GPS Speed
IdentifiedType	activity type inferred by the application collecting the data
IdentifiedConfidence	confidence about identified type

Python programming language was used to build data preprocessing pipeline and perform analysis of movement data. Provided data was converted from multiple *csv* files into *parquet* format for efficient storage. Partitioning was used to split dataset into smaller subsets for efficient calculations using parallel processing where applicable. This made it possible to run computations with a full dataset with limited compute resources.

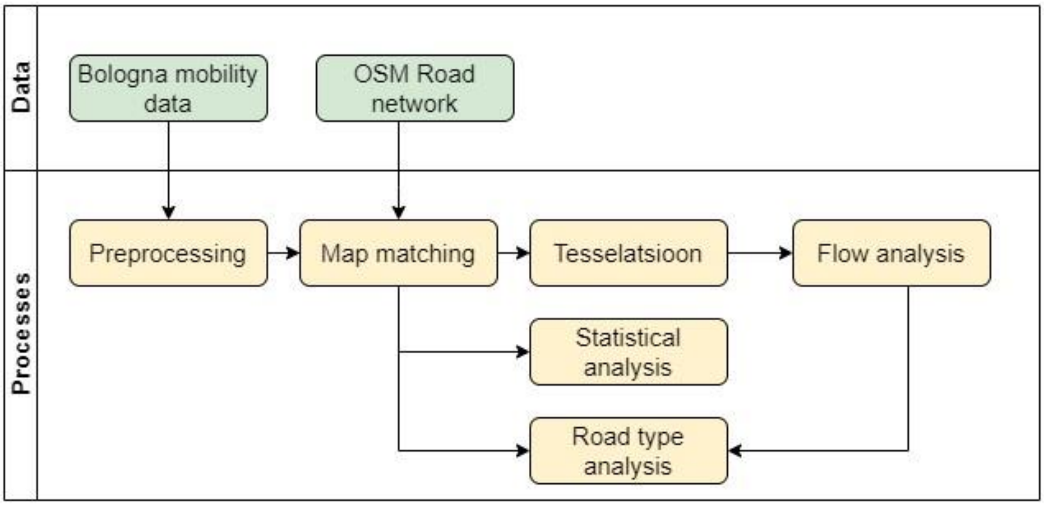

Figure 3. The framework for the analysis.

3.2 Data Preprocessing

Data preprocessing includes activities necessary to transform raw data into suitable format for further analysis. The data was provided in multiple files in csv format and they were converted to parquet format and various processes were done to correct or clean the data from incomplete or inaccurate activity records.

GPS data was sorted by ActivityID and time to have a better overview of data and activities. The analysis revealed that some activities were extremely long, containing several movements with long pauses in between them. For example, during the trip, the user made long stops and continued later, while the entire journey was saved as one. To correct this, activities where data recording is stopped for longer than 6 minutes or cyclist stayed in one location longer than 6 minutes were split into multiple activities. Activities that are longer than 2 hours or above 30 km after segmenting were removed since most likely contained more than one segment or are made for other purposes than transport (sport etc.). Activities with distance below 100 m or time less than 3 minutes were removed since most likely to not contain any reasonable information. Data also included records outside study period and study area which were removed.

The data preprocessing was conducted in following order:

1. removed activities that contained less than 2 recorded GPS points;
2. split activity into multiple activities if time gap between two GPS points is more than 6 minutes;
3. split activity into 2 if GPS points stay in 30 m radius location more than 6 minutes;
4. removed activities with distance above 30 km or time more than 2 hour;
5. removed activities with distance below 100 m or time less than 3 minutes;
6. removed activities whose all points were outside Bologna city;
7. removed activities with timestamps outside the study period.

The number of data points and unique activityID-s representing unique activities or movements, respectively, before and after data preprocessing are outlined in Table 3. The total number of unique activities was 290 177 and included a total of 60 414 481 GPS points with unique timestamp. Most of the GPS points were recorded with intervals of 3-10 seconds, although longer gaps between datapoints were also present.

Table 3. Data characteristics.

Characteristics	**Before data preprocessing**	**After Data Preprocessing**
Number of unique activityID-s	320 109	290 117
Number of GPS points	72 396 179	60 414 481

3.3 Map matching

Preprocessed data was prepared for further analysis using map matching. Map matching algorithm assigns original GPS coordinates to corresponding road network nodes in the most plausible way. This identifies the actual travel routes of cyclists and allows making more accurate route analysis based on actual road network. This was used for 2 main purposes - to improve the accuracy of GPS data locations and to link GPS coordinates to corresponding road network [59].

Map matching was done using Open-Source Routing Machine (OSRM) which uses OSM street network as data source and HMM with Viterbi path finding algorithm [74]. OSRM maps all possible combinations of nodes and transitions for consecutive GPS points, determines measurement and transition probabilities and finds optimal route. This process is repeated until gaining the full GPS trajectory [75]. The time intervals between consecutive GPS points and usual speed of transportation mode can be considered to determine the optimal route [76]. The algorithm also removes outliers if they cannot be matched and splits traces in case of large gaps in timestamps of improbable transitions [75].

OpenstreetMap street network data of Nord-east Italy was sourced from Geofabrik [77] download server using *pyrosm* library. Since the mobility project was conducted in 2017, the data dated to 1.01.2018 was chosen. The OSM road network is provided as a directed graph where edges refer to road segments and nodes refer to intersections. With directed graph, each driving direction is represented by a separate edge [78]. Since many streets are one-way streets, the edges were added to street network also in reverse. Adding one-way edges in both directions enables to take driving in opposite direction into account during map-matching. Although driving in the opposite direction is not allowed it is very often done using bicycles. The final street network of Bologna was in a total of 106 552 edges with a total length of 3545,12 km.

Local OSRM installation using Docker containers and downloaded street network data was used. Profile *bike* was used for map matching and list of GPS coordinates, timestamps and radiuses/accuracy as input parameters. The radius limits the search to given radius in meters, while timestamp enables to remove outliers since otherwise it is assumed that the GPS coordinates have equal intervals [74]. Although map matching is faster when only GPS coordinates are used, time and accuracy parameters were also used since more accurate results were gained with the expense of longer processing time. For each trajectory - list of GPS points, the output response of map matching is a *json* object that includes tracepoints object and matchings object. The tracepoints object include a list of waypoint objects which include all matched coordinates. The matchings object includes legs attribute with matched OSM nodes that assemble the ways. The Figure 4 illustrates the map matching process and its output data.

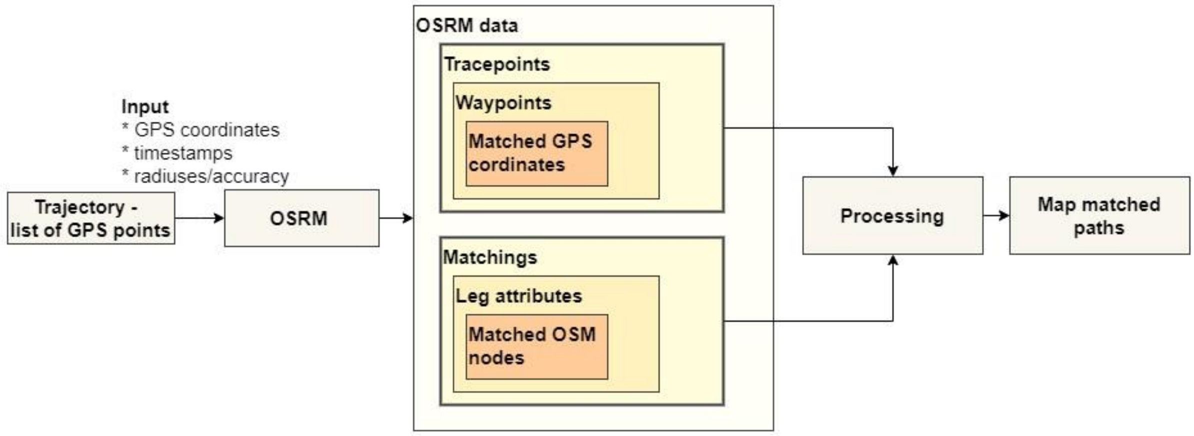

Figure 4. Map-matching process.

Additional processing was done to construct paths from map matching output data. The path is a sequence of OSM ways. For this, both tracepoints and matching objects were joined together to single pandas dataframe and timestamp column was restored from input data as it was not present in OSRM output. The way column contained a list of ways associated with the point which were split to individual rows and duplicate ways were removed. Example of map matched trajectory compared to initial trajectory is presented in Figure 5. The final dataframe contained mapmatched GPS points and information about the node of each point.

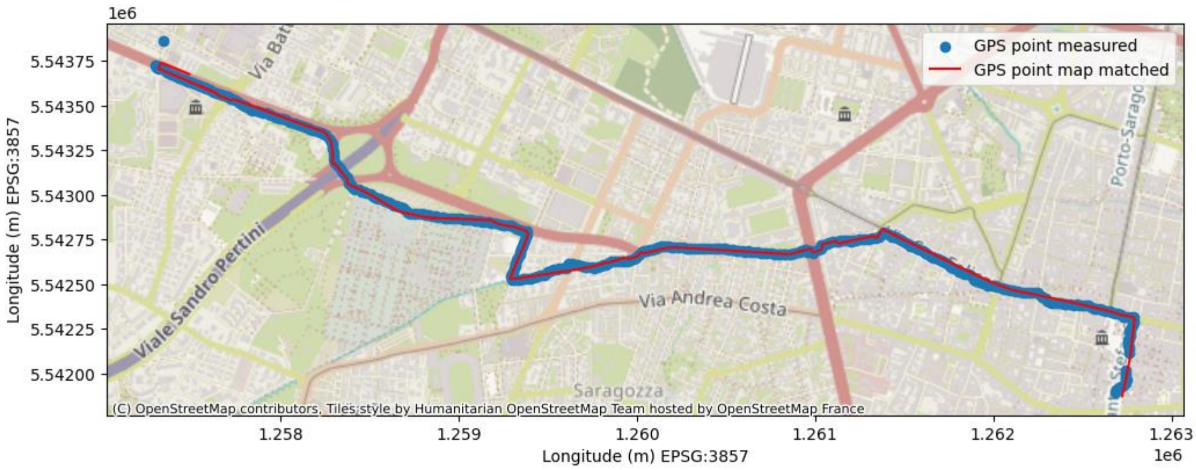

Figure 5. Example of GPS points of one activity before and after map matching.

3.4 Descriptive analysis

Descriptive analysis was done using aggregated data after map matching. Data was aggregated based on ActivityID. The information about aggregated dataset variables have been added to Table 4. The activity duration was calculated based on time difference of first and last GPS point of each activity. The travelled distance was calculated as a sum of distances of consecutive road segments. Additional variables from time were prepared indicating different time periods (day of week, hour of day etc.).

Table 4. Variables included in the aggregated dataset.

Variable	Variable description
ActivityId	Unique identifier of activity assigned during processing
geometry	Multilinestring of mapmatched trajectory
start_t	Start time of activity
end_t	End time of activity
OrigActivityId	Original unique identifier of activity
duration	Trip duration
distance	Trip distance

3.5 Road type analysis

Road type analysis was done to investigate what type of roads are mostly used for commuting using bicycle in Bologna city. To optimize calculations, the street type analysis was done using only 1 month data (April 2017). Data from entire program period was used when road type analysis was done as a part of the flow analysis. Google street view images were used to identify and compare street type information when necessary.

The dataset with map matched GPS data was merged with OSM street network data. The entries not matched with OSM node and activities that had only one matched GPS point were removed. Highway, bicycle and cycleway tags and their values were used to characterize the usage of different road types.

Additionally, cycling road classification as described in Table 1 was adapted and used to gain deeper insight of road usage. The custom classification rules are outlined in Code snippet 1. To estimate the usage intensities of different streets and street types, the edges from OSM data and their occurrence were counted.

The proportion of road type usage was calculated as the lengths of all road types in activity divided by the total length of the activity. The length in one activity represents the total length of this road type in the activity.

```
1. data['category_cycle_track'] = (data['highway']=="cycleway") |
(data['cycleway']=="track")
2. data['category_bicycle_lane'] = (data['cycleway'].isin(["lane",
"opposite_lane"]))
3. data['category_path'] = ((data['highway']=="path") &
(data['bicycle'].isin(["designated", "yes"]))) | ((data['highway']=="footway") &
(data['bicycle']=="yes")) | ((data['highway']=="service") &
(data['bicycle'].isin(["designated", "yes"])) & (data['motor_vehicle']=="no"))
4. data['category_local_bikeway'] = (data['cycleway'].isin(["share_busway",
"opposite_share_busway"])) | (data[' bicycle']=="designated")
```

Code snippet 1. Custom classification rules for road type analysis.

3.6 Spatial tessellation

Spatial tessellation was used to divide the city into polygons with no overlaps and no gaps based on GPS points and point densities. Irregular tessellation method with Voronoi diagrams was used since this method uses POI data and results in more realistic tiles compared to regular tessellation [79]. The Voronoi tiles are prepared using POI-s as seeds. The tiles in a plane are defined by assigning each area to its nearest POI using Euclidean distance. Areas in equal distance from POI are shared, forming region boundaries [80]. In this way, each point will have a dedicated area, Voronoi cell or tile, defined by these individual points – POI-s and Euclidean distance metric. According to Adhinugraha et al a "Voronoi cell can be defined as the region where any point x located in a Voronoi cell $C(f_i)$ will consider f_i as the nearest facility point" [81]:

$$C(f_i) = \{x \mid d(x, f_i) \leq d(x, f_j), i \neq j\}$$

A Voronoi Diagram [82] can be defined as the set of Voronoi cells [81]

$$VD = \{C(f_1), C(f_2), \ldots, C(f_n)\}$$

Voronoi diagrams are often used to define the catchment areas for each point from a group of points in a two-dimensional space [60] and the method is used in various ways in spatial analysis [81,82].

The clustering algorithm selects randomly *k* initial cluster centres (c_1, c_2, …, c_k) from all available *n* points $\{x_1, x_2, \ldots, x_n\}$, $k \leq n$. Each point is assigned to cluster C_j corresponding to its cluster centre c_j which is closest to it:

$$j = 1, 2, \ldots, k \text{ iff } \lVert x_i - c_j \rVert \leq \lVert x_i - c_p \rVert, p = 1, 2, \ldots, k \text{ and } j \neq p.$$

The new cluster centres c_1^*, c_2^* , …, c_k^* are calculated according to

$$c_i^* = \frac{1}{n_i} \sum_{x_j \in C_i} x_i \text{ for i= } 1, 2, \ldots, k$$

Where n is the number of data points in cluster C_i. The process is repeated from assigning points to clusters until $c_i^* = c_i$ [83]. The K-means clustering algorithm assigns all points/tiles to predetermined number of clusters in a way to minimize the intracluster distances, maximize the intercluster distances [29].

Tesspy library was used to create Voronoi diagrams and GPS coordinates of origin points of all activities were used as POI-s to create tessellation tiles. The algorithm created individual tessellation tile for each activity's point of origin. To decrease the number of tiles, K-means clustering was used to merge similar tiles.

Different numbers of *k* were applied, and the final *k*-value was set to 300. If the *k* value is small, large tiles were large in suburban areas and trajectories with origin and destination in same tile could not be considered to be in the same area. If the number of tiles were set too large, then number of activities in clusters is too small for further analysis. The created tiles do not overlap and cover the entire Bologna city area.

3.7 Flow analysis

Flow analysis was used to describe cyclists' movement patterns and routes in the city of Bologna. Flow refers to the movement from one spatial unit to another. In the scope of this study, flow is defined as the movement of cyclists from a specific origin tile to a designated destination tile.

Flow analysis was conducted using Voronoi tiles. Origin and destination tiles for each activity were determined based on the location of first and last GPS point of each activity, respectively. Subsequently, the data was grouped by origin and destination tiles and number of activities from each origin-destination tile pairs were counted. The resulting table contained data about number of movements between all Voronoi tile pairs.

Detail trajectory analysis was done using 20 most numerous flows. The flow analysis distinguished flows from opposite directions as different flows. However, the results showed that if flow from one direction is among top 20, then the opposite direction is also numerous. Therefore, to increase the number of activities in flows, the flows from both directions were considered as one flow. However, movements from one direction were turned around so all movements would be in the same direction to simplify distance calculation and clustering.

Trajectory clustering analysis was done to gain the most common routes used to commute between different locations represented by tessellation tiles. The route is a trajectory for moving between two certain location which are defined as tiles in this study. A flow from one tile to another can have more than one possible trajectory - route. Frechet distance [84] was used as a metric for clustering and DBSCAN (density based spatial clustering for applications with noise) as clustering method.

DBSCAN is a widely used density-based approach for urban travel pattern analysis. This is a nonparametric method that can determine the number of clusters, handle clusters with varying densities and it identifies outliers as noise [85]. Method identifies regions with high density in data by computing distances between points based on a selected metrics. The minimal parameters specified are the maximum distance radius (Epsilon) and the minimum number of points in clusters (MinPts) [86,87].

The Fréchet distance is a metric for measuring the similarity between two curves, commonly explained through an analogy of a person and a dog connected by a leash. Both are moving along their respective curves from start to end, with the freedom to control their speed but without the ability to backtrack. The Fréchet distance is defined as the minimum length of the leash required for both the person and the dog to traverse their curves in this manner [88,89].

The Fréchet distance between two directed curves A and B is determined as the maximum distance reached between the two curves at any point on that curve. The Fréchet distance, denoted as $\delta_F(A, B)$, is the minimum possible cost over all curves. Formally, it is expressed as [90]:

$$\delta_F(A, B) = \inf_{\mu} \max_{a \in A} \operatorname{dist}\big(a, \mu(a)\big),$$

where $\operatorname{dist}(\cdot, \cdot)$ represents Euclidean distance, and $\mu: A \to B$ is a function that maps each point $a \in A$ to a point $\mu(a) \in B$ [90].

The input of the clustering algorithm was the distance matrix of trajectories for one specific flow, minimal number of points in clusters and eps value. The eps value was set to eps=0.001 and the minimal number of points in clusters was 1. For each flow, the distance matrix was calculated, and clustering algorithm was applied individually.

The result of clustering was all trajectories in one flow divided into clusters according to route taken. The cluster size varied from 1 to more than 200 trajectories. For further analysis clusters with less than 5 trajectories were discarded since these contained trajectories which can be considered isolated instances or outliers. The Figure 6 illustrates the flow analysis where the tessellation tiles are marked with blue and different trajectories between two tiles are marked in red and after clustering different routes are marked with different colours.

Street type analysis using similar approach as described in previous chapter were applied to trajectory clusters to determine if street type has influence in cyclist route choice.

Figure 6. Sample of trajectory clustering results from one flow analysis: Origin and destination tiles are indicated with blue. Left figure illustrates all trajectories in this flow with red. Right figure includes only trajectories which were clustered into groups of 5 or more and form different routes.

4. Results and discussion

4.1 Descriptive analysis

Figure 7 describes the number of daily activities during the study period – the count of activities in each day and sliding average over 7 days. Largest number of daily activities were recorded in April and May – 58 033 and 78 892 activities in month, respectively. After that the number of activities decreased and reached a minimum in August – 21 291 activities during the month. The decrease starting from June can be explained by starting of school breaks and therefore less activities due to school and University. Also, in Italy, the summer holiday is usually scheduled in August and many people leave cities which may be the cause of minimum records. The increase in number of activities in September compared to August can be explained by the end of the holiday season and start of the school year. However, the recorded activities did not reach to same level when the program was started – April and May. Probably many people who started the mobility recording program quit, thereby the recorded numbers decreased during the program.

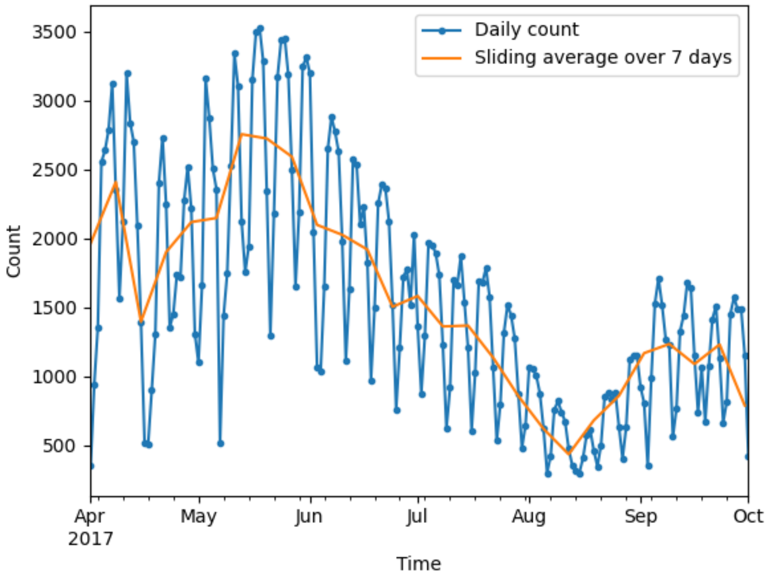

Figure 7. Number of activities in a day and 7 day sliding average during study period.

In addition to seasonal variations, the number of activities exhibits weekly fluctuations. Figure 8 presents a heatmap of the frequencies of movements in different weekdays and hours throughout the day. In a weekly basis, higher volume of movements was recorded during workdays and less during weekends. However, on a daily scale, the peak hours were morning between 6-8 AM and evenings at 16-18 PM. These peaks hours correlate with the activities of people commuting to work and school in the morning and back home in the evening. The lunch peaks around 12-13 can also be seen.

The weekend and workdays are different in terms of number of movements and daily routine. The number of activities was smaller during the weekend and the peak hours, where number of activities is significantly higher, are barely detected. The morning peaks hour is 2 hours later during the weekend, compared to a working day, but the evening peak hour remains at the same time.

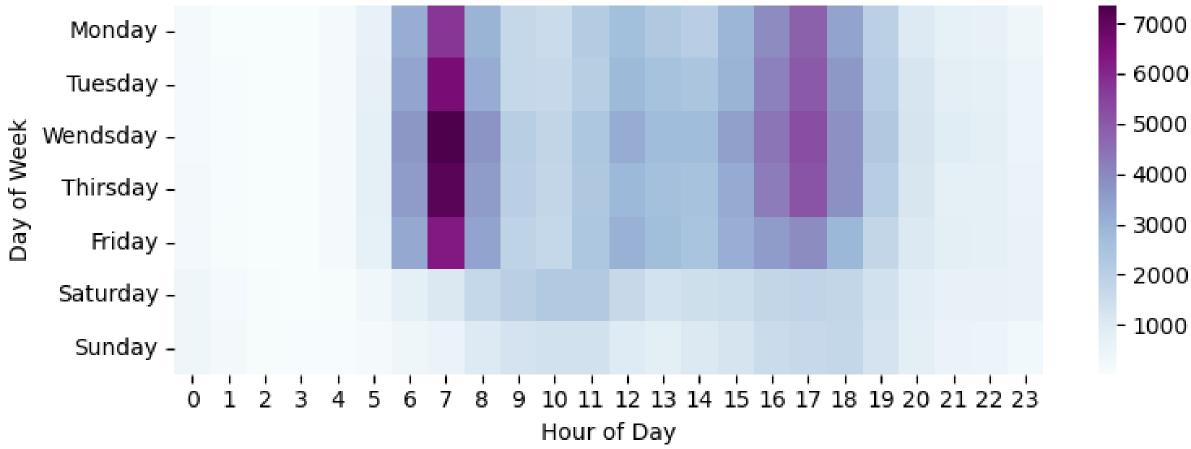

Figure 8. The Count of activities across weekdays and different hours of the day.

From Figure 9, it can be seen that most activities are less than 8 km long and take less than 40 minutes. The mean distances of weekend and workday activities are 2985,5 m and 3080,0 m and mean durations are 17:33 minutes and 16:56 minutes, correspondingly. Cycling is used mainly for short or medium distance tips in the city. The analysis is based on Bologna city and during preprocessing many long activities were split into multiple smaller ones and additionally limits were set on duration of activity.

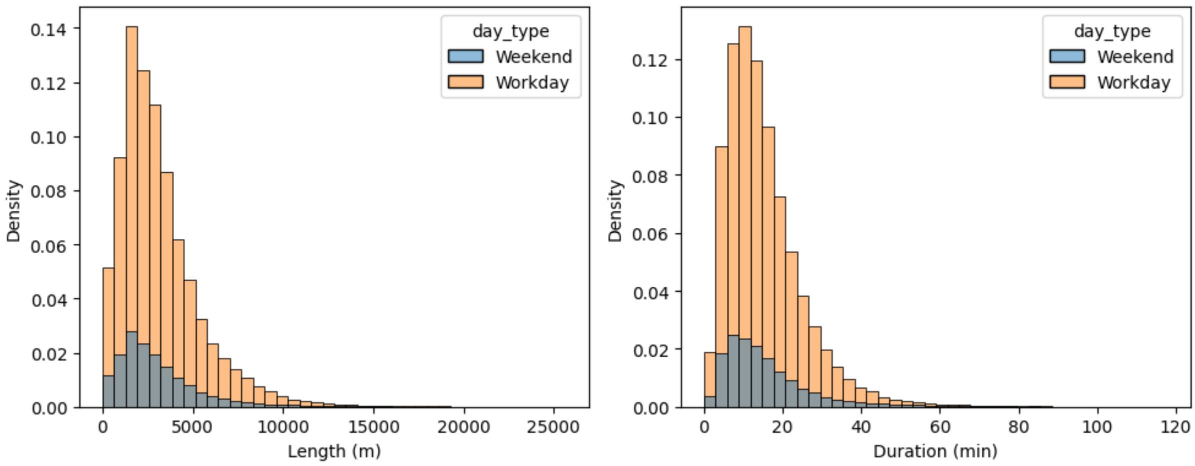

Figure 9. Histograms of distances and durations of activities in workdays and weekends.

4.2 Spatial tessellation

Tessellation divides the space into subspaces or tiles with no overlaps and no gaps which enables to gain better understanding of geographic space and analyse geospatial data [79]. Irregular tessellation with Voronoi diagrams was used to divide city into tiles and analyse where activities start and end (Figure 10).

The tiles where the highest number of activities started and ended was in the city centre. Compared to suburban areas, city centre has denser road network, more activities start and end there and this also leads to smaller and darker tiles in the city centre. Less activities were starting and ending in suburban areas, but also in the southern and western areas of the city. Compared to other districts, the southern area of the city is a sparsely populated residential area with hilly landscape and bad public transport network. However, western area is primarily a residential area with tall apartment buildings and good public transportation [91]. Due to low population density in southern areas and good public transportation network in other areas, the number of cycling activities in these suburban areas were lower.

The colour scheme in Figure 10, illustrating the spatial distribution of origin and destination points within the tessellation tiles, showcases a similarity in patterns. This suggests that there is a similarity in the patterns of places where activities start and end, with comparable occurrences across the tiles. Similarly, no significant difference was found when workday and weekend mobility pattern maps were compared.

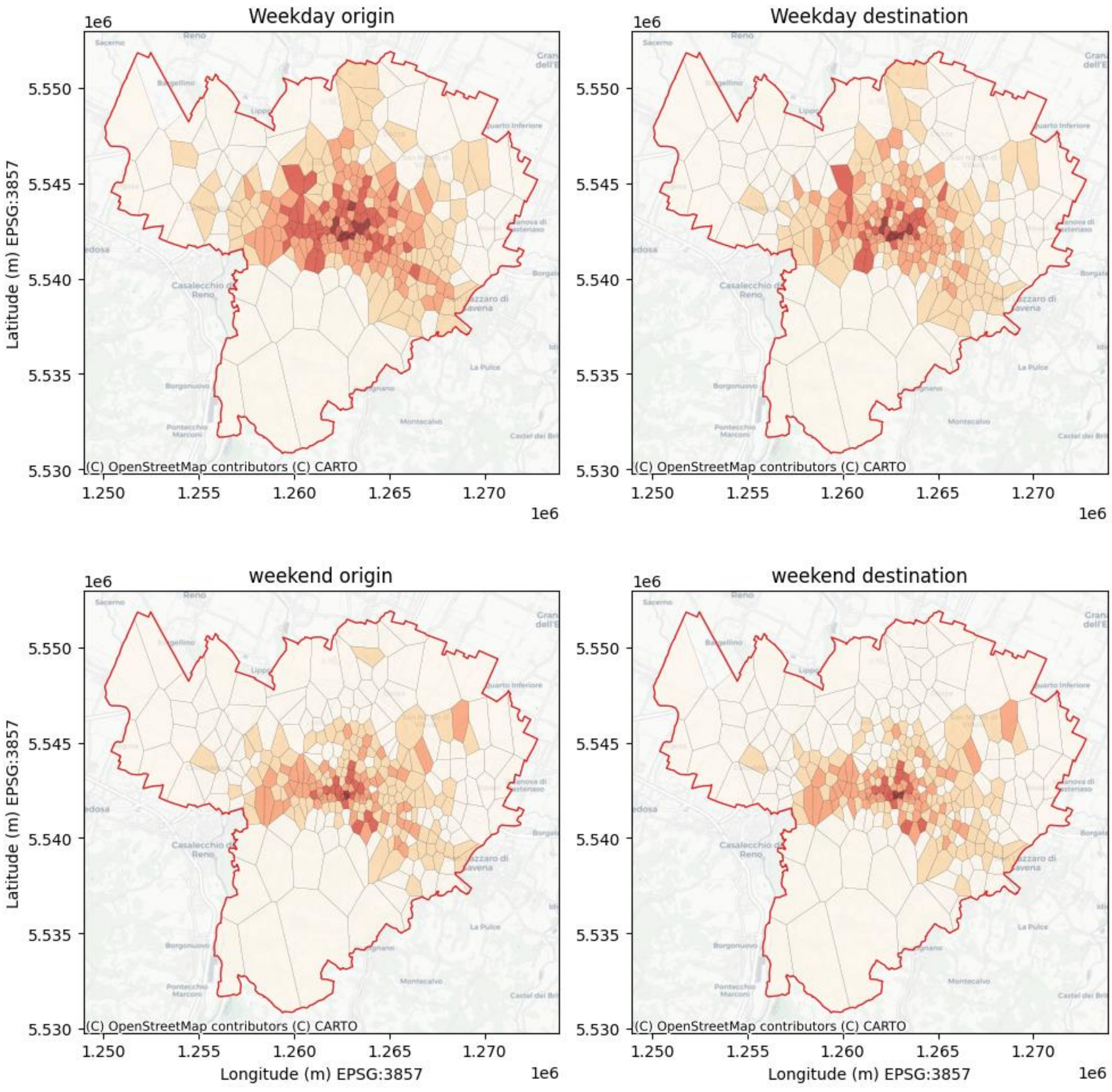

Figure 10. A visualization of the spatial distribution of origin or destination points using Voronoi polygons in workdays and weekend. K-means (300 tiles) is used for clustering and the tile colour is representing the number of activities in the tile.

4.3 Road usage analysis

Road usage analysis was conducted to study the urban infrastructure and its usage preferences for cycling. The heatmap in Figure 11 illustrates the road network of Bologna and its usage intensities for cycling during 1 month period (1.04.2017-30.04.2017).

The most frequently used roads and streets are coloured using a darker tone of red. These are bigger roads, multi-lane highways and avenues with more intense cycling traffic. In suburban areas, mostly highways which are used to enter the city and lead towards the city centre. In the city centre, the circular route which enables cyclist to have a ride around the old town is heavily used. This route is a two-way cycling road separated from cars and thereby a very comfortable and safe way to move near city centre. Major streets lead to city centre with a historic old town and very narrow and often one-way streets. In this region, cycling tends to converge to the streets that are wider, in some sections two directional and may contain on-street bicycle lanes (via Ugo Bassi, via dell'Indipendenza, via Riva di Reno, etc.). These streets enable to drive through old town faster and more convenient way than narrow and winding historic streets.

During the weekends, the cycling infrastructure is used less than on workdays. However, the most heavily used streets are still the same as during workdays. The smaller usage intensity during weekend can be seen also in Figure 8 where number of movements and daily routine is visualized. It is also important to point out that the number of workdays in a week is 5 and weekends is 2 therefore this will also accentuate higher road usage intensities during workdays.

Figure 11. Count of usage events of different road segments describing Bolognas street usage intensities during the period of 1 month in workdays (left subplot) and weekends (right subplot).

Additional analysis was done to investigate what type of streets are preferred by cyclists. The street type information was gained from OSM data using different tags and their values. Although the information contained in OSM is not always correct it holds reasonably good quality to give good overview of street infrastructure [78,92].

Figure 12 illustrates the usage of different street types in Bologna. Figure 12 A shows the relative length of different road types and Figure 12 B the length of each street type in activities. The results show that cyclists prefer streets and roads with smaller traffic intensities and dedicated cycling roads. The most used street types, considering both relative length and length of different road types, are cycleways and residential streets. These streets are designated cycleways and smaller streets which serve as an access to houses, respectively [93]. The average relative length of these street types in activities are 30% and 23%, respectively, which is an average of 870 m and 841 m in total length, respectively.

The largest and assumably busiest street types in Bologna are primary and secondary streets. These tags are used to label major traffic arteries and boulevards within cities [94]. Averages of 569 and 686 meters in an activity, which is 15% and 20% of the length of activity is cycled on these street type. Additionally, tertiary roads which facilitate connections between local centres and linking minor streets to more major roads and activity includes an average of 20% or 668 m of this road type.

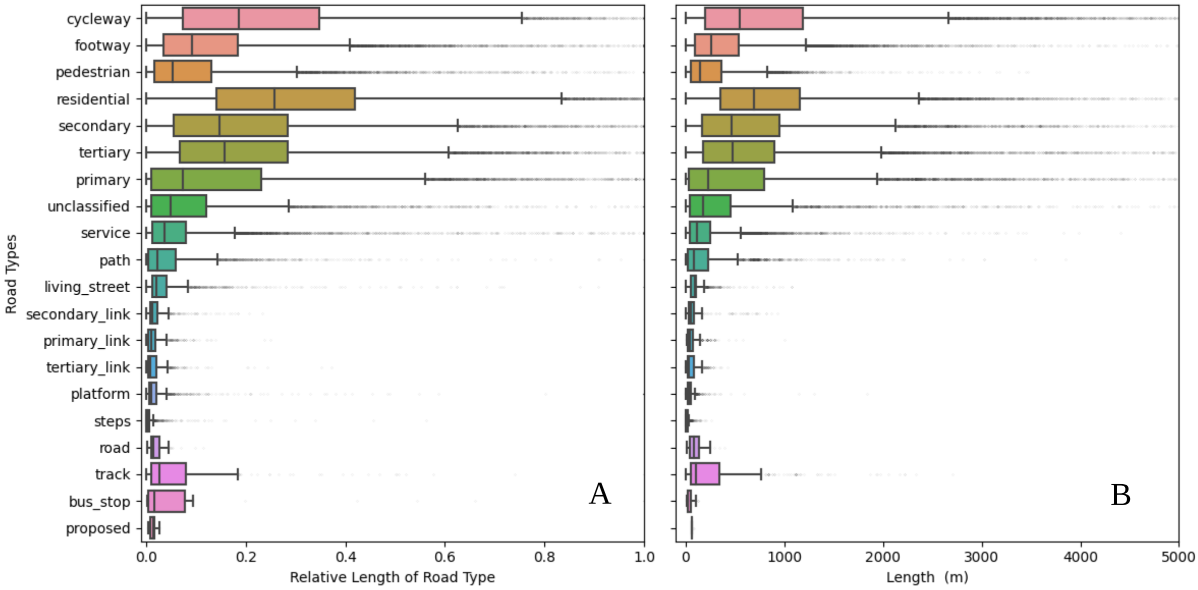

Figure 12. The relative length of different road types (A) and length of different road types (B) based on highway tag and its values.

Similar analysis was done using only cycling infrastructure (Figure 13). Custom classification was used to group all streets with cycling infrastructure to see how much each type of infrastructure is used. The results show that amongst different bicycle roads, local bikeways are most preferred, followed by cycle tracks. These street types are mainly roads dedicated for public transport on which cyclists are also allowed to bike and specially designated paths for bicycle use [95,96]. Streets where cycling road are painted to streets and can be mixed with parking places are labelled as bicycle_lanes and are less preferred. Although paths are separated bicycle roads shared with pedestrians, the usage of these is small. The higher usage of cycle tracks and local bikeways are caused by longer length of these streets. The total length of these street types in street network data was 137,0 km and 107,3 km, respectively. The total length of bicycle lanes and paths in street network data was 2,9 km and 57,3 km, respectively.

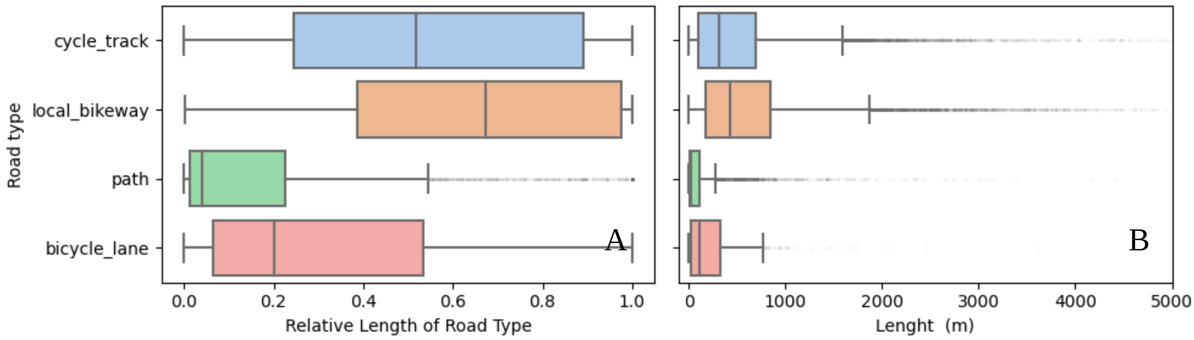

Figure 13. The relative length of different road types (A) and length of different road types (B) based on custom tags describing cycling road types.

4.4 Flow analysis

Flow analysis was done using tessellation tiles and all activities were assigned corresponding origin and destination tiles. The activities were grouped by origin and destination tiles and all activities with same tile set were counted and considered as one flow. The trajectory analysis was done using 20 of the most numerous flows between two tiles. Trajectory clustering was used for each flow to group similar trajectories and thereby identify different routes used to move between two locations. The most popular route from of 20 most used flows with corresponding origin and destination tiles are visualized in Figure 14. Most of the origin-destination tiles are in the city centre or its nearby areas. The trajectories are mostly between the city centre and its surrounding areas or from farther suburban areas towards city centre.

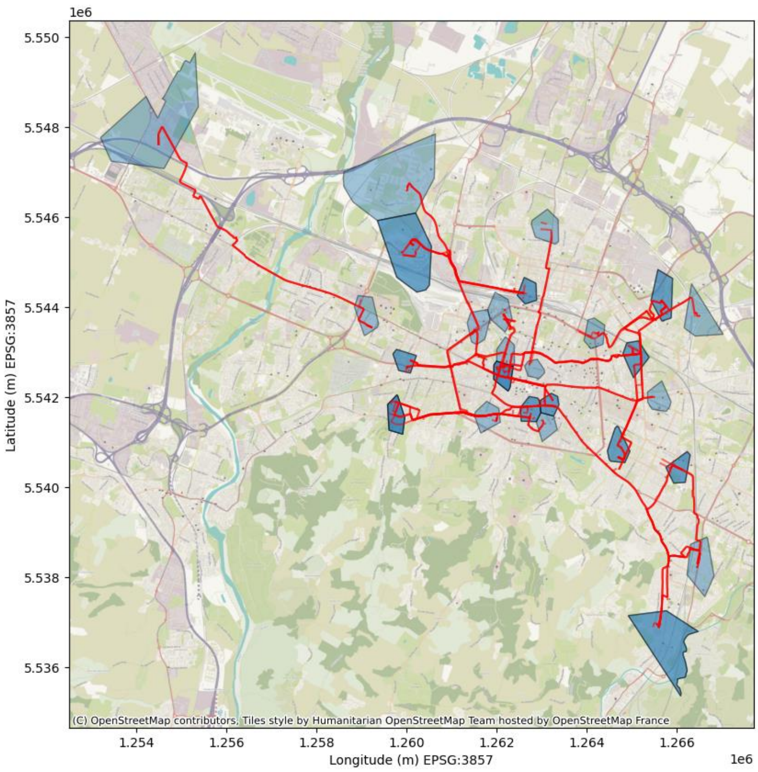

Figure 14. The origin-destination tiles of 20 most popular flows (in blue) and most used trajectories to commute between them (in red).

Each flow was analysed individually to see different routes between two tiles. The DBSCAN clustering algorithm was applied to cluster all trajectories into groups of similar routes. Street type analysis was done to see if street type impacts the route choices. The maps illustrating 20 most used flows and results of street type analysis are added to Appendixes I.

The analysed flows had between 1 to 5 clusters and the number of trajectories in a cluster ranged from 5 to more than 200. Clusters with 5 or less trajectories were considered isolated instances or outliers and were removed from further analysis. Only one cluster was gained when two tessellation tiles were too close to each other and trajectories with short distances were dispersed over a narrow area causing too small differences in distance matrix (Appendix I, Figures 25, 26, 28 and 33). One cluster was also results if only one main route for commuting was present and possible alternative routes were used less than 5 times (Appendix I, Figure 24 and 27). The clustering resulted in several clusters if multiple actively used and distinctive routes from two tiles was possible.

The main reason for having multiple active routes is the occurrence of one-way roads. Due to this, many flows are clustered into distinctive clusters based on movement direction in one-way street (Appendix, Figures 15, 17, 23, 29, 33). The activities in one or many clusters move in one direction using one-way street and trajectories in other cluster or clusters move in the opposite direction using alternative routes. For example, in Appendix I Figure 15 the trajectories in clusters 0 and 1 move all in one direction and in clusters 9 and 10 in the opposite direction. Similar effect can be seen in Appendix Figure 17 where clusters 0 is majority one way and clusters 6 and 8 the opposite way due to narrow one-way street segment in historic old town. If the route choice is influenced by one-way streets, then these streets are very narrow. Often even tighter due to parking cars and therefore not convenient for moving in the opposite direction. Although cycling in opposite direction in one-way street is not allowed, the data indicates that in some sections this is still done (Figure 32, cluster 1).

The influence of street type, distance, and duration of activity to cyclist route choice preference were additionally analysed. When multiple routes are available then streets with cycling infrastructure are preferred and when possible, narrow streets in the old town are avoided. For example, in Figure 32 cyclists prefer to use longer routes (clusters 0 and 22) to avoid narrow street sections in old town (in cluster 1). The longer route includes wider streets and enables people to arrive at their destination faster. In Figure 30, the most used path takes more time and is longer but includes long sections of cycling and tertiary roads while the significantly less used alternative route includes high traffic intensity primary roads. Similarly in Figure 16 travel time and duration for activities in clusters 0 and 9 are similar but majority of cyclist use route in cluster 0 since streets include two way cycling path on the sidewalk. The street analysis also reveals that the preferred route has more cycleways and residential street types and less secondary street types. If no significant difference in street types can be found, the cyclist prefers the routes that enable shorter distance and/or smaller duration of activity (Figure 21). To better estimate the influence of infrastructure and other parameters to route choice, modelling could be applied where more influencing factors can be considered. Current work only analyses and compares routes that are actively used by cyclist. However, for more profound analysis, the routes that are not used should be also analysed to see why some routes are not used at all.

5. Conclusions

Due to environmental concerns and urbanisation, increased attention has been paid to alternative transport modes like cycling and walking. Numerous studies been done to investigate factors influencing cyclist route choices, and various models for different cities have been published. Most previously reported studies use localized data and applying diverse infrastructure variables to examine their impact on route selection.

In this work, the mobility campaign data from city of Bologna was used to study cyclist mobility and route choices. This thesis presents a framework of analysing mobility flows in urban areas by origin-destination tiles and studying the people's routes between them. Spatial tessellation was used to divide city into tiles and divide all trajectories into flows according to their origin and destination tiles. Additionally, clustering with DBSCAN was applied to group all trajectories in flows into characteristic routes. Finally, data embedded in downloaded OSM data was used to see how street type and it attributes influence the route choice made by cyclist on Bologna.

The descriptive analysis revealed general fluctuations in cyclist activity. More people cycled when the Bella Mossa mobility campaign started, and the activity decreased gradually. Activity was lower in August, during the vacation period, but increased slightly when schools reopened in September. Additionally, cycling was more active on workdays and less active during weekends. In a daily schedule, cycling activity followed the daily routine – higher when people commuted to school and work in the morning or returned home in the evening. The weekend routine was slightly different compared to workdays and only one peak in activity was detected. Cycling served as a mobility method for shorter or medium-length distances. The mean distance and duration of cycling activities were around 3 km and 17 minutes, respectively.

Spatial tessellation revealed areas with lower and higher cycling activities. More activities start and end in city centre while suburban areas have lower cycling activities. Due to regional specifics, the south and east areas of city have lower cycling activity compared to other suburban areas. Also important is to further study the reasons of lower cycling activity in certain areas and use the results of this study as an input to improve the cycling infrastructure to increase cycling in these areas.

The road preference is also different is different areas of city. In suburban areas, cyclist prefer mostly highways and wider streets which are convenient to enter the city and lead towards the city centre. On the other hand, in the city centre, the circular road which avoids old town or larger roads in the old town which enable to drive through old town faster and more convenient way. The road type analysis revealed that cyclists prefer streets and roads with smaller traffic intensities and with dedicated cycling roads. The most used street types are cycleways and residential streets.

The 20 most used flows between origin-destination were clustered into distinctive routes and the analysis revealed that route choice was mainly influenced by the presence of one-way streets. Cyclists prefer not to drive one-way street in the opposite direction even if it is possible in the sidewalk. In addition, streets with cycling infrastructure are preferred and when possible, narrow streets in the old town are avoided. Consistent with findings from previous studies, cyclists in Bologna prefer shorter and faster routes but are willing to take longer routes to avoid certain cycling infrastructure types. In Bologna, cyclist prefer streets with good cycling infrastructure and avoid narrow streets in old town.

Current work included only routes which were used and included in the dataset. The route preference can be further studied by adding routes generated by dedicated algorithm. This enables to see if suggested (for example shortest or fastest) routes are used by cyclist or if not, then what might cause the discrepancies. The information about cyclist route choices is necessary and can be used to improve cycling infrastructure to make cycling more convenient and thereby increase the cycling activities. The research could be further elaborated by adding other mobility modes to find possibilities to increase usage of other alternative transport modes and thereby decrease the usage of motorized transport.

6. References

[1] L. Pappalardo, F. Simini, G. Barlacchi, R. Pellungrini, Scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data, J Stat Softw. 103 (2019). https://doi.org/10.48550/arxiv.1907.07062.

[2] V.A. Paz-Soldan, R.C. Reiner Jr, A.C. Morrison, S.T. Stoddard, U. Kitron, T.W. Scott, J.P. Elder, E.S. Halsey, T.J. Kochel, H. Astete, G.M. Vazquez-Prokopec, Strengths and Weaknesses of Global Positioning System (GPS) Data-Loggers and Semi-structured Interviews for Capturing Fine-scale Human Mobility: Findings from Iquitos, Peru, PLoS Negl Trop Dis. 8 (2014) e2888-. https://doi.org/10.1371/journal.pntd.0002888.

[3] R. Wang, Q. Wang, N. Li, Percolation transitions in urban mobility networks in America's 50 largest cities, Sustain Cities Soc. 91 (2023) 104435. https://doi.org/10.1016/J.SCS.2023.104435.

[4] A. Roy, D. Fuller, T. Nelson, P. Kedron, Assessing the role of geographic context in transportation mode detection from GPS data, J Transp Geogr. 100 (2022) 103330. https://doi.org/10.1016/J.JTRANGEO.2022.103330.

[5] W. Oueslati, S. Tahri, H. Limam, J. Akaichi, A systematic review on moving objects' trajectory data and trajectory data warehouse modeling, Comput Sci Rev. 47 (2023) 100516. https://doi.org/10.1016/J.COSREV.2022.100516.

[6] L. Antón-González, M. Pans, J. Devís-Devís, L.M. González, Cycling in urban environments: Quantitative text analysis, J Transp Health. 32 (2023) 101651. https://doi.org/10.1016/J.JTH.2023.101651.

[7] H. Charreire, C. Roda, T. Feuillet, A. Piombini, H. Bardos, H. Rutter, S. Compernolle, J.D. Mackenbach, J. Lakerveld, J.M. Oppert, Walking, cycling, and public transport for commuting and non-commuting travels across 5 European urban regions: Modal choice correlates and motivations, J Transp Geogr. 96 (2021) 103196. https://doi.org/10.1016/J.JTRANGEO.2021.103196.

[8] M. Jordi-Sánchez, M. Hernández-Ramírez, M. Cabillas, A.M. Pérez-Flores, V.M. Muñoz-Sánchez, Deconstructing the categories of urban cycling: beyond transport, leisure and sport, Mobilities. 17 (2022) 795–813. https://doi.org/10.1080/17450101.2022.2057811.

[9] T. Yanar, Understanding the choice for sustainable modes of transport in commuting trips with a comparative case study, Case Stud Transp Policy. 11 (2023) 100964. https://doi.org/10.1016/J.CSTP.2023.100964.

[10] J. De Vos, H.T.K. Le, M. Kroesen, Does commute duration attenuate the effect of travel mode choice on commute satisfaction?, Travel Behav Soc. 28 (2022) 13–21. https://doi.org/10.1016/J.TBS.2022.02.004.

[11] X. Zhao, Y. Guo, Planning bikeway network for urban commute based on mobile phone data: A case study of Beijing, Travel Behav Soc. 34 (2024) 100672. https://doi.org/10.1016/J.TBS.2023.100672.

[12] L. Mussone, F. Changizi, A study on the factors that influenced the choice of transport mode before, during, and after the first lockdown in Milan, Italy, Cities. 136 (2023) 104251. https://doi.org/10.1016/J.CITIES.2023.104251.

[13] C. Ding, D. Wang, C. Liu, Y. Zhang, J. Yang, Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance, Transp Res Part A Policy Pract. 100 (2017) 65–80. https://doi.org/10.1016/J.TRA.2017.04.008.

[14] G. Reggiani, T. Verma, W. Daamen, S. Hoogendoorn, A multi-city study on structural characteristics of bicycle networks, Environ Plan B Urban Anal City Sci. 50 (2023) 2017–2037. https://doi.org/10.1177/23998083231170637.

[15] C. Ferster, J. Fischer, K. Manaugh, T. Nelson, M. Winters, Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities, Int J Sustain Transp. 14 (2020) 64–73. https://doi.org/10.1080/15568318.2018.1519746.

[16] L. Meng, Political economy and cycling infrastructure investment, Transp Res Interdiscip Perspect. 14 (2022) 100618. https://doi.org/10.1016/J.TRIP.2022.100618.

[17] L. Wysling, R.S. Purves, Where to improve cycling infrastructure? Assessing bicycle suitability and bikeability with open data in the city of Paris, Transp Res Interdiscip Perspect. 15 (2022) 100648. https://doi.org/10.1016/J.TRIP.2022.100648.

[18] Urban Bikeway Design Guide, (n.d.). https://nacto.org/publication/urban-bikeway-design-guide/bicycle-boulevards/ (accessed December 14, 2023).

[19] O. Codina, M. Maciejewska, J. Nadal, O. Marquet, Built environment bikeability as a predictor of cycling frequency: Lessons from Barcelona, Transp Res Interdiscip Perspect. 16 (2022) 100725. https://doi.org/10.1016/J.TRIP.2022.100725.

[20] B. Guirao, D. Gálvez-Pérez, N. Casado-Sanz, The impact of the cyclist infrastructure type on bike accidents: the experience of Madrid, Transportation Research Procedia. 71 (2023) 403–410. https://doi.org/10.1016/J.TRPRO.2023.11.101.

[21] H. Mahfouz, R. Lovelace, E. Arcaute, A road segment prioritization approach for cycling infrastructure, J Transp Geogr. 113 (2023) 103715. https://doi.org/10.1016/J.JTRANGEO.2023.103715.

[22] D.M. Scott, W. Lu, M.J. Brown, Route choice of bike share users: Leveraging GPS data to derive choice sets, J Transp Geogr. 90 (2021) 102903. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2020.102903.

[23] A. Meister, M. Felder, B. Schmid, K.W. Axhausen, Route choice modeling for cyclists on urban networks, Transp Res Part A Policy Pract. 173 (2023) 103723. https://doi.org/10.1016/J.TRA.2023.103723.

[24] E. Desjardins, E. Apatu, S.D. Razavi, C.D. Higgins, D.M. Scott, A. Páez, “Going through a little bit of growing pains”: A qualitative study of the factors that influence the route choice of regular bicyclists in a developing cycling city, Transp Res Part F Traffic Psychol Behav. 81 (2021) 431–444. https://doi.org/10.1016/J.TRF.2021.06.005.

[25] F.L. Berghoefer, M. Vollrath, Cyclists’ perception of cycling infrastructure – A Repertory Grid approach, Transp Res Part F Traffic Psychol Behav. 87 (2022) 249–263. https://doi.org/10.1016/J.TRF.2022.04.012.

[26] M. Łukawska, M. Paulsen, T.K. Rasmussen, A.F. Jensen, O.A. Nielsen, A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset, Transp Res Part A Policy Pract. 176 (2023) 103834. https://doi.org/10.1016/J.TRA.2023.103834.

[27] R. Félix, P. Cambra, F. Moura, Build it and give ‘em bikes, and they will come: The effects of cycling infrastructure and bike-sharing system in Lisbon, Case Stud Transp Policy. 8 (2020) 672–682. https://doi.org/10.1016/J.CSTP.2020.03.002.

[28] F.L. Berghoefer, M. Vollrath, Motivational and deterrent effects of route attributes in cyclists’ route choice, Transp Res Part F Traffic Psychol Behav. 95 (2023) 343–354. https://doi.org/10.1016/J.TRF.2023.04.003.

[29] D. Ton, O. Cats, D. Duives, S. Hoogendoorn, How Do People Cycle in Amsterdam, Netherlands?: Estimating Cyclists’ Route Choice Determinants with GPS Data from an Urban Area, Transp Res Rec. 2662 (2017) 75–82. https://doi.org/10.3141/2662-09.

[30] W. Lu, D.M. Scott, R. Dalumpines, Understanding bike share cyclist route choice using GPS data: Comparing dominant routes and shortest paths, J Transp Geogr. 71 (2018) 172–181. https://doi.org/10.1016/J.JTRANGEO.2018.07.012.

[31] F. Rupi, M. Freo, C. Poliziani, M.N. Postorino, J. Schweizer, Analysis of gender-specific bicycle route choices using revealed preference surveys based on GPS traces, Transp Policy (Oxf). 133 (2023) 1–14. https://doi.org/10.1016/J.TRANPOL.2023.01.001.

[32] Z. Ma, P. Zhang, Individual mobility prediction review: Data, problem, method and application, Multimodal Transportation. 1 (2022) 100002. https://doi.org/10.1016/J.MULTRA.2022.100002.

[33] R. Pritchard, D. Bucher, Y. Frøyen, Does new bicycle infrastructure result in new or rerouted bicyclists? A longitudinal GPS study in Oslo, J Transp Geogr. 77 (2019) 113–125. https://doi.org/10.1016/J.JTRANGEO.2019.05.005.

[34] P. Sadeghian, J. Håkansson, X. Zhao, Review and evaluation of methods in transport mode detection based on GPS tracking data, Journal of Traffic and Transportation Engineering (English Edition). 8 (2021) 467–482. https://doi.org/10.1016/J.JTTE.2021.04.004.

[35] A. Cetateanu, A. Jones, How can GPS technology help us better understand exposure to the food environment? A systematic review, SSM Popul Health. 2 (2016) 196–205. https://doi.org/10.1016/J.SSMPH.2016.04.001.

[36] M. Lin, W.J. Hsu, Mining GPS data for mobility patterns: A survey, Pervasive Mob Comput. 12 (2014) 1–16. https://doi.org/10.1016/J.PMCJ.2013.06.005.

[37] F. Rupi, C. Poliziani, J. Schweizer, Analysing the dynamic performances of a bicycle network with a temporal analysis of GPS traces, Case Stud Transp Policy. 8 (2020) 770–777. https://doi.org/10.1016/J.CSTP.2020.05.007.

[38] K. Siła-Nowicka, J. Vandrol, T. Oshan, J.A. Long, U. Demšar, A.S. Fotheringham, Analysis of human mobility patterns from GPS trajectories and contextual information, International Journal of Geographical Information Science. 30 (2016) 881–906. https://doi.org/10.1080/13658816.2015.1100731.

[39] K. Lee, I.N. Sener, Strava Metro data for bicycle monitoring: a literature review, Transp Rev. 41 (2021) 27–47. https://doi.org/10.1080/01441647.2020.1798558.

[40] E. Berjisian, A. Bigazzi, Evaluation of methods to distinguish trips from activities in walking and cycling GPS data, Transp Res Part C Emerg Technol. 137 (2022) 103588. https://doi.org/10.1016/J.TRC.2022.103588.

[41] H.Y. Song, J.H. Lee, A map matching algorithm based on modified hidden Markov model considering time series dependency over larger time span, Heliyon. 9 (2023) e21368. https://doi.org/10.1016/J.HELIYON.2023.E21368.

[42] D. Wang, T. Miwa, T. Morikawa, Big Trajectory Data Mining: A Survey of Methods, Applications, and Services, Sensors 2020, Vol. 20, Page 4571. 20 (2020) 4571. https://doi.org/10.3390/S20164571.

[43] Y.U. Zheng, Trajectory data mining: An overview, ACM Trans. Intell. Syst. Technol. 6 (2015). https://doi.org/10.1145/2743025.

[44] J. Bian, D. Tian, Y. Tang, D. Tao, A survey on trajectory clustering analysis, ArXiv E-Prints. (2018) arXiv–1802.

[45] Y. Zheng, Trajectory Data Mining: An Overview, ACM Trans. Intell. Syst. Technol. 6 (2015). https://doi.org/10.1145/2743025.

[46] X. Zhou, K. Zheng, H. Jueng, J. Xu, S. Sadiq, Making Sense of Spatial Trajectories, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA, 2015: pp. 671–672. https://doi.org/10.1145/2806416.2806418.

[47] J.-G. Lee, J. Han, K.-Y. Whang, Trajectory Clustering: A Partition-and-Group Framework, in: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery, New York, NY, USA, 2007: pp. 593–604. https://doi.org/10.1145/1247480.1247546.

[48] P.C. Besse, B. Guillouet, J.-M. Loubes, F. Royer, Review and Perspective for Distance-Based Clustering of Vehicle Trajectories, IEEE Transactions on Intelligent Transportation Systems. 17 (2016) 3306–3317. https://doi.org/10.1109/TITS.2016.2547641.

[49] X. Kong, Q. Chen, M. Hou, H. Wang, F. Xia, Mobility trajectory generation: a survey, Artif Intell Rev. 56 (2023) 3057–3098. https://doi.org/10.1007/s10462-023-10598-x.

[50] E. Brusilovskiy, L.A. Klein, G. Townley, G. Snethen, B. McCormick, S. V. Hiremath, M.S. Salzer, Examining the relationship between community mobility and participation using GPS and self-report data, Soc Sci Med. 265 (2020) 113539. https://doi.org/10.1016/J.SOCSCIMED.2020.113539.

[51] S. Saki, T. Hagen, A Practical Guide to an Open-Source Map-Matching Approach for Big GPS Data, SN Comput Sci. 3 (2022) 1–13. https://doi.org/10.1007/S42979-022-01340-5/TABLES/8.

[52] P. Sadeghian, J. Håkansson, X. Zhao, Review and evaluation of methods in transport mode detection based on GPS tracking data, Journal of Traffic and Transportation Engineering (English Edition). 8 (2021) 467–482. https://doi.org/10.1016/J.JTTE.2021.04.004.

[53] C. Parent, S. Spaccapietra, C. Renso, G.L. Andrienko, N. V Andrienko, V. Bogorny, M.L. Damiani, A. Gkoulalas-Divanis, J.A.F. de Macêdo, N. Pelekis, Y. Theodoridis, Z. Yan, Semantic trajectories modeling and analysis, ACM Comput. Surv. 45 (2013) 42:1-42:32. https://api.semanticscholar.org/CorpusID:2359543.

[54] W. Huang, L. Wang, Towards big data behavioral analysis: rethinking GPS trajectory mining approaches from geographic, semantic, and quantitative perspectives, Architectural Intelligence. 1 (2022) 7. https://doi.org/10.1007/s44223-022-00011-y.

[55] D. Zhang, K. Lee, I. Lee, Mining hierarchical semantic periodic patterns from GPS-collected spatio-temporal trajectories, Expert Syst Appl. 122 (2019) 85–101. https://doi.org/10.1016/J.ESWA.2018.12.047.

[56] J.D. Mazimpaka, S. Timpf, Trajectory data mining: A review of methods and applications, J. Spatial Inf. Sci. 13 (2016) 61–99. https://api.semanticscholar.org/CorpusID:28075280.

[57] Z. Zhang, X. Zhao, Y. Zhang, J. Zhang, H. Nie, Y. Lou, Efficient Mining of Hotspot Regional Patterns with Multi-Semantic Trajectories, Big Data Research. 22 (2020) 100157. https://doi.org/10.1016/J.BDR.2020.100157.

[58] M.G.M. Pinho, B. Flueckiger, A. Valentin, M.I. Kasdagli, K. Kyriakou, J. Lakerveld, J.D. Mackenbach, J.W.J. Beulens, K. de Hoogh, The quality of OpenStreetMap food-related point-of-interest data for use in epidemiological research, Health Place. 83 (2023) 103075. https://doi.org/10.1016/J.HEALTHPLACE.2023.103075.

[59] C. Barrington-Leigh, A. Millard-Ball, The world's user-generated road map is more than 80% complete, PLoS One. 12 (2017) e0180698-. https://doi.org/10.1371/journal.pone.0180698.

[60] L.E. Olmos, M.S. Tadeo, D. Vlachogiannis, F. Alhasoun, X. Espinet Alegre, C. Ochoa, F. Targa, M.C. González, A data science framework for planning the growth of bicycle infrastructures, Transp Res Part C Emerg Technol. 115 (2020) 102640. https://doi.org/10.1016/J.TRC.2020.102640.

[61] M. Minaei, Evolution, density and completeness of OpenStreetMap road networks in developing countries: The case of Iran, Applied Geography. 119 (2020) 102246. https://doi.org/10.1016/J.APGEOG.2020.102246.

[62] W. Zhao, Y. Bo, J. Chen, D. Tiede, B. Thomas, W.J. Emery, Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM), ISPRS Journal of Photogrammetry and Remote Sensing. 151 (2019) 237–250. https://doi.org/10.1016/J.ISPRSJPRS.2019.03.019.

[63] C. Ferster, J. Fischer, K. Manaugh, T. Nelson, M. Winters, Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities, Int J Sustain Transp. 14 (2020) 64–73. https://doi.org/10.1080/15568318.2018.1519746.

[64] Tags, (n.d.). https://wiki.openstreetmap.org/wiki/Tags (accessed October 15, 2023).

[65] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, OpenStreetMap data in modelling the urban energy infrastructure: a first assessment and analysis, Energy Procedia. 142 (2017) 1968–1976. https://doi.org/10.1016/J.EGYPRO.2017.12.397.

[66] Y.L. Hsueh, H.C. Chen, Map matching for low-sampling-rate GPS trajectories by exploring real-time moving directions, Inf Sci (N Y). 433–434 (2018) 55–69. https://doi.org/10.1016/J.INS.2017.12.031.

[67] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, Y. Huang, Map-matching for low-sampling-rate GPS trajectories, Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. (2009). https://api.semanticscholar.org/CorpusID:11470200.

[68] Y. and H.W. and Z.X. Chao Pingfu and Xu, A Survey on Map-Matching Algorithms, in: J. and W.W. Borovica-Gajic Renata and Qi (Ed.), Databases Theory and Applications, Springer International Publishing, Cham, 2020: pp. 121–133.

[69] Z. Xiong, B. Li, D. Liu, Map-matching using hidden markov model and path choice preferences under sparse trajectory, Sustainability (Switzerland). 13 (2021). https://doi.org/10.3390/su132212820.

[70] D. Fiedler, M. Čáp, J. Nykl, P. Zilecky, M. Schaefer, Map Matching Algorithm for Large-scale Datasets, 2019.

[71] M. Nardino, L. Cremonini, A. Crisci, T. Georgiadis, G. Guerri, M. Morabito, E. Fiorillo, Mapping daytime thermal patterns of Bologna municipality (Italy) during a heatwave: A new methodology for cities adaptation to global climate change, (2022). https://doi.org/10.1016/j.uclim.2022.101317.

[72] Bologna's Bella Mossa, 2020. https://marketplace.eiturbanmobility.eu/best-practices/bologna-s-bella-mossa-app-provided-incentives-for-citizens-to-travel-more-sustainably-in-the-city/ (accessed October 10, 2023).

[73] J. Murphy, Y. Pao, A. Yuen, Map matching when the map is wrong: Efficient on/off road vehicle tracking and map learning, 2019. https://doi.org/10.1145/3357000.3366143.

[74] P. Newson, J. Krumm, Hidden Markov map matching through noise and sparseness, in: 2009: pp. 336–343. https://doi.org/10.1145/1653771.1653818.

[75] OSRM API Docs: Match Services, (n.d.). http://project-osrm.org/docs/v5.5.1/api/#match-service (accessed June 15, 2023).

[76] R. Pritchard, D. Bucher, Y. Frøyen, Does new bicycle infrastructure result in new or rerouted bicyclists? A longitudinal GPS study in Oslo, J Transp Geogr. 77 (2019) 113–125. https://doi.org/10.1016/J.JTRANGEO.2019.05.005.

[77] OpenStreetMao data for Nord-Est region, (n.d.). http://download.geofabrik.de/europe/italy/nord-est.html# (accessed June 15, 2023).

[78] J. Vuurstaek, G. Cich, L. Knapen, W. Ectors, A.U.H. Yasar, T. Bellemans, D. Janssens, GTFS bus stop mapping to the OSM network, Future Generation Computer Systems. 110 (2020) 393–406. https://doi.org/10.1016/J.FUTURE.2018.02.020.

[79] T. Hagen, J. Hamann, S. Saki, Discretization of Urban Areas using POI-based Tesselation, (2022) 26. https://doi.org/https://doi.org/10.48718/7jjr-1c66.

[80] M. Gahegan, I. Lee, Data structures and algorithms to support interactive spatial analysis using dynamic Voronoi diagrams, Comput Environ Urban Syst. 24 (2000) 509–537. https://doi.org/10.1016/S0198-9715(00)00007-7.

[81] K. Adhinugraha, D. Taniar, T. Phan, R. Beare, Predicting travel time within catchment area using Time Travel Voronoi Diagram (TTVD) and crowdsource map features, Inf Process Manag. 59 (2022) 102922. https://doi.org/10.1016/J.IPM.2022.102922.

[82] S. Park, T.M. Oshan, A. El Ali, A. Finamore, Are we breaking bubbles as we move? Using a large sample to explore the relationship between urban mobility and segregation, Comput Environ Urban Syst. 86 (2021) 101585. https://doi.org/10.1016/J.COMPENVURBSYS.2020.101585.

[83] D. Reddy, P.K. Jana, Initialization for K-means Clustering using Voronoi Diagram, Procedia Technology. 4 (2012) 395–400. https://doi.org/10.1016/J.PROTCY.2012.05.061.

[84] G. Yuan, P. Sun, J. Zhao, D. Li, C. Wang, A review of moving object trajectory clustering algorithms, Artif Intell Rev. 47 (2017) 123–144. https://doi.org/10.1007/s10462-016-9477-7.

[85] M.D. and V.C.G. Wilson Andrew T. and Rintoul, Exploratory Trajectory Clustering with Distance Geometry, in: C.M. Schmorrow Dylan D. and Fidopiastis (Ed.), Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience, Springer International Publishing, Cham, 2016: pp. 263–274.

[86] Y. Chen, X. Zhang, C. Wu, G. Li, Field-road trajectory segmentation for agricultural machinery based on direction distribution, Comput Electron Agric. 186 (2021) 106180. https://doi.org/10.1016/J.COMPAG.2021.106180.

[87] M. Mieczynska, I. Czarnowski, DBSCAN algorithm for AIS data reconstruction, Procedia Comput Sci. 192 (2021) 2512–2521. https://doi.org/10.1016/J.PROCS.2021.09.020.

[88] O. Filtser, Universal approximate simplification under the discrete Fréchet distance, Inf Process Lett. 132 (2018) 22–27. https://doi.org/10.1016/J.IPL.2017.10.002.

[89] K. Buchin, M. Löffler, T. Ophelders, A. Popov, J. Urhausen, K. Verbeek, Computing the Fréchet distance between uncertain curves in one dimension - ScienceDirect, Computational Geometry. 109 (2023). https://www.sciencedirect.com/science/article/pii/S0925772122000669?via%3Dihub (accessed November 24, 2023).

[90] M. de Berg, A.F. Cook, J. Gudmundsson, Fast Fréchet queries, Computational Geometry. 46 (2013) 747–755. https://doi.org/10.1016/J.COMGEO.2012.11.006.

[91] Neighborhood Guide, Https://Bolognainside.Iwfbologna.Com/Housing/Neighborhood-Guide/. (n.d.). https://bolognainside.iwfbologna.com/housing/neighborhood-guide/ (accessed October 31, 2023).

[92] G. Boeing, OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, Comput Environ Urban Syst. 65 (2017) 126–139. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

[93] Highway, (n.d.). https://wiki.openstreetmap.org/wiki/Key:highway (accessed November 6, 2023).

[94] Tag:highway=primary, (n.d.). https://wiki.openstreetmap.org/wiki/Tag:highway%3Dprimary (accessed November 6, 2023).

[95] Bicycle, (n.d.). https://wiki.openstreetmap.org/wiki/Bicycle (accessed November 10, 2023).

[96] Cycleway, (n.d.). https://wiki.openstreetmap.org/wiki/Key:cycleway (accessed November 10, 2023).

7. Appendix

I. Results of trajectory analysis for 20 most used flows

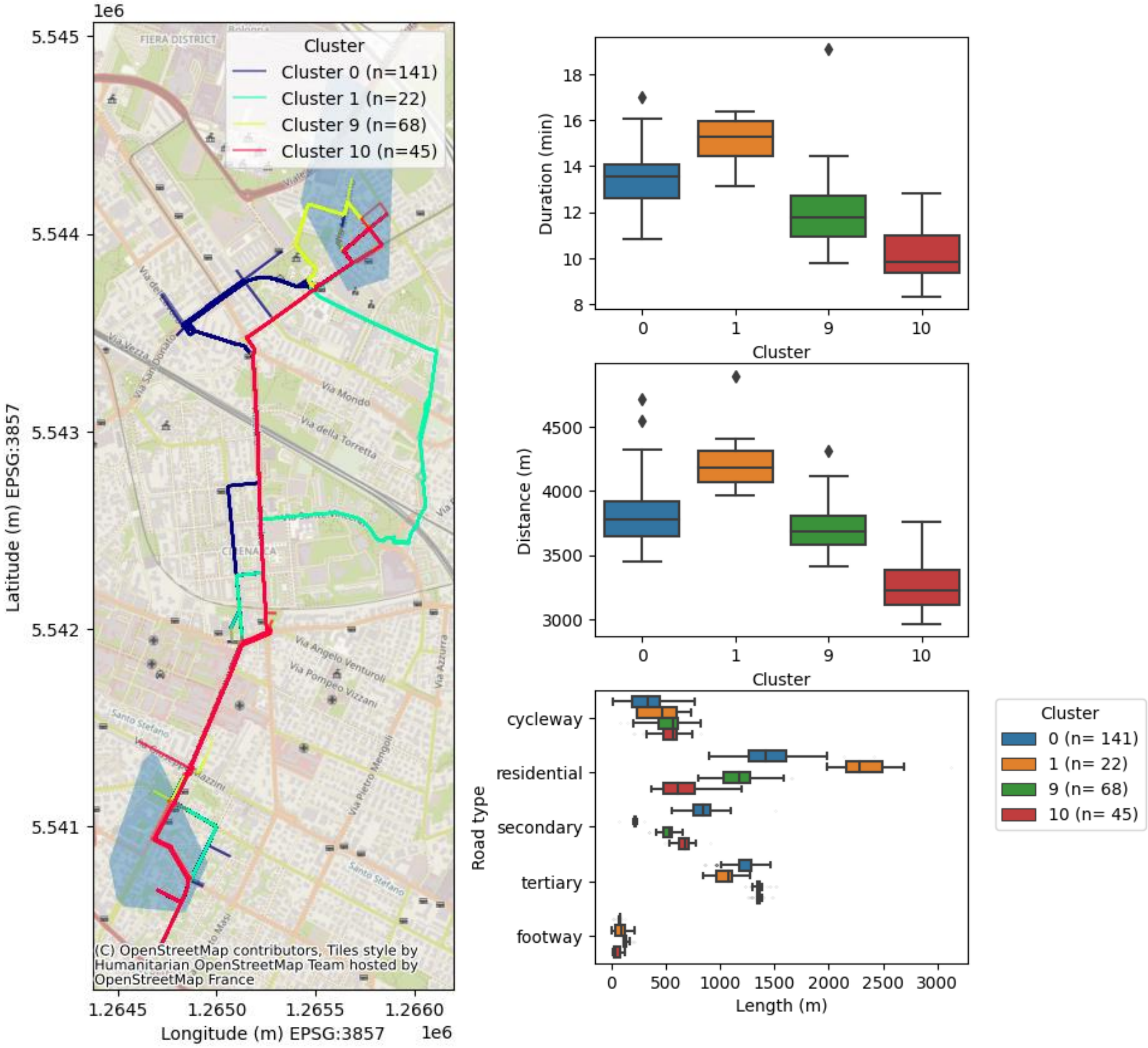

Figure 15. Trajectory analysis between tiles voronoiID138 and voronoiID213 (n= 293). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

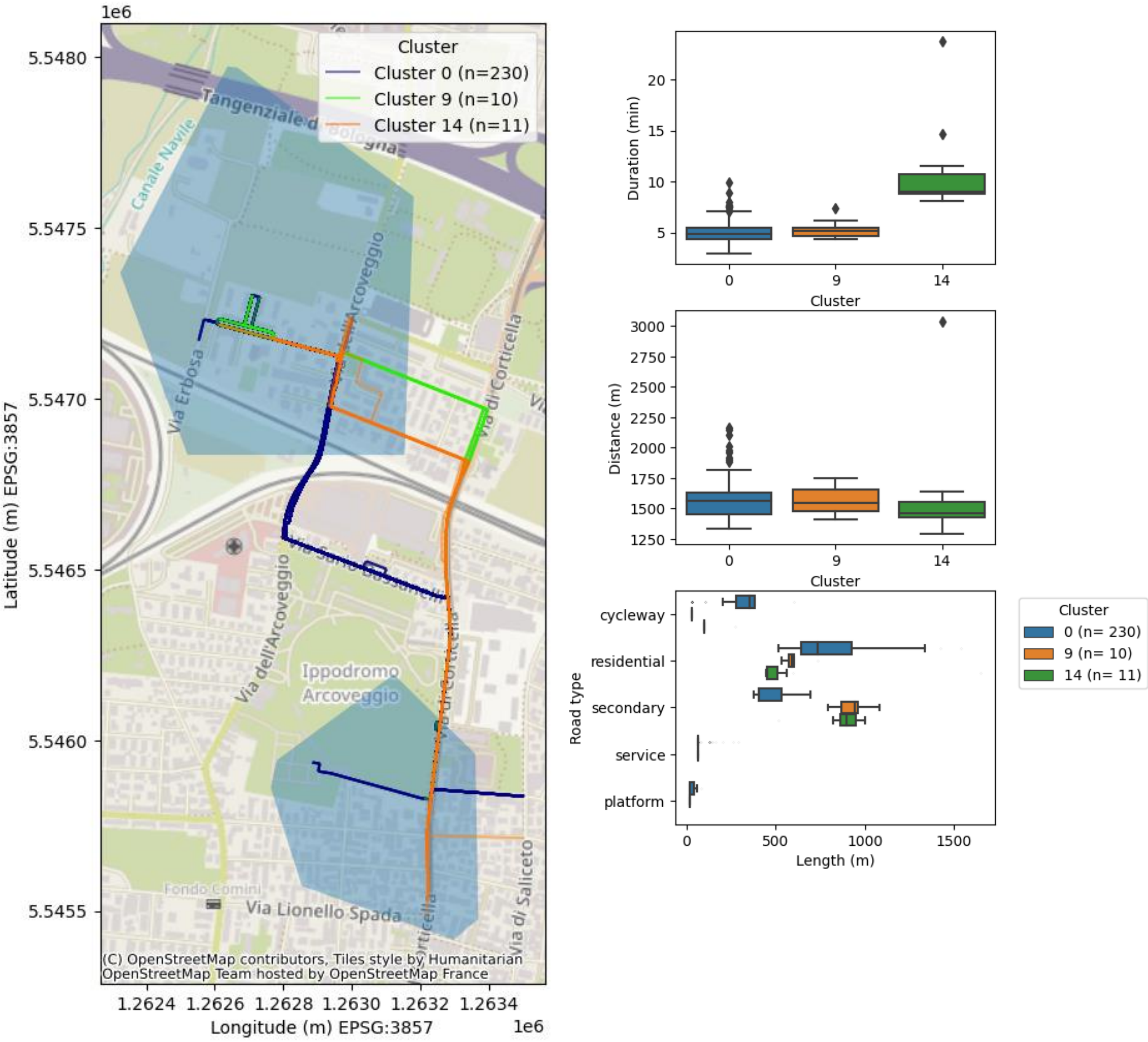

Figure 16. Trajectory analysis between tiles voronoiID37 and voronoiID142 (n= 279). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

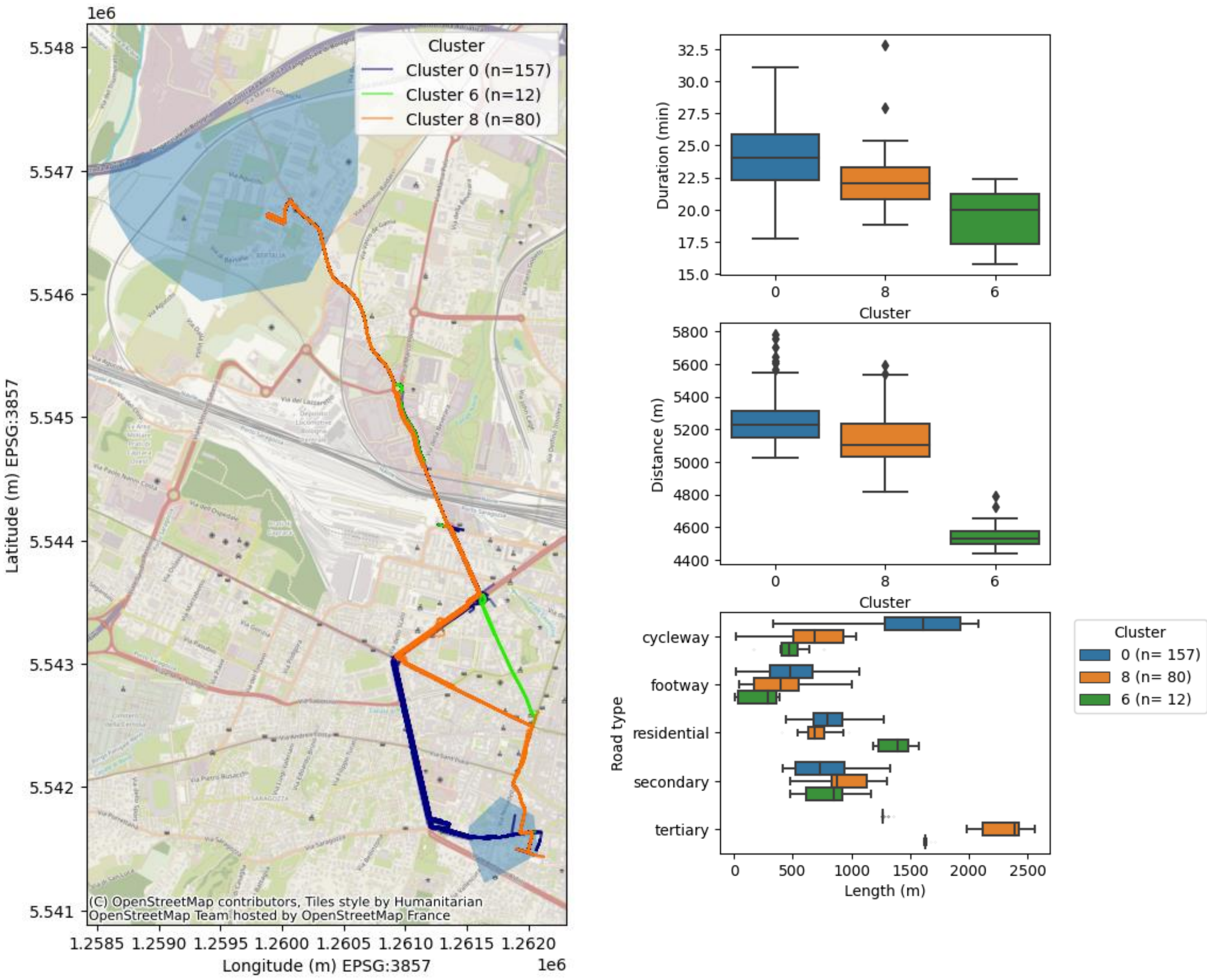

Figure 17. Trajectory analysis between tiles voronoiID166 and voronoiID124 (n= 273). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

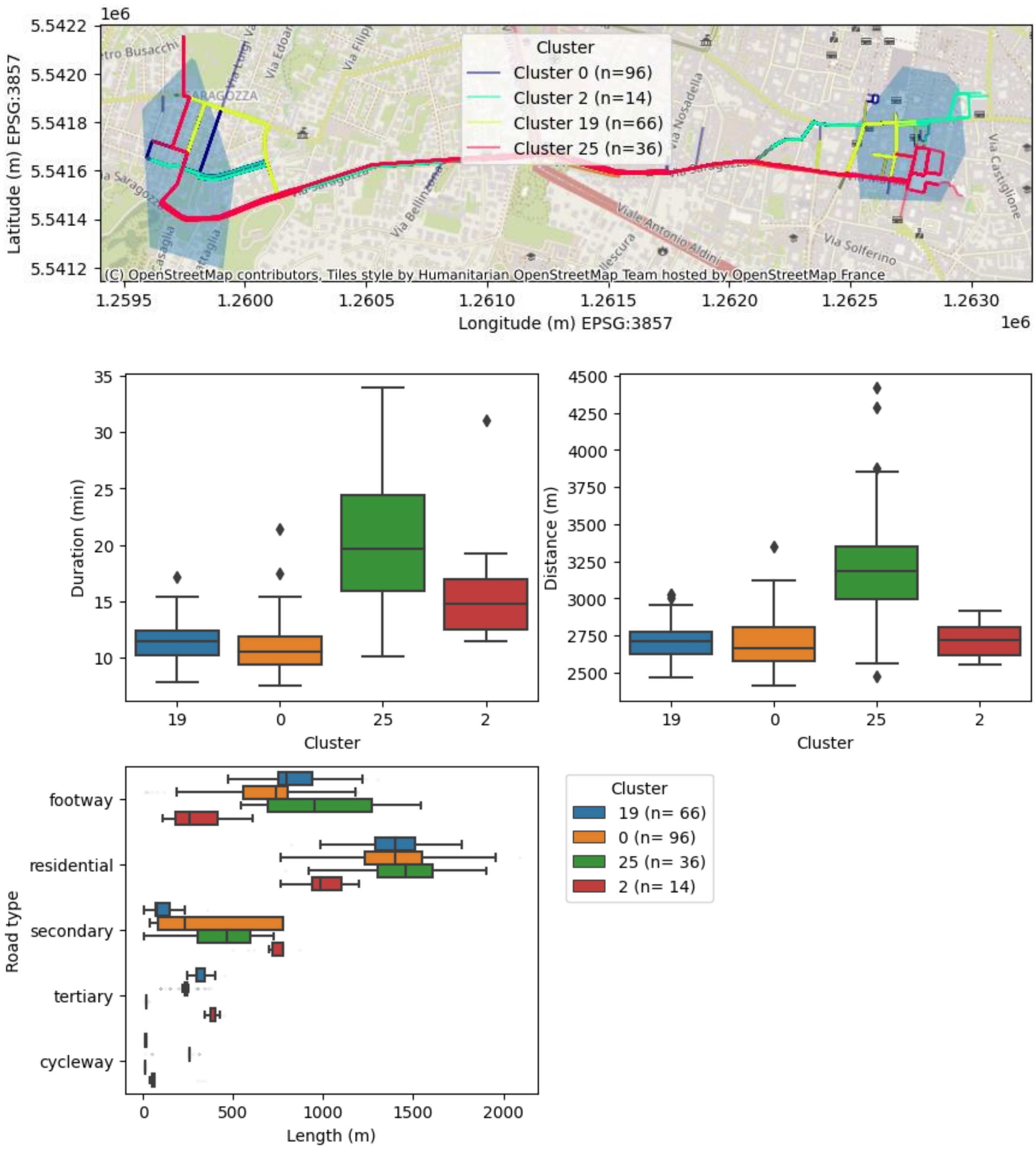

Figure 18. Trajectory analysis between tiles voronoiID161 and voronoiID154 (n= 267). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

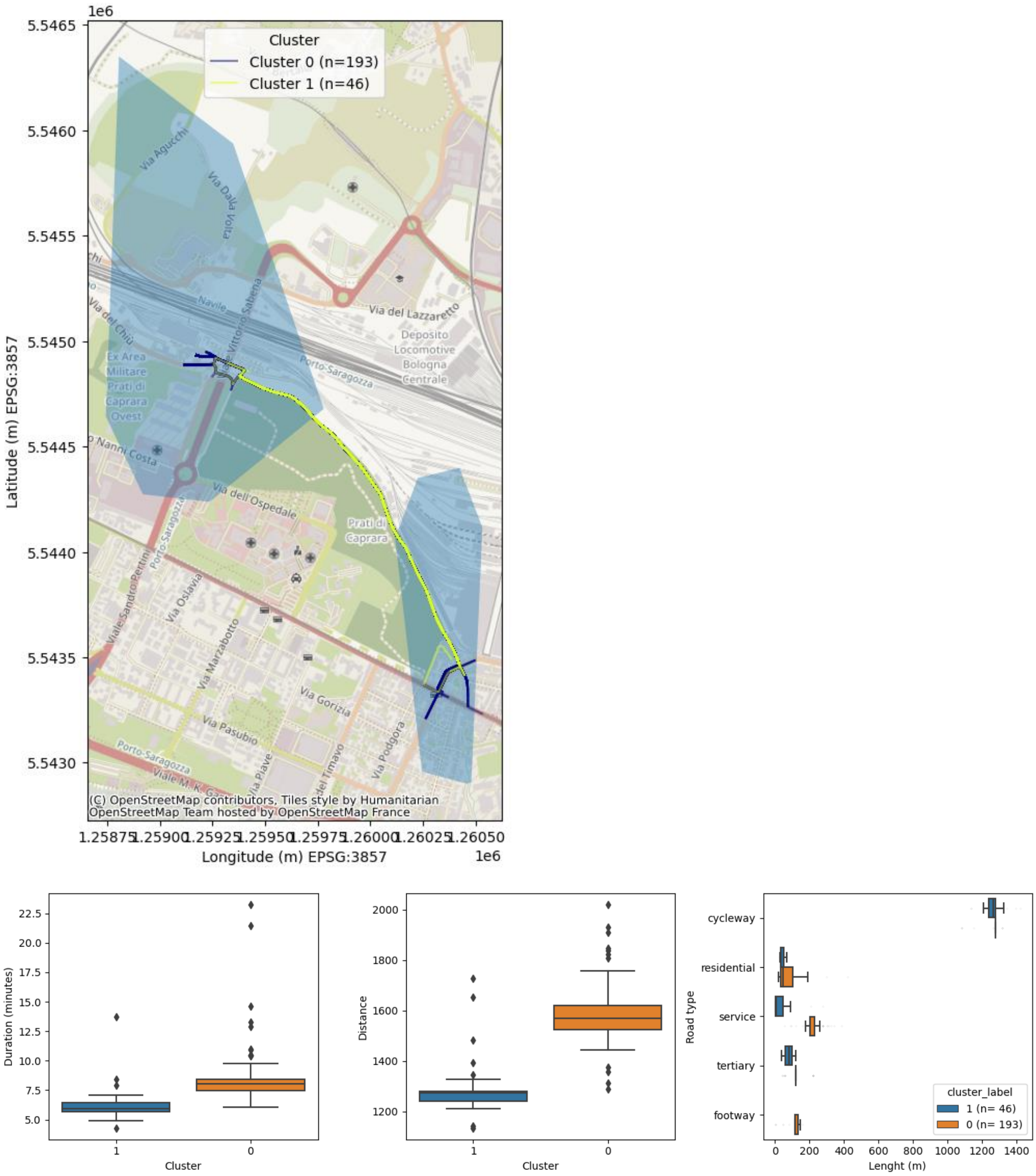

Figure 19. Trajectory analysis between tiles voronoiID76 and voronoiID201 (n= 254). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

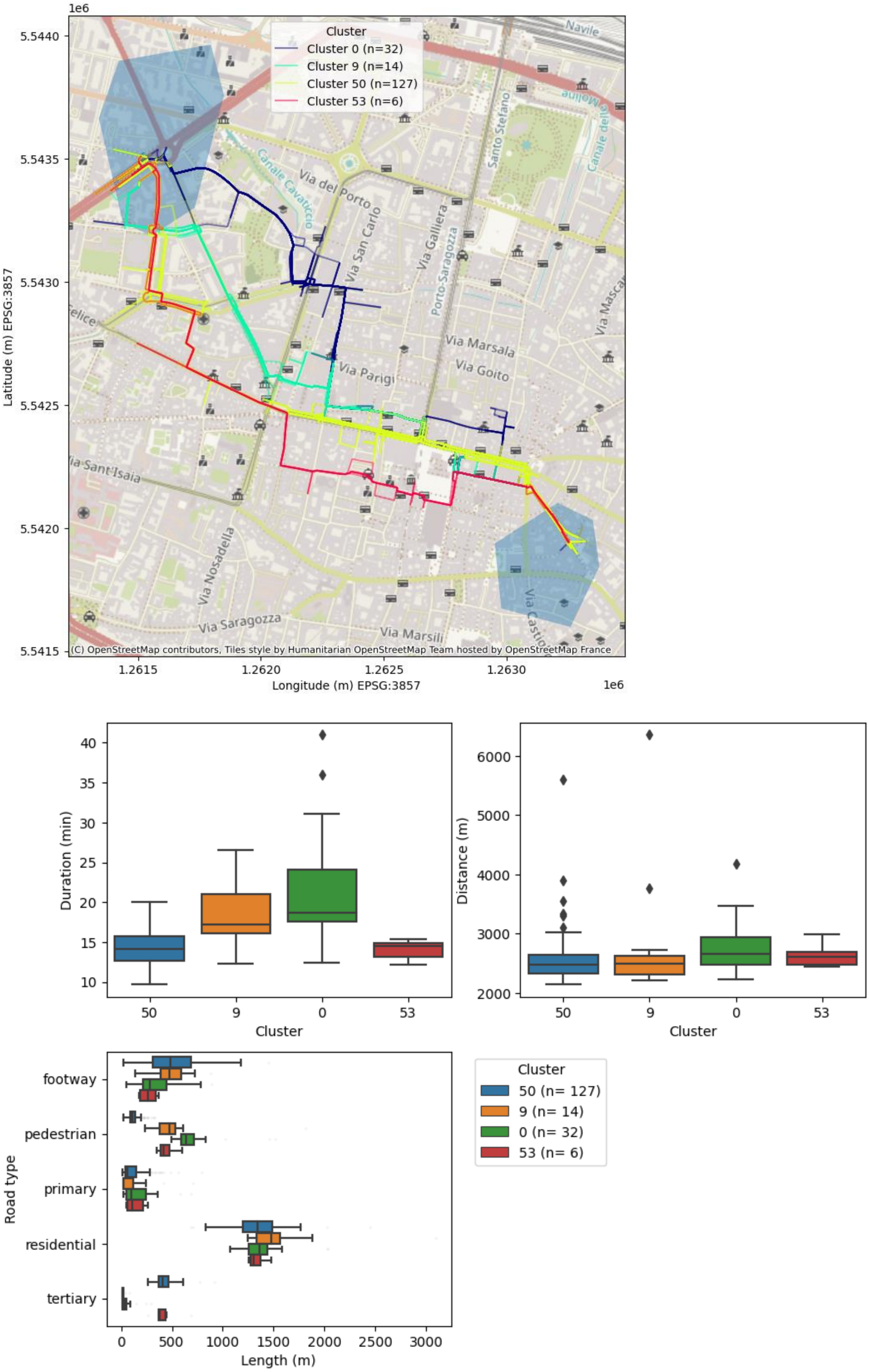

Figure 20. Trajectory analysis between tiles voronoiID44 and voronoiID97 (n= 250). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

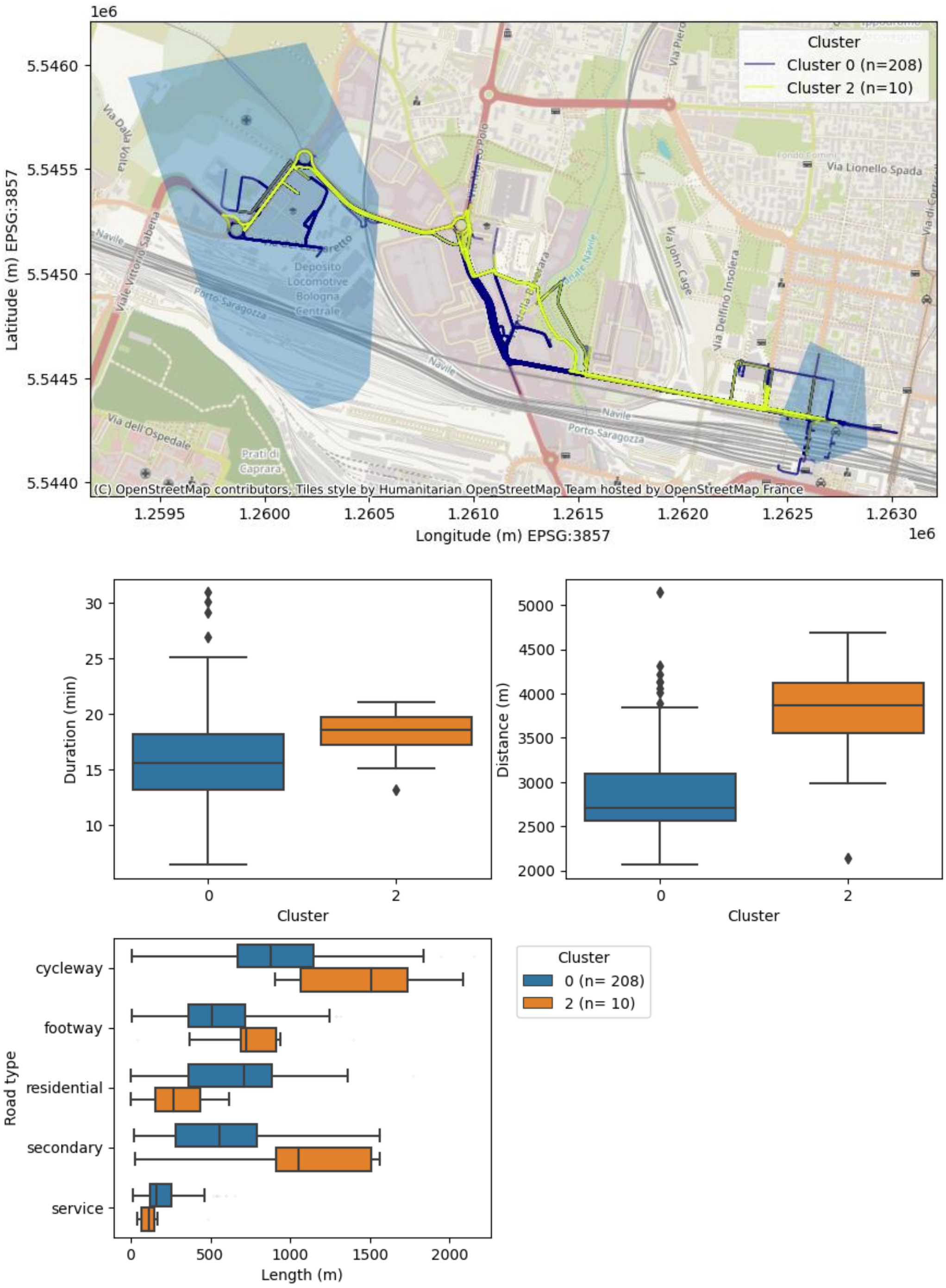

Figure 21. Trajectory analysis between tiles voronoiID297 and voronoiID25 (n= 238). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

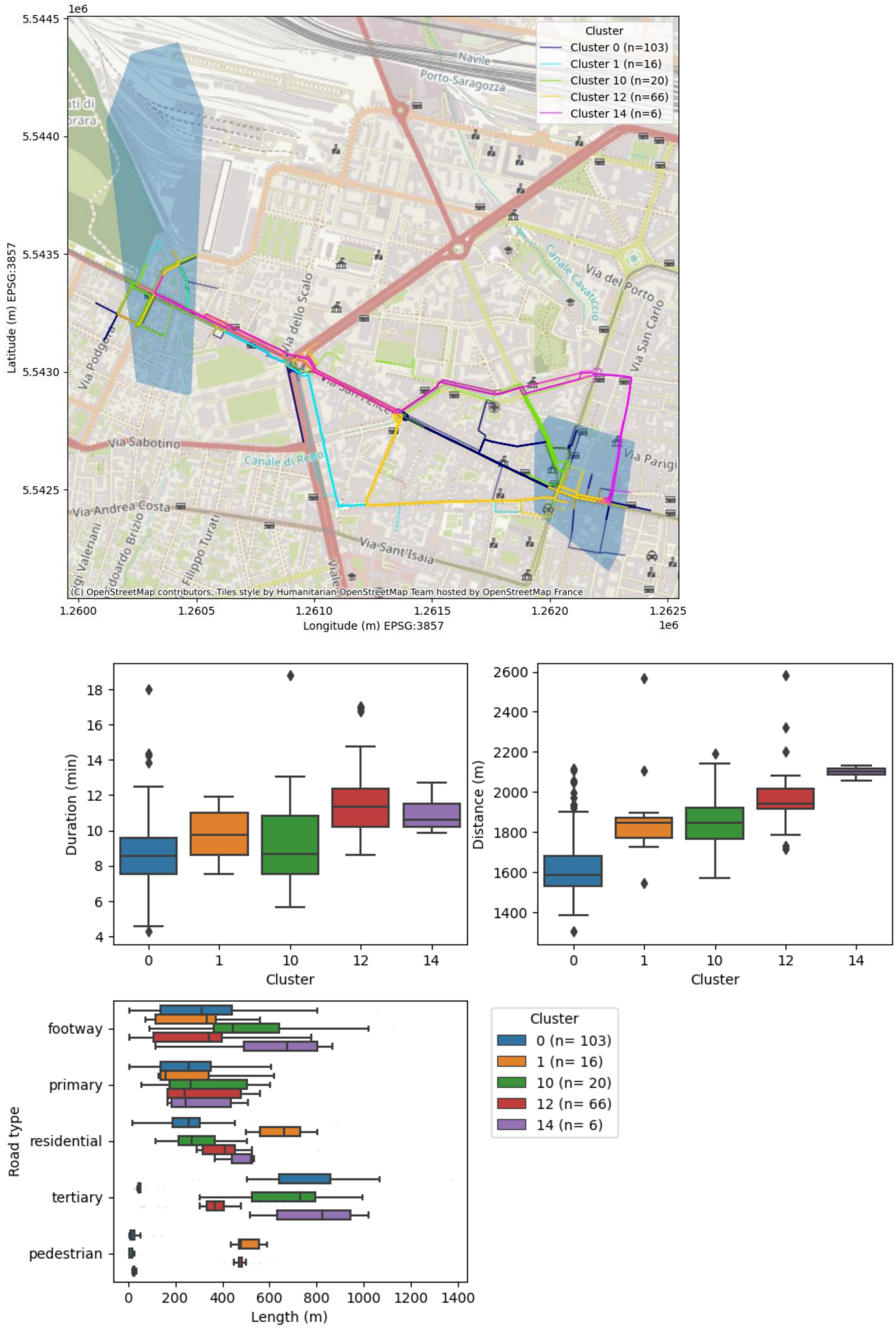

Figure 22. Trajectory analysis between tiles voronoiID76 and voronoiID181 (n= 237). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

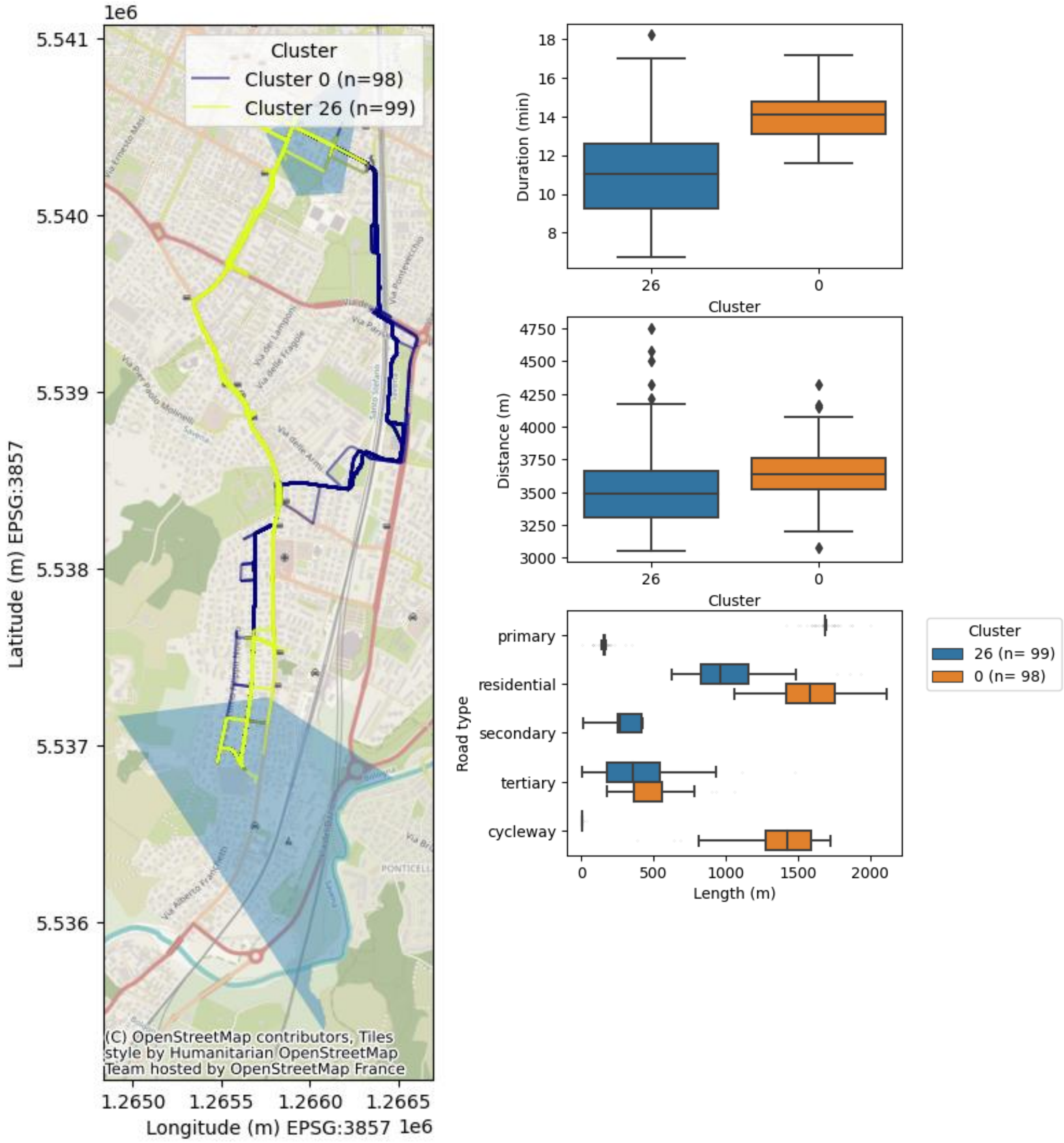

Figure 23. Trajectory analysis between tiles voronoiID60 and voronoiID89 (n= 236). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

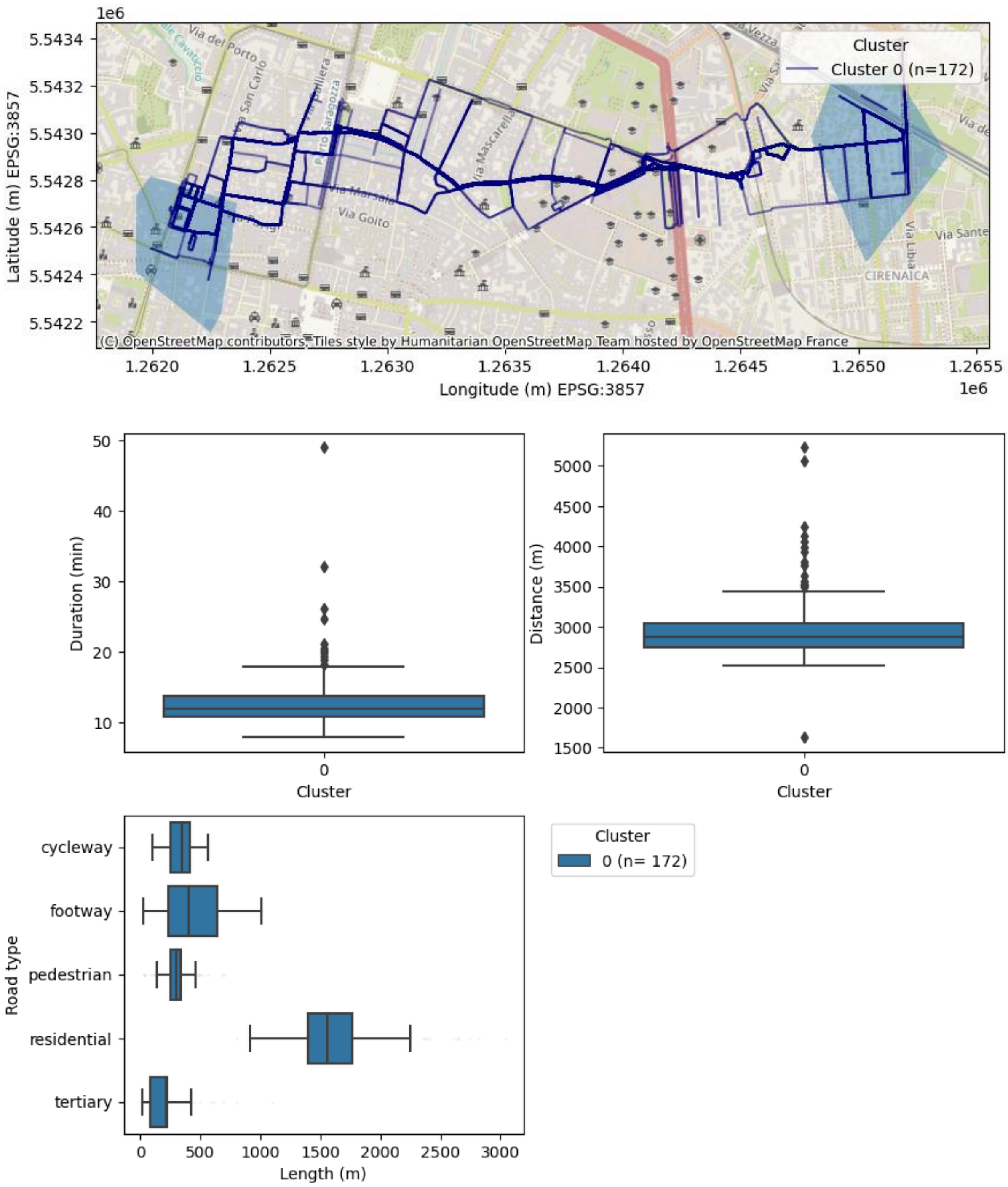

Figure 24. Trajectory analysis between tiles voronoiID210 and voronoiID181 (n= 230). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

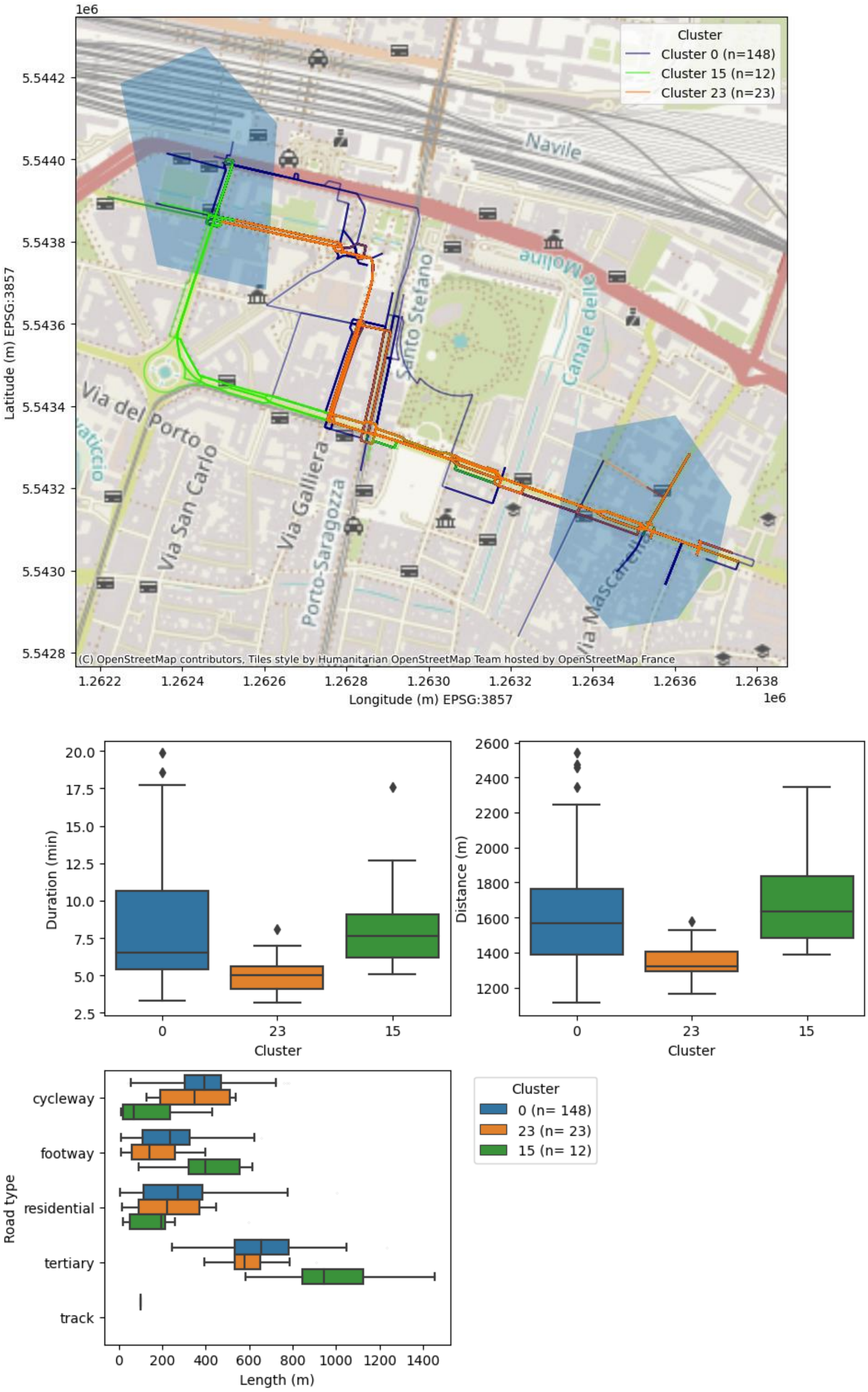

Figure 25. Trajectory analysis between tiles voronoiID62 and voronoiID185 (n= 225). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

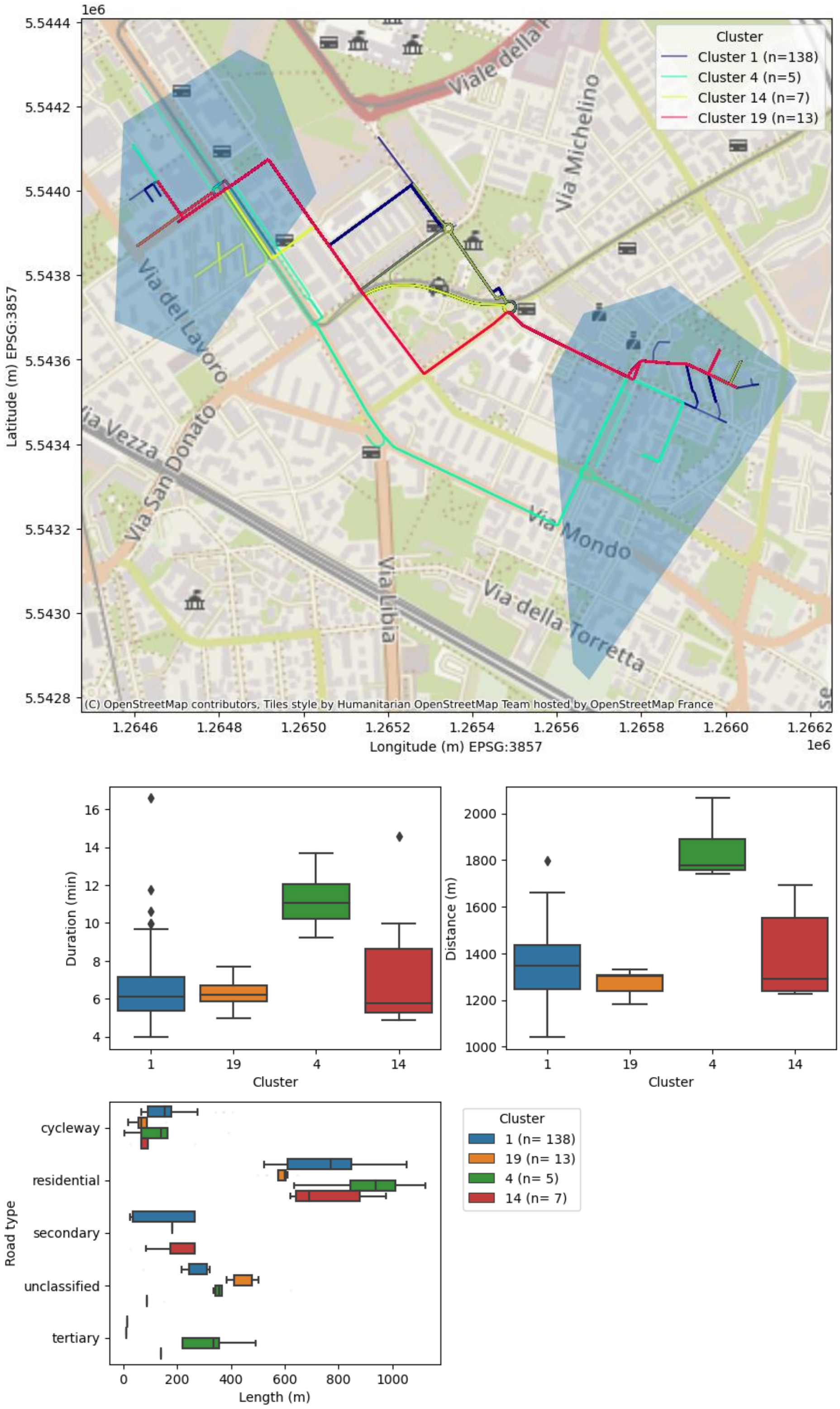

Figure 26. Trajectory analysis between tiles voronoiID204 and voronoiID193 (n= 218). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

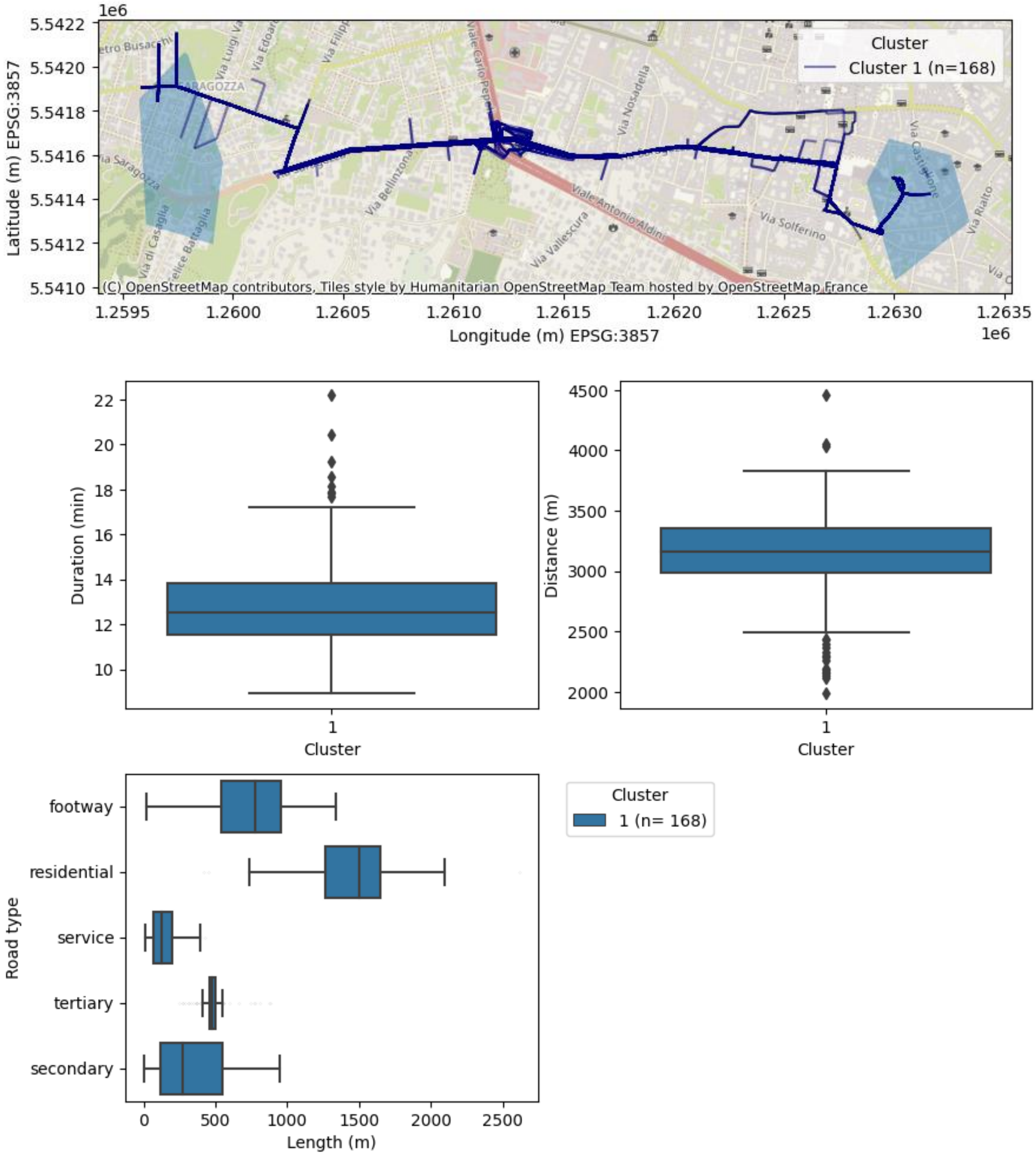

Figure 27. Trajectory analysis between tiles voronoiID18 and voronoiID154 (n= 209). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

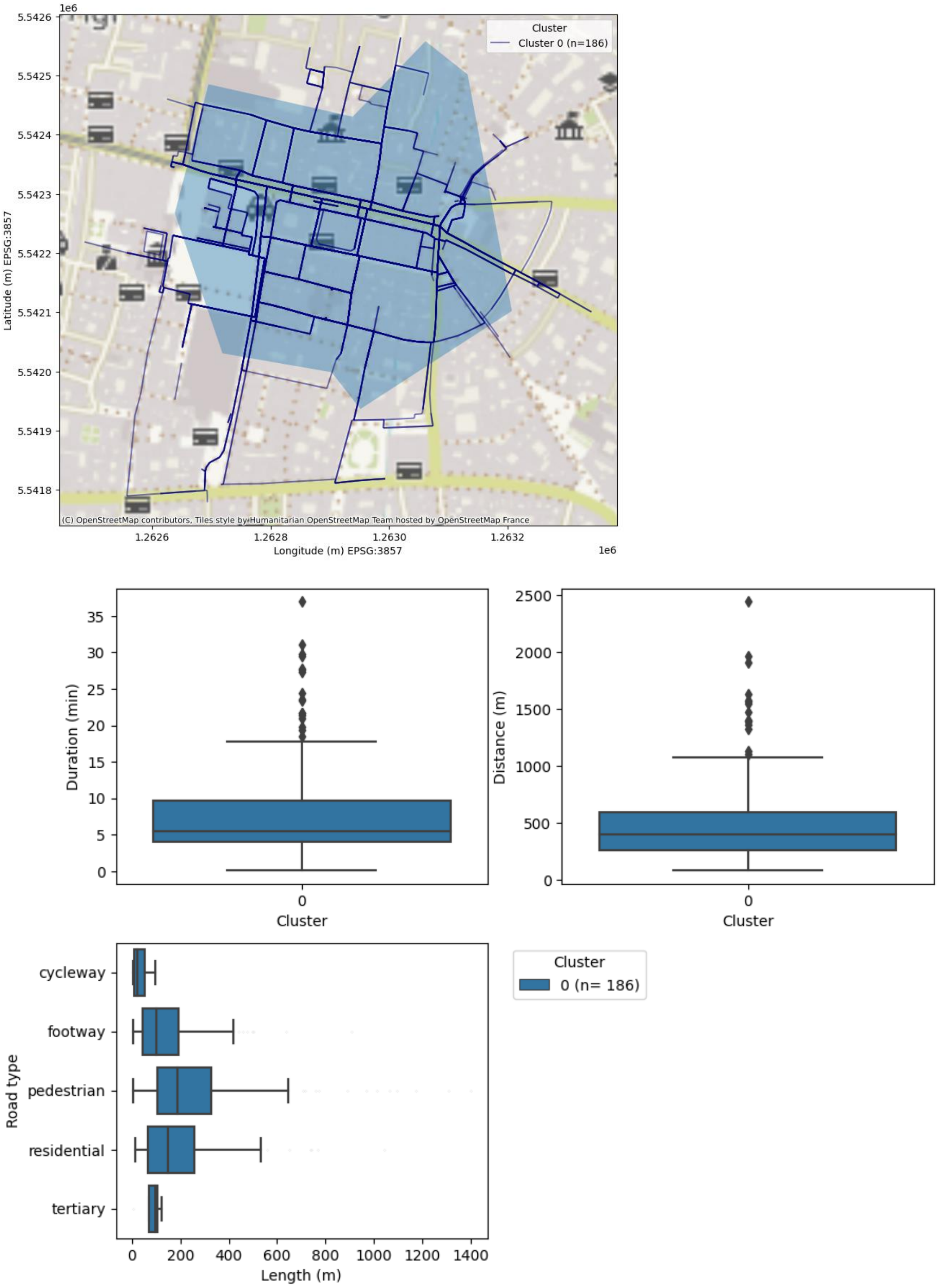

Figure 28. Trajectory analysis between tiles voronoiID68 and voronoiID120 (n= 207). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

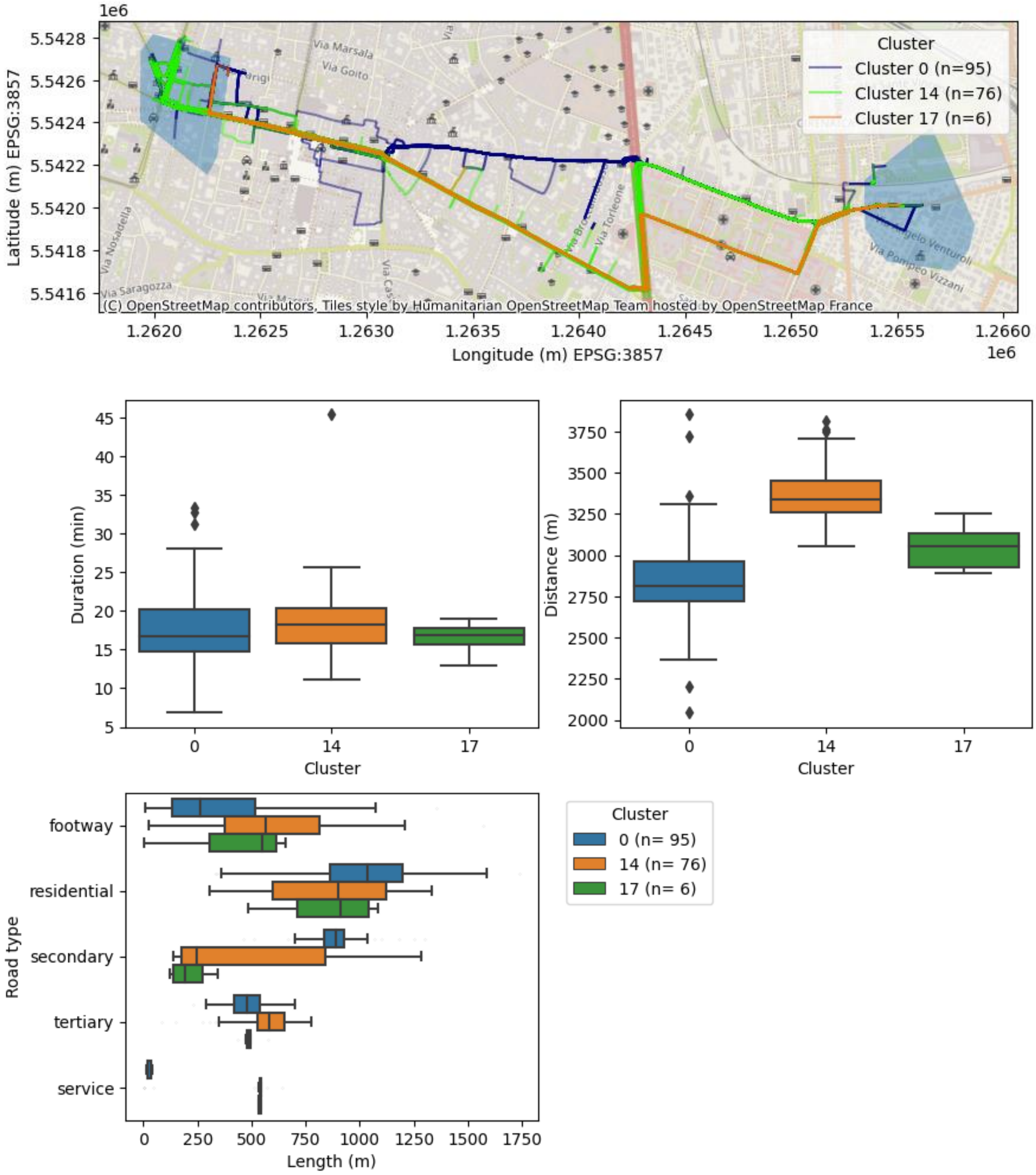

Figure 29. Trajectory analysis between tiles voronoiID239 and voronoiID181 (n= 207). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

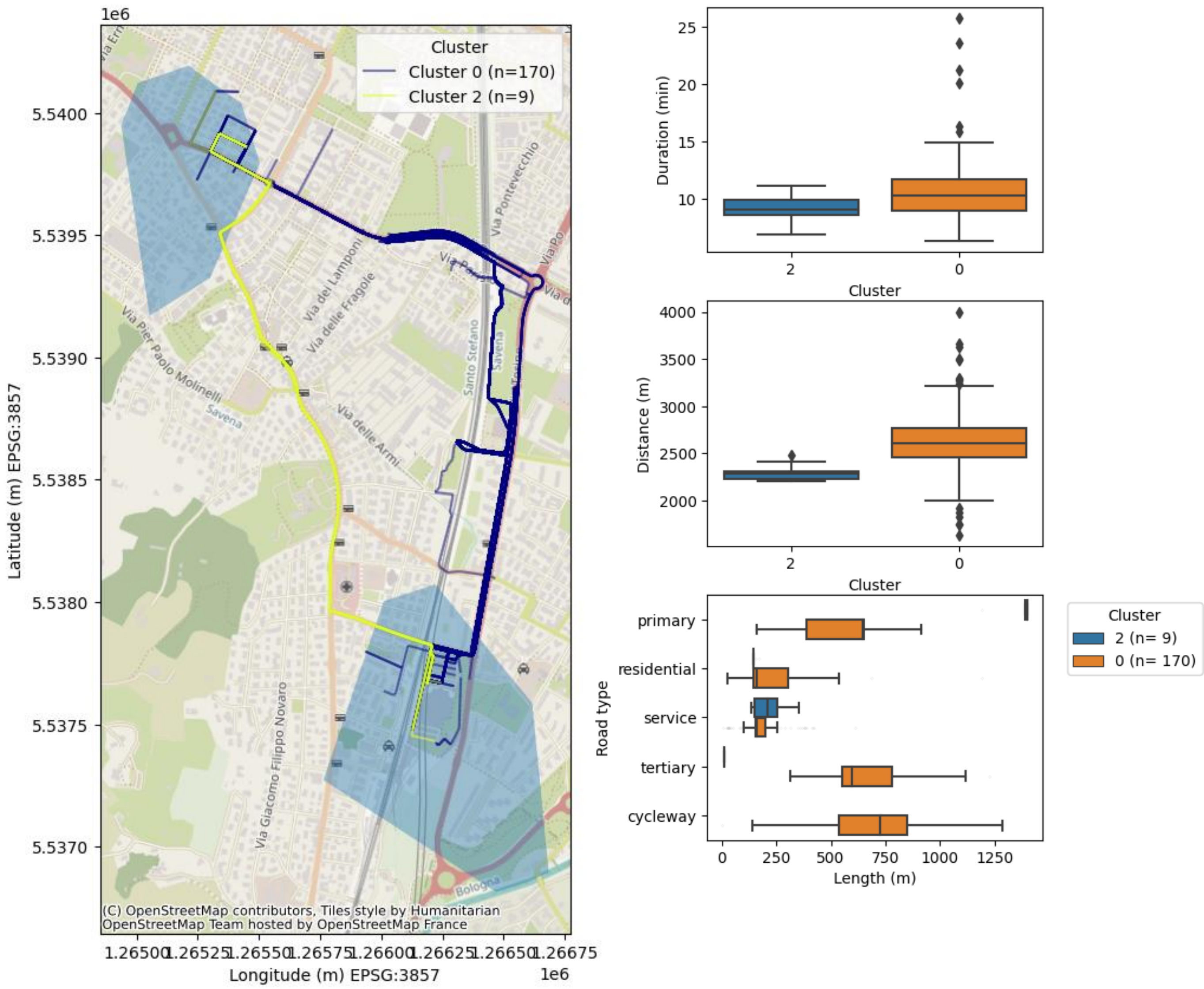

Figure 30. Trajectory analysis between tiles voronoiID26 and voronoiID136 (n= 200). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

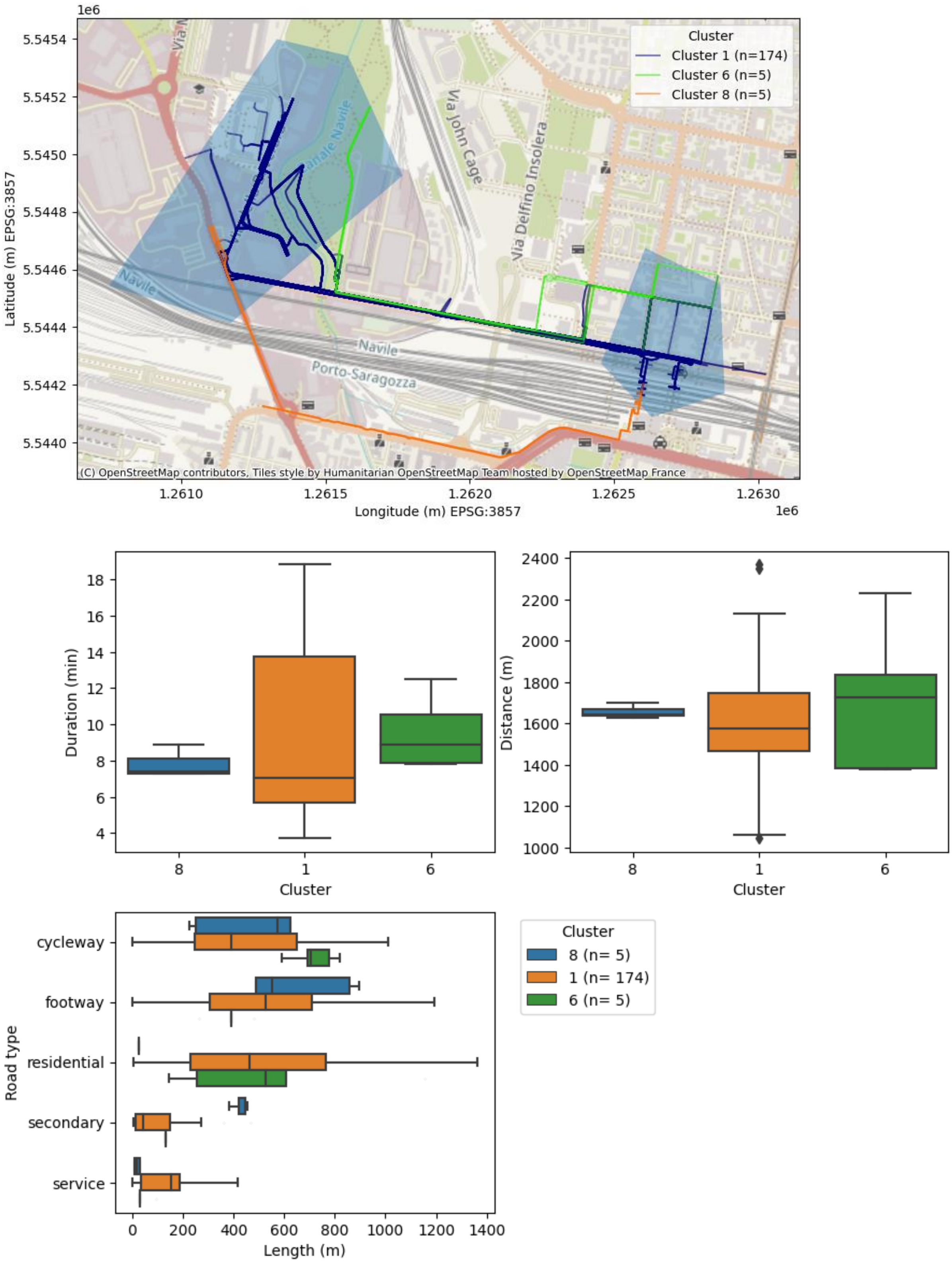

Figure 31. Trajectory analysis between tiles voronoiID246 and voronoiID297 (n= 200). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

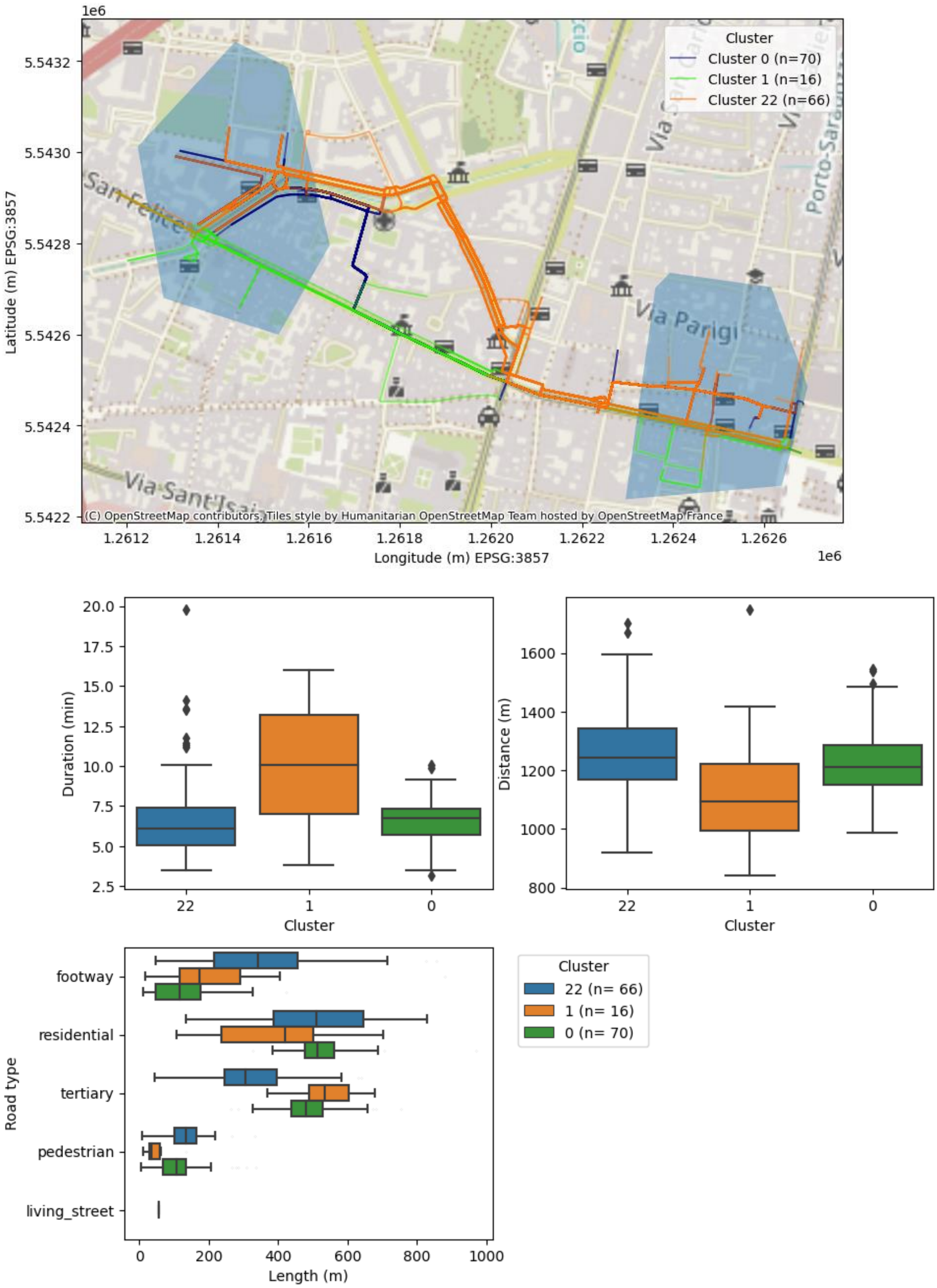

Figure 32. Trajectory analysis between tiles voronoiID248 and voronoiID6 (n=198). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

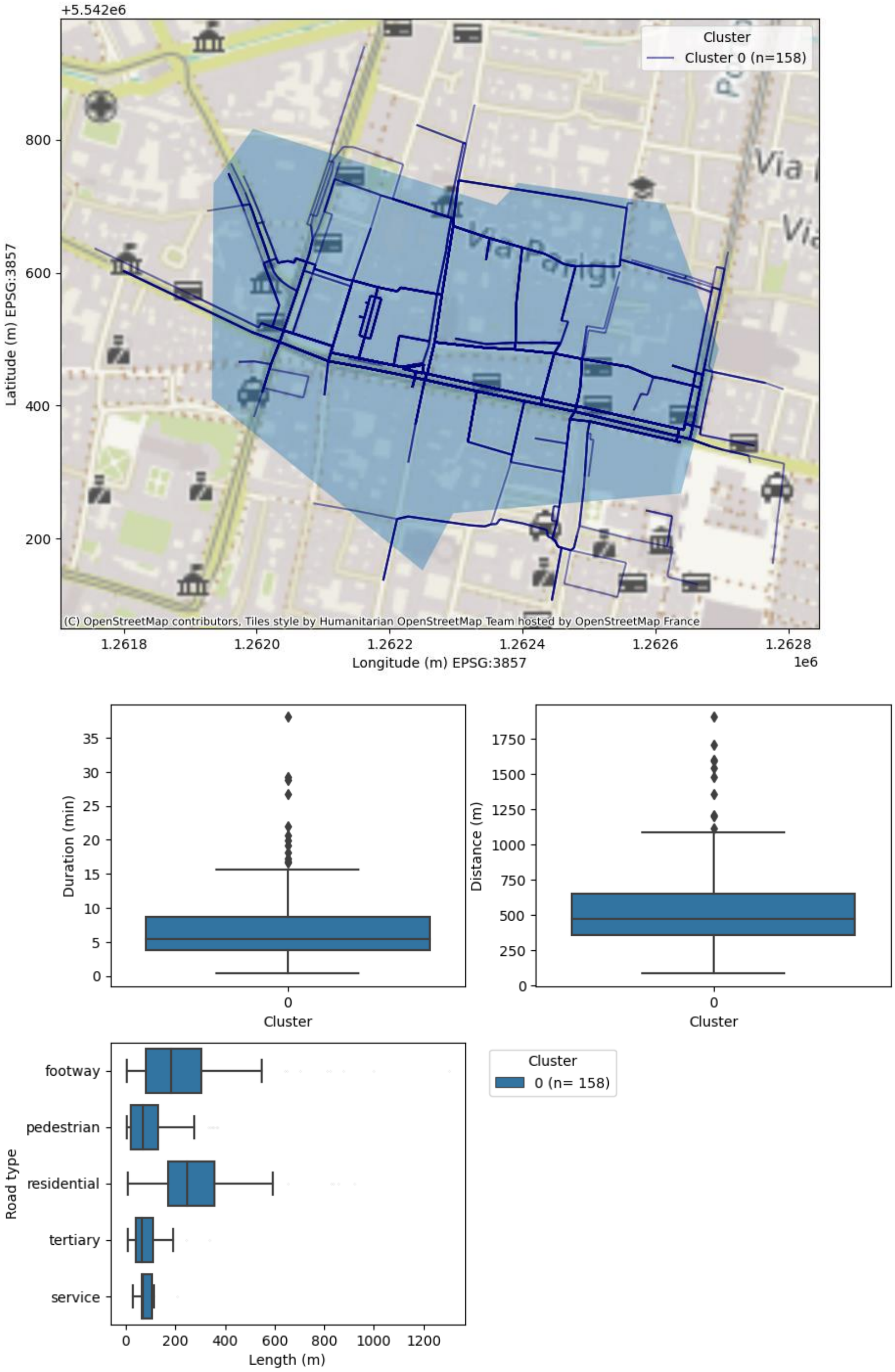

Figure 33. Trajectory analysis between tiles voronoiID248 and voronoiID6 (n=198). Upper figure illustrates different routes between tiles and bottom figures present the road type analysis results.

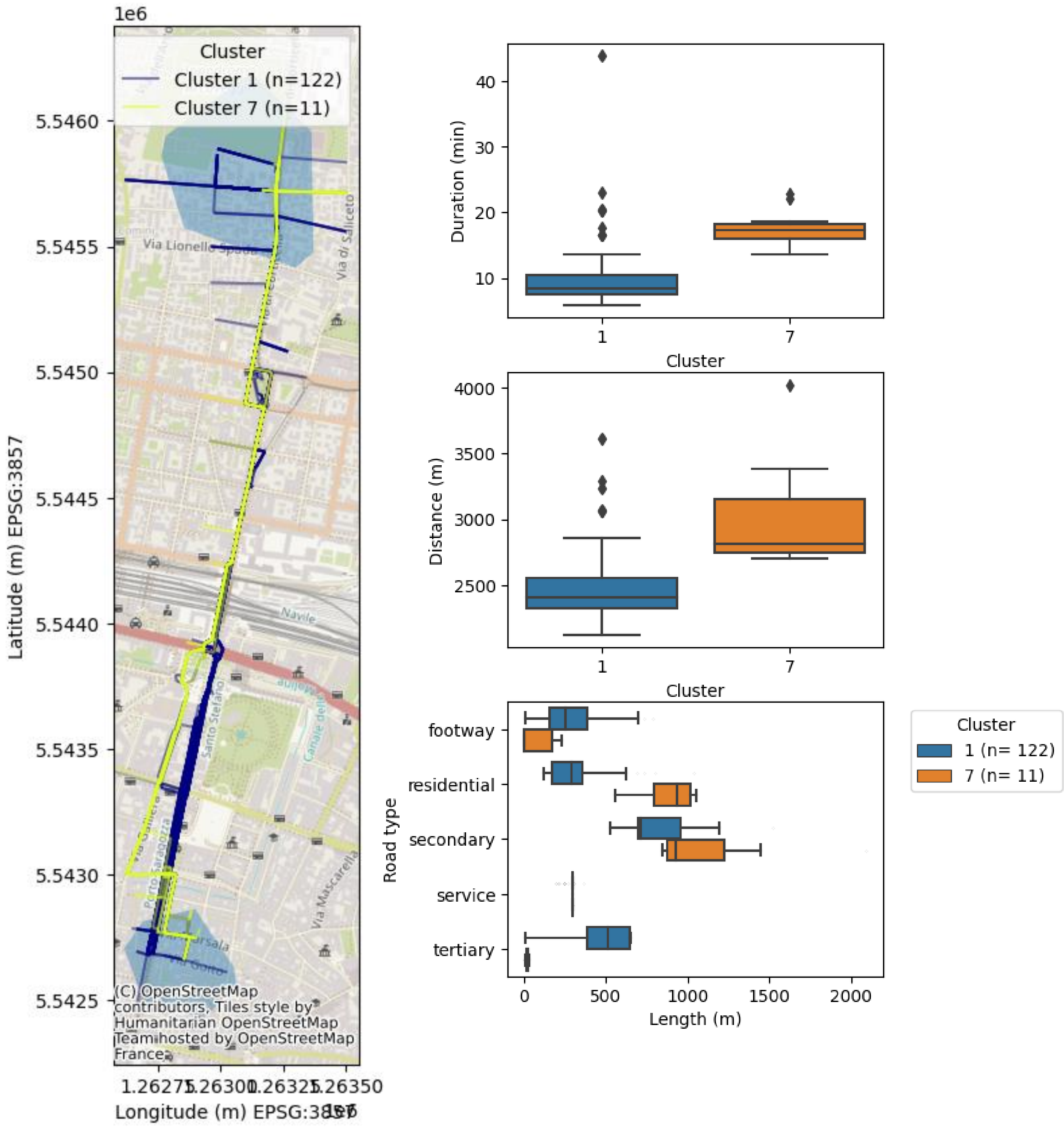

Figure 34. Trajectory analysis between tiles voronoiID142 and voronoiID222 (n=184). Left figure illustrates different routes between tiles and right figures present the road type analysis results.

II License

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Merlin Raud,

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

The influence of infrastructure to cyclist route choice in Bologna,

supervised by Rajesh Sharma, Flavio Bertini, Amnir Hadachi

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Merlin Raud

4.01.2024