Politecnico di Torino

Master's degree in civil engineering – Infrastructures and transportation systems
Academic year 2022/2023

Matching public transport offer and travel demand during night-time through open-source data

Supervisor:
Prof. Diana Marco

Candidate:
Benvenuti Giacomo Maria

Abstract

This thesis considers how the travel demand is matched by the public transport offer during the night time of weekends, developing a framework which can be used in any European city (and not), using open source data available online. The composition of the framework and its implementation is explained and the sources on which it is based are shown. Finally, the results of the framework on the eight cities are exposed, and correlation between the considered cities and the parameters involved are presented.

We collected data related to public transport lines in GTFS format from the related local agencies or from third part web data collectors. Population distributions were downloaded from the databases GHS-UCDB R2019A, which helped us delimitate the study area through the definition of Functional Urban Areas and Urban centres, and Urban Atlas 2018, which we used to actually populate the cities. We then downloaded information about the night time related amenities in each city trough Overpassturbo. The platform used for data analysis and treatment is QGIS, where downloaded data related to population distribution, night-life attractions and public transport services were displayed and correlated to each other.

We identified the locations within each city where the nightlife takes place (Night Life Areas, NLAs): to do so we clustered the night time related amenities with the DBSCAN algorithm, which is a clustering algorithm sensible to density. Later, using the plugin Traveltime, we computed how many inhabitants can be reached departing from these points using public transport for their movements, considering two different time thresholds of 30 and 45 minutes at nine different departure times spread over two hours studied time span. The number of inhabitants reached from the different NLAs, when related to the population inside a buffer area around each NLA to be considered as a neighbourhood, provide a service coverage percentage. We also considered another set of two values, which are the average number of inhabitants which are served by the network from at least one NLA considering each departure time separately and the number of inhabitants served at least by one NLA at least one departure time. The results of the percentages of inhabitants served by the night-time public service were then correlated numerically and graphically to other factors such as the public transport offer (deduced from the length of the night-line network and its lines' frequency) and the number of inhabitants in the study area.

The framework has been studied on 8 different European cities which were chosen to form a sufficiently heterogenous set related to the country they are placed in, their geographical position, their historical origins and therefore their size and structure; these cities are Budapest, Milano, München, Praha, Roma, Torino, Valencia and Wien.

Abstract (Italiano)

La presente tesi esamina come la domanda di viaggio si concilia con l'offerta di trasporto pubblico durante le ore notturne dei fine settimana, sviluppando un framework che può essere utilizzato in qualsiasi città europea (e non), utilizzando dati open source disponibili online. Dopo un'introduzione della composizione del framework e la sua implementazione, vengono mostrate le fonti open source a cui attinge. Infine, vengono esposti i risultati del framework sulle otto città e vengono presentate le correlazioni tra le città considerate e i parametri coinvolti.

Abbiamo raccolto i dati relativi alle linee di trasporto pubblico in formato GTFS dalle relative agenzie locali o da database web di parti terze. Le distribuzioni di popolazione sono state scaricate dai database GHS-UCDB R2019A (con il quale abbiamo delimitato l'area di studio seguendo la definizione di Functional Urban Areas e centri urbani) e Urban Atlas 2018 che abbiamo utilizzato per popolare numericamente le città. Abbiamo poi scaricato le informazioni sulle amenities riferite alle ore notturne di ogni città attraverso Overpassturbo. La piattaforma che abbiamo utilizzato per il trattamento dei dati è QGIS, dove i dati scaricati relativi alla distribuzione della popolazione, alle attrazioni della vita notturna e ai servizi di trasporto pubblico sono stati visualizzati e correlati tra loro.

Abbiamo identificato i luoghi all'interno di ogni città in cui si svolge la vita notturna (Night Life Areas, NLAs): per farlo, abbiamo raggruppato le attrazioni notturne con l'algoritmo DBSCAN, che è un algoritmo di clustering sensibile alla densità. Successivamente, utilizzando il plugin Traveltime, abbiamo calcolato quanti abitanti possono essere raggiunti partendo da questi punti utilizzando il trasporto pubblico, considerando due diverse soglie temporali di 30 e 45 minuti a nove diversi orari di partenza distribuiti su un arco di tempo di due ore. Il numero di abitanti raggiunti da diversi NLA fornisce, se rapportato alla popolazione all'interno di un'area intorno a ciascun NLA, una percentuale di copertura del servizio. Abbiamo considerato anche un'altra serie di due valori, che sono il numero medio di abitanti serviti da almeno un NLA considerando separatamente ogni orario di partenza e il numero di abitanti serviti da almeno un NLA ad almeno un orario di partenza. I risultati delle percentuali di abitanti serviti dal servizio pubblico notturno sono stati poi correlati numericamente e graficamente ad altri fattori quali l'offerta di trasporto pubblico (dedotta dalla lunghezza della rete di linee notturne e dalla frequenza delle linee) e il numero di abitanti dell'area di studio.

Il framework è stato studiato su 8 diverse città europee che sono state scelte in base a criteri eterogeneità legati al Paese in cui si trovano, alla loro posizione geografica, alle origini storiche e quindi alle loro dimensioni e strutture; queste città sono Budapest, Milano, München, Praha, Roma, Torino, Valencia e Wien.

List of Figures

Figure 1- Studied cities delimited by the Urban Centres borders with information about number of inhabitants and their distribution. a) Budapest, b) Milano, c) München, d) Praha, e) Torino, f) Roma, g) Valencia, h) Wien 28
Figure 2 - Geographical distribution of the studied cities 29
Figure 3 - Diagram of the data model (Davis, 2011) 32
Figure 4 – Add a delimited text layer. 33
Figure 5- Points layer "shapes" obtained by uploading on QGIS the file shapes.txt (Turin) 33
Figure 6 - Points to path 34
Figure 7- Results of Points to path (Torino) 34
Figure 8 - Join layer. 35
Figure 9 - Processing model for Offer Network visualization. 36
Figure 10 - Three steps of the urban centre definition of Toulouse, France (Dijkstra, 2019) 38
Figure 11 - Functional Urban Area of Torino. Each colour represents a different land usage. 40
Figure 12 - Comparison of FUA (blue), Urban Core (Green), Urban Centre (Red) for each studied city: a) Budapest, b) Milano, c) München, d) Praha, e) Roma, f) Torino, g) Valencia, h) Wien 45
Figure 13 – In red the urban centre border is visible; only the blocks falling inside it will be considered. 45
Figure 14 - Extract by location tool and parameters setting. 46
Figure 15 - Building blocks-based data falling inside the urban centre borders. 46
Figure 16 - Centroids tool 46
Figure 17 - Model for the extrapolation of population density concerning the urban centre. 47
Figure 18 - Comparation between population distributions: GHSL 201, Urban Atlas 2018 and Official last updated municipal population 48
Figure 19 - Graph showing the surfaces in km² of each city and their density in inhabitants per km². 49
Figure 20 – Overpass Turbo interface 52
Figure 21 - Interactive map showing the individuated POIs for the city of Budapest 53
Figure 22 – Points of Interest (POIs) on QGIS for the city of Milano 53
Figure 23 - Point of Interest falling inside the Urban Centre of Budapest. Only these points will be considered in the creation of the Clusters. 54
Figure 24 – Different Kernel Shapes compared by Brian Amberg 55
Figure 25 - Parameters set of the Heatmap creation. 56
Figure 26 – Heatmap of Torino a visual impression of how the amenities are distributed in Torino 57
Figure 27 - Explanation of DBSCAN function – Example for Wien 58
Figure 28 - Explanation of how the results change according to the parameters settings – Example for Wien 59
Figure 29 - DBSCAN clustering input example 60
Figure 30 - Minimum bounding geometry input example 61
Figure 31 – Polygons enclosing the POIs forming the different clusters and their centroids (red dots) – Example for the city of Rome 61
Figure 32- Division in clusters of the amenities 61
Figure 33 - Comparation between heatmap of all the amenities points and the clusters boundaries (orange polygons) and their centroids (black stars) for Torino 62
Figure 34 - Time Filter function, input window 66
Figure 35 - Example of how isochrones appear, Torino. 66
Figure 36 - "Buffer" input window; here a 7500 meters radius is inserted, signifying this computation regard a time threshold of 30 min; we did the same one for the 40 min one, inserting a radius of 11250 meters. 67
Figure 37 - "Join attributes by location" tool input window. 68
Figure 38 - "Extract by expressions" has been used to extract those points falling inside an isochrone and a buffer area related to the same NLA; "index_area" is the code meaning the belonging to an isochrone (reached domain), "CLUSTER_ID" indicates the belonging to a buffer area (reachable domain). 68
Figure 39 - Simplified example with 3 NLAs, Torino. 69
Figure 40 - Aggregation over same Night Life Areas 70
Figure 41 – results of the simplified example for two of the 9 departing times only. 70
Figure 42 - Extract by location inputs parameters 71
Figure 43 - Aggregate tool 72
Figure 44 - Result layer of the aggregation 72
Figure 45 - Comparison between networks lengths (the total ones the ones inside the study area) and offer measures. 74
Figure 46 - Visualization of NLAs distribution through clustering and heatmap: a) Budapest, b) Milano, b2) detail from Milano, c) München, d) Praha, e) Torino, f) Roma, f2) detail from Roma, g) Valencia, h) Wien 81

Figure 47 – Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 30 minutes. ____ 86
Figure 48 - Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 45 minutes. ____ 87
Figure 49- Correlation between length of the networks and number of inhabitants in the study area. The dimension of the points is given by the value of service coverage for the 30 minutes time threshold.90
Figure 50- Correlation between length of the networks and number of inhabitants in the study. The dimension of the points is given by the value of service coverage for the 45 minutes time threshold.90
Figure 51- Correlation between service Intensity and service coverage for the 30 minutes time threshold. a) service coverage computed as the mean of all the coverage percentages of each NLA at each departure time considered without weight; b) service coverage is weighted over the number of amenities inside each cluster as explained in 4.3.____ 92
Figure 52- Correlation between service intensity and service coverage for the 45 minutes time threshold. a) service coverage computed as the mean of all the coverage percentages of each NLA at each departure time considered without weight; b) service coverage is weighted over the number of amenities inside each cluster as explained in 4.3____ 93
Figure 53 - Average service coverage for each city for each departure slot for a time threshold of 30minutes; a) with not weighted values, with NLA weighted values. ____ 95
Figure 54 - Average service coverage for each city for each departure slot for a time threshold of 30minutes; a) with not weighted values, with NLA weighted values. ____ 95
Figure 55 - Correlation between offer and percentage of inhabitants served by at least one NLA for the 30 minutes time threshold. a) percentage of inhabitants served at least once computed as an average of the percentages at each departure time (Average Once Served); b) percentage of inhabitants served at least once in the two hours study time (Total Once Served). ____ 98
Figure 56 - Correlation between offer and percentage of inhabitants served by at least one NLA for the 45 minutes time threshold. a) percentage of inhabitants served at least once computed as an average of the percentages at each departure time (Average Once Served); b) percentage of inhabitants served at least once in the two hours study time (Total Once Served). ____ 99
Figure 57 – Average service coverage for each city considering all the NLAs jointly. Time threshold of 30 minutes. ____ 100
Figure 58 - Average service coverage for each city considering all the NLAs jointly. Time threshold of 45minutes. ____101
Figure 59 - Evolution of covered inhabitants for each NLA in Budapest during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 102
Figure 60 - Evolution of covered inhabitants for each NLA in Milano during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 103
Figure 61 - Evolution of covered inhabitants for each NLA in München during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 104
Figure 62 - Evolution of covered inhabitants for each NLA in Praha during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 104
Figure 63 - Evolution of covered inhabitants for each NLA in Roma during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 105
Figure 64 - Evolution of covered inhabitants for each NLA in Torino during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 105
Figure 65 - Evolution of covered inhabitants for each NLA in Valencia during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 106
Figure 66 - Evolution of covered inhabitants for each NLA in Wien during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold. ____ 106

List of Tables

Table 1 – Studies on accessibility, recap with accessibility definition and goal and methodology of the study expressed for each study. ____ 14
Table 2 – Recapitulation of studies about areas identification ____ 17
Table 3 - Resume of articles regarding the concept of Travel Time Budget and its criticalities ____ 19
Table 4 -indicators proposed by (Salado García, Díaz Muñoz, Bosque Sendra, Carvalho Cantergiani, & Rojas Quezada, 2006) ____ 21
Table 5 - Recapitulation of articles regarding coverage rate ____ 22
Table 6 – Sources of GTFS for each studied city ____ 30
Table 7 – Description of files contained in a GTFS format file, (Google, 2022) ____ 31
Table 8 - Service and peak hours for each city ____ 37
Table 9 - Comparison between Municipal official population, population inside the Urban Centre borders for GHSL 2015 and for Urban Atlas 2018. ____ 47
Table 10 - Surface and population density for each Urban Centre; computed with Urban Atlas 2018 data ____ 48
Table 12 - List of selected amenities with their definition (wiki.openstreetmap, 2022) ____ 50
Table 13 - DBSCAN parameters used for each city. ____ 60
Table 14 – Public transport agencies covered in each studied city; ____ 64
Table 15 - Length of night-time public transport lines (km) for the total network and for the study area; offer measures computed as service intensity and service intensity per thousand inhabitants; average frequency for each line operated in the two hours study time; Network intensity. ____ 73
Table 16 – Number of night-time related amenities, the ones being part of a cluster, found clusters and virtual number of inhabitants for each amenity for each city. ____ 82
Table 17 – Collection of data about the weights of each NLA and the number of inhabitants inside the buffer area. Example for Budapest with a 30 minutes time threshold. ____ 83
Table 18 - Example of data treatment with the results of the Traveltime computation. Example for Budapest, with 30 minutes time threshold. ____ 83
Table 19 – Coverage aggregated at city level. Example for Budapest, with 30 minutes time threshold. 84
Table 20 – Inhabitants served at least once for each departure time and percentage over the total population. Example for Budapest, with 30 minutes time threshold. ____ 84
Table 21 – Inhabitants served at least once at any departure time and percentage over the total population. Example for Budapest, with 30 minutes time threshold. ____ 85
Table 22 - Results of the traveltime computation for each city ____ 85
Table 23 – Values of the parameters indicating the percentage of inhabitants served at least once for both the two given definitions and the two considered time thresholds. Followed by the variation between one definition and the other for both time thresholds (Variation - 30 and Variation – 45) and the variation between the two time thresholds following the same definition(Variation – Av and Variation – Tot). The conditional formatting shows with the lightest green the highest variation, while the deepest red the lower ones. ____ 99

Table of contents

Abstract ... 3

Abstract (Italiano) ... 4

List of Figures ... 5

List of Tables ... 7

1. Introduction ... 10

2. State of the art ... 10

2.1. Studies on Accidents and model choice ... 11

2.2. Studies on Accessibility ... 12

2.3. Studies on hotspots areas identification ... 14

2.4. Studies on travel time budget and walking time ... 18

2.4.1. Travel time budgets to/from a bus station ... 19

2.5. Studies on coefficients and indices to measure public transport coverage. ... 20

2.5.1. Coverage rate ... 20

2.5.2. Cumulative opportunities index ... 22

3. Tools and data ... 24

3.1. Selection of the cities ... 24

3.2. Extraction of the public transport lines from GTFS data ... 30

3.2.1. Inserting GTFS in QGIS ... 33

3.2.2. Study time ... 37

3.3. Population density distribution ... 38

3.4. Night -life areas identification ... 49

3.4.1. Downloading the POIs and inserting them on QGIS ... 51

3.4.2. Hotspots identification ... 55

3.5. Travel time computation ... 62

3.5.1. Computation input ... 65

3.5.2. Identification of reachable and reached centroids of each NLA ... 67

3.5.3. Once served ... 71

4. Computation results ... 73

4.1. Public transport night lines length and network intensity ... 73

4.2. Night-life areas identification ... 76

4.3. Public transport coverage of the different Night-life areas ... 83

4.4. Benchmark analysis ... 89

4.4.1. Number of inhabitants versus service offer ... 89

4.4.2. Service offer versus service coverage, average over all Night-life areas and departure times ... 91

4.4.3. Average service coverage, average over all Night-life areas ... 94

4.4.4. Service offer versus service coverage, all Night-life areas jointly considered ... 97

4.4.5. Average service coverage, all Night-life areas jointly considered ... 100

4.4.6. Time variations due to the impact of service frequencies 102
5. Conclusions 108
References 110
APPENDIX A: Parameters' set for NLAs clustering. 114
APPENDIX B – Results overview 135

1. Introduction

When talking about transport in urban contexts, it is the paradigm to invoke a decrease of private motorization in favour of an increasing usage of public transport means or non-motorized solutions. Most of the times this axiom is meant to fight congestions phenomena and more and more in the last years, environmental pollution. But when we consider the weekend night time, this axiom can make the difference between social inclusion and exclusion or between life and death, at least for specific social groups such as young people, considering transport safety issues. European statistics recognise how night hours between Friday and Saturday and between Saturday and Sunday are the hours of the week with the higher percentage of death over road accidents, and the young age population, namely the range 18-24, is the most involved in such fatalities. The reason of these data is well known and rely on the alcohol consumption during the weekend nights by European young inhabitants. Many cities in Europe (and not) have introduced in the last decade a night-time service of public transport in order to offer a fast, safe and convenient option to private means for the ones who enjoy the night life of the city, aiming to provide safer roads.

The goal of this paper is to create a framework to consider European metropolitan cities on the basis of their night-time public transport service, studying their performance based on service coverage. To do so, open-source data were used to identify the spatial location and distribution of night-time related amenities such as bars, pubs, biergartens and nightclubs and other night time-related amenities, which we clustered in an attempt of identifying the cities' hotspots of night life (named NLAs: Night Life Areas). To assess the computation of service coverage we used block-level aggregated data concerning the population distribution to understand the demand. We finally made use of open data related to night-time public transport lines, in order to get a length extension of the offer, which linked with the scheduled frequencies, allowed us to quantify the offer in terms of service coverage.

As for the structure, the thesis is composed by an Introduction and 4 following chapters; two appendixes are included for a better comprehension.

- In chapter 2. State of the art, we presented a literature review related to the concepts that shaped the thesis redaction, namely regarding studies on accessibility, on accidents and model choice, on hotspots areas identification, on travel time budget and walking time and on coefficients and indices to measure efficiency.
- In chapter 3. Tools and data we presented the sources from which we took the needed data, the way we extrapolated and manipulated them and the methodology we followed with the results. The chapter is divided in sub chapters, each one focusing on a different type of data, namely the public transport line, the population density distribution, the NLA identification and the service coverage computation trough Traveltime.
- In chapter 4. Computation results the results of the previous manipulations and computations are shown through tables and graphs for each type of data as before. In addition, a sub chapter showing the correlation, and related considerations, between the eight cities considering different parameters is presented.
- In chapter 5. Conclusions the work and the results are resumed, and we also exposed the found criticalities during the redaction of the thesis and possible developments of the same.
- In Appendix A, the reasons why we choose a set of parameters or another for the NLA clustering processes in each city are shown.
- In Appendix B, an overview of all the results related to population distribution, public transport offer, amenities and computational result is given for each city separately.

2. State of the art

In the following chapter we went through a literature review about the concepts which relate the most to the goal of this work. Since Night – time public transport offer is not a high trending topic, separated reviews regarding partial portions of the matter were necessary: for this reason, the paragraph is divided into several sub-paragraphs, each of them tackling a part of the matter. Namely, we reported studies about the concept of accessibility, which can help frame the matter we faced in our work and understand the importance of it; studies on the relation between night time public transport service and alcohol consumption and so road safety; studies where hotspots and specific areas are individuated and studied; studies on travel time budget which helped us choosing those mostly subjective parameters related to travel times; and finally studies on different efficiency measurements and related coefficients and indices.

2.1. Studies on Accidents and model choice

Both Italian and European studies agree that young people are more likely to be engaged in car accidents: the age range 18-24 years old has the highest car accidents - related mortality rate of all age groups (European Road Safety Observatory, 2021) and the most common time window for these fatalities to happen is weekend night time, and this is most probably due to weened night alcohol consumption. So, the matter of night time public transport is tightly related to road safety since it furnishes a fundamental option to private means. In this sub paragraph, articles related to the alcohol consumption – modal choice – accidents reduction relationship are reported, so that the relevance of the topic can be assessed.

The article "***Can public transportation reduce accidents? Evidence from the introduction of late-night buses in Israeli cities***" (Lichtman-Sadot, 2019) has been written on this purpose: in the article the authors study the reduction of car accidents involving young drivers, aged from 15 to 29 years, in Israeli cities following the introduction of night-time bus services from 2007. To do so, a difference in differences and a triple differences frameworks were used, and the results show a reduction of the 37% of fatalities and a 24% decrease of the resulting injuries, implying the effectiveness of the common notion that public transportation benefits road safety.

Another article on the matter is "***Quantifying the social impacts of the London Night Tube with a double/debiased machine learning based difference-in-differences approach***", (Zhang, Li, & Ren, 2022). The article considered how the new night-time service offered through the late-night opening of a series of underground changed factors as the price of households, creation of workplaces, crime rates and road safety in London. However, the post-launch causal impacts are found to be insignificant with a car accidents reduction of -2.39 %. The explanation other authors gave in previous works for this counter-intuitive discovery is that the reduction of drunk driving reflects in a limited buffer zone around the metro stop (Jackson & Owens, 2011). Indeed, limiting the buffer to a radius of 400 m from each underground stop, the reduction in road accidents is more significant. Another parameter related to safety is the one of the crash-related casualty rates, which reduced of 9.91 % indicating road safety improvement.

In the article "***Young drivers' night-time mobility preferences and attitude toward alcohol consumption: A Hybrid Choice Model***" (Scagnolari, Walker, & Maggi, 2015) the authors focused on the matter of mobility demand considering the attitude toward alcohol of young people and how to prevent those same people to use the car in order to have safer roads. The work insisted on the preferences of young people toward new modes of public transport, through a Stated Preferences Experiment. The model design covered the attributes of the mode, the characteristics of young drivers and alcohol related psychological variables. The mode choice must be made between three possibilities: private means (car, motorcycle, car as a passenger), public means (night bus/train) and no choice (staying home). The experiment results were shown for different scenarios, starting from a base one going to the implementation of new means (such as mini-buses or shared taxi) and new policies (such us more police checkpoints, harder consequences or lower price of public means and others). The results showed that with those countermeasures, *people are more incline to choose a public mean and that especially the sensitivity to this countermeasure is higher for the people with a high alcohol attitude than for the low alcohol attitude group* and so *the relationship between alcohol attitude and public more secure transport modes in young drivers are weak, but consistent.*

In the article “***Off the rails—Evaluating the nightlife impact of Melbourne, Australia’s 24-h public transport trial***” (Curtis, et al., 2019) the behaviour of people during night-time of the weekend before and after the introduction of a 24 h public transport service in the city of Melbourne, Australia, has been studied. To do so the authors used Covert venue observations (pre-post) and a convenience sample of nightlife participants interviews (post-only). The results showed as the behaviour about alcohol and drug consumption didn’t change but 44% indicated spending more time in the night-time economy, 27% reported spending more money, and 56% reported increasing their train use.

The above paper shows well the relevance of the topic of this thesis, namely the provision of a viable alternative to private car use for late night trips.

2.2. Studies on Accessibility

A concept which gained vital importance in designing and evaluating the transit system in terms of mobility and sustainability, is accessibility (Saif, Zefreh, & Torok, 2019). Since the debate about accessibility in public transport is a long term one, it involved many authors trying to give their idea on what does the concept mean for them and debating between each other about it. It follows that it doesn’t exist a single definition of the term, but many, and none of them can be the absolute best or most right, but each one has its place in specific researches; indeed the authors of works which rotate around the concept of accessibility use or give a definition or another based on the goal of their project.

A good point to start looking around in the multitude of definitions and debates is the review article “***Public Transport Accessibility: A Literature Review***” (Saif, Zefreh, & Torok, 2019). In this paper the authors focus on the previously available literature and make their consideration about some related themes such us the perceived accessibility, the correlation between accessibility and public health and employment rates and finally focus on how the public transport accessibility is also to be considered as accessibility to social life (which is the core of our work) and the consequences of its lack. Then the article focuses on relation and difference between accessibility and mobility, the relation between accessibility and sustainability and finally how accessibility can be evaluated considering economical, spatial and temporal efficiency. The conclusion of the article is that *not just the performance of public transportation but its impact on other social aspects should be considered while planning the public transport facilities*.

In order to approach the measurement of accessibility in, we can firstly refer at the study “***Strategic analysis of public transport coverage***” , (Murray, 2001) where the concept of accessibility is seen as merely objective distance (a spatial one in this case). The case study is Brisbane, Australia, in which the goal is to ensure to the 90% of the population of Brisbane the opportunity to have a bus stop at least 400 m from home. To minimise the number of stops they use a mathematical optimization method called *Location Set Covering Problem (LSCP)* which was originally used to identify a *minimum number of emergency service facilities*, (Toregas, Swain, ReVelle, & Bergman, 1971).

In the article ***“A planning tool for maximizing transit services”*** (Ma, Li, & Han, 2016) the authors developed a planning tool to aid the decision making process about service area using optimisation techniques, namely the Prize collecting Salesman Problem (PCTSP) and the Prize collecting Steiner Tree.

- Prize collecting traveling salesman problem: A salesman who travels between pairs of cities at a cost depending only on the pair, gets a prize in every city that he visits and pays a penalty to every city that he fails to visit. He wishes to minimize his travel costs and net penalties, while visiting enough cities to collect a prescribed amount of prize money. (Balas, 1989).
- Prize collecting Steiner tree: The Prize-Collecting Steiner Tree Problem (PCST) is an extension of the Steiner Tree Problem where each vertex left out of the tree pays a

penalty. The goal is to find a tree that minimizes the sum of its edge costs and the penalties for the vertices left out of the tree (Nagarajan, 2017).

Accessibility in this study is defined as the ability of the targeted population to move across space based on travel needs. The target population is determined by the specific planning or policy objective(s). The better the target population is served and/or the larger the area that a transit system can cover, the greater the accessibility the transit system is able to offer.

Another attempt to measure accessibility is in the study "***Measuring accessibility of urban scales: A trip-based interaction potential model***" (Liu & Jiang, 2021), where the authors *formulate and estimate attraction choice models (*a logit model and a gravity one are used*) that provide measurements of accessibility on various scales reflecting the choice of people to travel to facilities or activity places and characterize the interaction between land use patterns and transportation facilities.* In this study the accessibility is measured analysing four dimensions:

- travel desire
- travel mode
- spatial distribution
- attractiveness of activity points

Another article where different definition of accessibility are exposed and where one of them is investigated so that it was possible to measure it, is the article "***Measuring City-Level Transit Accessibility Based on the Weight of Residential Land Area: A Case of Nanning City, China***" (Le & Ye, 2022) where the authors try to develop an overall index to measure the transit accessibility with weighted residential areas. This index is developed for Nanning City in China but has the purpose to be applicable for other cities as well. It considers the *transit accessibility as the ease to travel by transit from origin to destination* but putting more interest in the role of the destinations; indeed, the authors want to highlight the spatial connection between residential communities and relevant destinations connected by transit. In the article is also present a literature review about the *definition of transit accessibility* in which the available literature is divided in four categories:

- Accessibility measures on objective journeys to specific destination in terms of distance or time.
- Measures which also include the difference between activities in terms of quantity and quality; gravity models are used to give different values to different destinations.
- Measures which consider the impact of different socio-economic factors on transit accessibility.
- Measures which consider the mobility offer including the frequency, so the waiting time for the users; these measures get close to the real time analysis.

The calculations consider for each residential area a set of different kind of facilities (such as medical centres, shopping retail, employment centres etc.) and the coverage of residential area within the transit stop service area. It considers the distance (time and space separately) between RCs (residential communities) and relevant facilities. Then each RC is weighted based on his surface and two indices which represent the overall city-level transit accessibility based on the weight of residential land area expressed by distance and time are computed. The results show that in twenty years the urban weighted average distance of trip is increased while the time is shortened which indicates how the overall transit accessibility at city level has increased.

Another parameter, which meaning is like the one of accessibility but with some relevant differences, is the one of **mobility**. It can be summarised as the ability to move from one place to another, (Costa, Neto, & Bertolde, 2017). An article which gives a good overview on what are the definitions and the meanings given to the concept of mobility and the one of Indexes is ***"Urban Mobility Indexes: A Brief Review of the Literature"***, (Costa, Neto, & Bertolde, 2017). A definition contained in this work which summarizes well the two concepts is: *Facing the challenges related to planning of mobility, it is necessary to use indexes and indicators for monitoring the conditions of mobility in urban areas* (Miranda, 2010). At the end of the article there is a table with all the cited authors, the Indexes they created, their purposed and their study area: an input from this work is indeed that, as stated before for the concept of accessibility, the definition change in relation to the study area and the foal we work for.

In this paragraph, different definitions of accessibility were explored, as well as different methodologies to measure it. In the following thesis, accessibility has been stripped of its social aspects, appearing as mere temporal distance, so close to how intended and studied in (Murray, 2001). The methodology that is showcased, on the other hand, embraces the idea of inhabitants coverage explored by (Le & Ye, 2022). In Table 1 follows a recap of the studies considered in the paragraph, with a short resume of what concept of accessibility is assumed and which methodology is used to measure it (when applicable).

Table 1 – Studies on accessibility, recap with accessibility definition and goal and methodology of the study expressed for each study.

Study	Accessiblity definition	Goal and methodology of the study
"Public Transport Accessibility: A Literature Review" (Saif, Zefreh, & Torok, 2019)	Not just the performance of public transportation but its impact on other social aspects should be considered when considering accessiblity	Relation between accessibility and social life
"Strategic analysis of public transport coverage", (Murray, 2001)	Accessibility as a spatial objective distance	reaching a certain percentage of inhabitants with a stop nearby, minimising the number of stops using a mathematical optimization method called Location Set Covering Problem (LSCP)
"A planning tool for maximizing transit services" (Ma, Li, & Han, 2016)	Accessibilitty as the ability of the targeted population to move across space based on travel needs; The better the target population is served and/or the larger the area that a transit system can cover, the greater the accessibility the transit system is able to offer.	Aid decision makers to increase accessibility using optimisation techniques, namely the Prize collecting Salesman Problem (PCTSP) and the Prize collecting Steiner Tree.
"Measuring accessibility of urban scales: A trip-based interaction potential model" (Liu & Jiang, 2021),	Accessibility is influenced by the choice of people to travel to facilities or activity places and is characterised by the interaction between land use patterns and transportation facilities	measuring accessibility through attraction choice models (logit model and gravity model)
"Measuring City-Level Transit Accessibility Based on the Weight of Residential Land Area: A Case of Nanning City, China" (Le & Ye, 2022)	Transit accessibility as the ease to travel by transit from origin to destination	developing of an index to measure transit accessibility, based on the land areas of residential communities.

2.3. Studies on hotspots areas identification

Because of the nature of our work, we are also interested in knowing how other research before this one tackled the matter of identifying the night life areas. In the following sub paragraph, researches on new ways to obtain data to create cluster are shown, specifically social networks and taxi services data were used. Also researches regarding distribution patterns of clusters and their scaling properties are reported.

An interesting approach in this sense in the one given in two articles, "***B-Planner: Night bus route planning using large-scale taxi GPS traces***" (Chen, et al., 2013) and "***When Taxi Meets Bus: Night Bus Stop Planning over Large-Scale Traffic Data***" (Xiao, et al., 2016) in which the intent was to address the night-bus route planning issue by leveraging taxi GPS crowd-sourced GPS data. Knowing the pick-up and drop-off locations of each taxi travel, a cloud of points was created; using an improved version of the algorithm DBSCAN (explained and used in 3.4.2), it was possible to point a series of hot spots. Finally, in one case (Chen, et al., 2013) the locations of the hot spots and the actual bus stops of the night time public transport service were compared, and an adjustment of the offer was proposed. In the other case (Xiao, et al., 2016) the public transport lines were created anew, trying to collect the maximum amount of hot spots. The methodology used in these two articles for hotspots identification, namely the DBSCAN clustering algorithm, is similar to the one used in our thesis, with the difference we changed the way to identify attractiveness.

A correlation between the studies using data from taxi rides and our choice to consider activity dense areas as origin points, is assessed by the study "**Exploring the influence of built environment on Uber demand**" (Sabouri, Park, Smith, Tian, & Ewing, 2020). These authors examine how the built environment affects demand for ride-sourcing services like Uber in 24 diverse U.S. regions, after controlling for socioeconomic factors. The research shows that Uber demand is positively correlated with population, employment, **activity density**, land use mix, and transit stop density, but negatively correlated with intersection density and destination accessibility.

The clustering algorithm DBSCAN, which we actually used in our thesis, is a density-based clustering algorithm which requires two parameters: the minimum size of the cluster in term of points and the maximum distance epsilon for a point to be inserted in a cluster; these two parameters are chosen so that the results look similar to the one of the other two methods.

Choosing the parameters for DBSCAN is somehow a subjective procedure based on how good the results look like. The CrimestatWorkbook 3 (Levine, 2007) proposes, in order to choose these two parameters correctly, in the context of fight against crime, to ask ourselves two questions:

- How large, in area terms, will be the effect of each point of interest (in our case, how much area will be influenced by the presence of a leisure activity).
- How much of this effect will focus on the source point and how much of it will be dispersed throughout the bandwidth interval.

Also, a table with different kinds of crimes and the suggested distribution is proposed.

While an article in the web portal ECOVIEW, (Frate, 2021) proposed three criteria for choosing the search radius:

- use rules, as the Silverman function (is a statistical method used to estimate the optimal bandwidth parameter for kernel density estimation)
- use cross-validation algorithms (assess the model's ability to generalize to new data)
- evaluating the quality from a visual perspective

Staying on the topic of input data from social networks, the authors of the article "***Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS***" (García-Palomares, Gutiérrez, & Mínguez, 2015), downloaded from Panoramio, a social network about photography, data about the stored photographs for eight different European cities, containing geographic coordinates and other information. These data have been aggregated in uniform hexagons which have been used to produce density map and so to define the intensity and degree of concentration of the photographs. Two indicators are calculated to determine global location patterns: the Getis-Ord General G statistic and the Global Moran's I statistic. The General G index measures the degree of clustering for either high or low values, while Moran's I index measures spatial autocorrelation based on feature locations and attribute values. Following this path, it is possible to visualize the location of the most visited areas of each city are and through the computed indicators to gain information related to the hotspot distribution within the city boundaries.

Another relevant work in this direction is the one in the article "***Mapping the popularity of urban restaurants using social media data***" (Xu, Yang, Zhou, Zhang, & Qiu, 2015) in which the authors tried to quantify the popularity of urban restaurant through the data downloaded from a Consumer Review Website. Based on the previous attributes they managed to develop a popularity index. This index depends both on the location of the restaurant and on the presence of other relatable activities in the neighbourhood.

Another innovative way to collect input data is explained in the work "***Generating demand responsive bus routes from social network data analysi***s" (Sala, Wright, Cottrill, & Flores-Sola, 2021) the authors extrapolated data from a social network, namely Twitter, from which they identify a spatially distributed picture of the demand to attend a musical event in Barcelona; to do so, all the interactions with contents related to the event were considered, along with the location of the user that interacted. Thanks to this information, an *influence score* was assigned to each municipality in the Barcelona region, allowing to understand which municipalities contained more potential attendants Thanks to these data they managed to design an extraordinary service of viable buses to connect different rural areas to the venue location. This work took inspiration from different previous articles in which a correlation between big data reachable through social media and participation to big events was made. This study is not regarding night time but another *extraordinary situation* for public transport, where is again required a knowledge of relevant hotspots. In this case, however, the production is studied (the households of the attendants) and not the attraction (the festival).

Another way to solve the problem we were facing, namely to get data on where people spend their night time during the weekend, could have been to analyse not geographical points (as the pick-up and drop off points from taxis, or the twitter interactions about a topic or the pictures from a social network about photography or the location of related amenities, as we used in our study) but individuals movements. The study "***Identifying the city centre using human travel flows generated from location-based social networking data***" (Sun, Fan, Li, & Zipf, 2016) aims to use motion data obtained from location-based social networks in order to identify the city centre (or multiple city centres in case of polycentric cities). The research employed three distinct clustering techniques and compared them in terms of their ability to recognize urban centres and define their limits. The used techniques were the local Getis-Ord G statistical method (which measure the spatial clustering or spatial autocorrelation of a variable within a geographic area), the DBSCAN (already explained) and the Grivan-newman algorithm (a graph-based algorithm used for detecting community structures within complex networks).

In the article "***The amenity mix of urban neighbourhoods***" (Hidalgo, Castañer, & Sevtsuk, 2020) the authors studied the collocation patterns of more than a million amenities in 47 different cities (in the USA) which were downloaded from the Google Places API as data points containing the latitude, longitude and type of amenity: this procedure is similar to the one we adopted, as similar as the obtained results are. They divided the cities in neighbourhoods using a simple home-made clustering algorithm to identify the boundaries of amenity dense neighbourhoods and which amenities belong to them. Once identified the neighbourhoods, they estimated the number of amenities of each type they expected in each neighbourhood by leveraging the principle of relatedness. The principle of relatedness (Hidalgo et al., 2018), is a statistical principle that can be used to predict the activities that a location is more likely to enter or exit in the future. They introduced the concepts of exogenous and endogenous (complementary and competitive) clustering:

- Exogenous clustering: the spatial clustering occurs due to external factors, such as geographical or political boundaries, rather than inherent characteristics of the objects or individuals being clustered. The authors made the example of a train station, around which multiple businesses collocate.
- Endogenous complimentary clustering: is the clustering of economic activities that are mutually beneficial and tend to co-locate in a particular area of a city based on shared characteristics and advantages; as it happens with a cinema and an ice cream shop for example.
- Endogenous competitive clustering: the spatial clustering of economic activities that are in direct competition with each other within a specific geographic area of a city. That happens for example with Outlets.

To test the accuracy of the model, they compared the predicted number of amenities of each type in a cluster with a simple benchmark where only the total number of amenities in a cluster were used as a predictor. This benchmark has been inspired by the literature on urban scaling laws, which has shown that many important urban characteristics (such as GDP or crime) scale with city size. Overall, the findings suggested that the presence of an amenity in a cluster is connected more strongly to the presence of other amenities, than to the size of the cluster, except for the case of airports, aquariums, bus stations, casinos, convenience stores, gas stations, and zoos. The paper "***Growth, innovation, scaling, and the pace of life in cities***" (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007) shows how many diverse cities properties, related for example to economic wealth, infrastructures, social life etc., are power law functions of population size with scaling exponents. Through data the authors explain how cities that are quite different in shape and location are on the average the scaled version of one another following specific power law functions for specific matter in which the exponents fall into three categories defined by beta = 1 (linear), beta <1 (sublinear), and beta>1 (superlinear):

- Individual human needs such as job, house and water consumption will have beta circa = 1, since they remain always the same
- Material quantities will be displaying economies of scale associated to the presence of infrastructures and will have beta circa = 0.8 <1
- Quantities related to social currencies such as information, innovation or wealth will have beta = 1.1 / 1.3 >1, signifying an increasing value with population size.

In this paragraph we have explored different ways to get input data in order to start a clustering process and different clustering algorithms, as recapitulated in Table 2. As mentioned, the solutions adopted in our study are:

- Regarding the input data, the download of amenities points from an API service, similar to what happens in (Hidalgo, Castañer, & Sevtsuk, 2020), even if on another platform
- Regarding the clustering phase, the adoption of DBSCAN algorithm, as happen in (Chen, et al., 2013), (Xiao, et al., 2016) and (Sun, Fan, Li, & Zipf, 2016).

Table 2 – Recapitulation of studies about areas identification

Study	Input data	Goal of the study	Clustering method
"B-Planner: Night bus route planning using large-scale taxi GPS traces" (Chen, et al., 2013)	taxi gps data	determinating hotspots based on pick up and drop off points	DBSCAN
"When Taxi Meets Bus: Night Bus Stop Planning over Large-Scale Traffic Data" (Xiao, et al., 2016)	taxi gps data	determining hotspots based on pick up and drop off points	DC-DBSCAN
"Generating demand responsive bus routes from social network data analysis" (Sala, Wright, Cottrill, & Flores-Sola, 2021)	location of twitter user which interacted with the event	generate demand responsive bus routes for a music festival	Not clustered
" Identification of tourist hot	pictures from	Identification of	division of the city

spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS" (García-Palomares, Gutiérrez, & Mínguez, 2015)	social network Panoramio	tourist hot spots	in uniform hexagons to produce density map
"Mapping the popularity of urban restaurants using social media data" (Xu, Yang, Zhou, Zhang, & Qiu, 2015)	customers feedbacks on restaurant from the app Consumer Review	quantify the popularity of urban restaurants	not clustered, but a popularity index influenced by the presence of other activities nearby was assigned.
Identifying the city centre using human travel flows generated from location-based social networking data" (Sun, Fan, Li, & Zipf, 2016)	motion data obtained from location based social network	identify the city centre (or multiple city centres in case of polycentric cities)	Getis Ord G statistical method, DBSCAN and Grivan Newman algorithm
"The amenity mix of urban neighbourhoods" (Hidalgo, Castañer, & Sevtsuk, 2020)	amenities from Google Places API, points with information about kind of amenity and geolocation	predict the distribution of different kind of amenities departing from actual distribution patterns	simple homemade algorithm based on an accessibility index

2.4. Studies on travel time budget and walking time

When working with travel time computing tools, some parameters are required, such as the quantity of time on average devoted to travels. To set these parameters, an overview of people preferences is required; in the following sub paragraph we reported some articles which helped us in our choices.

The main point of ***"Travel-Time Budget"*** (Supernak, 1967) is that there is not a general travel time budget for everyone. The article is a critique to the concept of travel time budget stability; it shows as findings from different cities around the world can hardly support the concept of stability of travel budgets, showing how the range of travel time average obtained from different cities in both the western and developing countries is very wide even at country-aggregated level. Other travel-budget formulations are also considered likewise doubtful such as the money expenditure. The results of travel budgets are shown to be even more confusing at a disaggregated level: in developed countries, the travel time budget rises with incomes, while it happens the opposite in western countries. The stratification of household income is seen by many researchers as meaningful but in this case brings confusing findings: for this reason, the author asks himself and to the readers if the different travel budgets are comparable at all or not, if it the right thing to look at, and if the methodology of travel budget analyses is acceptable. Another disaggregation level which implies dramatic differences in travel time budgets is the one in family members. Any average value of travel time budget calculated for the family as a whole, strongly depend on the family size and structure.

In the article **"Travel Time Budget – Decomposition of the Worldwide Mean"** (Joly, 2004) the author wants to understand if Jahavi opinion is correct; the opinion of Jahavi is that a stable Travel Time Budget (TTB) in different places in different times is given and that vehicles speed improvements just provoke an extension of the reachable area. The author points out how this stability is reachable only at a global aggregation level, not at finer levels of disaggregation. Namely, the different trends between two different philosophies of urbanisation are shown: the extensive ones (America, Oceania) and the intensive ones (Europe, Asia). Finally, a study is conducted at city level in Lyon, which goal is to gain the travel time budget from a travel demand diary questionnaire.

The paper ***"The role of travel time*** " (Joly, 2007) observes travel time ratio of work, leisure and shopping activity type, and estimates a linear model for the travel time associated to each type of activity. The author considers the assumption of a constant travel time budget as irrelevant at a disaggregated level; he reports how analyses of travel time budget at individual level reveal numerous relationships between travel time budget and other variables, such as socio-demographic attributes, mobility characteristics, or urban contexts. The paper studies Swiss and French surveys focused on distinct travel purposes classifications. A look on the results of estimations indicates that the daily travel times associated to a purpose is not a fixed proportion of the daily time of the associated activity type. There is no proportionality at daily expenditure level between daily time budget of an activity and daily travel time associated. Travel time intensity resulted to be increasing and convex with the shopping and leisure daily time budgets. Hence, travel time budget increases more than proportionally with the time spent in leisure and shopping activity. Leisure appeared more intensive in terms of travel time than shopping. Conversely, the travel time intensity for work is constant. Travel time for work increases linearly with work duration. In conclusion, the proportional assignment of daily available time is only valid for leisure and transport activities. Tests about the proportional relationship between daily traveltime given a purpose and daily activity time indicated that proportionality is not valid, regardless of the type of activity. Finally, the proportionality hypothesis is rejected between trip time and activity time at the destination. The main points of the three articles here listed are resumed in Table 3.

Table 3 - Resume of articles regarding the concept of Travel Time Budget and its criticalities

Study	Main point
"Travel-Time Budget" (Supernak, 1967)	there is not a general travel time budget for everyone
"Travel Time Budget – Decomposition of the Worldwide Mean"(Joly, 2004)	a stable travel time budget can be reached only at global aggregation level
"The role of travel time "(Joly, 2007)	estimates a linear model for the travel time associated to each type of activity

2.4.1. Travel time budgets to/from a bus station

Another issue worth being considered is the travellers' time budget for walking to/from a bus station; this parameter was important in our study; we see in this paragraph articles which helped us to take a decision on its value.

The study "***Walking distances to public transport in smaller and larger Norwegian cities"*** (A.Tennøy, Knapskog, & Wolday, 2022) considers 4 different Norwegian cities. Through surveys, the walking trip times to local public transport stops and to railway stations is computed for each city: walking trips to local public transport stops last 4.1–6.0 min (328–520 m), on average, in different cities, and walking trips to railway stations last 6.6–8.6 min (528–688 m). Average walking distances found in studies relevant for this paper range from 170 to 549 m to local public transport stops and 805–882 m to railway stations. Walking trips to railway stations are significantly longer compared with those to local public transport stops and the 75th-percentile duration is 10 min for both types of trips in both cities. The authors believe that when applying results from studies like this in the planning and development of land use and public transport services, 75th percentiles might be more useful measures than average (mean) values, by indicating the distances 75% of respondents walk shorter than. A qualitative comparison of findings across the four cities showed increasing average duration of walking trips to local public transport stops with increasing city size.

Another interesting work is ***"Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany"*** (Sarker, Mailer, & Sikder, 2020): As written in (Chia, Lee, & Kamruzzaman, 2016) a one-size-fits-all solution to determine catchment areas is unlikely to be effective. There is no consensus with real evidence between a common walking distance/time and the actual one. The article focuses so on the following question: how far are people willing to walk to access public transport facilities? The answer changes based on the different area of the city (the study considers 3 kinds: an inner-city area – walking zone, service area – transit zone and low-density suburban area – automobile zone), the trip mode, and the purpose of the trip. The more we move outside the centre, the more the inhabitants are willing to walk more to reach the service station; they are willing to walk more for trains and underground than for bus or tram.

In the two studies regarding walking time home – station has been highlighted how also this travel time budget is subjective and not stable. We have seen how it is influenced by factors such as the kind of mean which serve the station and the distance from the city centre. This topic has been further considered in 653.5.1.

2.5. Studies on coefficients and indices to measure public transport coverage.

The matter of public transport system effectiveness evaluation in meeting travel demand has been debated for many years and got more complex year after year, adding new considerations, concepts, and parameters. In the following sub paragraph research regarding public transport assessment through coverage rate and cumulative opportunities index are reported.

2.5.1. Coverage rate

We can see a definition of coverage rate in the article ***"Do the population density and coverage rate of transit affect the public transport contribution?***" (Jasim, Al-Jaberi, Al-Mamoori, Al-Ansari, & K., 2022) was to investigate the impact of public transport coverage on usage within the neighbourhoods of Kut city. The concept of access coverage is studied and considered essential in public transit planning, since it shows how service is provided to riders. The research hypothesis is that increasing the coverage rate and the intensity of development increases the use of public transport. Quantitative analysis has been linked to descriptive analysis and uses SPSS and GIS to analyse spatial relationships and to explain the relationship between public transportation and the coverage rate. The BUFFER ZONE analysis was used to determine the coverage rate of transit in Kut city. The buffer zone was made around public transport routes with a 350 metres radius, which is the distance resulting from the average distance travelled to the public transport service in Kut city, which amounted to 356 metres, according to the questionnaire. Then, analysis using linear regression (global and local) and weighted geographical regression were conducted.

- Linear regression: used to predict the value of a variable (dependent variable) based on the value of another one (independent variable). This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values (IBM, s.d.).
- Weighted geographical regression: Geographically Weighted Regression (GWR) is one of several spatial regression techniques. GWR evaluates a local model of the variable or process to be predicted by fitting a regression equation to every feature in the dataset. GWR constructs these separate equations by incorporating the dependent and explanatory variables of the features falling within the neighbourhood of each target feature. The shape and extent of each neighbourhood analysed is based on the Neighbourhood Type and Neighbourhood Selection Method parameters. GWR should be applied to datasets with several hundred features. It is not an appropriate method for small datasets and does not work with multipoint data (esri, s.d.).

The paper ***“Serial formation and parallel competition in public transportation”*** (Edrisi, Barzegari, & Nourinejad, 2021) examines a duopolistic transit market considering two different network structures: the parallel one, where there are different operators and so the passengers can choose between the two, and the serial one, in which the passengers are dedicated to one or more operators. The authors consider a many-to-many and a many-to-one demand distribution using continuum approximation for the modelling. The results are that:

- In many-to-many demand patterns: the fleet size is minimum where there is full overlap of service regions.
- In many-to-one demand patterns: the serial formation is more favourable for the society than the parallel formation.

Many-to-many demand patterns: In the mobility demand field, a "many-to-many demand pattern" refers to a scenario where multiple individuals demand travel between multiple origin-destination pairs. This type of demand is characterised by the fact that there is not a direct one-to-one relationship between individuals and origin-destination pairs, but rather multiple individuals demanding travel between multiple origin-destination pairs. The scenario of our work is a Many to many. Many-to-one demand patterns: In the mobility demand field, a "many-to-one demand pattern" refers to a scenario where multiple individuals demand travel to a single destination. This type of demand is characterised by the fact that there is a direct many-to-one relationship between individuals and destinations, where multiple individuals are travelling to a single destination. Continuum approximation: It refers to a mathematical method used to model a system or process as a continuous, rather than discrete, entity. This approach involves using a continuous mathematical representation, such as a differential equation, to approximate the behaviour of the system or process being modelled. The goal of this approximation is to simplify the complex system or process, making it easier to analyse and understand. The accuracy of the approximation depends on the level of detail included in the mathematical representation and the quality of the data used to validate the model.

The article “***A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations”*** (Domènech & Gutiérrez, 2017) develops a methodology for evaluating the effectiveness and spatial coverage of public transport in touristic cities. The results of the study firstly allow the detection of unequal spatial accessibility and coverage in terms of public transport in the municipality, with significant differences between central neighbourhoods and peripheral urban areas of lower population density. An efficient spatial coverage of public transport is understood to be one that favours the access of citizens to transport services and ensures the connection between journeys at intra and inter-city scale. In order to study the level of achievement of these concepts in a touristic city, an evaluation system of the spatial coverage of public transport facilities in relation to the amount of potential population that can access them has been designed. Specifically, in the designed evaluation system, indicators have been incorporated (see

Table 4) to evaluate public transport conditions in terms of spatial coverage of the network for resident and tourist population (supply and quality), and to measure the degree of network accessibility (interconnectivity, intermodality).

Table 4 -indicators proposed by (Salado García, Díaz Muñoz, Bosque Sendra, Carvalho Cantergiani, & Rojas Quezada, 2006)

Typology		Indicator
Supply/Quality	1	Stops/1000 inhabitants
	2	Population percentage with bus stop/s at less than 200 m
Interconnectivity	3	Population percentage with multimodal stops at less than 200 m
	4	Connectivity between TA (Transport Areas): Drift Index
Intermodality	5	Population percentage at less than 500 m from an intercity bus stop

The authors define different areas of the city as Transport Areas (TA) and use the concept of centre of gravity: the centres of gravity defined for each TA allowed the finest calculation of overall connectivity between TAs, since they are strategically located to cover the maximum population in their surroundings. The indicator "Drift INDEX" measures the extra time that urban buses require to connect the different TAs in comparison to the shortest possible journey if using a private vehicle. For its calculation the authors used, for both the public transport network as well as the road network, the minimum possible time between the centres of gravity of the TAs.

The main points of the previous three studies regarding different ways to consider and study coverage rate, are resumed in Table 5.

Table 5 - Recapitulation of articles regarding coverage rate

Study	Main point
"Do the population density and coverage rate of transit affect the public transport contribution?" (Jasim, Al-Jaberi, Al-Mamoori, Al-Ansari, & K., 2022)	Coverage considered as a geometrical buffer zone
"Serial formation and parallel competition in public transportation" (Edrisi, Barzegari, & Nourinejad, 2021)	Two demand patterns (many to many and many to one) are considered, such as two different network structure of a duopolistic transit market: the combination for better coverage is researched,
"A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations" (Domènech & Gutiérrez, 2017)	an evaluation system of the spatial coverage of public transport facilities in relation to the amount of potential population that can access them has been designed

2.5.2. Cumulative opportunities index

The cumulative opportunities index, as intended in our study, is a way of measuring the extent to which different groups of people have access to public transport services that enable them to take advantage of mobility opportunities. This matter is tackled in the following articles.

In the study ***"Cumulative versus gravity-based accessibility measures: which one to use? "*** (Palacios & El-geneidy, 2022) the authors suggest that estimating access to jobs by public transport using the cumulative opportunities approach highly approximates the best performing gravity-based access measures promoted in transport planning and geography scholarship. The authors agree that, having no weights, cumulative opportunity accessibility measures are easier to compute and interpret than gravity-based ones, and so preferable. In another paper ***"Making accessibility work in practice"*** (El-Geneidy & Levinson, Making accessibility work in practice, 2022) the authors agree that even simple measures, such as the cumulative opportunities measures of accessibility, can be an important help in complex studies. And so, they send a call for researchers to explain these measures (ndr accessibility) in a way that makes them easily adoptable in practice and to work more on highlighting the added value from such complexity.

In order to mitigate the impacts of arbitrary choices of trip duration on cumulative accessibility analyses while keeping its computation and communicability advantages. ***"A time interval metric for cumulative opportunity accessibility"*** (Tomasiello D. B., 2022) the authors introduce a new accessibility metric , the time interval cumulative accessibility measure., that The author describes the Cumulative accessibility as the measures which allow one to estimate the number of opportunities that can be reached within a given travel time threshold. They have become the most commonly used metric in transport research and planning because of how simple it makes to calculate and communicate accessibility results; nonetheless the authors states that an important limitation of cumulative opportunity measures is that they require an arbitrary choice of a single travel time threshold To mitigate the negative impacts of choosing ad-hoc travel time thresholds when conducting accessibility analyses, they proposed a new time interval cumulative accessibility metric. Whilst in the standard approach each travel time cut-off had a correspondent accessibility estimate, in the proposed one, each time interval also returns a single accessibility value, but which is calculated as the average (or median) of several accessibility estimates (one per minute) within the interval.

In this paragraph we explored studies focusing on coverage rate and cumulative opportunity indices. In our study we considered these concepts, defining a coverage rate based on the ratio between number of inhabitants reached from an origin point through public transport vehicles following their scheduled paths and the number of inhabitants living in a geometrical buffer area drawn around these points. The trip duration had been chosen arbitrarily.

3. Tools and data

In this work we essentially processed open-source data about public transport night lines, amenities distribution and population distribution of a set of cities in a specific study time of two hours at late night. The goal of these initial manipulations was to identify a population distribution, a set of amenities density-based clusters (which we called Night Life Areas - NLA) and information about the length and frequency of the public transport offer. In this chapter all the data sources and the subsequent manipulations needed to reach the efficiency percentage result will be explained. Specifically, the following issues will be illustrated:

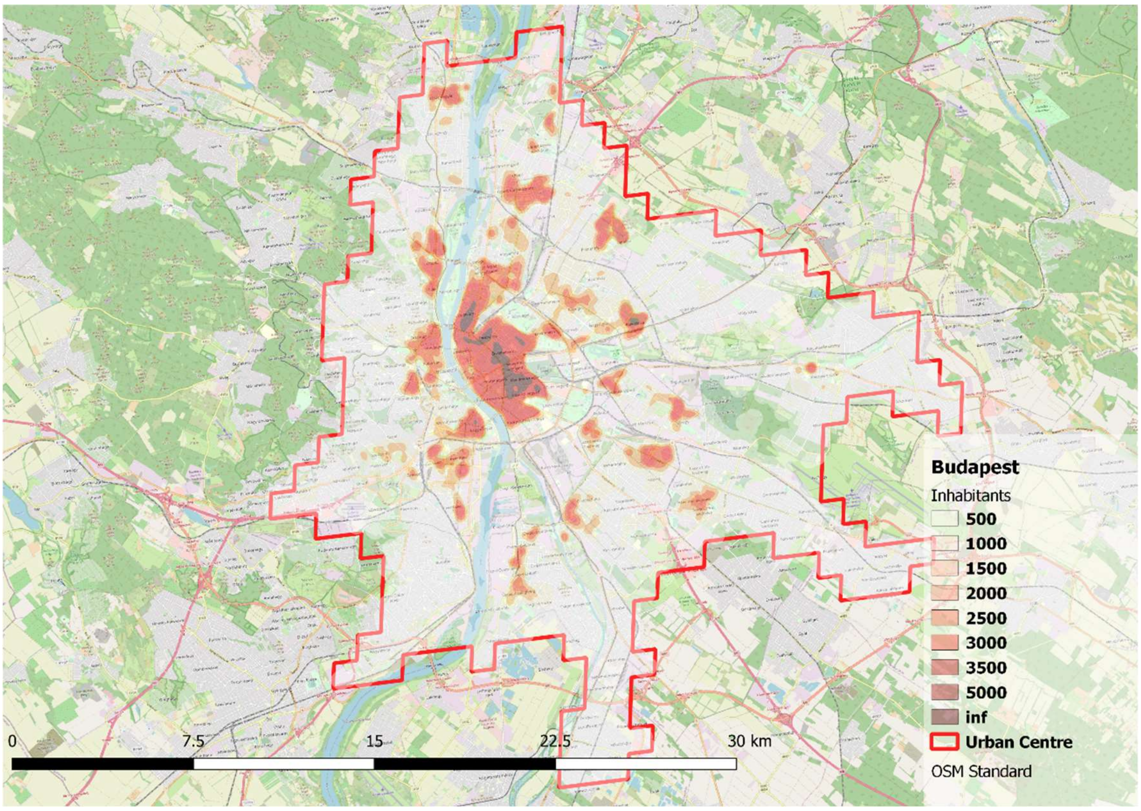
- In 3.1 the selection of cities that are considered in this work is presented.
- In 3.2 the extrapolation of Public Transport Networks information in GTFS format from agencies or third part data collectors and the successive visualization in QGIS is shown
- In 3.3 the adopted study area delimitations are presented, and the procedures of download and manipulation on QGIS of the population density distributions inside those borders are explained.
- Chapter 3.4 shows how it was possible to pass from Points Of Internets representing the leisure related activities for each city (downloaded on open sources) to a set of clusters representing the Night Life Areas of a city.
- In 3.4 we entered in the details of the so-called reachability computations on QGIS with the plugin Traveltime which allowed us to compute the efficiency percentages,

3.1. Selection of the cities

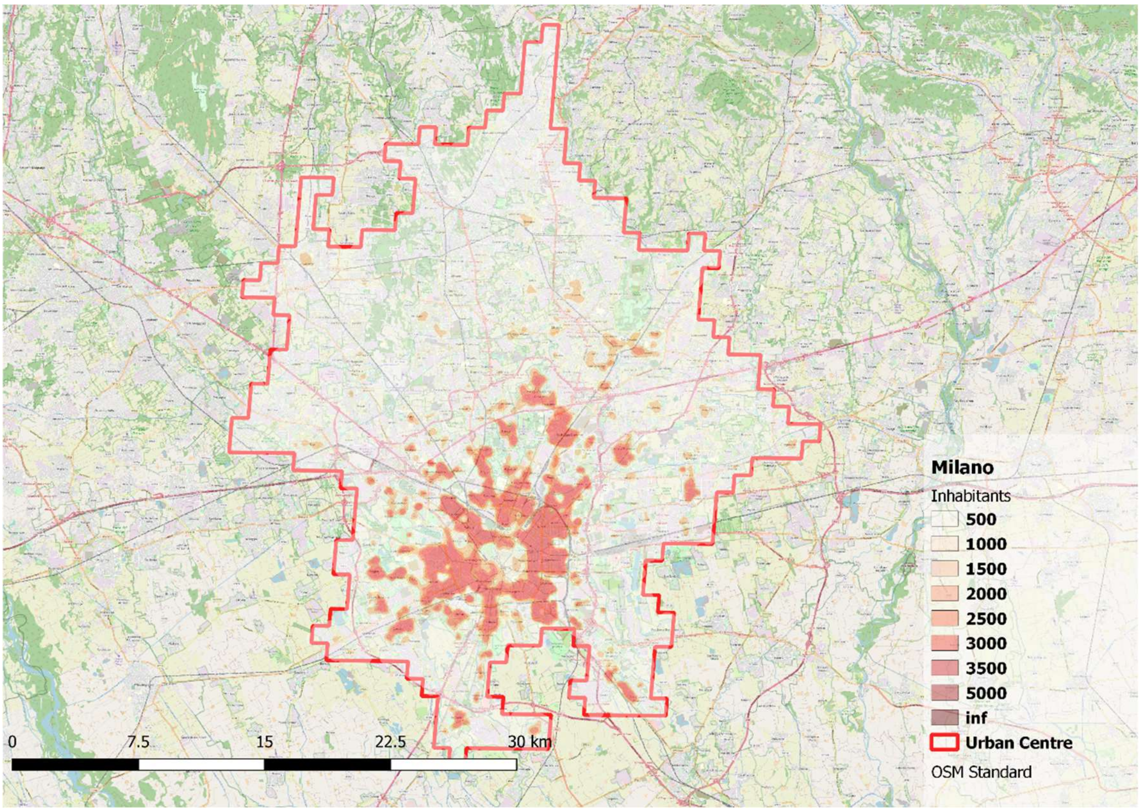
The first step of the work has been to define a sample of cities on which the proposed analysis could be showcased.
Their selection is based on the following criteria:

- Availability of General Transit Feed Specification (GTFS) data updated at least once after 2021 and with available file “shapes.txt”, so that it was possible to have information about the length of the lines. These data will be later introduced in section 3.2.
- Availability of information about night public transport lines through the Traveltime plug-in defined running the command “Isochrone” of such plug in at a night time hour of a weekend day (Sunday early morning for example). As better discussed in section 3.5 from the created isochrone it is possible to visually state the availability or absence of night lines data.

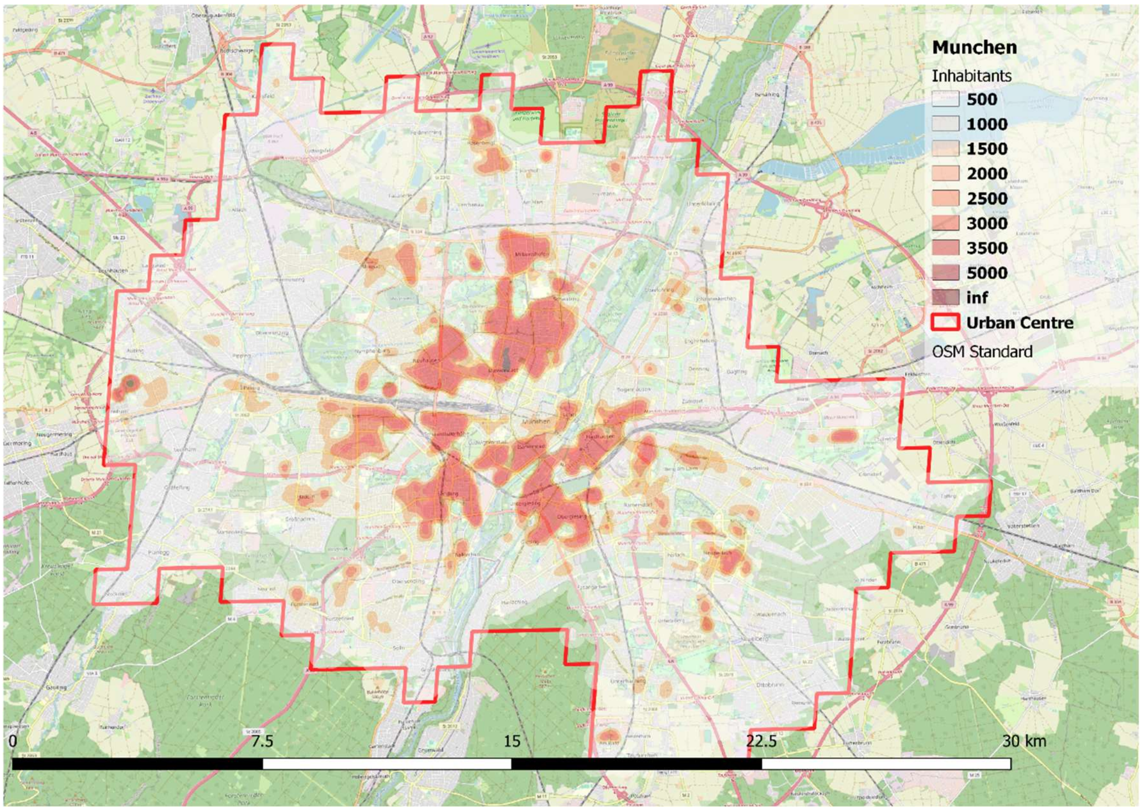
The study could have been made on every city respecting the above requisites; we considered a sample composed of eight cities, chosen trying to give heterogeneity in terms of country provenience, inhabitants, and urban conformation. The considered cities are the ones in Figure 1a-h, where heatmaps showing the density of each population are shown in order to have an idea of the shape of each city and the urban sprawl of it. The name of each city and the number of inhabitants within the considered Urban Centre (which is not the territory of the municipality or any other administrative entity, as it will be discussed in section 3.3) are also specified in the corresponding captions. The eight studied cities are also shown in a continental map in Figure 2.



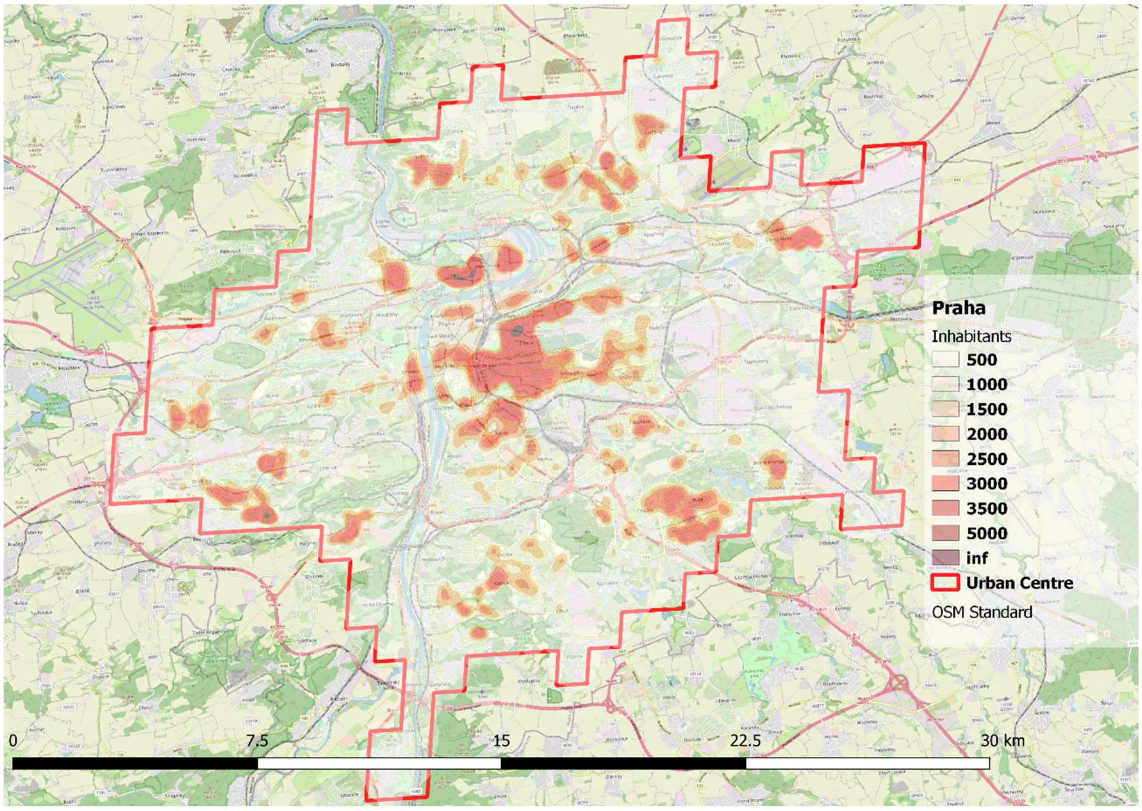
a) Budapest: 1,723,314 inhabitants



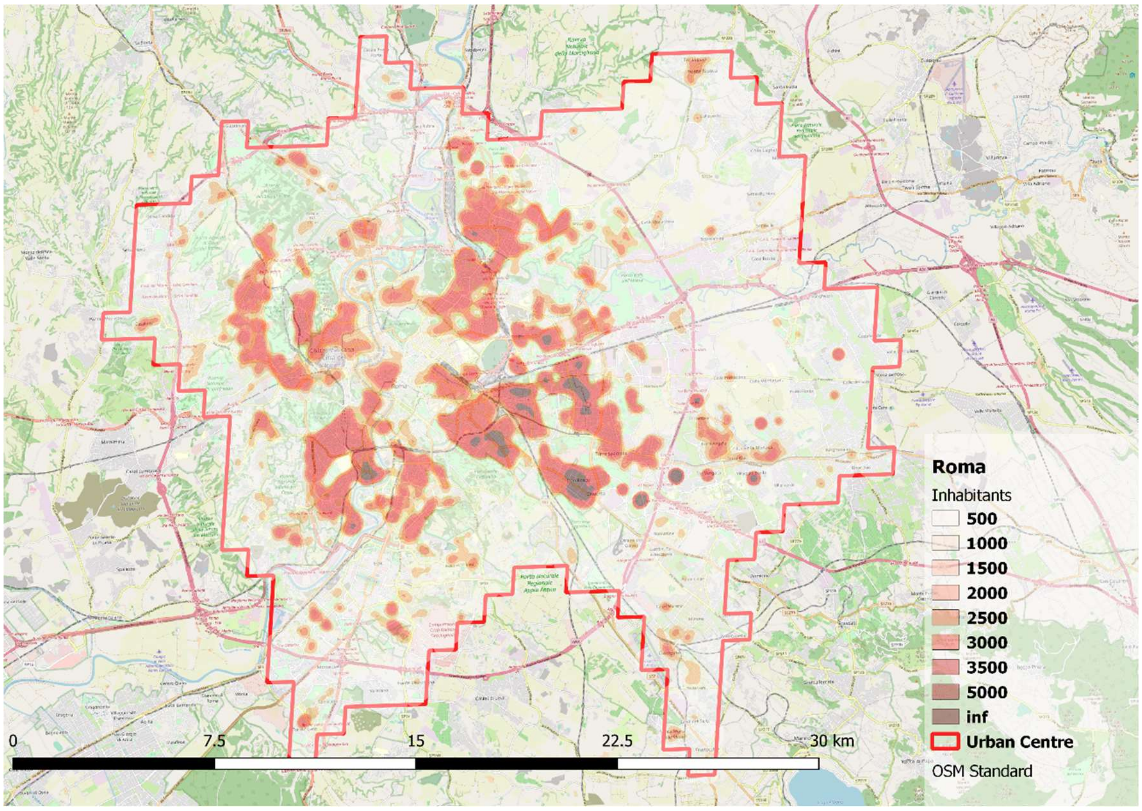
b) Milano: 3,157,550 inhabitants



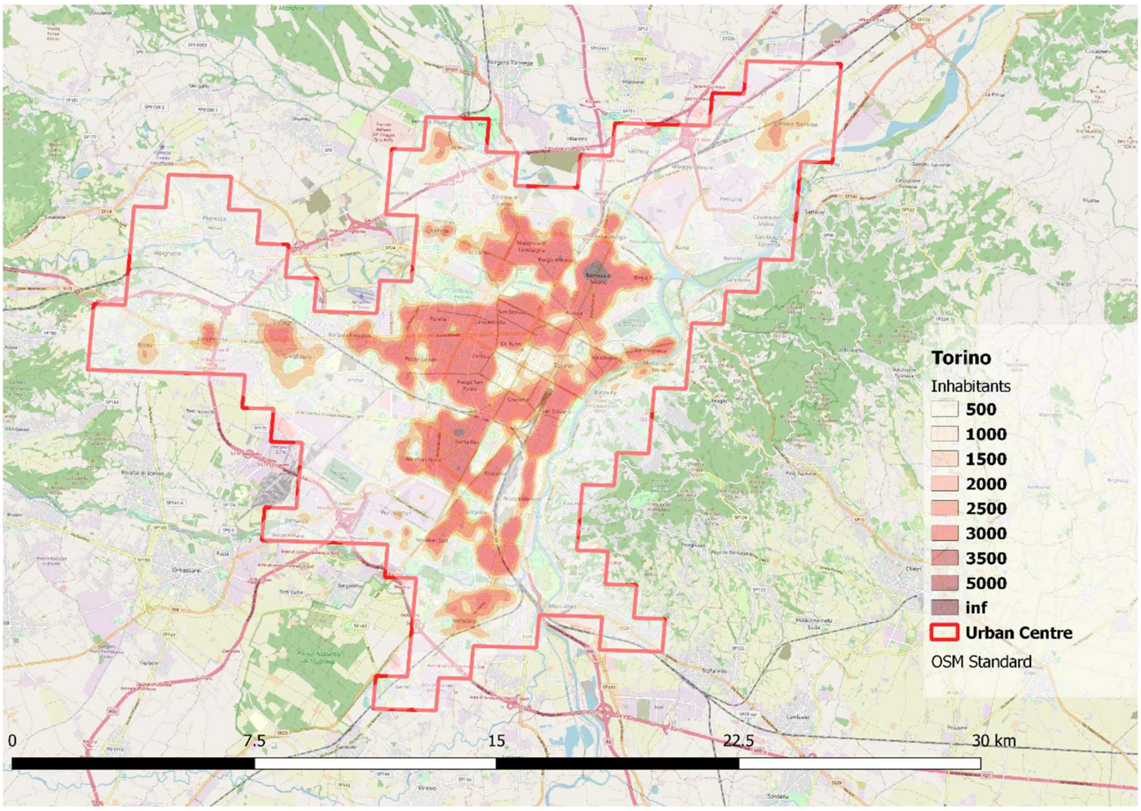
c) München: 1,582,533 inhabitants



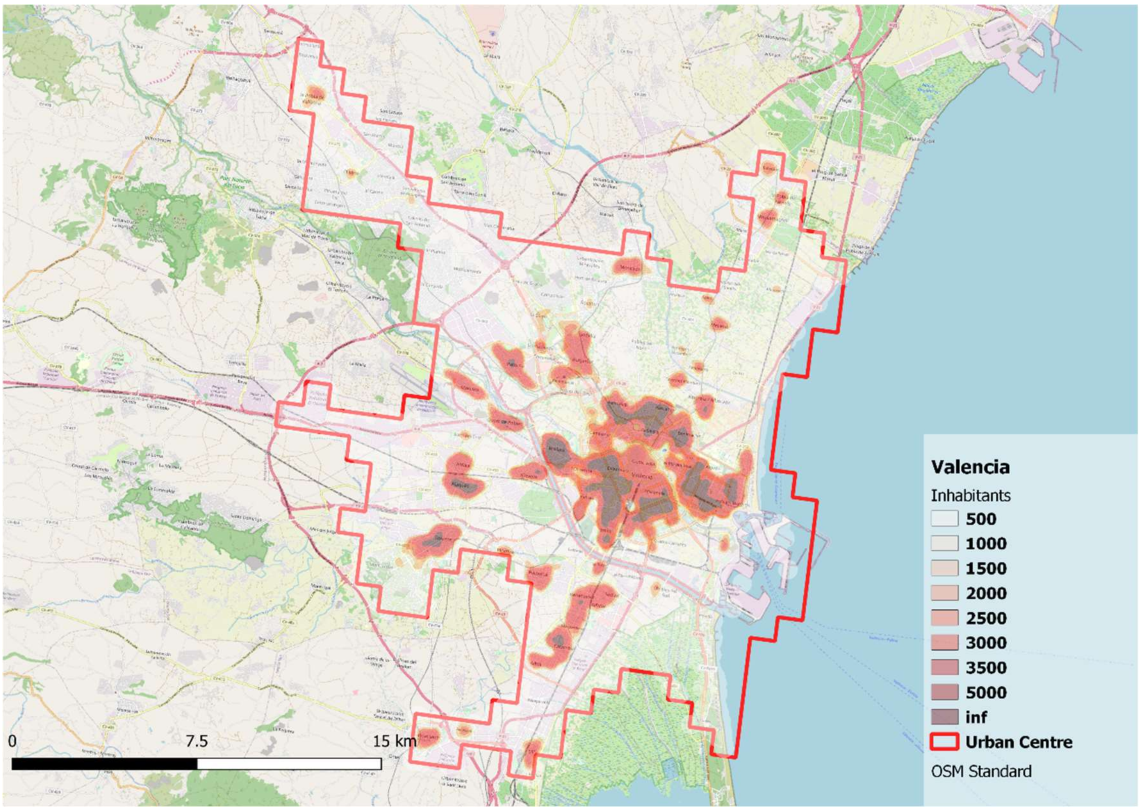
d) Praha: 1,194,604 inhabitants



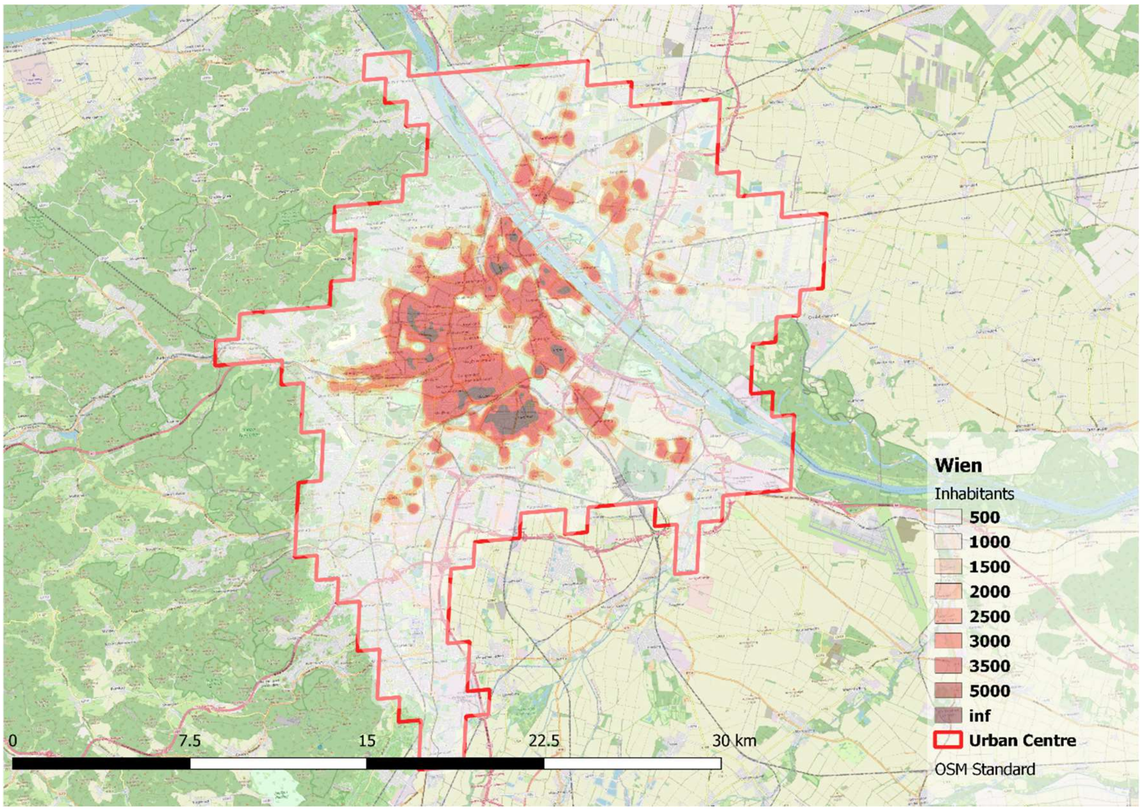
e) Roma: 2,493,124 inhabitants



f) Torino: 1,207,163 inhabitants



g) Valencia: 1,514,210 inhabitants



h) Wien: 1,984,266 inhabitants

Figure 1- Studied cities delimited by the Urban Centres borders with information about number of inhabitants and their distribution. a) Budapest, b) Milano, c) München, d) Praha, e) Torino, f) Roma, g) Valencia, h) Wien

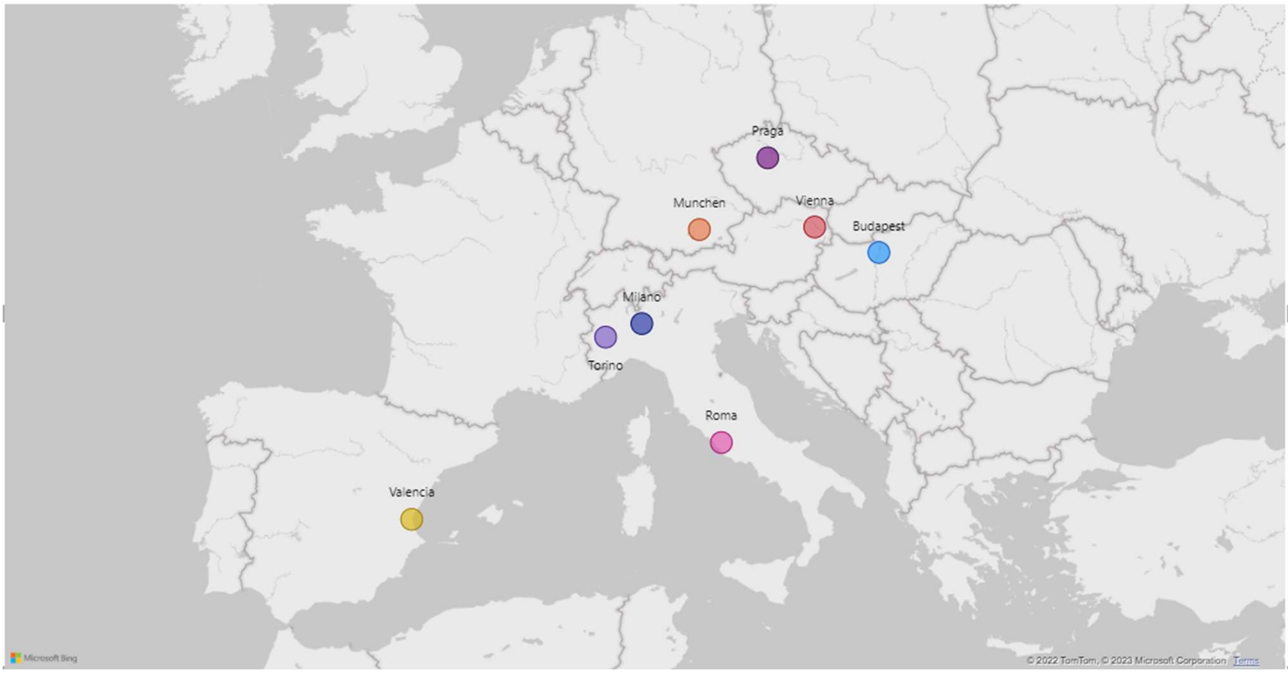


Figure 2 - Geographical distribution of the studied cities

-

3.2. Extraction of the public transport lines from GTFS data

We collected data about the public transport agencies in the GTFS format (General Transit Feed Specification) for each city mentioned in section 3.1, to have a visual overview and to gain information on the extension of the night lines. GTFS is a standardized format for describing public transit schedules and geographic information. It allows transit agencies to publish their data, making it easier for developers to build applications that can use this data. GTFS presents itself as a collection of text files which describe routes, stops, and schedules, and it is widely used as a data exchange format for public transit information. We researched and downloaded the dataset for each city from the related public transport company website or an associated data collector agency when available, but in most cases, it was necessary to download data from a third part website which contains different GTFS for different cities, namely Transitfeed[1]. When data were available from different sources, the most recent one (as of January 2023) has been considered. Table 6 contains the sources of each studied city's GTFS.

Table 6 – Sources of GTFS for each studied city

City	GTFS source	Link	Release date
Budapest	Third part	https://transitfeeds.com/p/bkk/42?p=2	12/01/2023
Milano	Third part	https://transitfeeds.com/p/agenzia-mobilita-ambiente-territorio/341	08/09/2021
München	Local transport agency	https://www.mvg.de/services/fahrgastservice/fahrplandaten.html	22/12/2022
Praha	Third part	https://openmobilitydata.org/p/praha/801	06/01/2023
Roma	Municipality	https://dati.comune.roma.it/catalog/dataset/c_h501-d-9000	30/12/2022
Torino	Municipality	http://aperto.comune.torino.it/gl/dataset/feed-gtfs-trasporti-gtt/resource/88035475-b429-40de-ba64-99de232d6327	3/12/2022
Valencia	Third part	https://transitfeeds.com/p/emt-valencia/719	10/08/2021
Wien	Country goverment	https://www.data.gv.at/katalog/dataset/wiener-linien-fahrplandaten-gtfs-wien#resources	22/06/2020[2]

The GTFS dataset presents itself as a zip folder containing the csv files described in

Table 7, each of them carrying one or more identification codes which allowed us to join the files together, in order to add information step by step.

[1] https://transitfeeds.com/

[2] Release date in summer 2020, right after COVID 19 crisis. Nonetheless the lines included in this file are the same lines working nowadays (April 2023).

Table 7 – Description of files contained in a GTFS format file, (Google, 2022)

File	Identification code(s)	Description	Priority
agency	agency_id	Transit agencies with service represented in this dataset.	Required
routes	route_id agency_id	Transit routes. A route is a group of trips that are displayed to riders as a single service.	Required
shapes	shape_id	Rules for mapping vehicle travel paths, sometimes referred to as route alignments.	Optional
stop_times	trip_idstop_id	Times that a vehicle arrives at and departs from stops for each trip.	Required
stops	stop_id zone_id	Stops where vehicles pick up or drop off riders. Also defines stations and station entrances.	Required
timetables	timetable_id route_id direction_id timetable_page_id	Information about the passage times of means at each stop	Required
trips	route_idservice_id trip_iddirection_id block_idshape_id	Trips for each route. A trip is a sequence of two or more stops that occur during a specific time period.	Required

In Figure 3 a diagram which explains how these different files can be linked with each other through their identification codes is given. The diagram shows with blue sky boxes the Entities, which are the .txt files containing information exposed in

Table 7. The green boxes, renamed Spatial entities, are those .txt files containing geographical information, which allow the network visualization on a GIS platform; the ones showed in the diagram, namely "Shapes" and "Stops" are both point layer: the first one contains a massive series of points from which it is possible to deduce the lines composing the routes and so the network, the second one contains a point for each bus/ tram/ train/ underground/ etc stop. The white box, pointed out as a virtual entity, describes a .txt file which has no information and just allows the connection to other entities. The connections are determined, as explained, by identification codes which allow the joint of different entities. They are expressed in the diagram as arrows renamed "Foreign key" and "Foreign key – optional" and show which entities are connected between each other.

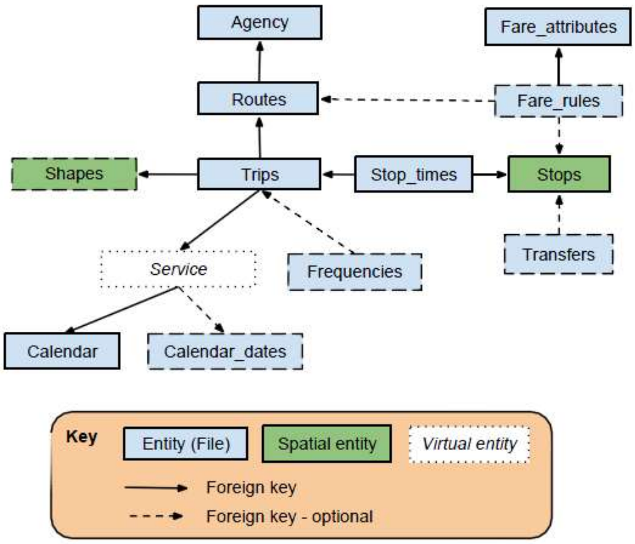


Figure 3 - Diagram of the data model (Davis, 2011)

Once we collected the data about the public transport offer of a city in GTFS format, we can translate it in visual representation through a GIS platform, which is, in our case, QGIS.

3.2.1. Inserting GTFS in QGIS

To use the GTFS dataset in QGIS we started uploading the file shapes.txt through the *add a delimited text layer* tool: since this file contains geographic information, we gave a geometry definition as shown in Figure 4 and we obtained a point layer well positioned in space (Figure 5).

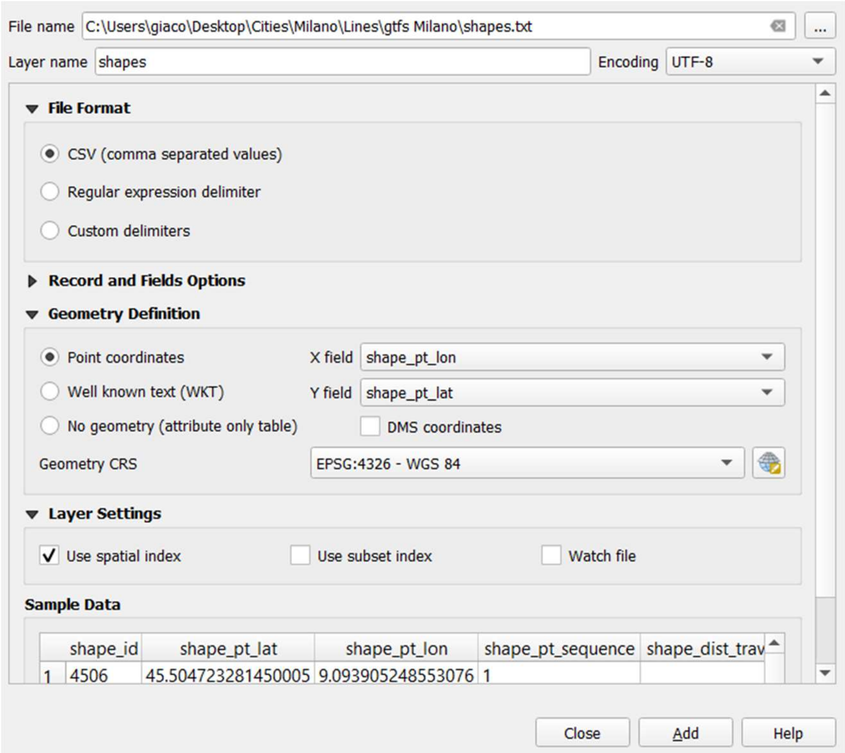


Figure 4 – Add a delimited text layer.

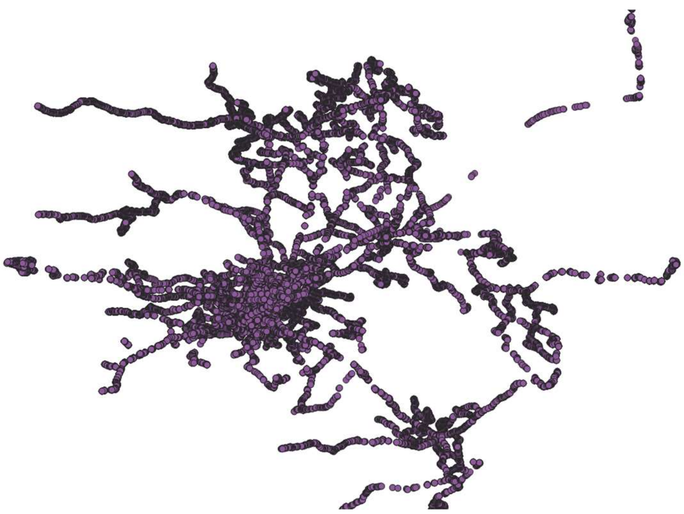


Figure 5- Points layer "shapes" obtained by uploading on QGIS the file shapes.txt (Turin)

These points identify all the different trips, according to the definition of trips within the GTFS (as reported in Table 7) which is not the same definition of trip generally given in transport planning. Indeed in transport planning a Trip is defined as a one way person movement by a mechanized mode of transport, having two trip ends (Sakshat virtual lab , s.d.), while in our specific case, as defined in

Table 7, is meant as a series of movements over an established route; taking as example an hypothetical bus number 10, each bus will follow the specific route number 10; every time the route number 10 is travelled (the bus goes from an headway to another, in one of the two possible directions) a trip has been made. So, every day multiple trips are conducted over a route. To visualize these trips as lines, we used the tool *points to path* and created the layer "paths" (Figure 6 and Figure 7). In QGIS, *points to path* is a vector geoprocessing operation that converts a series of point features into a single line or polygon feature. The operation connects the points in the specified order based on a chosen attribute field.

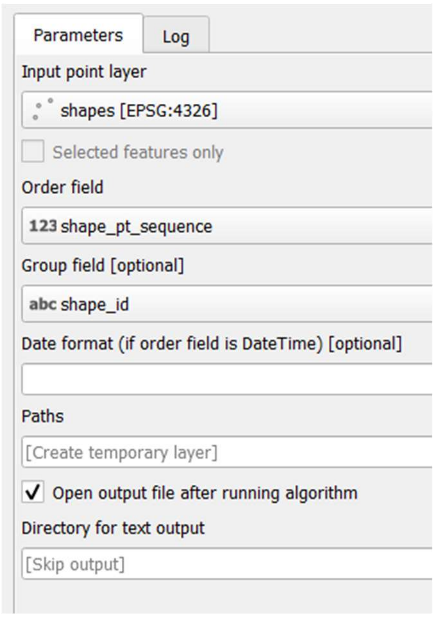


Figure 6 - Points to path

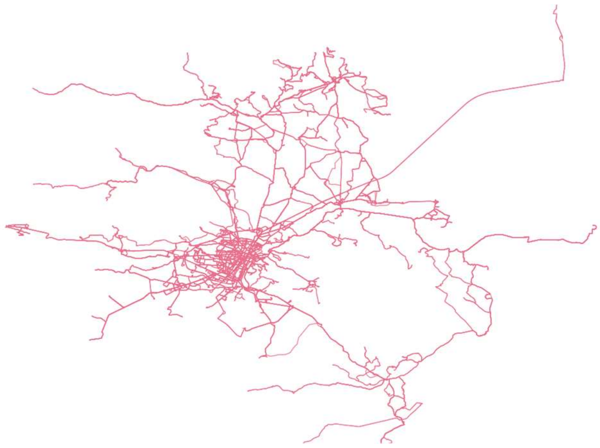


Figure 7- Results of Points to path (Torino)

At this point we could visualize the trips, but we didn't have actual information about them. So, we added the vector layer trips.txt as we did for shapes.txt and then performed a *join* through the field "shape_id". In QGIS, *join* refers to the process of combining two or more datasets based on a common attribute, such as a shared ID or name. Since trip.txt has no geographical references, we didn't specify a geometry definition during the adding procedure (Figure 8). We *joined* the layer "trips" to the existing layer "paths" (which we created aligning the points of the layer shape).

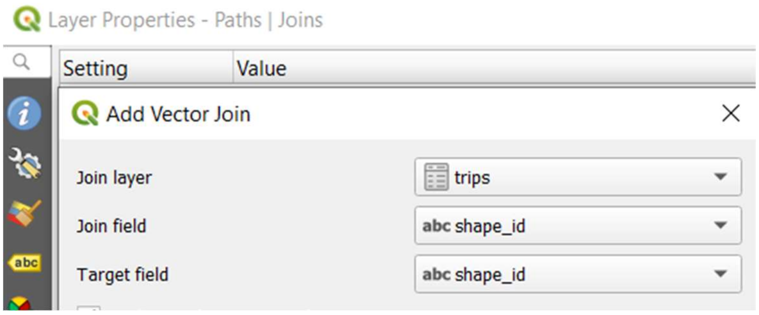


Figure 8 - Join layer.

To see the routes and have information about them we added the file routes.txt as before and *joined* it to the layer "paths" as well, through the identification code "route_id". Since we wanted to see the routes only and not all the trips, we performed a *dissolve for* the layer "paths" through the attribute "route_id": doing so we dissolved all the trips contained in a route in order to show the routes only. Specifically, *dissolve,* in QGIS, is a vector geoprocessing operation that aggregates features in a vector layer based on a specified attribute field. The *dissolve* operation merges features with the same attribute value into a single feature, reducing the number of features in the output layer. Doing so, we lose information about the single trips, which was anyway not of our interest, since we wanted to see just the routes. The only reason why we used the file trips_txt (with related information over the trips) was that we needed an intermediate identification code to connect the information about the shapes of the lines to the travel informations. At this point, the goal was to show only the routes related to the night lines to do so we checked online in the service providers agencies which lines were running in the night time service, and we extracted them from the totality of the network.

All of this can be implemented in a ***Processing Model*** as shown in Figure 9. In QGIS a processing model is a set of procedures and algorithms, so we used it to automatize the workflow previously explained. The yellow boxes define the inputs, which had to be selected time by time between the uploaded layers. The white boxes define the algorithms which worked on the chosen inputs to which they are linked. The inputs can be a new inserted one (yellow box) or can be the result of a previous algorithm (coming from a white box) When the result of an algorithm is a definitive (or even a partial) one that the user wants to export as an output, the creation of an output layer is demanded, and it will be shown as a green box. In Figure 9 we can finally see, following the explained notations, a resume of all the manipulations related to the public transport lines extraction.

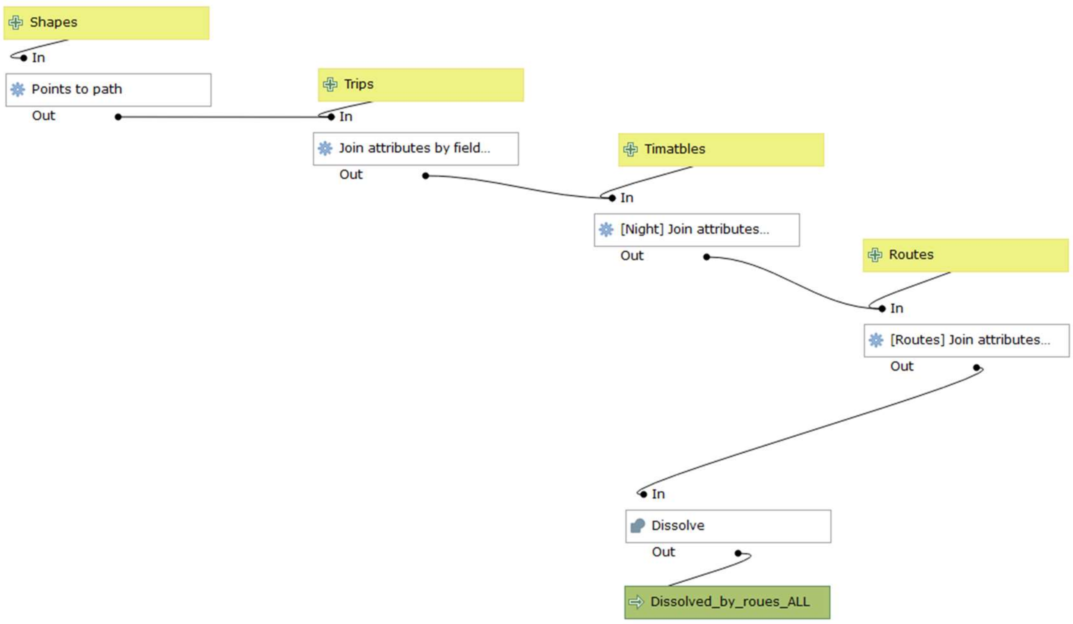


Figure 9 - Processing model for Offer Network visualization.

3.2.2. Study time

For each city we defined a study time, i.e., a range of 2 hours falling inside the operating hours of the night time service when headways are minimum. With the term headways, we refer to the time interval between successive vehicles along a particular route or line. The choices were made separately for each city in order to avoid, when possible, the influence of day-time services or the absence of any. In Table 8 the service hours of the eight night time services and the selected peak hours are displayed.

Table 8 - Service and peak hours for each city

City	Service hours	chosen two hours
Budapest	23:00- 05:00	01:00 – 03:00
Milano	00.30 – 06.30	03:00 – 05:00
München	01:30 - 04:30	02:00-04:00
Praha	23.30 – 05.00	02:00-04:00
Roma	00:00 - 05:00	02:00 – 04:00
Torino	00:00 - 05:00	02:00 – 04:00
Valencia	00:00 – 3:30	01:00 – 03:00
Vienna	00.30 – 05.00	02:00 – 04:00

For each city we estimated through reachability computations (explained in 3.5) the number of inhabitants whose residential location can be reached from each NLA within a specific travel time by public transport during the above mentioned study time. We performed a reachability computation every 15 minutes during the two hours' time range, resulting in a total of 9 computations. Having a time window of two hours with a resolution of 15 minutes allowed us to visualise all the reachability patterns due to different frequencies of different means. Indeed, in our study sample of eight cities and their eight services we noted frequencies going from a minimum of 15 minutes to a maximum of 2 hours. We also noticed as in some cities the frequencies vary from time to time: taking a two-hour span allows us to take a general picture of the service so that we do not under- or overestimate it.

3.3. Population density distribution

In order to define the study area of each city in a consistent way throughout different European countries, we used the EU-OECD Functional Urban Area definition for Urban Centres, which we later populated with high - resolution (at block level) geospatial data on population distribution. As from the EU-OECD definition, an urban centre is a cluster of contiguous 1 km^2 population grid cells with a density of at least 1500 inhabitants / km^2 and a total population of at least 50 000 inhabitants (Dijkstra, Poelman, & Veneri, 2019). A Population Grid is a Geospatial tessellation of regular squared cells with population counts or estimates which gives realistic depictions of actual population distribution, and it is a convenient format for overlaying with other spatially continuous variables (e.g., environmental) (Silva, Poelman, & Dijkstra, 2021). An explanation on the urban centres definition process is reported. Once the population grid with 1 km^2 grids is obtained, it is possible to generate the urban centre of a specific city in three steps (Figure 10):

1. All the grid cells with a density of more than 1500 inhabitants are selected.
2. The cells answering the previous criteria, which can be defined as high-density cells are clustered together: only the clusters with a population of more than 50000 inhabitants are kept. Two cells have to share at least a face to be considered contiguous, sharing a corner is not enough.
3. The gaps within the cluster are filled and the edges are smoothed.

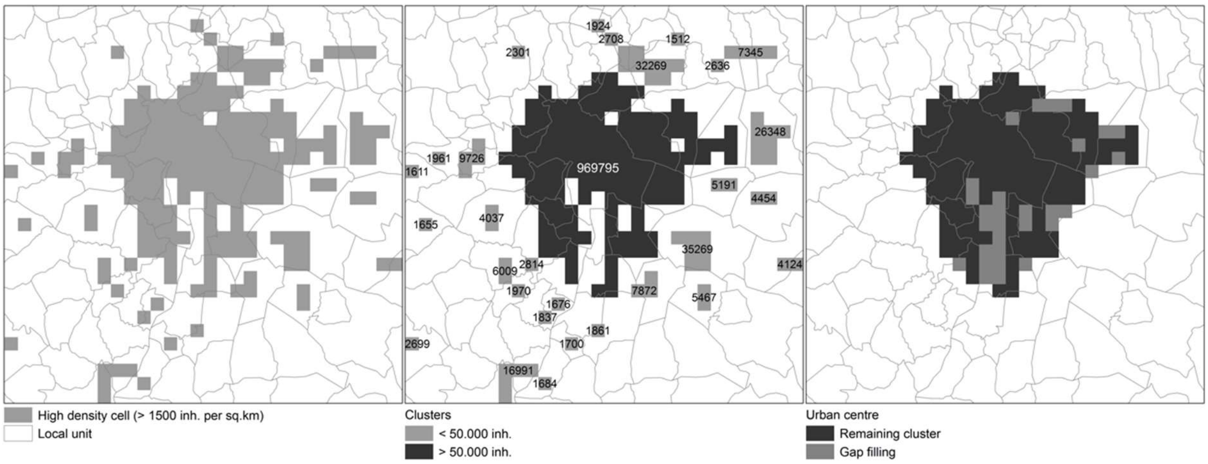


Figure 10 - Three steps of the urban centre definition of Toulouse, France (Dijkstra, 2019)

We didn't compute the urban centres by ourselves but downloaded them from the Urban Centre Database GHS-UCDB R2019A, which collect more than 10 000 urban centres identified following the above definition. They are available in shapefile for the download at the Global Human Settlement Layer (GHSL) project website. It can be read there that *the **Global Human Settlement Layer (GHSL)** project produces global spatial information about the human presence on the planet over time. This in the form of built-up maps, population density maps and settlement maps. This information is generated with evidence-based analytics and knowledge using new spatial data mining technologies. The GHSL processing framework uses heterogeneous data including global archives of fine-scale satellite imagery, census data, and volunteered geographic information. The data is processed fully automatically and generates analytics and knowledge reporting objectively and systematically about the presence of population and built-up infrastructures.* The project is supported by the Joint Research Centre (JRC) and the DG for Regional and Urban Policy (DG REGIO) of the European Commission, together with the international partnership GEO Human Planet Initiative (GHSL - Global Human Settlement Layer, 2023)

The shapefile we downloaded comes from the **Urban Centre Database UCDB R2019A**; in the same webpage is explained how *the database represents the global status on Urban Centres in 2015 by offering cities location, their extent (surface, shape), and describing each city with a set of geographical, socio-economic and environmental attributes. Urban Centres are defined* *in a consistent way across geographical locations and over time, applying the "Global Definition of Cities and Settlements" developed by the European Union to the Global Human Settlement Layer Built-up (GHS-BUILT) areas and Population (**GHS-POP R2019A**) grids* (Florczyk, et al., 2019).

Within the relation from (JRC - ISPRA EC, 2019) the "**GHS-POP R2019A** – GHS population grid multitemporal (1975-1990-2000-2015)" *is a spatial raster dataset that depicts the distribution and density of population, expressed as the number of people per cell. Residential population estimates for target years 1975, 1990, 2000 and 2015 provided by **CIESIN GPWv4** were disaggregated from census or administrative units to grid cells, informed by the distribution and density of built-up as mapped in the **Global Human Settlement Layer (GHSL)** global layer per corresponding epoch. The Urban Centres are delineated from spatial grids in equal area World Mollweide geographical projection at 1x1 km resolution. The visualisation of this map is provided in Web Mercator projection, resulting in an apparent visual distortion of the original grid shape.*

Finally, in SEDAC website is explained how the **CIESIN GPW v4** *is the Gridded Population of the World (GPW) collection from* SEDAC, *now in its fourth version (GPWv4). It models the distribution of human population (counts and densities) on a continuous global raster surface. Since the release of the first version of this global population surface in 1995, the essential inputs to GPW have been population census tables and corresponding geographic boundaries* (SEDAC, s.d.)

Since the aggregation level of 1 km^2 population grid cells was too high for our purposes, we populated the defined urban centres with higher precision population distributions. Namely, we used the population distribution at building blocks level offered by Urban Atlas, a land monitoring service offered by the European union programme Copernicus, a joint initiative of the Directorate-general for Regional Policy and the Directorate-general for Enterprise and Industry, with the support of the European Agency (European Environment Agency , 2017). This service was the first one to create harmonised land cover and land use maps over several hundreds of cities and their surroundings in the European Union and EFTA countries. The data about the population density distribution are updated at 2018. We downloaded from the Urban Atlas dataset the vector data in OGC GeoPackage SQLite format (reference system EPSG: 3035) for each of our studied cities. The downloaded data regarding each city are comprehensive of the entire Functional Urban Area (an example for the city of Turin is shown in Figure 11), which is an entity containing both the city and the urban centre.

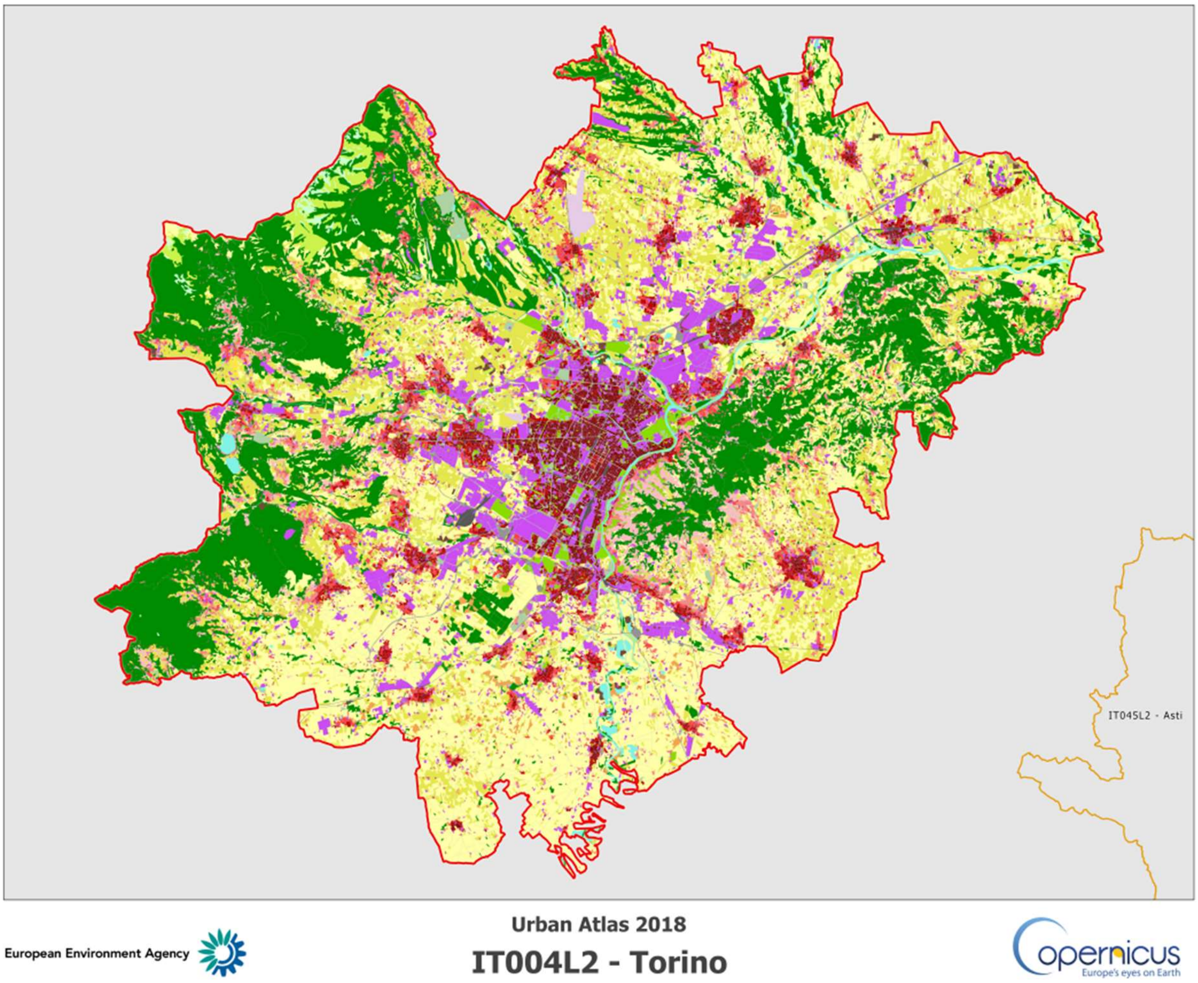
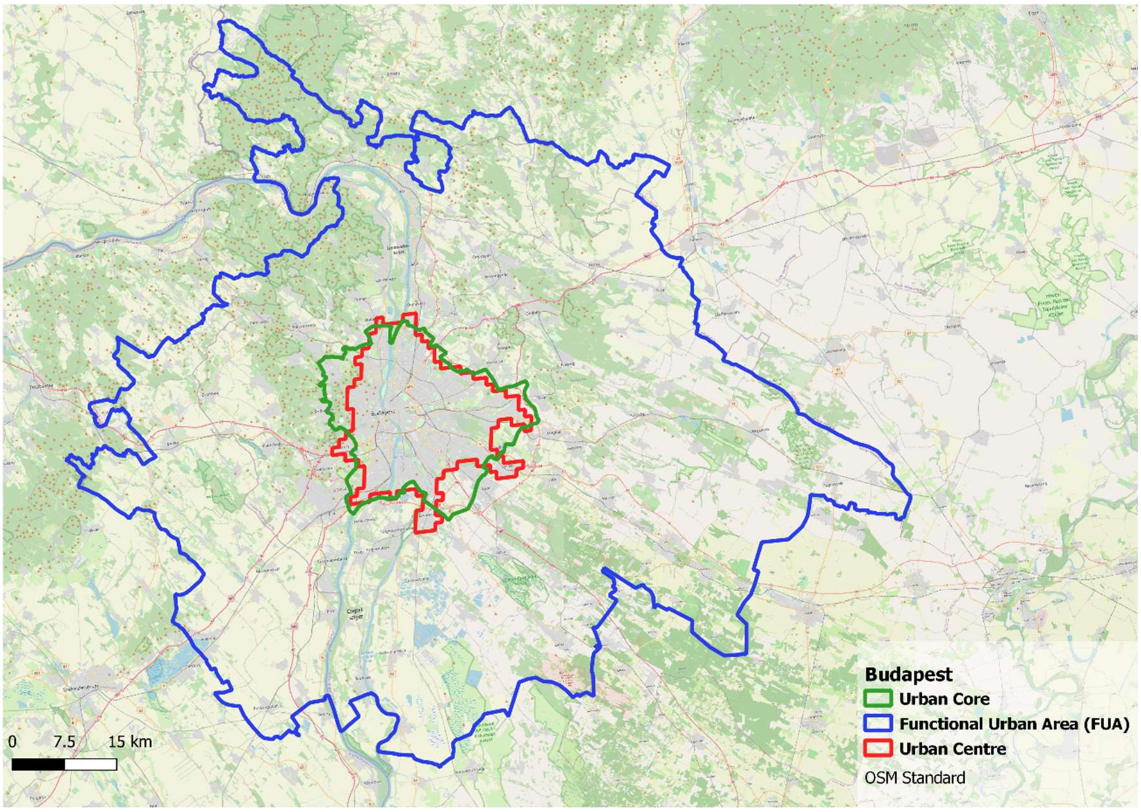


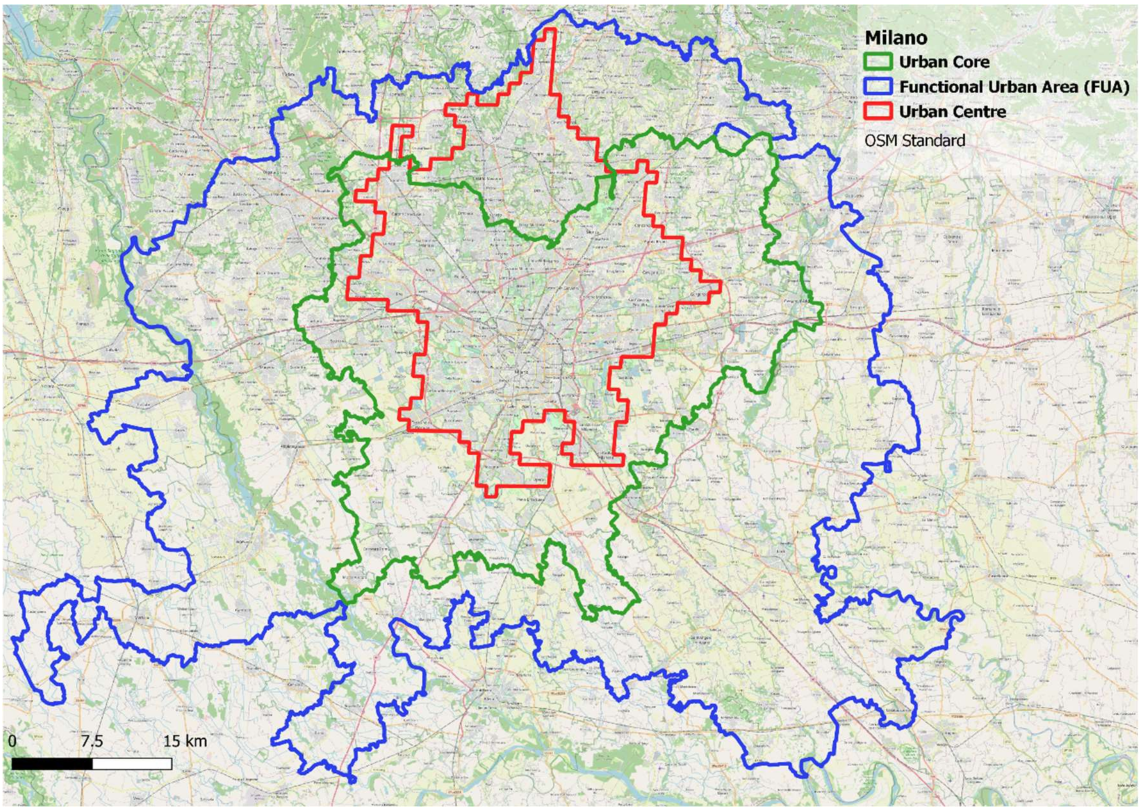
Figure 11 - Functional Urban Area of Torino. Each colour represents a different land usage.

Urban Atlas provides data not on urban centre scale but on Functional Urban Area one. Functional Urban Area (FUA), as defined by the EU-OECD definition, are composed of a city, to be intended as an area of high population density with a minimum size of population, and the commuting zone, a less densely populated surrounding which is still part of the city labour market. FUAs are defined in several steps: the first one is the creation of the urban centre, as already explained. Then the urban centre is adapted to the closest local units to define a city. Finally, commuting flows are used to identify which surrounding are part of the commuting zone. With the implementation of the commuting zone, FUAs aim to be more functional than administrative areas in capturing agglomeration economies.

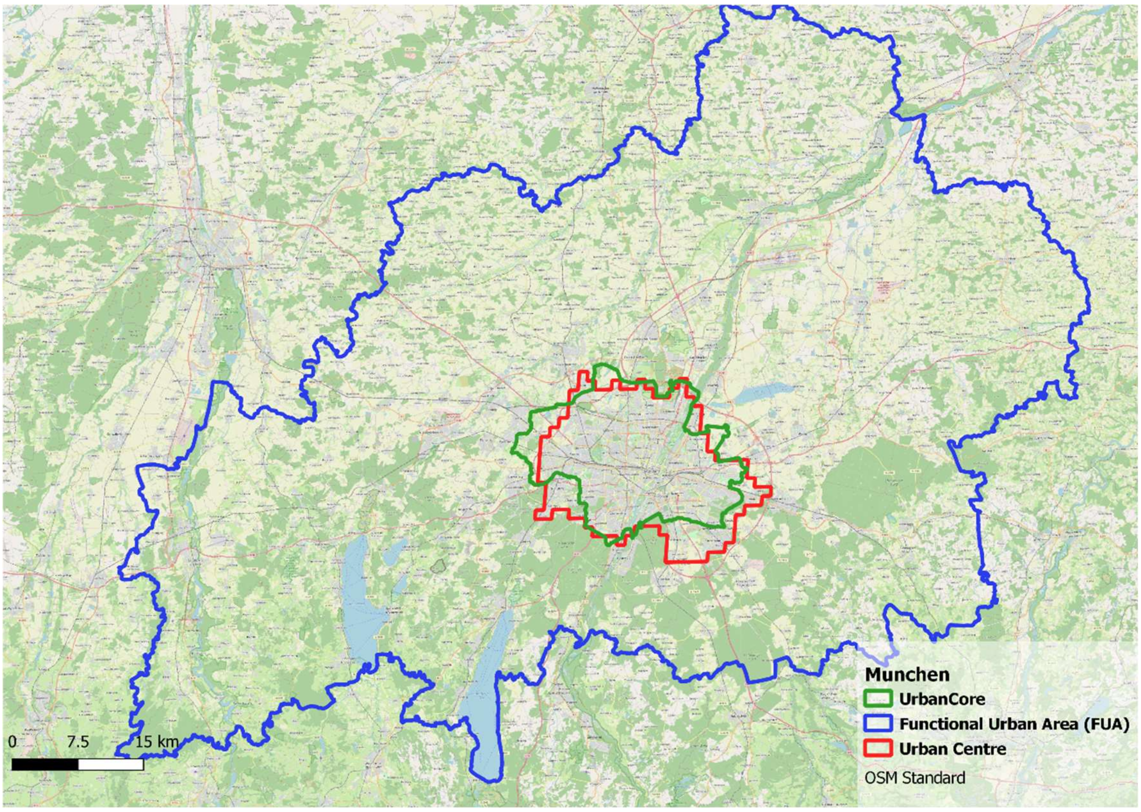
We can see in Figure 12 how the FUAs, the Urban Cores and the Urban Centres compare to each other for the eight considered case studies. Urban Core is a term used for "city", which is defined by the EU-OECD as one or more local units with at least 50% of their population in an urban centre, where a local unit can be both a statistical or an administrative one. In the majority of the eight studied cases, it corresponds to the municipality of the specific city but in two cases, namely Milano and Valencia, the urban core doesn't follow the administrative boundaries of the municipality, but it is designed over statistical local units differently individuated, such as census blocks.



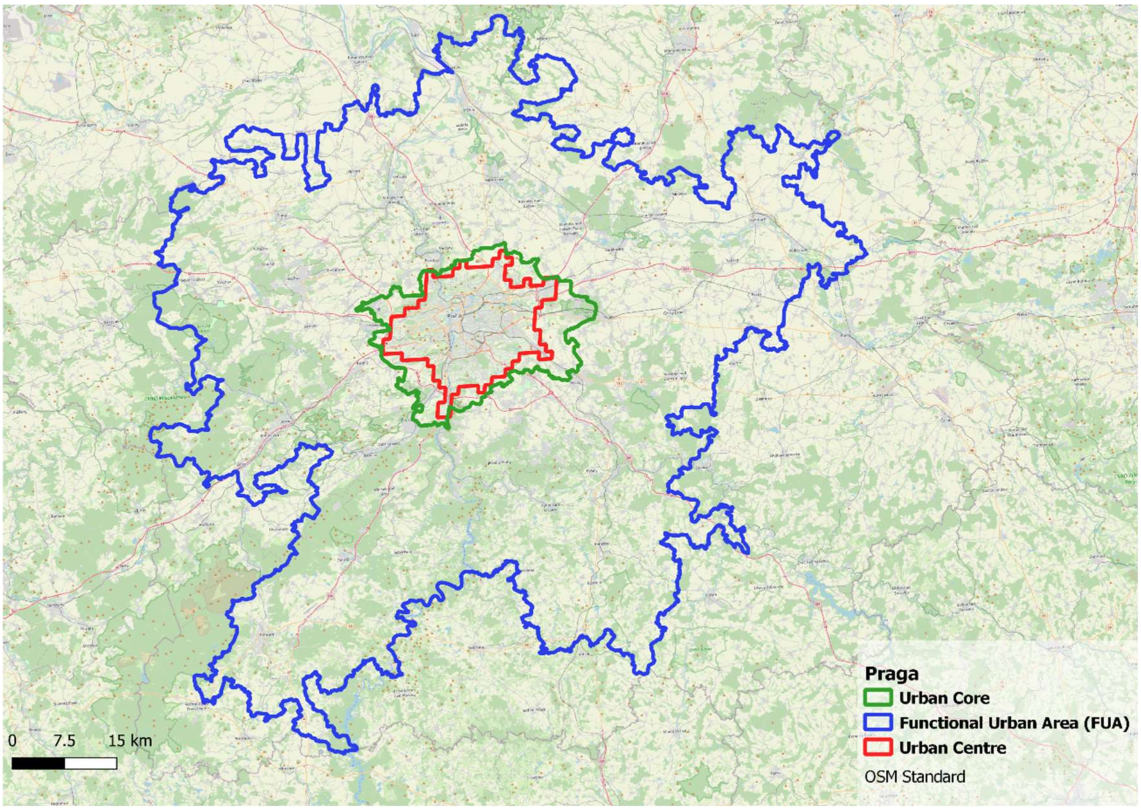
a) Budapest



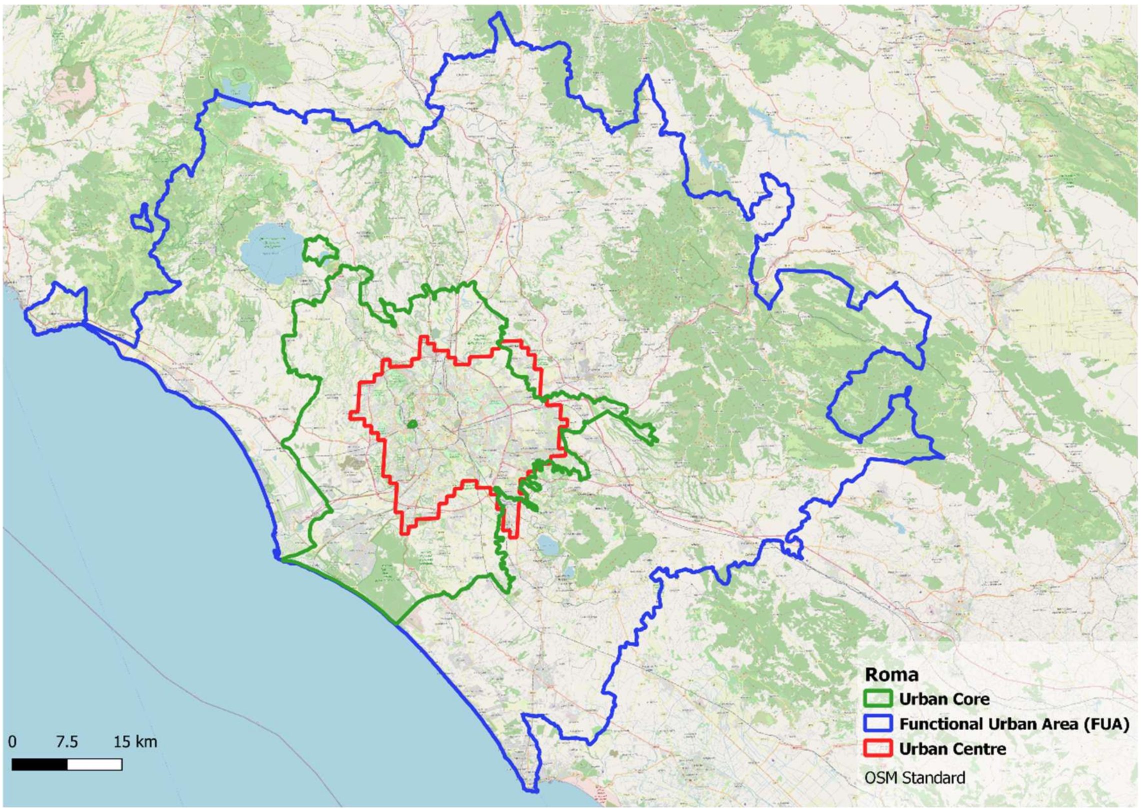
b) Milano



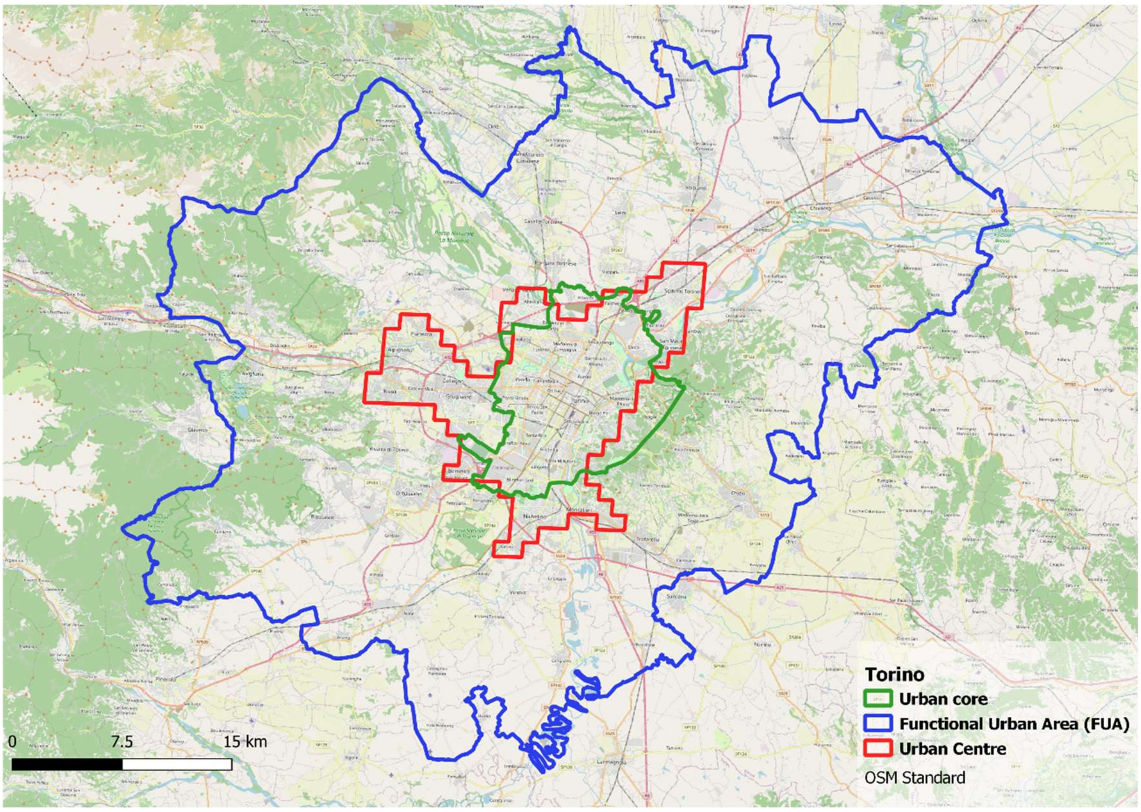
c) München



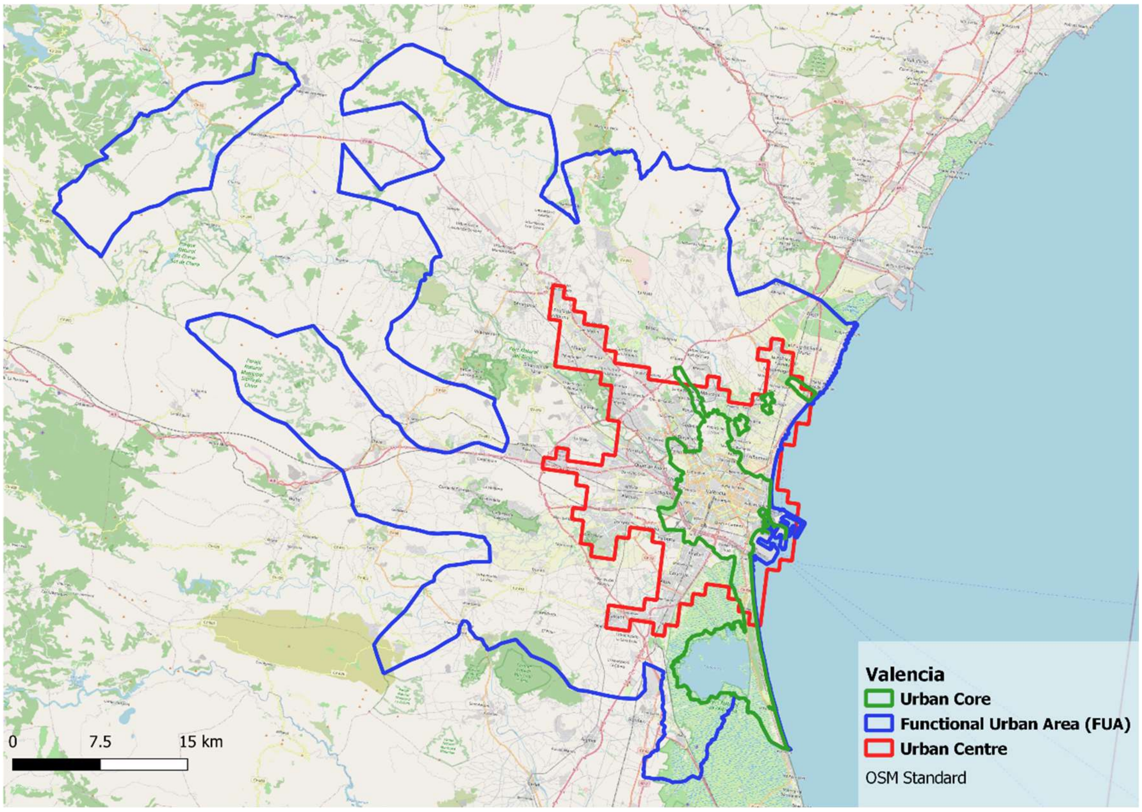
d) Praha



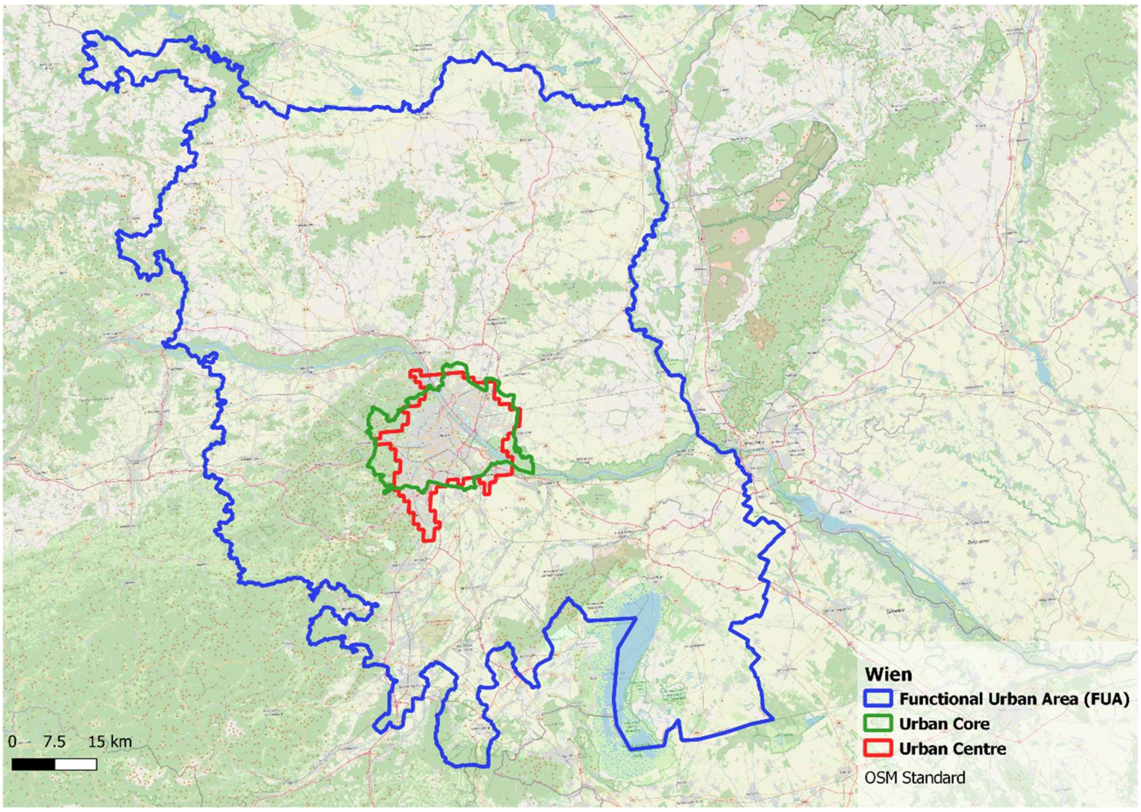
e) Roma



f) Torino



g) *Valencia*



h) *Wien*

Figure 12 - Comparison of FUA (blue), Urban Core (Green), Urban Centre (Red) for each studied city: a) Budapest, b) Milano, c) München, d) Praha, e) Roma, f) Torino, g) Valencia, h) Wien

Since we are interested in the urban centre only, we cut and kept the building blocks falling inside the urban centre borders (Figure 13 shows this for the city of Torino). To do so we used the tool *Extract by location* (Figure 14), so that a layer with only blocks falling inside the urban centre border has been created. The "Extract by Location" function in QGIS is a tool that allows users to extract features from one layer based on their spatial relationship with features in another layer (in this case the imposed relationship was *within*). We finally deleted the elements with a null population, which referred to streets and natural areas. The result only for the city of Torino is shown in Figure 15.

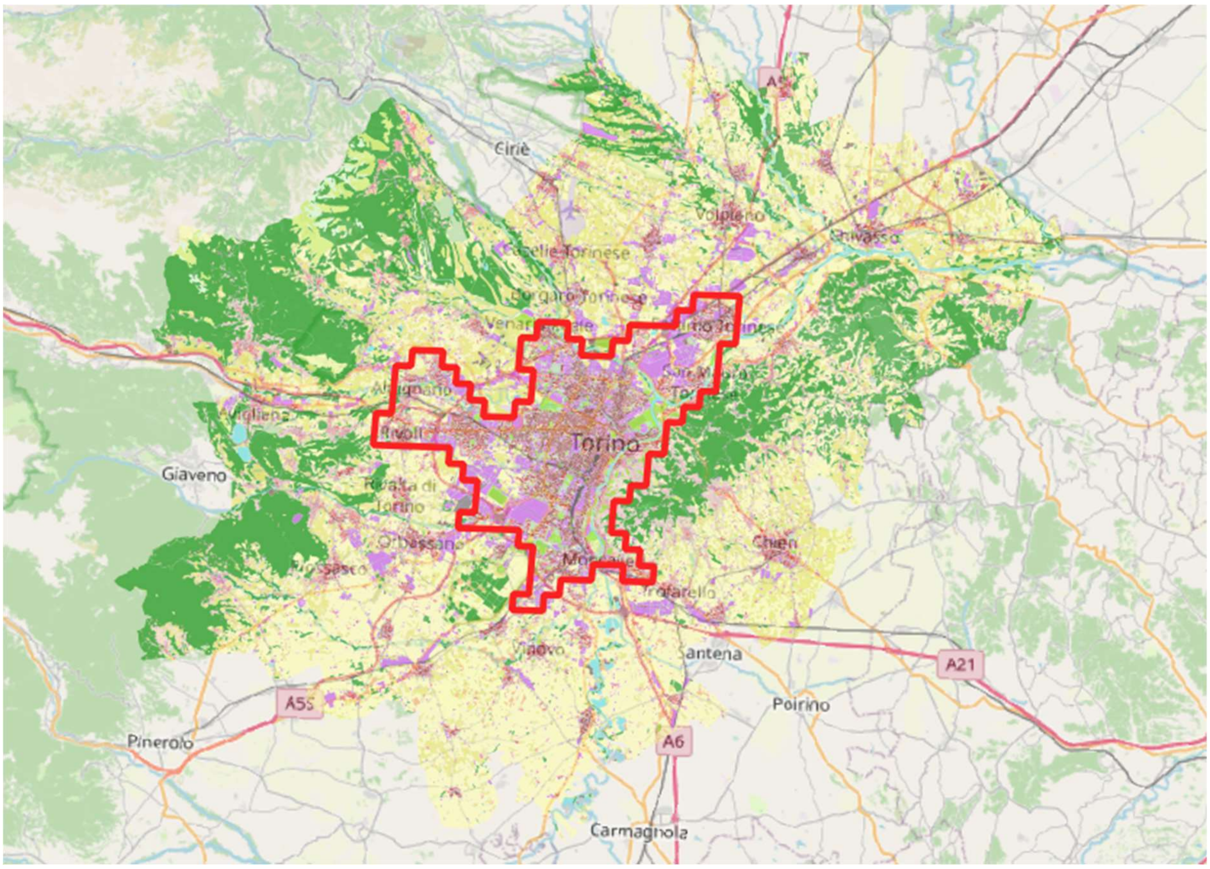


Figure 13 – In red the urban centre border is visible; only the blocks falling inside it will be considered.

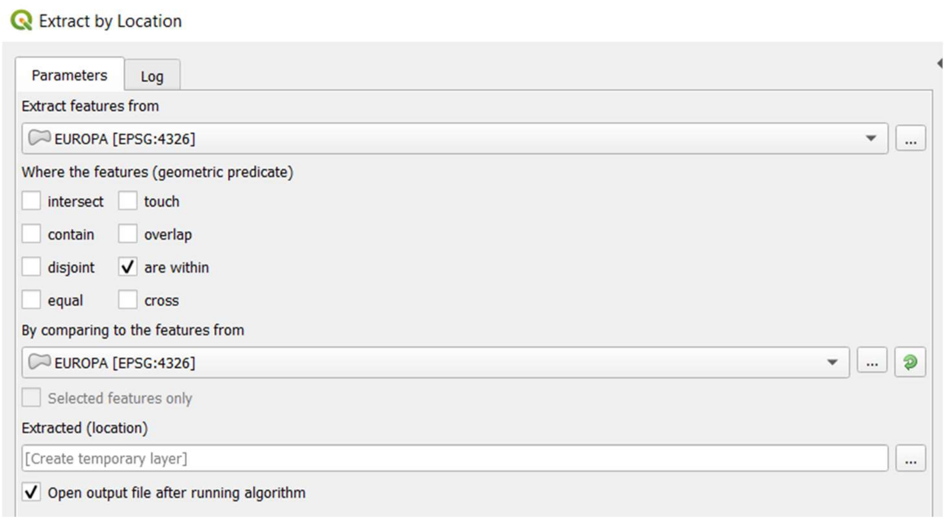


Figure 14 - Extract by location tool and parameters setting.

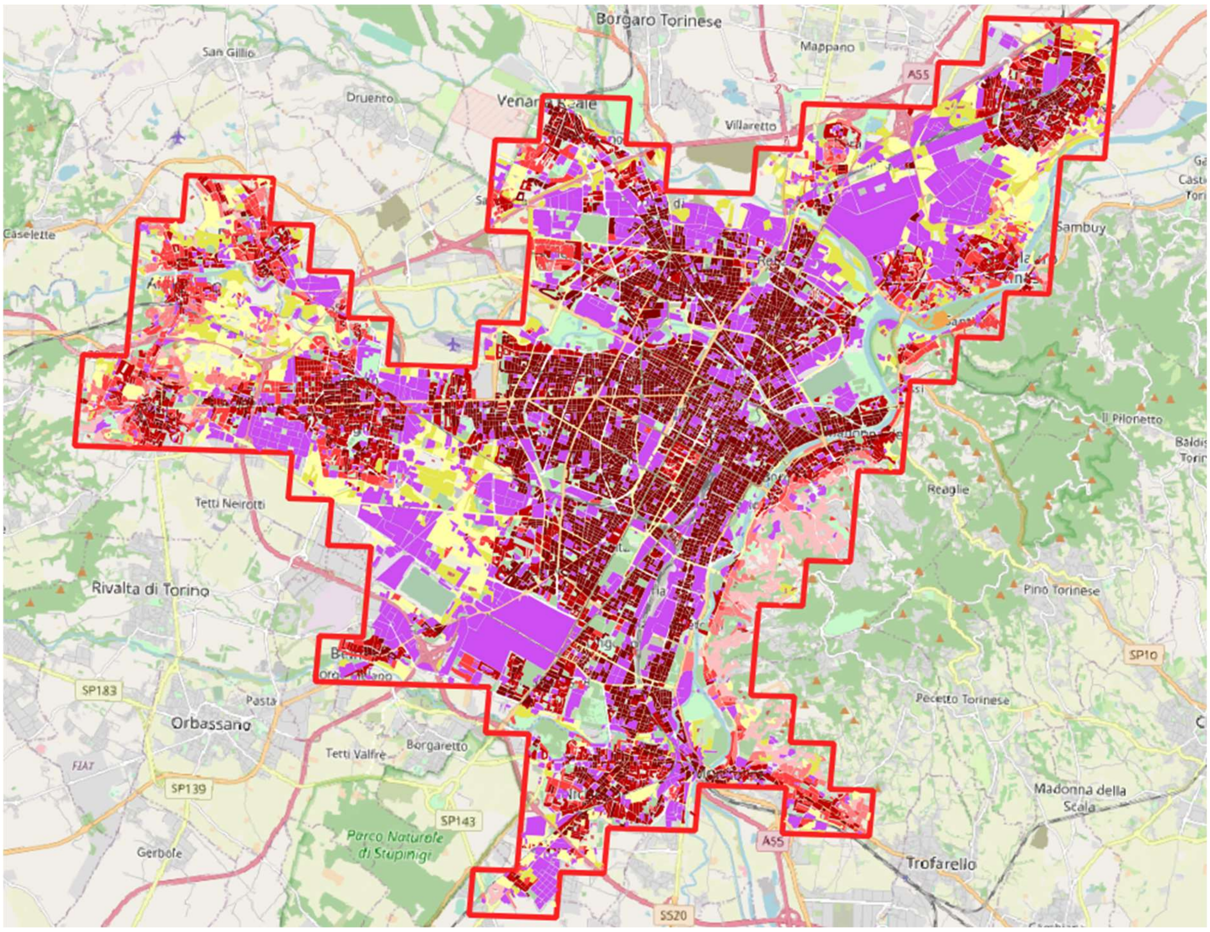


Figure 15 - Building blocks-based data falling inside the urban centre borders.

The last step of the data manipulation to make population distribution data easily accessible, is the creation of a centroid for each building block. A centroid is the geometric centre of a polygon and can be used as a representative location for the polygon. The *Centroids* command individuate the centroid of each polygon of the input layer and creates a new point feature at that location (Figure 16). The output is a points layer in which each point contains the attributes of the original polygon features. We used the generated centroids, and the information they contain, to compute the number of inhabitants reached from the Night Life Areas (NLAs) by night public transit.

Figure 16 - Centroids tool

The entire procedure can be implemented in a model as in Figure 17. The inputs are the Urban Atlas population distribution referred to the entire FUA (City blocks) and the layer containing the borders of all the European urban centres containing more than one million inhabitants (Europa). The outputs are two, one contains the blocks falling inside the urban centre (Urban Centre) while in the other these blocks are transformed into centroids (Centroids).

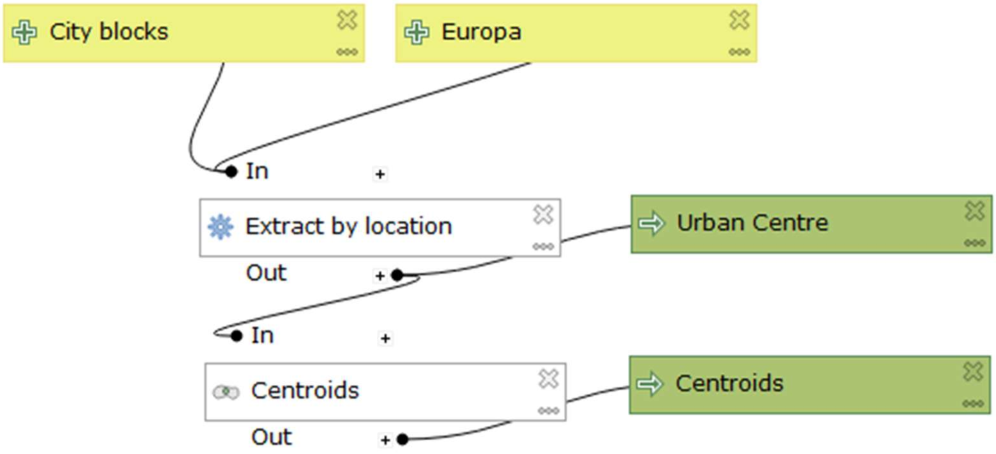


Figure 17 - Model for the extrapolation of population density concerning the urban centre.

To test the coherency between the population distributions offered by GHSL-POP in 2015 as 1 km x 1 km grid cells and by Urban Atlas in 2018 as building blocks we computed the number of inhabitants falling inside the Urban Centre borders for the two different distributions. In Table 9 we can see how these two values compare to each other and with the official municipal population; the sources and the reference years of the municipal population data of each city are also shown. The highest variation between 2015 and 2018 data can be seen in Valencia, where it sees an increase of 4.16%. These results are also visually shown in the graph of Figure 18.

Table 9 - Comparison between Municipal official population, population inside the Urban Centre borders for GHSL 2015 and for Urban Atlas 2018.

City	Municipal population	Reference year	Source	GHSL 2015	Urban Atlas 2018 - Urban centre	Variation GHSL/U.A.
Budapest	1 752 286	2019	United Nation	1 758 468	1 723 314	-1.01%
Milano	1 352 000	2023	ISTAT	3 011 030	3 157 550	2.38%
München	1 488.202	2021	Bayerisches Landesamt für Statistik	1 573 652	1 582 533	0.28%
Praha	1 275 406	2022	Český statistický úřad	1 126 681	1 194 604	2.93%
Roma	2 745 862	2023	ISTAT	2 342 860	2 493 124	3.11%
Torino	842 754	2023	ISTAT	1 205 385	1 207 163	0.07%
Valencia	789 744	2019	Ajuntament de Valencia	1 393 120	1 514 210	4.16%
Vienna	1 931 594	2023	Statistik	1 856 676	1 984 266	3.32%

			Austria			

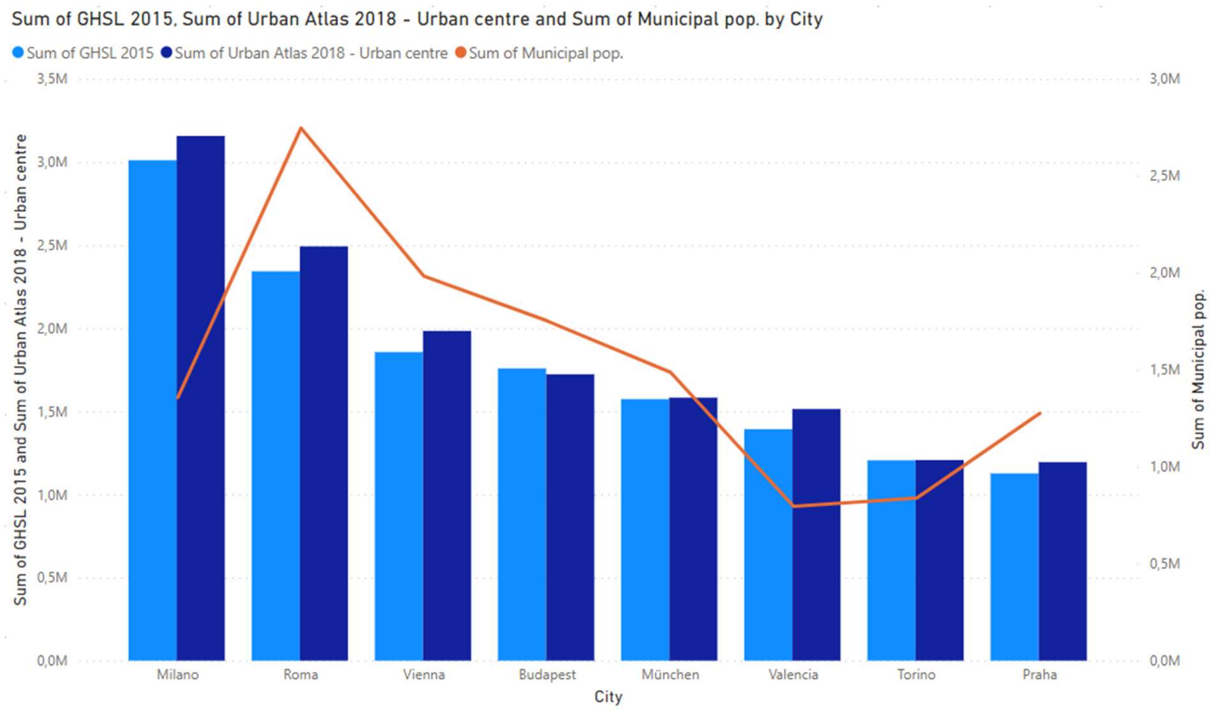


Figure 18 - Comparation between population distributions: GHSL 201, Urban Atlas 2018 and Official last updated municipal population

In Table 10 all the surfaces in km² for each urban centre are collected. With this information, related to the number of inhabitants considering Urban Atlas 2018 data, we could compute the population density of each city. The table is also shown as a graph in Figure 19. These values were used in the correlation at paragraph **Errore. L'origine riferimento non è stata trovata.** to better understand the computation results.

Table 10 - Surface and population density for each Urban Centre; computed with Urban Atlas 2018 data

City	Inhabitants Urban Centre	Surface [km^2]	Density [Inh/km]
Budapest	1,723,314	433	3980
Milano	3,157,550	754	4188
München	1,582,533	349	4534
Praha	1,194,604	295	4050
Roma	2,493,124	482	5172
Torino	1,207,163	207	5832
Valencia	1,514,210	333	4547
Wien	1,984,266	392	5062

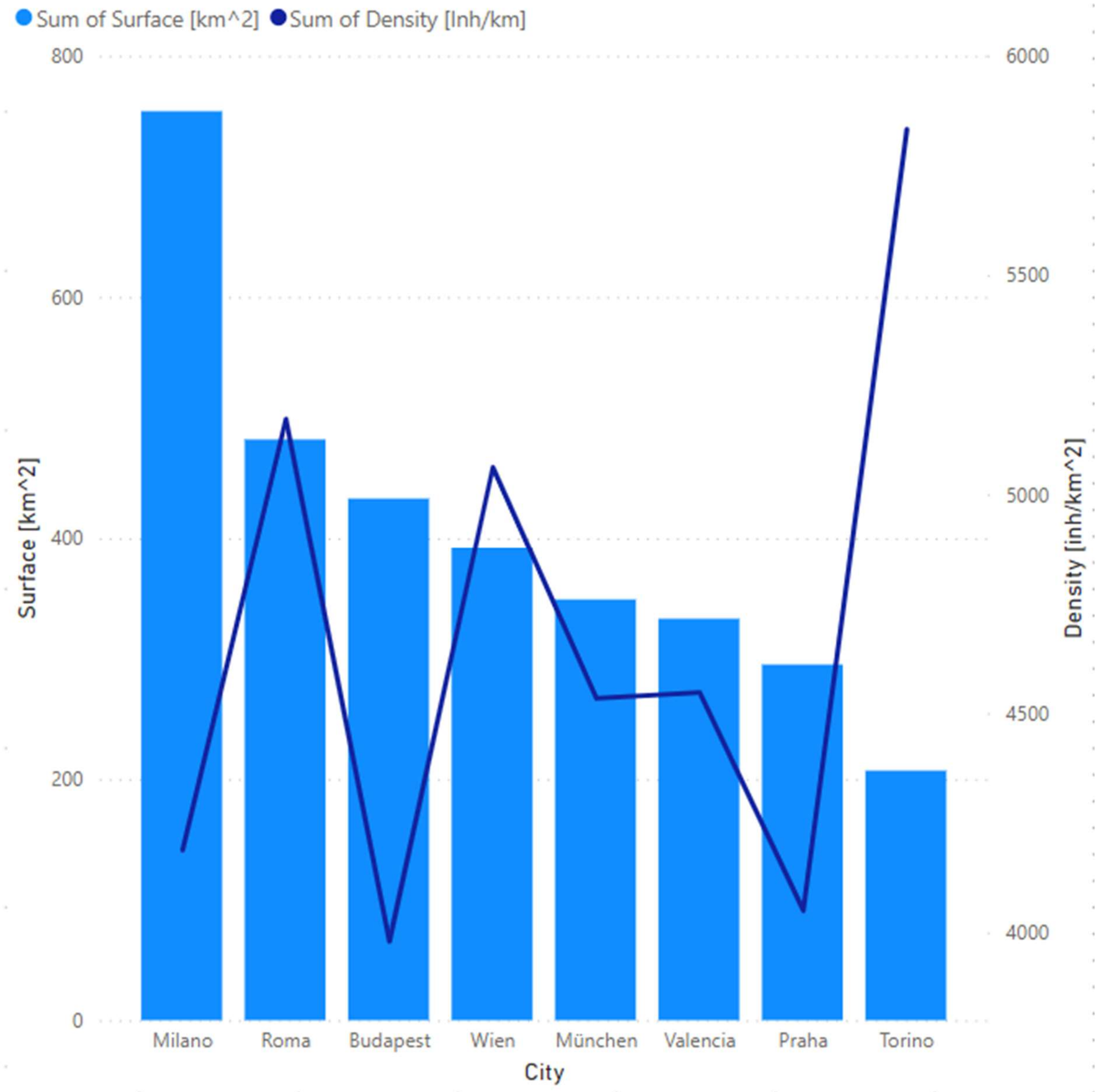


Figure 19 - Graph showing the surfaces in km² of each city and their density in inhabitants per km².

3.4. Night -life areas identification

Night Life Areas (NLAs) were created from the totality of the points of interest (POI) representing the night time leisure-related amenities of each city. Once downloaded for each city, these points have been clustered with the DBSCAN tool on QGIS, and the centroids of the resulting clusters were considered as the departure point of the related NLA. NLAs are defined as the areas of each city in which people gather during night-time, especially on weekends and public holidays. While in our reachability computations the population distribution represented the travel destinations at late night, Night Life Areas (NLAs) represented the origins since people are leaving such areas to come back home. In order to localize those areas, we explored the leisure-related amenities distributions in each city, considering an area with a massive presence of amenities as an appealing area for the weekend night-life. Amenities refer to the various services and facilities provided in a neighbourhood or community, including bars, pubs, restaurants, and other places of entertainment or leisure.

To do so, a georeferenced dataset of amenities was needed for each city; we downloaded these data through Overpass Turbo[3] which is a web-based data mining tool for OpenStreetMap. The data are indeed provided by OpenStreetMap, which is a free, editable map of the whole world that is being built by volunteers largely from scratch and released with an open-content license (Haklay, 2010). Overpass Turbo is an online platform for querying and retrieving OpenStreetMap data for various purposes, including geospatial analysis, mapping, and research (Neis & Zielstra, 2015) Asking in the dialogue window for a specific kind of amenity, the programme will provide an exportable shapefile of Points Of Interest (POI) with related information and will show the results on an interactive map. A point of interest (POI) is a specific location or feature of particular significance and it can be found on maps or navigation systems. Every POI available through Overpass Turbo has the following useful attributes:

- type of amenity
- name of the activity
- geographic location

Other secondary attributes such as payment methods, sub type of amenity, opening hours, which are not always uploaded for each POI and not so relevant to our study are contained on POI's attributes list.

OpenStreetMap represents physical features on the ground using tags; each tag describes a geographic attribute of the feature being shown. OpenStreetMap's free tagging system allows the map to include an unlimited number of attributes describing each feature. The community agrees on certain key and value combinations for the most commonly used tags, which act as informal standards. Short descriptions of tags that relate to particular topics or interests can be found on OSM wiki[4]. (wiki.openstreetmap, 2022). Between the whole set of available features, we selected the ones we could relate to night life. This selection of amenities is reported in Table 11.

Table 11 - List of selected amenities with their definition (wiki.openstreetmap, 2022)

Value	Definition
Bar	Bar is a purpose-built commercial establishment that sells alcoholic drinks to be consumed on the premises. They are characterised by a noisy and vibrant atmosphere, similar to a party and usually don't sell food.
Pub	A place selling beer and other alcoholic drinks;
Biergarten	Biergarten or beer garden is an open-air area where alcoholic beverages along with food is prepared and served.
Nightclub	A place to drink and dance . The Italian word for it is "Discoteca"; please don't confuse this with the Italian definition of "Nightclub" which is most likely described by the amenity "Stripclub"
Arts centre	A venue where a variety of arts are performed or conducted

[3] https://overpass-turbo.eu/ Attribution to Open Street Map, data are available under the Open Database License
[4] https://wiki.openstreetmap.org/wiki/Map_features#Amenity

Brothel	An establishment specifically dedicated to prostitution
Casino	A gambling venue with at least one table game (e.g. roulette, blackjack) that takes bets on sporting and other events at agreed upon odds.
Cinema	A place where films are shown
Community centre	A place mostly used for local events, festivities and group activities; including special interest and special age groups.
Conference centre	A large building that is used to hold a convention
Events venue	A building specifically used for organising events
Exhibition centre	A building specifically used for organising exhibitions and fairs
Gambling	A place for gambling, not being a bookmaker, lottery, casino, or adult_gaming_centre. An example of game covered by this definition is bingo.
Love hotel	A love hotel is a type of short-stay hotel operated primarily for the purpose of allowing guests privacy for sexual activities.
Music venue	An indoor place to hear contemporary live music.
Social centre	A place for free and not-for-profit activities.
Planetarium	A planetarium.
Strip club	A place that offers striptease or lapdancing (for sexual services use amenity=brothel).
Swinger club	A club where people meet to have a party and group sex.
Theatre	A theatre or opera house where live performances occur, such as plays, musicals and formal concerts. Use amenity=cinema for movie theaters.

We didn't consider food related features such as restaurants because they are only marginally related to the night life: indeed, they usually close when day-time public transport service is still available, thus before 1 am in most cities. In addition, the number of food related features is much higher than any other kind of considered feature, so they would disproportionately affect the definition of clusters.

Once defined the target features, we downloaded the related POIs and transferred them on QGIS, where manipulations were made in order to obtain an hotspots distribution through which the night life areas (NLA) were identified.

3.4.1. Downloading the POIs and inserting them on QGIS

Overpass Turbo interface presents itself divided in two as in Figure 20: on the left is a dialog window and, on the right, an interactive map. We typed the following query on the dialog window, asking to show the POIs referred to the amenities we are interested in for the city of Budapest. The scripting language used in the dialogue window of Overpass Turbo is Overpass QL, which is a custom query language designed specifically for querying OpenStreetMap data. After we run the query, the results became visible in the interactive map as in Figure 21, where an example on the city of Budapest is shown.

```
[out:json][timeout:25];
// gather results
(

node["amenity"="biergarten"]({{bbox}});
node["amenity"="bar"]({{bbox}});
node["amenity"="pub"]({{bbox}});
node["amenity"="nightclub"]({{bbox}});
  node["amenity"="arts_centre"]({{bbox}});
  node["amenity"="brothel"]({{bbox}});
  node["amenity"="casino"]({{bbox}});
  node["amenity"="cinema"]({{bbox}});
  node["amenity"="community_centre"]({{bbox}});
  node["amenity"="conference_centre"]({{bbox}});
  node["amenity"="events_venue"]({{bbox}});
  node["amenity"="exhibition_centre"]({{bbox}});
  node["amenity"="gambling"]({{bbox}});
  node["amenity"="love_hotel"]({{bbox}});
  node["amenity"="music_venue"]({{bbox}});
  node["amenity"="social_centre"]({{bbox}});
    node["amenity"="planetarium"]({{bbox}});
  node["amenity"="stripclub"]({{bbox}});
  node["amenity"="swingerclub"]({{bbox}});
  node["amenity"="theatre"]({{bbox}});

);
// print results
out body;
>;
out skel qt;
```

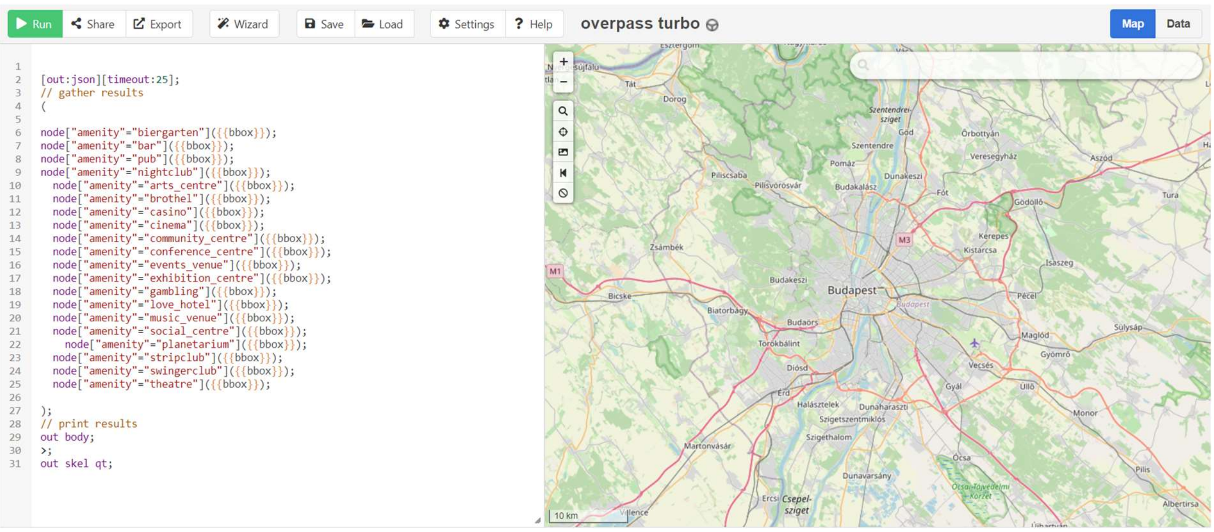


Figure 20 – Overpass Turbo interface

.

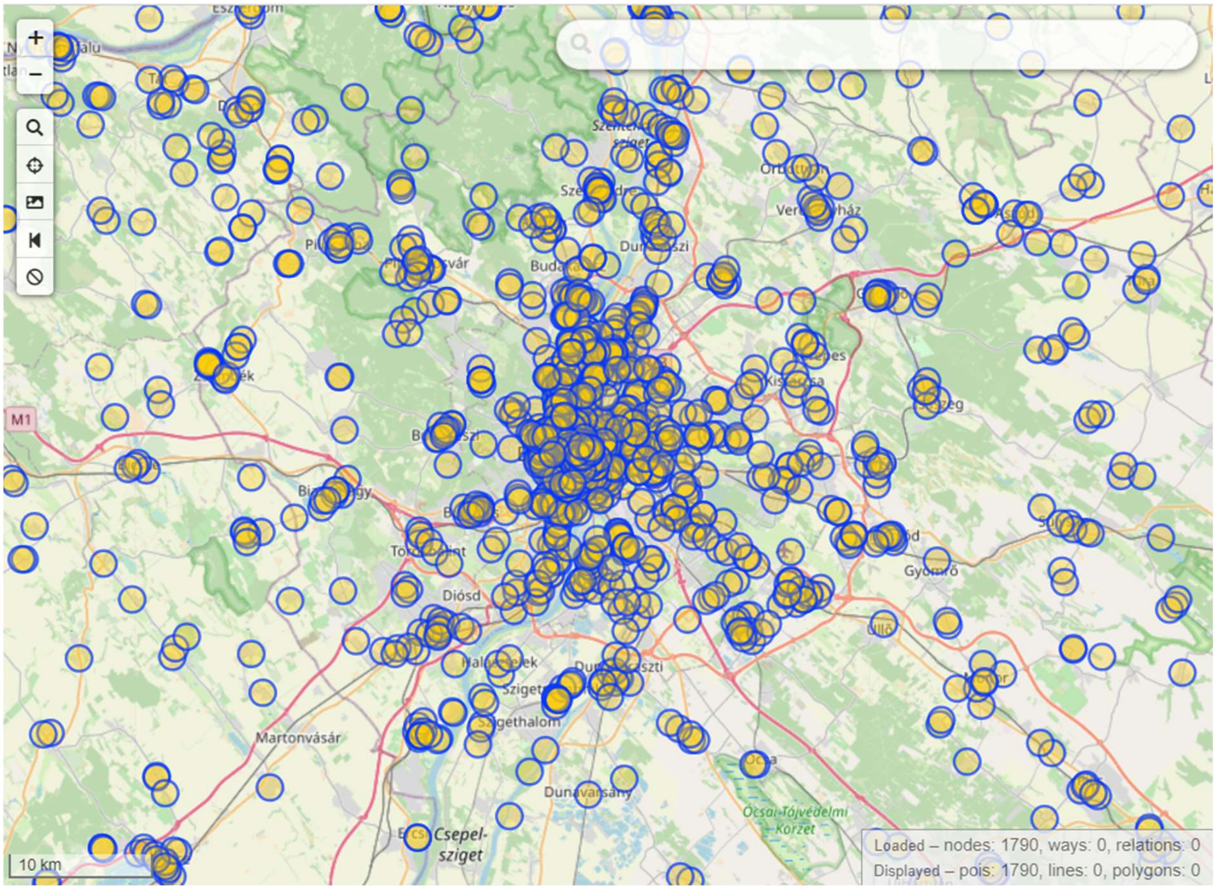


Figure 21 - Interactive map showing the individuated POIs for the city of Budapest

From the interactive map it is possible to see in the right bottom that 1790 POIs have been identified. The data about these points have been downloaded in KML format, ready to be used on QGIS. In Figure 22 the point layer uploaded on QGIS is shown. For the cluster computation we considered the points of interest falling inside the Urban Centre only (Figure 23). To do so we run an *extract by location*.

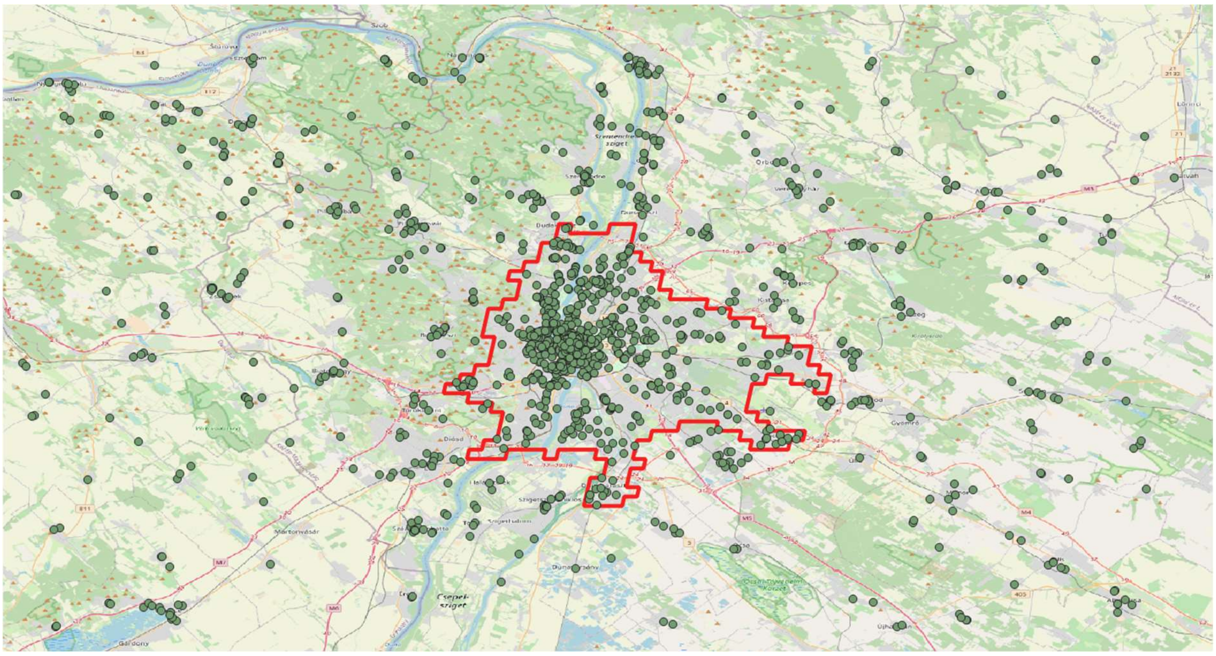


Figure 22 – Points of Interest (POIs) on QGIS for the city of Milano

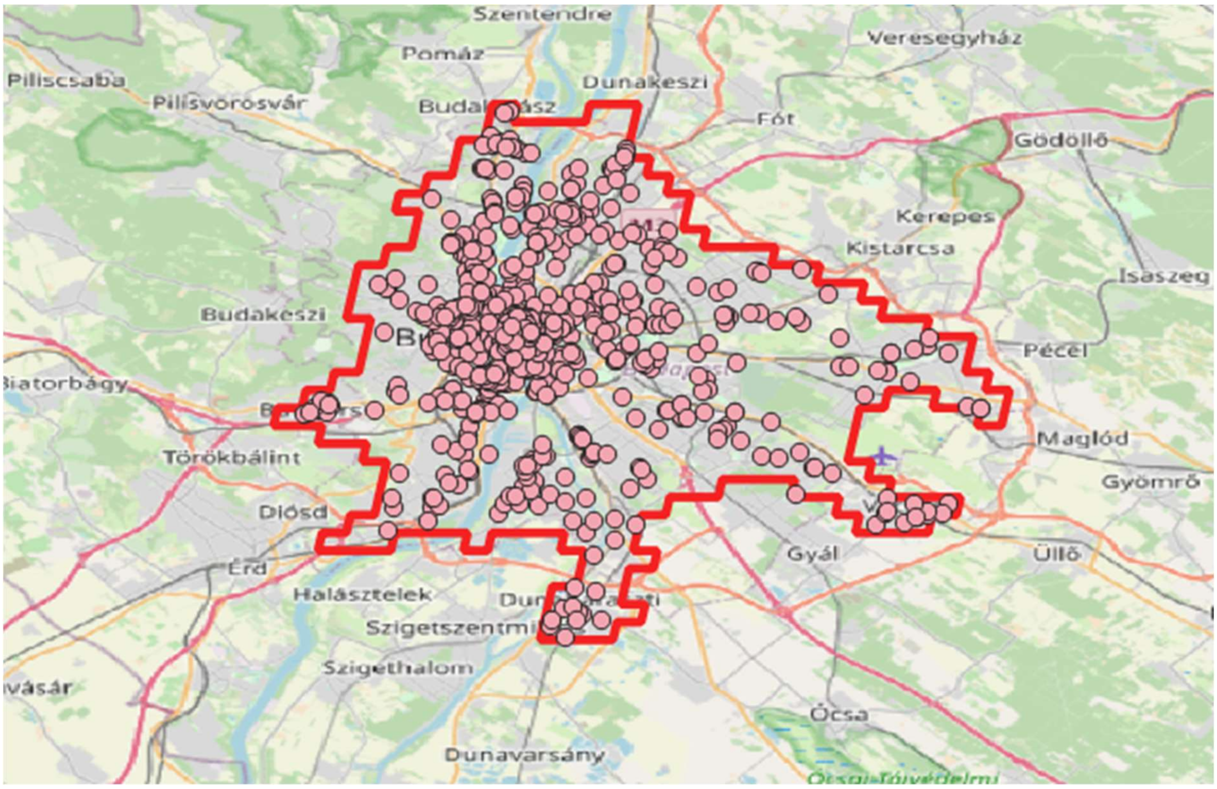


Figure 23 - Point of Interest falling inside the Urban Centre of Budapest. Only these points will be considered in the creation of the Clusters.

The data we downloaded from Overpassturbo in order to identify the amenities of each city, have their biases and limitations. Indeed, since these data are uploaded by volunteers rather than professionals, it happens these data could have coding issues (so for example a cafè is individuated as a bar). It could also happen that a closed business is still shown.

3.4.2. Hotspots identification

Through the POIs we could identify the hotspots, which are areas in which the POIs density is higher: these areas were identified as Night Life Areas (NLAs), which we later used as departure points for the successive reachability computations.

Once uploaded the POIs in QGIS, we performed a series of manipulations in order to identify the Night Life Areas. NLAs were identified as the areas with a major concentration of targeted amenities. To do so, we firstly displayed the POIs as a heatmap, which is a visual representation of data that uses color-coding to highlight the density of values in a dataset (Catalogue, The Data Visualization, 2023).On QGIS it is possible to show a heatmap through the *Heatmap (Kernel Density Estimation)* tool. Kernel density estimation (KDE) is a non-parametric technique used for estimating the probability density function (PDF) of a random variable. The basic idea of KDE is to estimate the PDF by placing a kernel function at each data point and then summing up the contributions of all the kernels. The resulting smoothed PDF provides an estimate of the underlying probability distribution that generated the data. To do so we had to specify two parameters, search radius and Kernel Shape which are defined in the QGIS User Guide[5] as follows:

- The **search radius** (expressed in meters) specifies the distance around a point at which the influence of the point will be felt. Larger values result in greater smoothing, but smaller values may show finer details and variation in point density.
- The **Kernel shape** parameter Controls the rate at which the influence of a point decreases as the distance from the point increases. Different kernels decay at different rates. There are many shapes available (Figure 24).

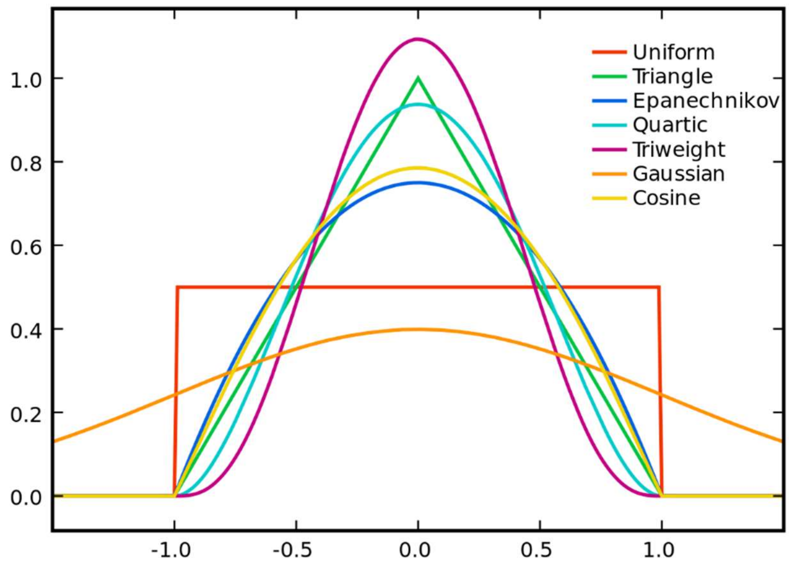


Figure 24 – Different Kernel Shapes compared by Brian Amberg[6]

For example, a triweight kernel gives features a greater weight for distances closer to the point than the Epanechnikov kernel does. Consequently, triweight results in “sharper” hotspots and Epanechnikov results in “smoother” hotspots (QGIS User Guide).

The limited range of studies containing documented parameters for density measurements means that the process of deciding the bandwidth and grid cell size is somewhat subjective (Anderson, 2009). In absence of any precedent study which links the presence of different kinds of amenities to specific kernel shapes or search radius and since the heatmap goal is to give an

[5] https://docs.qgis.org/3.22/en/docs/user_manual/index.html

[6] https://en.wikipedia.org/wiki/Kernel_(statistics)#/media/File:Kernels.svg

idea of how amenities are disposed in each city, we set the parameters in order to have a readable and intuitive visual representation of points of interest density.

For the computations on QGIS, the Coordinate Reference System WGS84 (EPSG: 4326) is the adopted standard. WGS84, or World Geodetic System 1984, is a global reference system used to model the shape of the earth and define the positions of objects on its surface. It was developed by the US Department of Defense and National Geospatial-Intelligence Agency as a successor to previous reference systems and is now widely used in various applications including mapping, surveying, and navigation,and it was used for all the layers introduced until now. WGS84 is based on the ellipsoid. To specify a location on the earth's surface in WGS84, geographic coordinates are used, which consist of a latitude and a longitude value in degrees. Measurement in geographic degrees works by dividing the earth's surface into a grid of longitude and latitude lines. This system allows for precise location of objects on the earth's surface. Regarding the clustering procedure, a radius of search was asked by the clusteing algorithm (as explained in the following of this paragraph). Since we prefer to work with metres rather than geographical degrees, we decided to reproject to a projected local coordinate system, namely the ETRS89-extended / LAEA Europe - EPSG:3035. The 3035 projection is a projected coordinate system used in Europe that represents the earth as a two-dimensional plane and its coordinates are expressed in units of metres. The 3035 projection is based on the Lambert Azimuthal Equal Area projection, which accurately preserves area but distorts shape and distance. The 3035 projection is converted from the WGS84 by using mathematical formulas to transform the coordinates from latitude and longitude degrees to projected coordinates in metres, but these formulas are not explicit in QGIS, where the conversion is done automatically. The conversion involves adjusting for the curvature of the earth's surface and the distortion caused by projecting the 3D surface onto a 2D plane. We set the search radius at 300 metres, the pixel sizes at 10 m and a Quartic kernel shape (Figure 25). The results are shown in Figure 26.

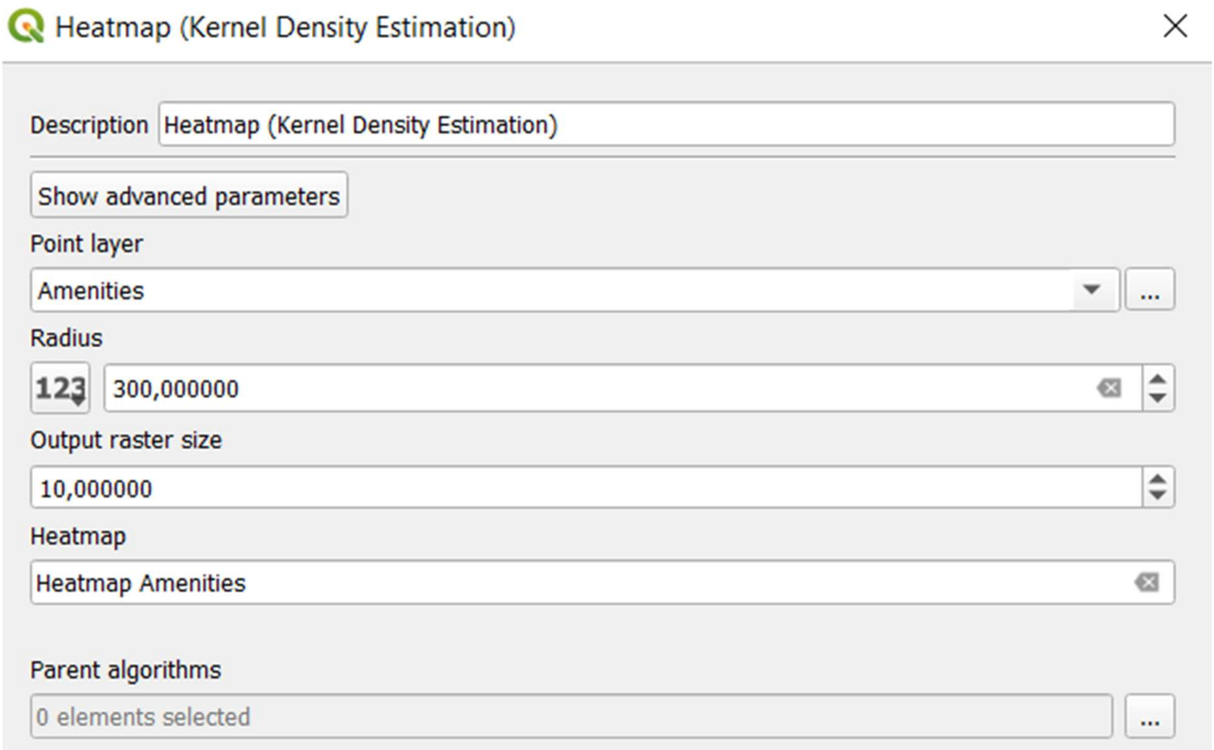


Figure 25 - Parameters set of the Heatmap creation.

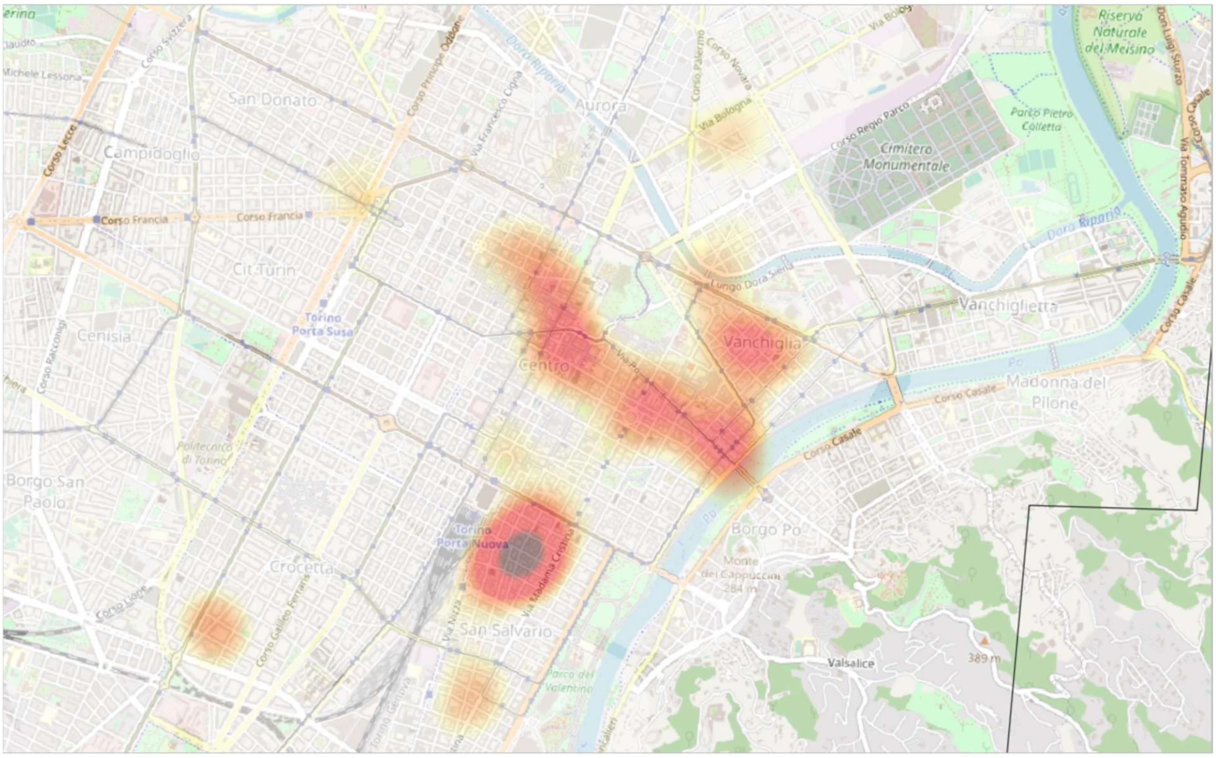


Figure 26 – Heatmap of Torino a visual impression of how the amenities are distributed in Torino

A visual impression gives a first idea of how the NLAs are distributed. But to make a comparison between different cities a standard measure to define the high-density hotspots with specific points was needed. Since we couldn't create a standard procedure by working with images only, we used DBSCAN. We clustered the POIs with the QGIS function *DBSCAN cluster* which uses the Density-based spatial clustering of applications with noise algorithm (DBSCAN) which is designed to discover the clusters and the noise in a spatial database (Yeran, 2015), (Bäcklund, Hedblom, & Neijman, 2011), (Ester, Kriegel, Sander, & Xu, 1996). This algorithm requires two parameters as inputs, namely the maximum distance between clustered points and the minimum size of cluster size:

- Minimum cluster size is the minimum number of features to generate a cluster (minPts).
- Maximum distance between clustered points is the distance beyond which two features cannot belong to the same cluster (epsilon).

By first choosing an epsilon and a minPts we define a neighbourhood around each data point. For each data point, the number of other data points in its epsilon neighbourhood is counted. If this count is less than minPts, the point is marked as a "noise" point, otherwise it is marked as a "core" point. For each core point, a new cluster is created and added to the cluster set. Then, all other core points that are directly or indirectly reachable from the initial core point (i.e., within epsilon distance) are recursively added to the cluster. The algorithm continues until all data points have been assigned to a cluster or marked as noise. The resulting clusters are dense regions in the data space, while the noise points are isolated points or points on the fringes of clusters. In Figure 27 we can see an example, over the city of Wien of how DBSCAN works: in the example we set maximum distance as 125 m and minimum cluster size as 5. The polygons (which we created here to assess a visual explanation and are not created by DBSCAN tool) represent the found clusters; we can see how their dimensions change in relation to their density.

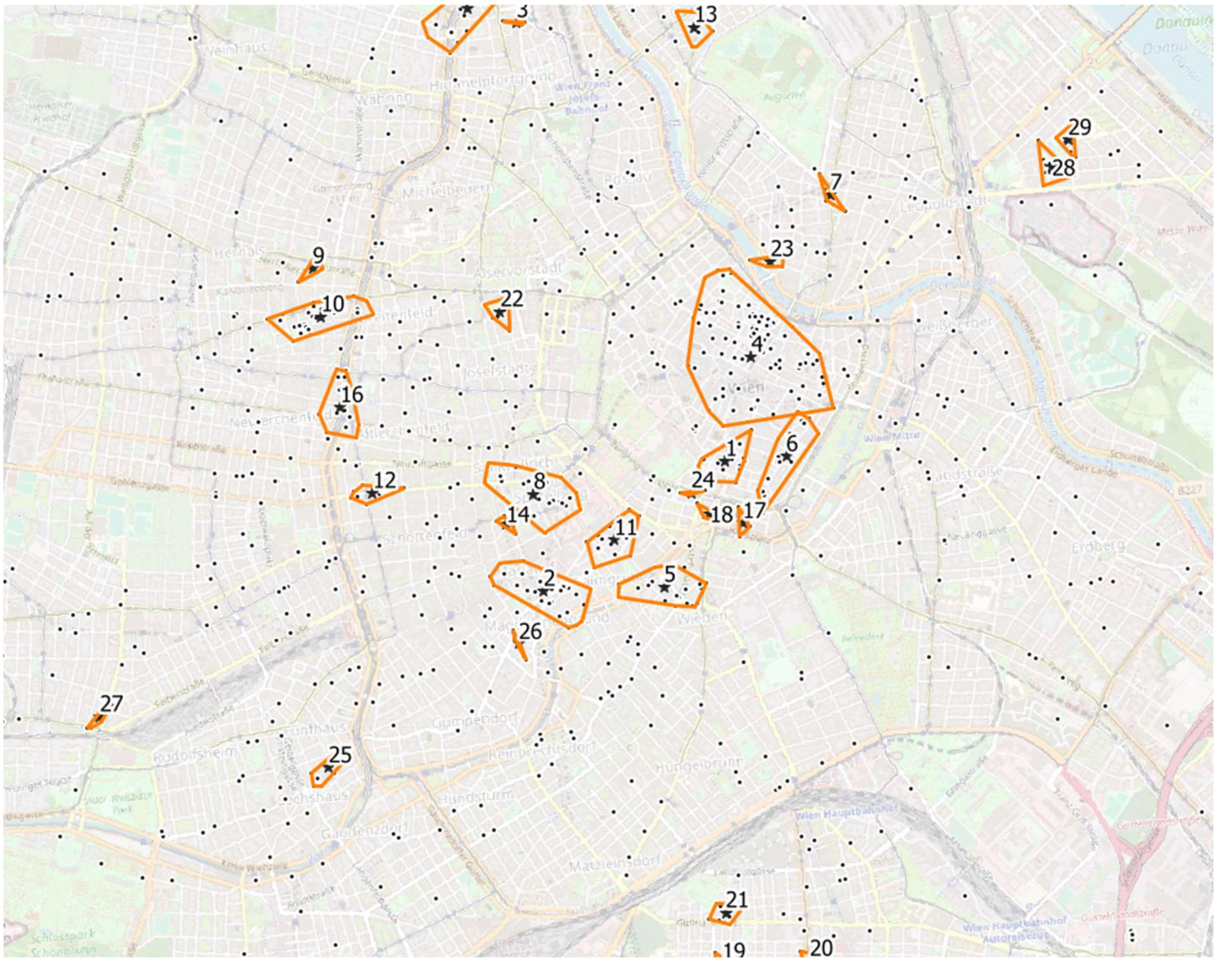


Figure 27 - Explanation of DBSCAN function – Example for Wien

In
Figure 28 is a comparison between the two cases in which we choose as maximum distance 125 m (in orange) and 150 m (in blue). The clusters created with a 150 m epsilon are bigger and sometimes involving multiple clusters created with 125 m epsilon.

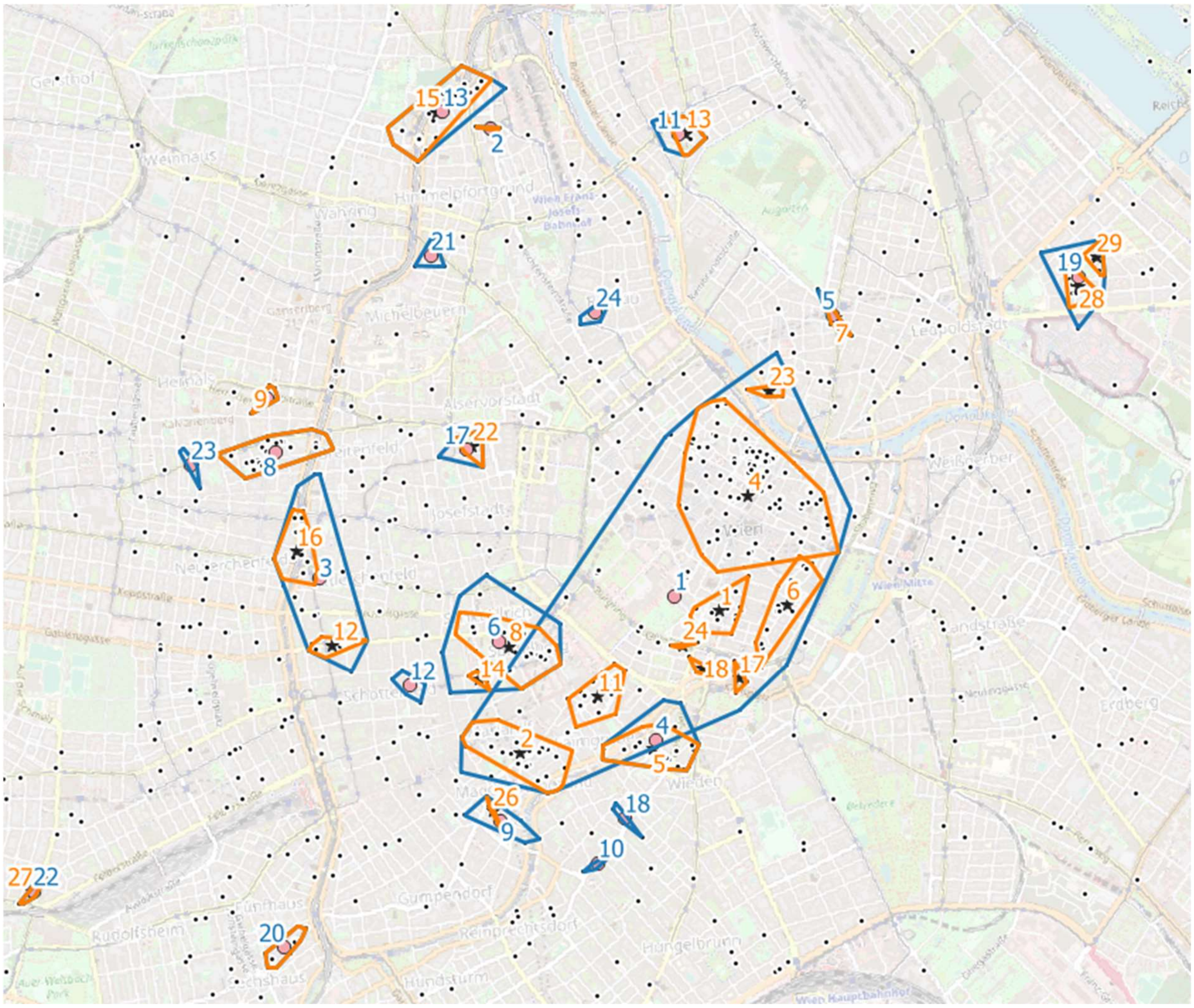


Figure 28 - Explanation of how the results change according to the parameters settings – Example for Wien

We identified 5 amenities as the minimum cluster size, and 150 metres as the maximum distance. It has been shown during the different analyses that the parameters are not univocally reliable for each city but have to be slightly changed from case to case. Indeed in some cities, namely Budapest, München, Praha and Wien, the Points of Interest in the city centre were so dense that they led to the creation of a wide central cluster. In a specific case, Wien, the distance between the centroid of the central cluster and the perimeter points of the same, tended to be over 1.5 km, while in the other three cases this distance was always lower or equal to 1 km. In order not to lose homogeneity on the cluster creation, we reconsidered the parameters for the city of Wien only and set 125 m as maximum distance (the difference was shown in Figure 28). In

Table 12 we can see the parameters used for each city. The reasons behind the choice of the parameters for each city can be found in Appendix A.

Table 12 - DBSCAN parameters used for each city.

City	min cluster size	max distance between cluster points	Found clusters
Budapest	5	150 m	14
Milano	5	150 m	35
München	5	150 m	30
Praha	5	150 m	26
Roma	5	150 m	23
Torino	5	150 m	12
Valencia	5	150 m	18
Wien	5	125 m	29

The parameter input window looks like in Figure 29. To create singular points for each cluster we performed the following procedure. In order to have a surface enclosing all the points of the cluster we used the function *Create a bounding geometry* for each cluster (Figure 30, Figure 31). Then we created a centroid for each polygon with the function *centroids*. All of this can be summarised in a model, as in Figure 32.

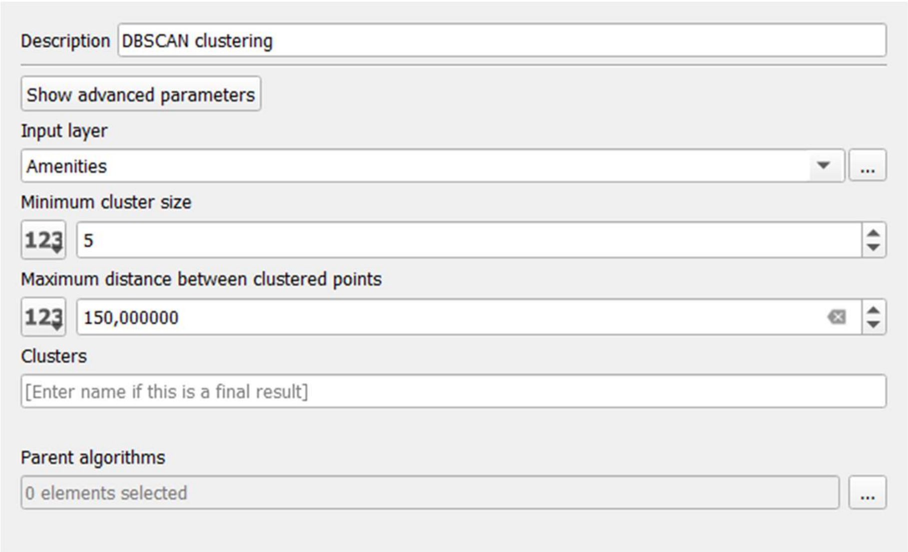


Figure 29 - DBSCAN clustering input example

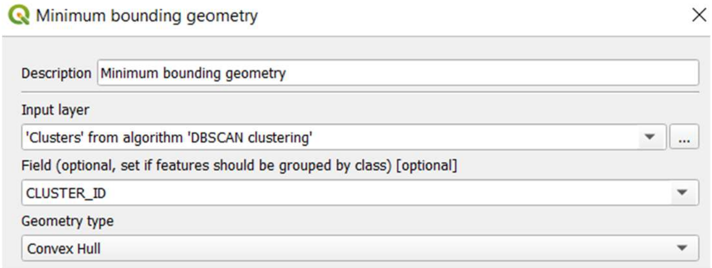


Figure 30 - Minimum bounding geometry input example

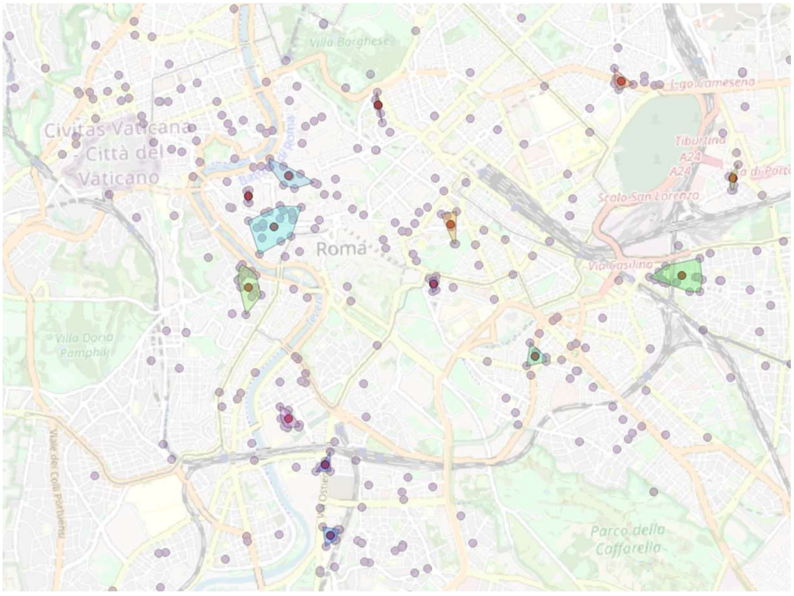


Figure 31 – Polygons enclosing the POIs forming the different clusters and their centroids (red dots) – Example for the city of Rome

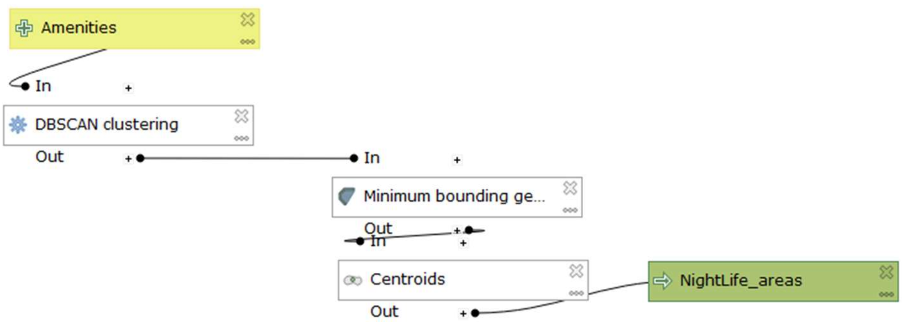


Figure 32- Division in clusters of the amenities

In Figure 33 is an example of visual comparation of computed clusters (each one identified by a centroid) with a heatmap. Since the Heatmap tool must serve the visualization, it needed different parameters values than the DBSCAN tool in order to be readable and visually significant. In Figure 33 the comparison between the visual results obtained with Heatmap tool (radius: 300 metre) and the BSCAN tool (radius: 150 meter) is shown.

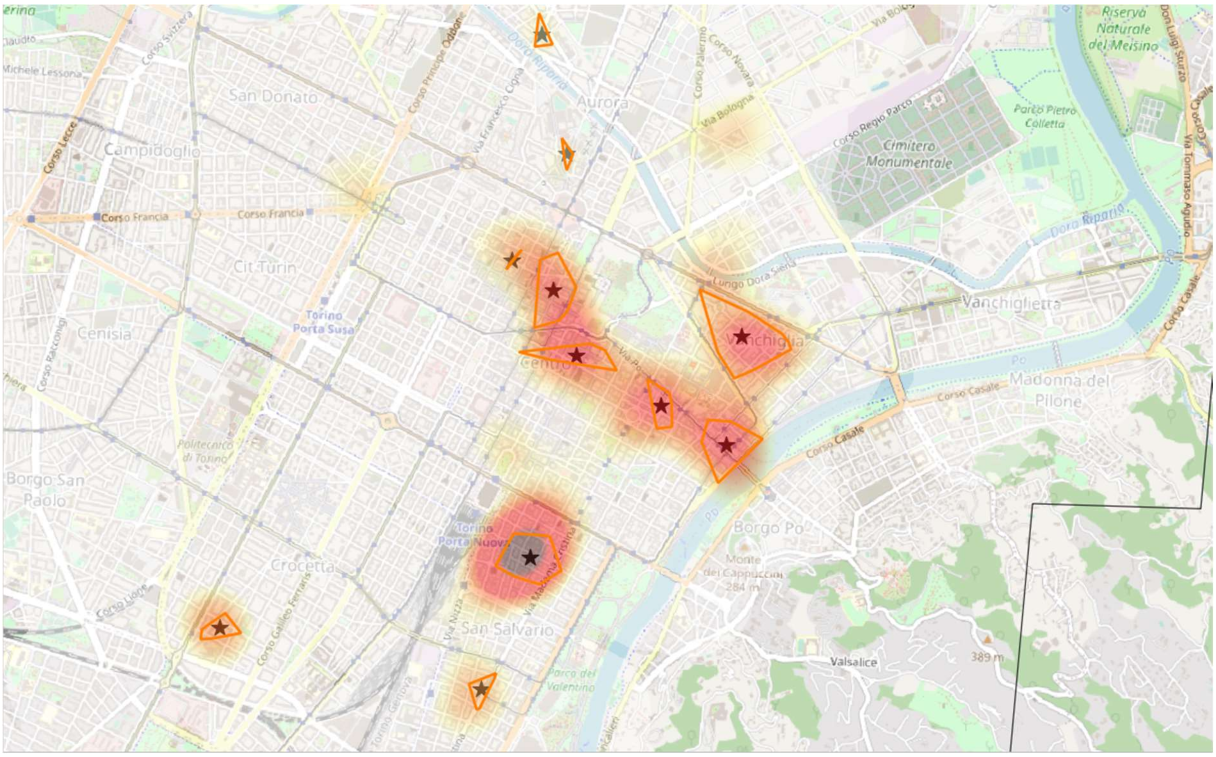


Figure 33 - Comparation between heatmap of all the amenities points and the clusters boundaries (orange polygons) and their centroids (black stars) for Torino

3.5. Travel time computation

The following step was to estimate the number of inhabitants whose residential location can be reached from each NLA within a specific travel time by public transport during the above mentioned study time. This information allowed us to define a level of service coverage, identified as a percentage of reached inhabitants over the totality of inhabitants living inside the above defined study area around each NLA. This metric is also an estimation of the amount of opportunities that can be reached within a given travel time threshold and so in our case the amount of people which can be reached from each NLA within a fixed maximum time. This method is also known as cumulative opportunity index, where examples of opportunities are jobs, education, and services or, as in our case, night-time related amenities These methods are easily adoptable in practice and, having no weights (the reached individuals have all the same weight, each individual has the same value in our study), they are easier to compute and interpret than gravity-based models, which are called so since, as in physics, the attractiveness of a location is represented by a given weight, and the cost of interaction between two locations is represented by their distance. We refer readers to section 2.5.2 for additional discussion on these definitions.

An important limitation of cumulative opportunity measures is that they require an arbitrary choice of a single travel time threshold (Tomasiello, Herszenhut, Oliveira, Braga, & Pereira., 2022); in our case we considered 30 minutes and 45 minutes as time thresholds. Findings from different cities around the world can hardly support the concept of stability of travel budgets, showing how the range of travel time average obtained from different cities in both the western and developing countries is very wide even at country-aggregated level (Supernak, 1967) and how a stability is reachable only at a global aggregation level, but not at finer levels of disaggregation (Joly, Travel Time Budget – Decomposition of the Worldwide Mean, 2004). Further tests of the proportional relationship between daily travel time given a purpose and daily activity time indicated that proportionality is not valid, regardless of the type of activity; that leads to the conclusion that, the proportionality hypothesis between trip time and activity time at the destination is inconsistent (Joly, 2007). Considering the lack of consistently suggested values, we took the two maximum times of 30 and 45 minutes to reflect the study boundaries adopted in “How many people can you reach by public transport, bicycle or on foot in European cities” (Poelman, Dijkstra, & Ackermans, 2020). Since different cities have different shapes and different spatial distributions, namely more compact or more sprawled we compared for each NLA the reached population with the population living in the NLA neighbourhood. For each NLA we created a buffer circular area to identify the neighbourhood of the NLA itself; the radius of the circular area is identified separately for the two different time thresholds of 30 and 45 minutes; indeed the concept of “neighborood” stands for the area around each NLA that can be reached at raven-flight when travelling at a fixed speed. Specifically, we considered an average straight-line speed of 15 km/h for the public transport means and we draw a radius out of this consideration for each time span. As in Poelman, Dijkstra & Ackermans, 2020 we considered 15 km/h to be the average speed of public transport: in Poelman, 2020 the straight-line speed of vehicle trips for bus, tram, metro and train are computed for each of the 42 studied cited separately. 15 km/h is the rounding of the average bus speed, which was 14.5 km/h. The authors claim how the average speeds change for different cities conformations. Straight line speeds of the vehicles have been computed comparing traveling lengths and trip times from GTFS files, namely as a ratio between the linear distance between two subsequent stops (the linear spatial distance, not the network one) and the time spent by a mean between two subsequent stops. To summarise:

- When we considered a maximum time of 30 min, the study area was a circumference of 7.5 km radius around the considered NLA
  (30 min/60 min) * 15 km/h = 7.5 km
- When we considered a maximum time of 45 min, the study area was a circumference of 11.25 km radius around the considered NLA
  (45 min/60 min) * 15 km/h = 11.25 km

.

. Considering a comparison between reached population and the population which is hypothetically reachable on a geographic standard base allowed us to assess these differences. The service coverage percentage for a specific NLA in a specific time span is finally the ratio between the inhabitants of the neighbrood of a certain NLA reached from the same NLA and the totality of inhabitants living in this neighboorod. This parameter has been calculated for each NLA of each city through **reachability computations,** which is the procedure within we estimate the number of inhabitants whose residential location can be reached from each NLA within a specific travel time by public transport during the above mentioned study time**.** carried using the Time Map tool furnished by Traveltime plug-in on QGIS. The Time Map tool allows to identify which building blocks can be reached from each NLA in a given time threshold, using a specific travel means, in a given day and hour. The reachability computation was performed every 15 minutes for a study time range of 2 hours. Since for some cities the frequencies change during the whole service hours, we choose for each city two hours which could depict the “peak hours”, meant as the two hours with most vehicles involved. Finally, the reachability computation was performed 9 times for each of the two time spans (30 minutes and 45 minutes) for each city, for a total of 9 x 2 x 8 = 144 computations. The results of each city have been later manipulated to obtain an overview of the efficiency percentage for each city over the two hours for the two different time spans.

Once we downloaded and manipulated data about population distribution, amenities distribution (and its clusters) and public transport offer at night time, the goal is to combine them and to gain information on the dwelling location of how many inhabitants inside the study area of each city area can be reached by public transport, in a specific time threshold, from each specific Night Life Area in a weekend night. To do so we performed a Reachability computation with the plug in Traveltime[7].

---

[7] https://traveltime.com/

The traveltime plug in, which can be used on QGIS uses APIs and data from various sources, such as OpenStreetMap and Google Maps, to provide travel time estimates for driving, public transportation, cycling, and walking. These estimations take into account real-time traffic conditions and other relevant data to provide accurate and up-to-date information on travel times between two points. We used the function "Time Map" which allowed us to understand which areas of the cities are reachable from each Night Life Area (NLA) in a certain amount of time by public transport. With this tool we created isochrones through we could assess which building blocks were reachable and which weren't.

The data about public transport offer are collected by Traveltime in GTFS format in accordance with the cities' Public transport agencies, which give GTFS files or data in order for them to create the GTFS anew. A list of the covered areas and the helping agencies can be found on Traveltime website. In Table 13 is shown the list of the covered agencies running the public transport services in the studied cities.

Table 13 – Public transport agencies covered in each studied city;

City	GTFS source
Budapest	BKK: budapesti közlekedési központ
Milano	ATM S.p.a.: Azienda Trasporti Milanesi
Munchen	MVG: Münchner Verkehrsgesellschaft
Praga	PID: Pražská integrovaná doprava
Roma	ATAC Azienda per la mobilità di Roma Capitale S.p.A
Torino	GTT: Gruppo Torinese Trasporti
Valencia	EMT: Empresa Municipal de Transportes de Valencia
Wien	Wiener Linien

3.5.1. Computation input

In order to start the computation, the function "Time Map" asked us for a series of inputs (Figure 34). First of all, a layer containing the departure points of the computation was asked: this layer must be the one containing the NLAs' centroids. Then, the transport mean we plan to use was asked, which was *public transport* in our case; then we inserted the travel time we intended to spend in the commuting, which we set in two different computation at 30 and 45 minutes, and the date and hour of the computation: since we are working on night time public transport we had to choose a night time hour in a weekend day: we choose the 12 of March 2023 as the date and a time range of two hours for each city (from x:00 to x+2:00) with a resolution of 15 minutes. so that it was possible to consider the differences in service frequencies: indeed, we expect that with a low frequency the results in different time frame will be very different from each other, while with an higher frequency they will be more similar (this evidence will be observed in **Errore. L'origine riferimento non è stata trovata.**) .

In Traveltime's computations *walking* is included in the public transport mode, as most journeys don't start and finish on public transport stops. The length of time allowed for walking at the start and end of the journey (i.e., from the origin to the first stop/station, and from the final stop/station to the destination) can be set by the user. From previous studies concerning the walking time from origin to a service stop (as a bus stop, train station etc.) it can be concluded that an overall solution (ndr, a standard walking time) is with high probability not going to be effective (Chia, Lee, & Kamruzzaman, 2016). The *walking time* that users are willing to walk to reach a transit stop depends on many factors such as the kind of area of the city where the origin is placed (city centre, suburbs, etc), the kind of mean to catch (bus, tram, train, etc.) and the purpose of the trip (Sarker, Mailer, & Sikder, 2020). Walking trips to railway stations are significantly longer compared with those to local public transport stops and in (A.Tennøy, Knapskog, & Wolday, 2022) the 75th-percentile duration is 10 min for both origin - station and station - destination trips. In (Poelman & Dijkstra, 2015) a 10 min walking time is assigned for train stations against a 5 min for bus stations; the willingness of walking more for a train than for a bus shows how some means are considered more relevant than others; we can state that during night time, because of the limitations of the network, each running line is to be considered as a relevant one. For this reason and for the evidence that longer walking distance was mostly reported for trips to home and related to shopping/recreational activities (Sarker, Mailer, & Sikder, 2020) we set for both origin - stop and stop - destination a 10 minutes walking time. The waiting time is limitless, within the boundaries of the set total travel time. Indeed, given the kind of activity at the origin, people could always stroll about the NLA until the departure time of the public transport service is coming.

The algorithm will run an isochrone from each NLA; the inhabitants identified by each centroid are considered to be reached from a specific NLA in a specific time threshold when the respective centroid falls inside that specific NLAs isochrone. The isochrone will be defined by the travel time we choose, and the available public transport offer at the hour and date we choose as shown in Figure 34 The results appear as red coloured polygons, indicating the reached areas in the given time threshold (Figure 35).

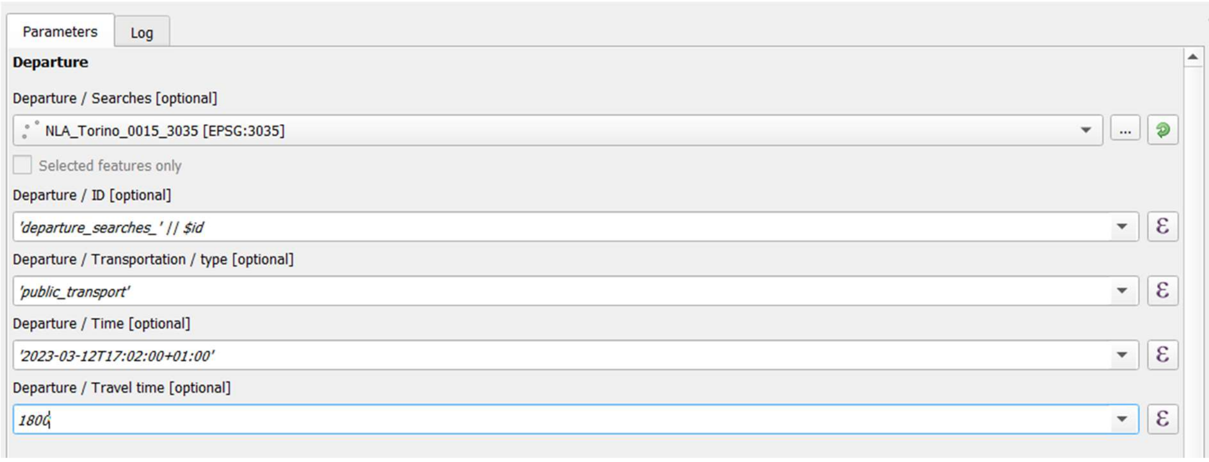


Figure 34 - Time Filter function, input window

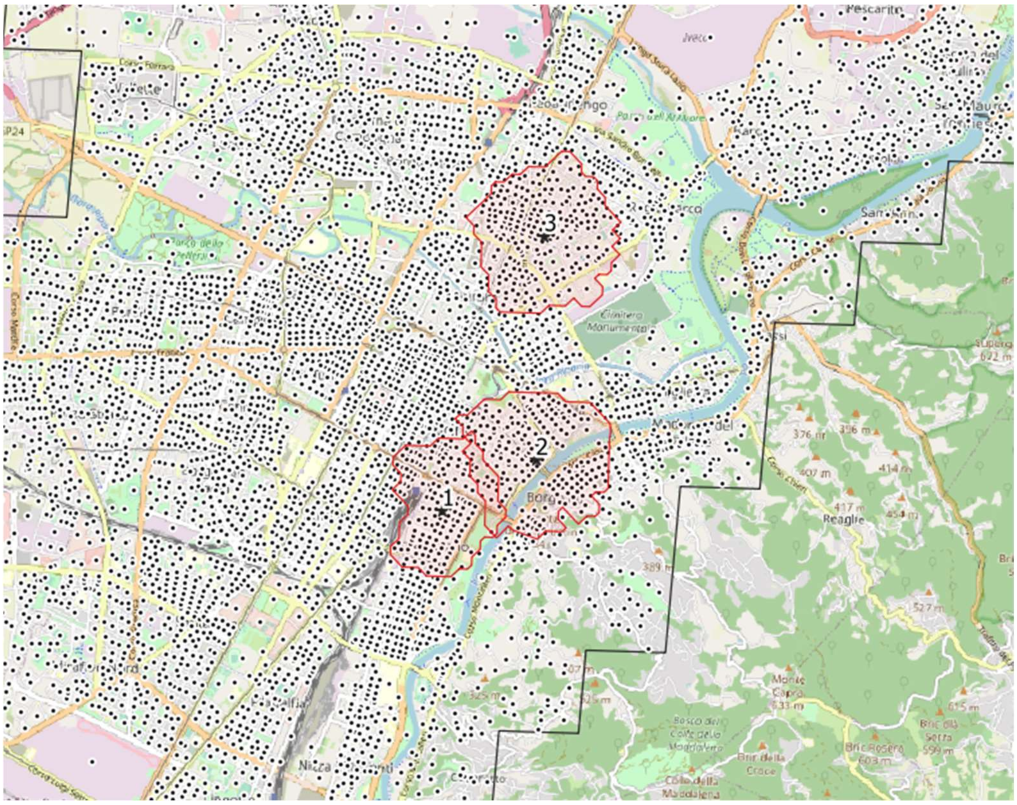


Figure 35 - Example of how isochrones appear, Torino.

We further manipulated isochrones and population grid cells to obtain useful results. Specifically, we extracted the centroids falling in both the isochrone and the time threshold related buffers (7.5 km or 11.25 km radius) of each NLA.

3.5.2. Identification of reachable and reached centroids of each NLA

As anticipated in Chapter **Errore. L'origine riferimento non è stata trovata.**, we considered, for each NLA, a circular buffer area with radius based on each of the two time thresholds (30 and 45 minutes) and average speed of the transport mean used. Since we considered the average speed of public transport to be 15 km/h, we had circular areas of radius 7.5 km when the maximum time was 30 minutes, and of radius 11.25 km when the maximum time was 45 minutes. Inside these circular areas were the **reachable blocks** of each NLA, The *reachable block* concept is based on geographical location, indeed all the points falling in the circular buffer are to be considered reachable and will be used against the **reached blocks** (the blocks within the buffer area actually reached in the selected time threshold) to collect the **service coverage** out of each NLA. In order to define them we used the tool "Buffer" (Figure 36) and with "Join attributes by location" (Figure 37) we assigned a belonging index to each centroid falling inside a buffer; by setting the option "one-to-many" we could consider multiple times the blocks falling in multiple NLAs. That means that a centroid falling in the circular buffer area of multiple NLAs will be considered multiple times, so that it will appear as a reachable block of each NLAs which covers it under its buffer area. Discarding the records we can't join, only the blocks which falls inside at least one NLA's buffer area will be considered and assigned a belonging index.

Figure 36 - "Buffer" input window; here a 7500 meters radius is inserted, signifying this computation regard a time threshold of 30 min; we did the same one for the 40 min one, inserting a radius of 11250 meters.

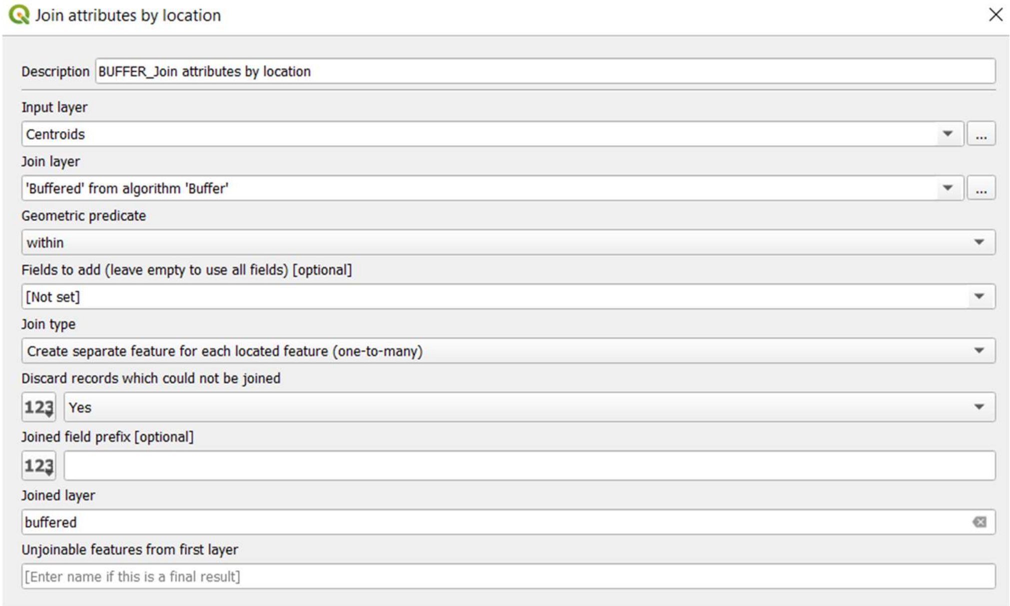


Figure 37 - "Join attributes by location" tool input window.

On the other side, we could identify the **reached** blocks of each NLA following a similar procedure on the isochrones. Indeed, the reached blocks of an NLA are the blocks falling within both the isochrone and the buffer area. Based on the source NLAs, we assigned an index to both the isochrones (index_area) and the buffers (CLUSTER_ID): these identification codes were lately passed to the points falling in their domain. This passage was needed, since we computed all the NLAs of a city together, and so needed to recognize which blocks belong to which NLA reachability or reached domain. Indeed, we finally extracted (Figure 38) from the whole set of blocks only the one belonging to both the isochrone and the buffer area of the same NLA (and so having "index_area" = "CLUSTER_ID").

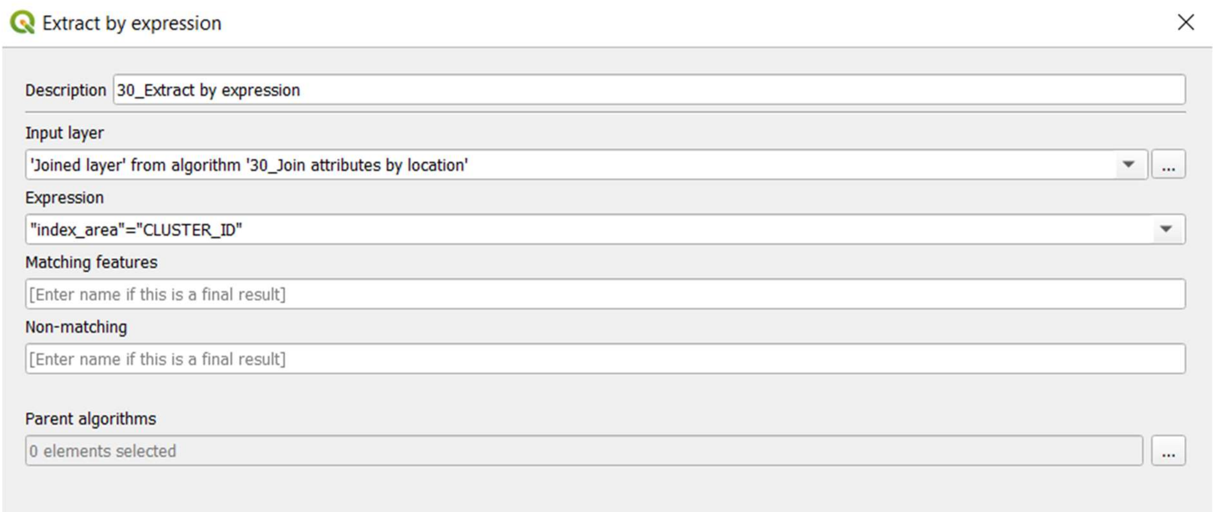


Figure 38 - "Extract by expressions" has been used to extract those points falling inside an isochrone and a buffer area related to the same NLA; "index_area" is the code meaning the belonging to an isochrone (reached domain), "CLUSTER_ID" indicates the belonging to a buffer area (reachable domain).

To better understand why a set of belonging codes was needed, a simplified example with 3 NLAs only is shown in Figure 39. The black dots are all the building blocks falling inside the three 7 km circular areas departing from each of the three NLAs (the reachable blocks). The red polygons are the three isochrones departing from the three NLAs (reached domain); with three different colours are indicated the building blocks falling inside each of them (the reached blocks). Since we want to make a comparison between reachable and reached blocks from a same NLA, we were not interested to all the points but only to the ones with the same identification codes (“index_area”=”CLUSTER_ID”).

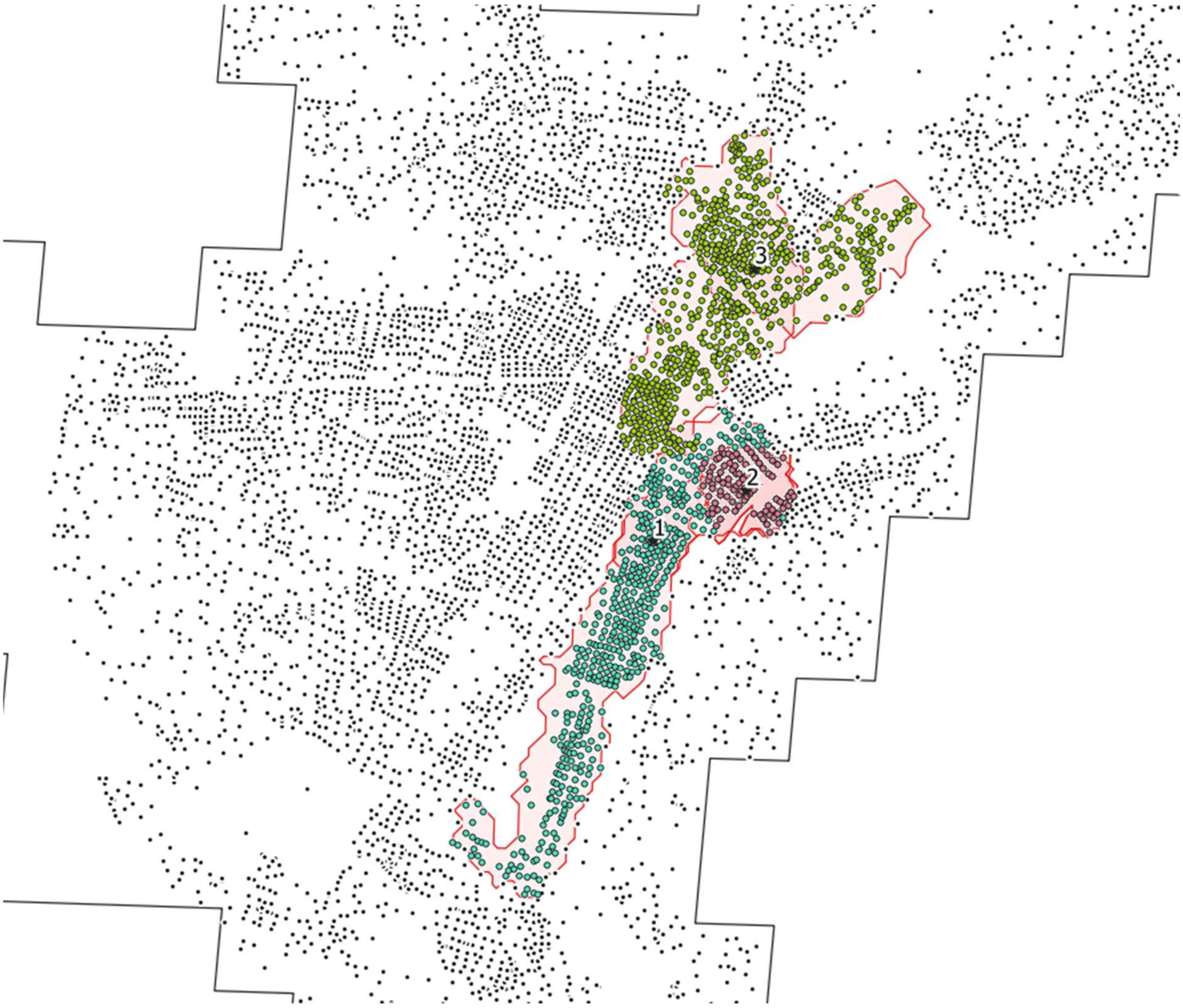


Figure 39 - Simplified example with 3 NLAs, Torino.

After it, we *aggregated* all the blocks which were both reached and reachable from a specific Night Life Areas asking the blocks populations values to be summed as shown in Figure 40. The *Aggregate* tool refers to a geoprocessing operation that creates a new layer by grouping features from an input layer based on a specified attribute and applying a function to the attribute values of the grouped features. This was made for each of the nine departing times and the results were merged, resulting in a layer containing n rows (where n is the number of NLAs) for each of the nine departure hours, each one reporting the number of inhabitants reached from each NLA (Figure 41).

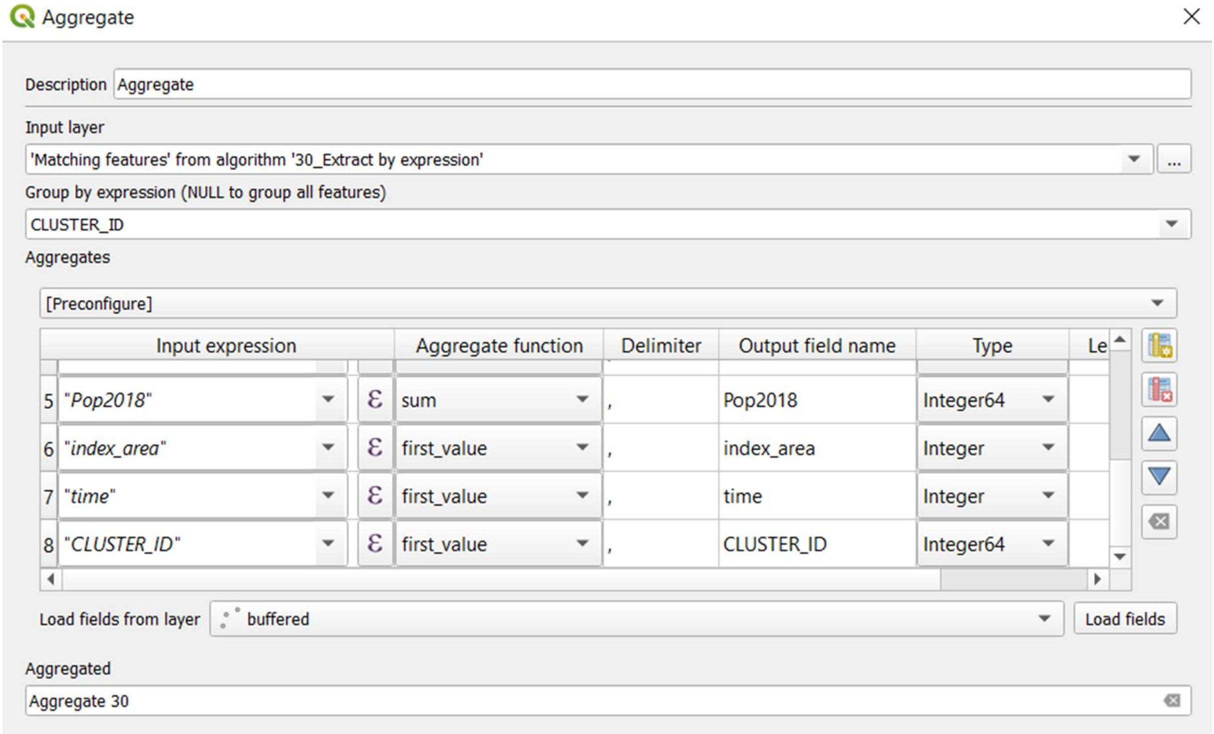


Figure 40 - Aggregation over same Night Life Areas

As results we obtained a table that looks as in Figure 41, which is the result of the merge of two of the nine aggregated layers obtained as described above; so, we have for each Night Life Area the number of inhabitants reached in each of the two departure times (02:00 and 02:15). Pop2018 is the population reached from the specific NLA considered, which can be identified through the number of "index_area" and "CLUSTER_ID" which are the same one for the reason explained above.

Merged :: Features Total: 6, Filtered: 6, Selected: 0

	country	fua_name	fua_code	perimeter	area	Pop2018	index_area	time	CLUSTER_ID
1	IT	Torino	IT004L2	148325,49526	3624905,10879	63611	1	200	1
2	IT	Torino	IT004L2	33471,16172	641329,30455	11734	2	200	2
3	IT	Torino	IT004L2	129913,96872	2825516,68170	61064	3	200	3
4	IT	Torino	IT004L2	94319,80552	1918161,62908	37253	1	215	1
5	IT	Torino	IT004L2	33471,16172	641329,30455	11734	2	215	2
6	IT	Torino	IT004L2	134830,33812	2600748,71165	57954	3	215	3

Figure 41 – results of the simplified example for two of the 9 departing times only.

3.5.3. Once served

Until now we considered the blocks reached by each NLA. That way, a single centroid could appear as reached from multiple NLA (this is the case of blocks located close to several NLAs or particularly well connected by the public transport network), or reached by a single NLA or never reached by any of the NLA (which is the case of blocks located in peripherical areas not connected to the network and with no NLAs around). That depends on the different spatial distribution of population and NLAs in different cities. However, on a transport policy viewpoint it could also be interesting to find out how many inhabitants of a given city are connected by public transport offer at late night to any NLA inside the city, rather than to a specific one. To consider this matter, we collected in a layer all the blocks which were reached at least once by any NLA at each computation. To extrapolate these blocks and their information we performed an "Extract by location" with the parameters set as in Figure 42. This time, since we are interested in the totality of the reached blocks, we didn't consider them multiple time when reached by multiple NLAs. Instead, we considered all of them only once: the join type has to be *one-to-one* and not *one-to-many*. To see an overview of how this value changes over the nine different departure times, we can aggregate (Figure 43) each of the nine results and merge them together at the end. We grouped by the attribute *fua_name* which is a string identifying the name of the considered FUA (in this case was "Torino"): we used this specific attribute just because being equal for all the considered features allowed us to create a layer with a feature only (Figure 44), we could have used another attribute with such characteristic, such as "country" (in this case Italy for all the blocks) or "fua_code" (in this case was IT004L2).

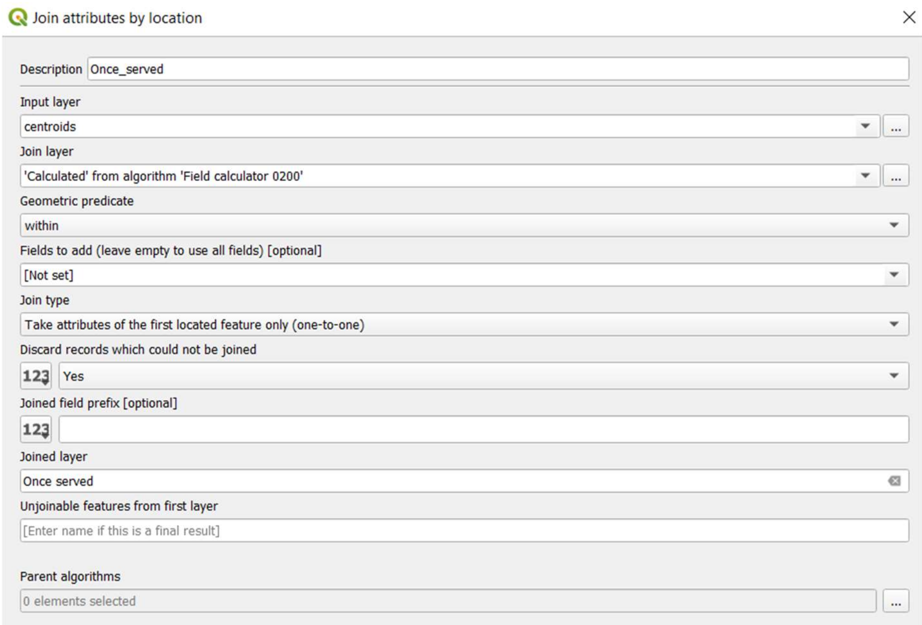


Figure 42 - Extract by location inputs parameters

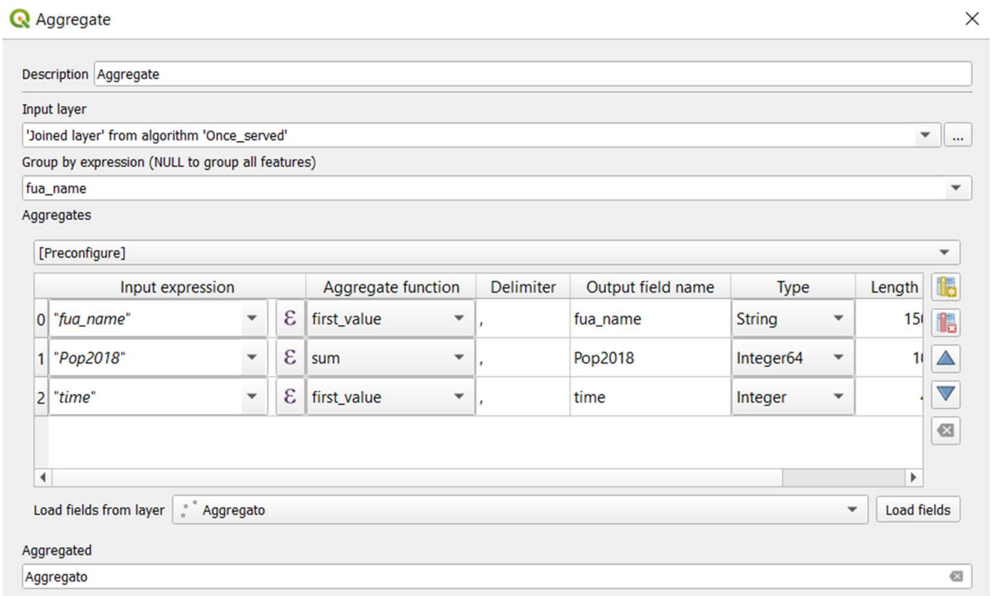


Figure 43 - Aggregate tool

	fua_name	Pop2018	time
1	Torino	101687	200

Figure 44 - Result layer of the aggregation

We performed the above analysis considering two different time aggregation levels:

1. As explained above, by considering it for each of the 9 departure time separately, so extrapolating the number of inhabitants served at least once by any NLA at each departure time.
2. Just taking the nine isochrones together at once and, following the same described procedure, extrapolating the number of inhabitants served at least once in the whole two hours study time range, without any specification regarding the departure time.

4. Computation results

In the following chapter an overview of the results gained applying the methods described in chapter 4 is given. Specifically, we will investigate the lengths of the night line networks, the number of amenities for each city, their grouping in clusters, their spatial distribution and the reachability computation results obtained with the Traveltime plugin. Finally, correlations analyses involving some of the quantitative results will be proposed.

4.1. Public transport night lines length and network intensity

A measure of the offer of public transport services are the vehicle per km travelled in the two hours, inferred thanks to network information acquired in GTFS format, specifically the lengths in km of each line and the number of vehicles running the route in the two hours study time. This measure is strictly related to the frequency of the service: the higher the frequency of the service and the length of the network the better the performance. It takes in consideration a specific time range, which was for us two hours. We were interested in the total number of vehicles operating on each line in the totality of the two hours and in the length of each line; we extrapolated this information using the following formula, which gave us a parameter we can call **service intensity**:

$$Length\ [km] * Vehicles\ in\ 2h\ [\frac{Veh.}{h}]$$

Since we wanted information related also to the population, we divided the product of these two measures per the thousandth part of the total population. In that way, we extrapolated the number of vehicles per km available for each thousand inhabitants in the two hours, which is a parameter we can call **service intensity per 1000 inhabitants**. The new formula looks like:

$$\frac{Length\ [km] * Vehicles\ in\ 2h\ [\frac{Veh.}{h}]}{inhabitants/1000}$$

The considered vehicles were all the vehicles whose departure time from one of the two terminus fall inside the two hours study time chosen for the specific city.

Following the previously explained procedures on QGIS (section 3.2) we can compute the length of the public transport network for each city, considering either the whole network or more incisively the portion of it falling inside the Urban Centre.

In Table 14 we can see for each city the length of the night-time public transport service, considering firstly the totality of the line and then the portion which falls inside our study area, delimited by the Urban Centre borders. The two described offer measures are shown, namely the service intensity and the service intensity per thousand inhabitants. Finally, it is shown, what was the average frequency for each network. In most cases we can see a rounded number, signifying that in the specific network the frequency is fixed at a same value for every line. So for example for Milano, where every line line has a 30 minutes frequency, we can see an average frequency of 30 minutes-; for München, where each line has its own schedule, a rounded value is not reached. This value is to be intended as a partial indication of what to expect in terms of measured offer, indeed, the higher the frequency the higher the offer (at least in the way we consider it). At the last column is presented the parameter called *Network Intensity,* obtained as the ration between the Urban Centre surface and the network length inside of it. Cities with an higher Network Intensity, have a longer network if compared to the related Urban Centre surface.

Table 14 - Length of night-time public transport lines (km) for the total network and for the study area; offer measures computed as service intensity and service intensity per thousand inhabitants; average frequency for each line operated in the two hours study time; Network intensity.

City	Total length	Length_Urban centre	Service efficiency	Service efficiency per 1000 inh	Average Frequency	Network intensity
	km	Km	Veh*km*2h	Veh*km*2h / (Inh/1000)	minutes	Km/km²
Budapest	1276.2	1078.7	3228.2	1.87	40	2,49
Milano	322.5	317.4	1269.6	0.40	30	0,42
München	773.1	726.0	3404.2	2.15	26	2,08
Praha	1159.2	896.9	2738.5	2.29	39	3,04
Roma	1104.1	849.7	3398.7	1.36	30	1,76
Torino	159.3	130.7	261.4	0.22	60	0,63
Valencia	411.2	366.3	1465.3	0.97	30	1,10
Wien	573.0	570.2	2678.7	1.35	26	1,45

In Figure 45 the different lengths (the total one and the one considering the urban centre only) are shown, together with a line indicating service intensity (the orange one) and another line indicating the service intensity per 1000 inhabitants (purple line): this value has been multiplied per 1000 to facilitate the visual impact of the data.

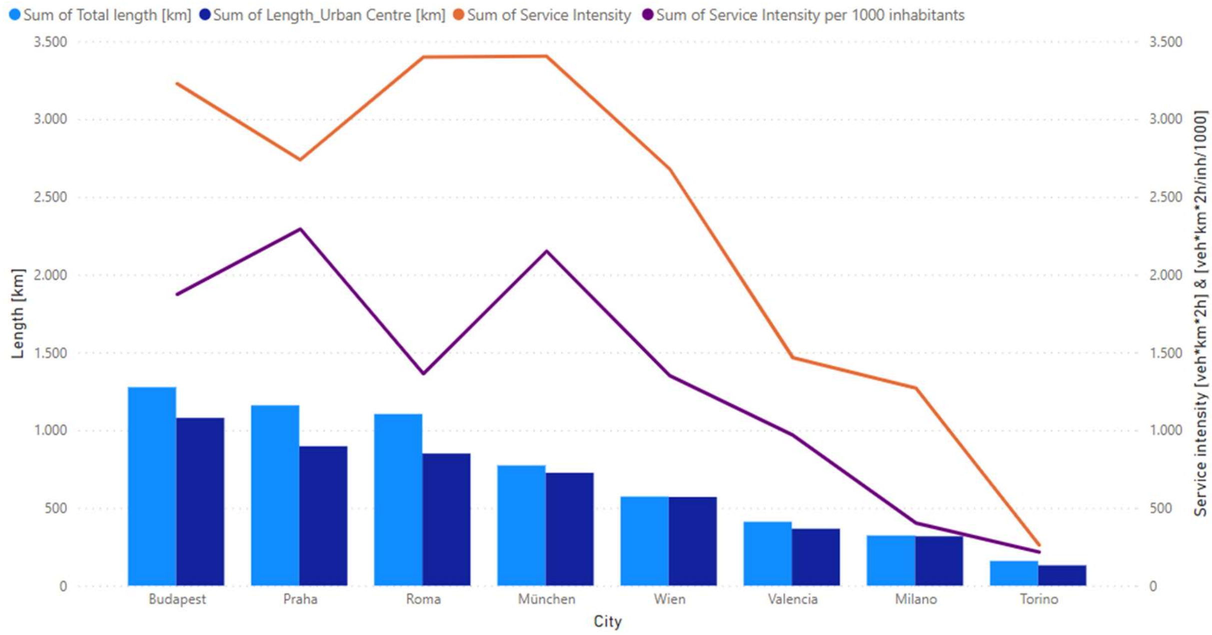
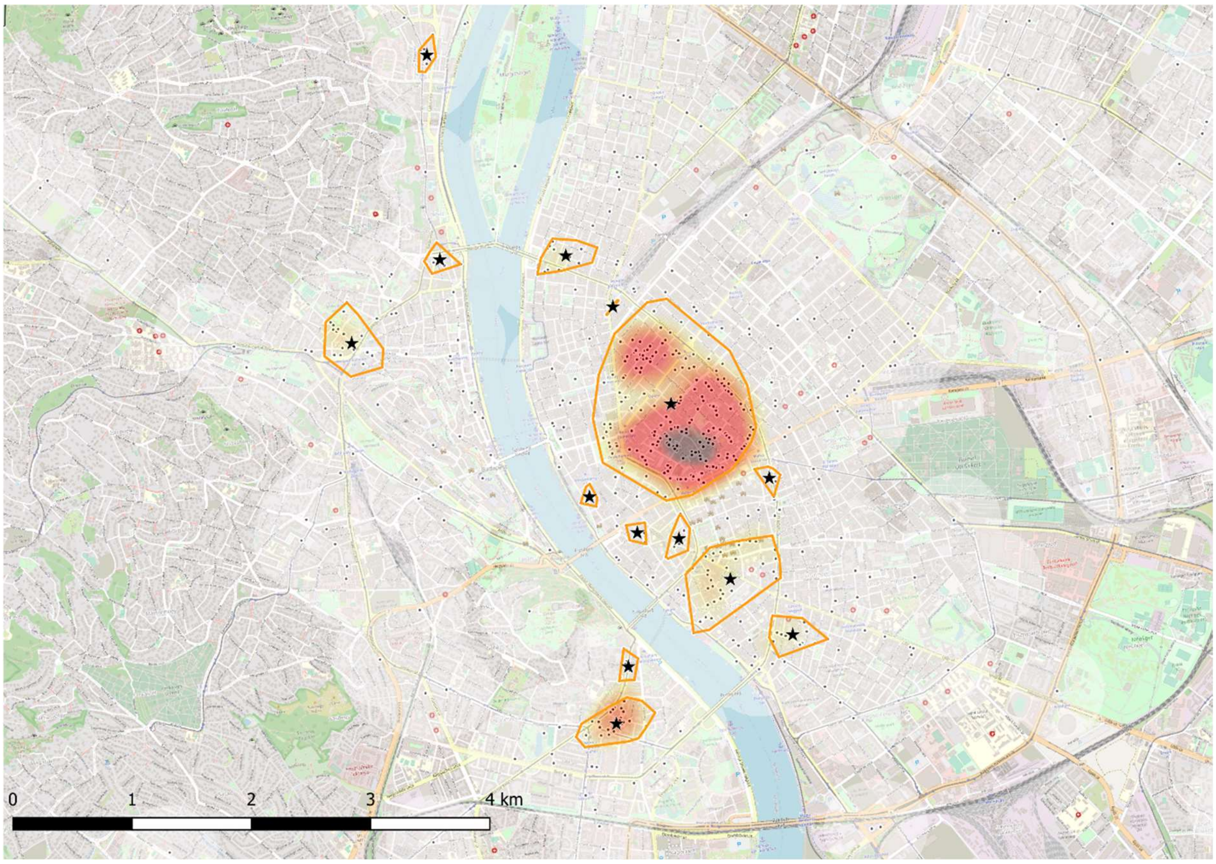


Figure 45 - Comparison between networks lengths (the total ones the ones inside the study area) and offer measures.

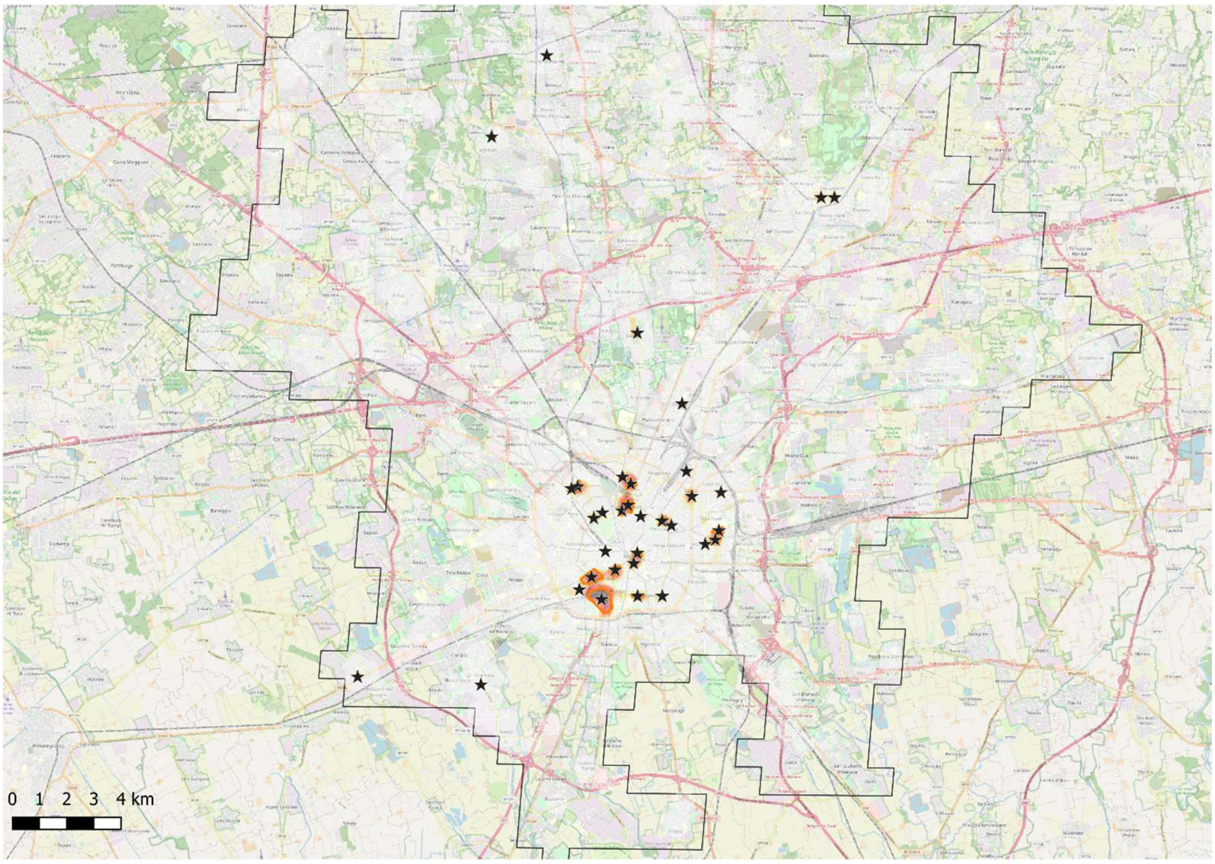
In *blue sky* we see the total lengths for each city, while in *dark blue* the lengths of the networks falling inside the study area delimited by the urban centre borders. The cities are disposed following the total length order, from the longer to the shorter. It can be seen as the order remains the same for the lengths inside the urban centres. The orange line, representing the offer measure expressed as vehicles per km per 2 hours (service intensity) is expected to show higher values for longest networks and higher frequencies.. Looking at the average frequencies in Table 14, we expect the orange line to be much higher than the columns for cities with an high frequency: this is particularly true for München, which register the best offer, still not presenting the longest network but with the highest average frequency. Torino presents the shortest network and the lowest frequency, which, added that it shows the second worst Network intensity after Milan, inevitably bring to the worst offer. Budapest average frequency is slightly lower than Praha one, but this evidence is not encountered in the offer measure, considering a similar length of the two networks: this shows how incorrect it is to completely lean on the average frequency concept, since the lines lengths are not considered: what could have happened in the Budapest – Praha comparison is that Budapest presented higher frequencies on the longest lines and/or vice versa for Praha. The purple line, representing the offer measure considered for thousand inhabitants (service intensity per 1000 inhabitants), allows to compare the offer measure of each network to the actual population of the city: for more populated citied an higher offer is desirable; the value is multiplied for 1000 in order to have a better visual impact. We can see how the value of this new offer measure, compared to the one of the previous (service intensity), drops dramatically in Roma and Milano's cases, which having the highest population inside the urban centre, require a better offer. Indeed, Roma's result for this parameter is not the worst one of the sample, but it has the highest difference compared to the same parameter not considering the inhabitants; that can be assessed by the vertical spacing between the two values and the following change of trend followed by the two lines of the two parameters. On the other side, Praha, having the smaller population came out to have the highest value of offer per 1000 inhabitants.

4.2. Night-life areas identification

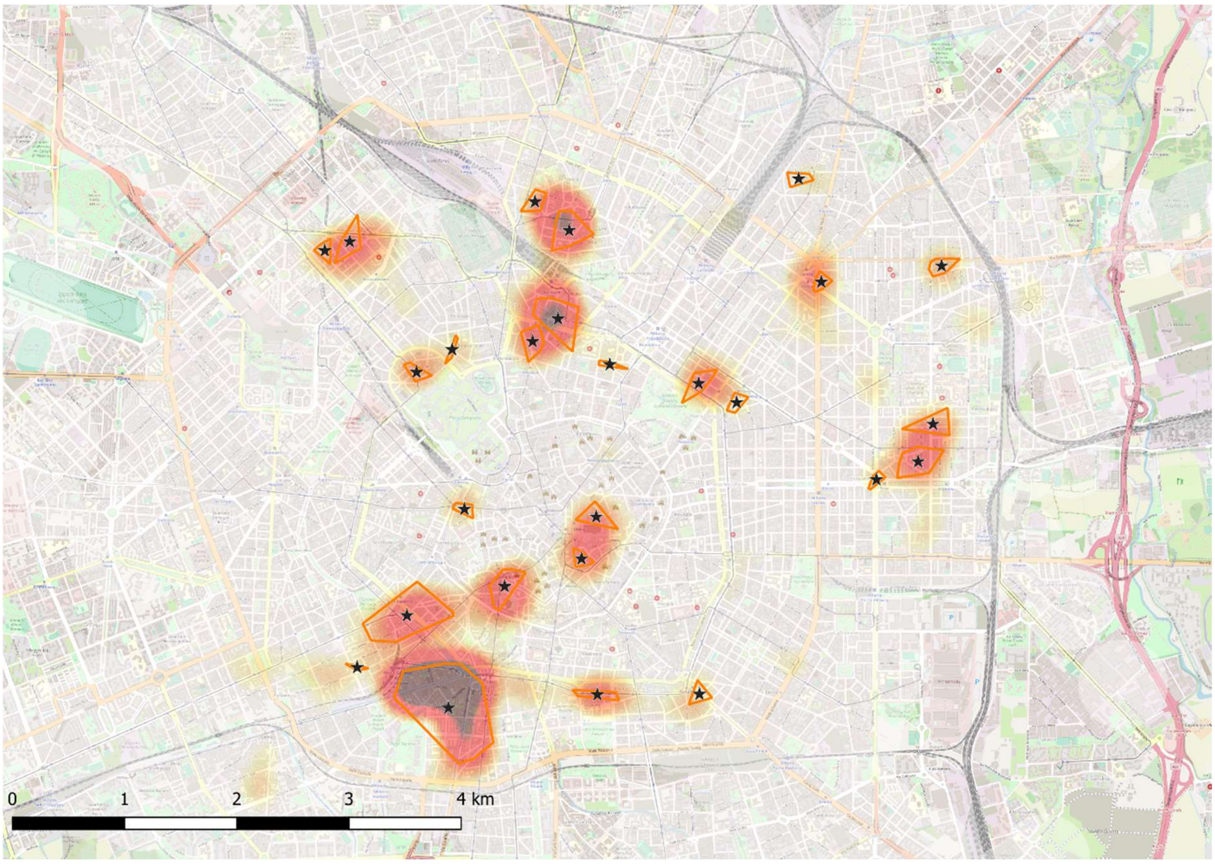
Each city has a different number of amenities relatable to night life and a different way in which they are distributed. We can see in Figure 46 how they are distributed looking at both an heatmap and a cluster representation, where the orange polygon represent the minimum surface enclosing all the amenities composing a cluster and the black star is the centroid of this polygon, from which the public transport offer was evaluated.



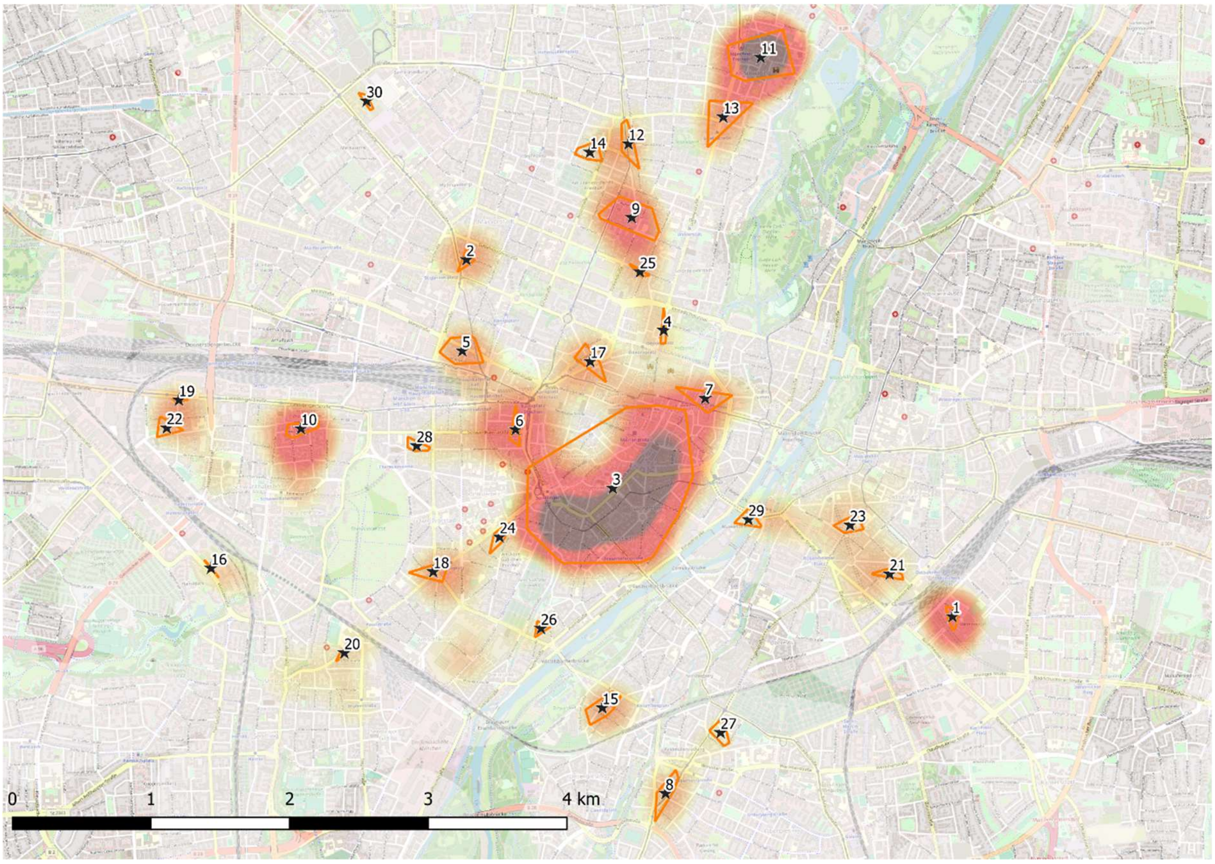
a) Budapest



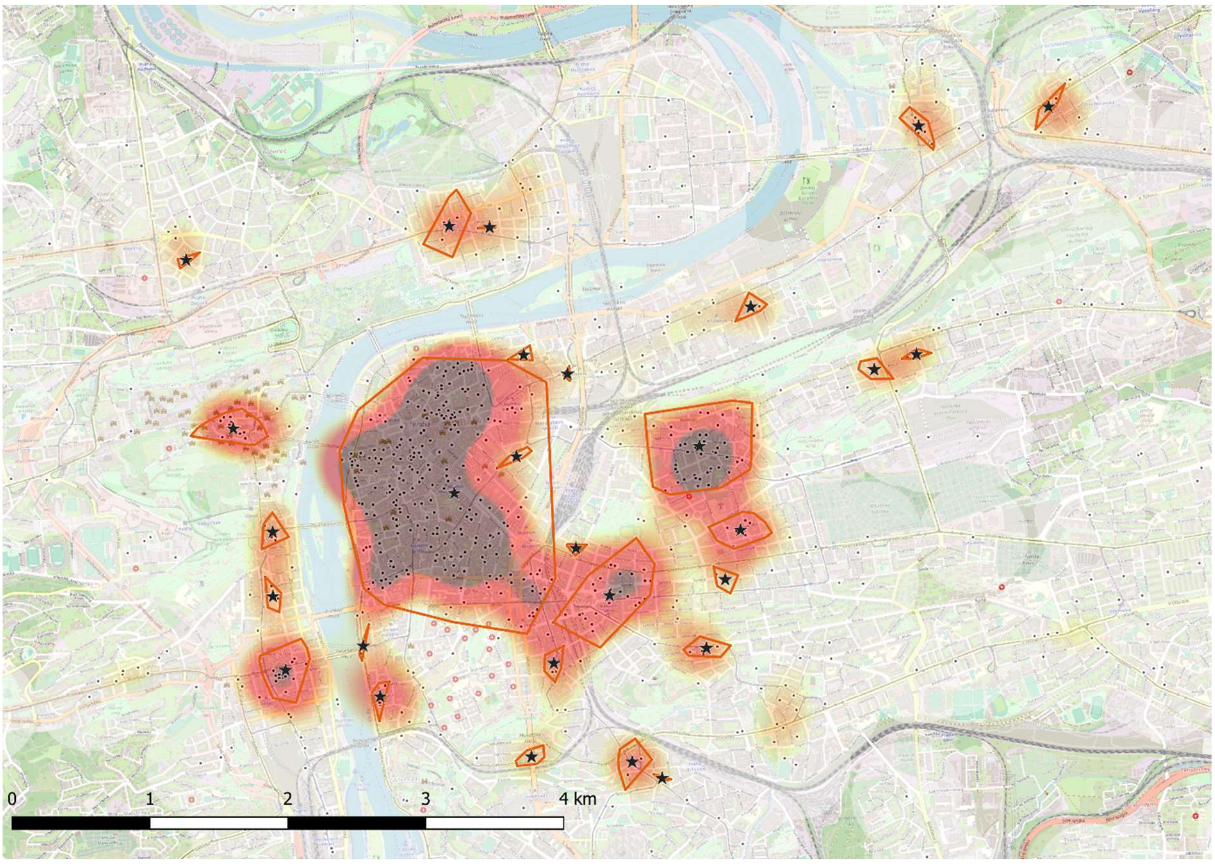
b) Milano



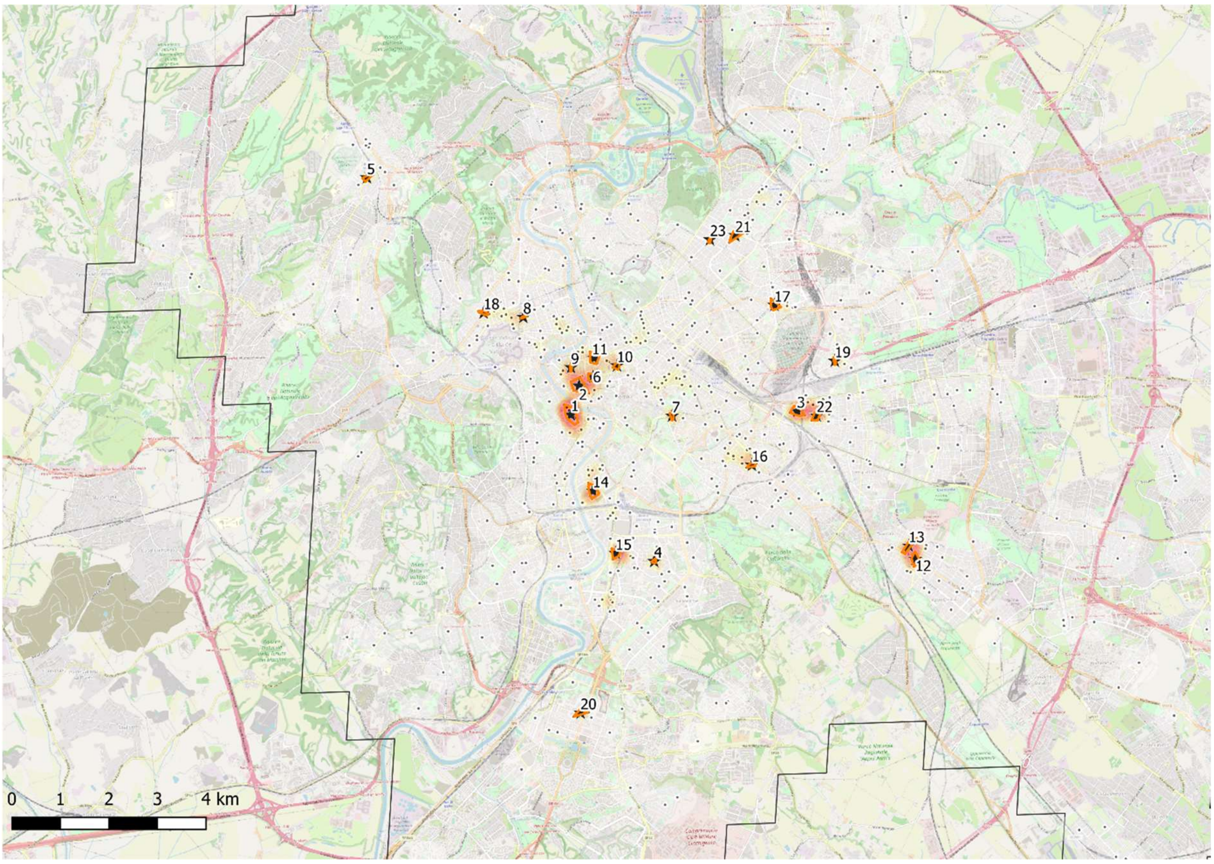
b2) Milano detail



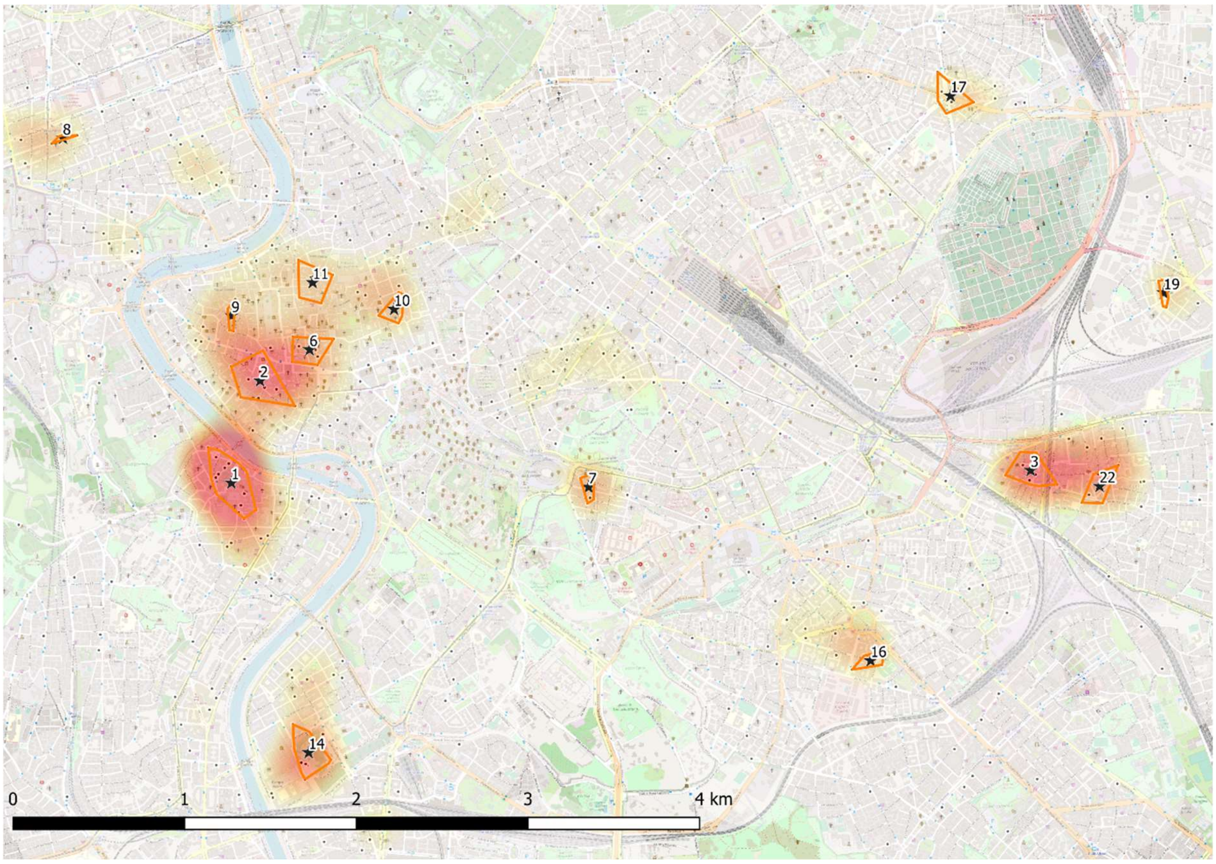
c) München



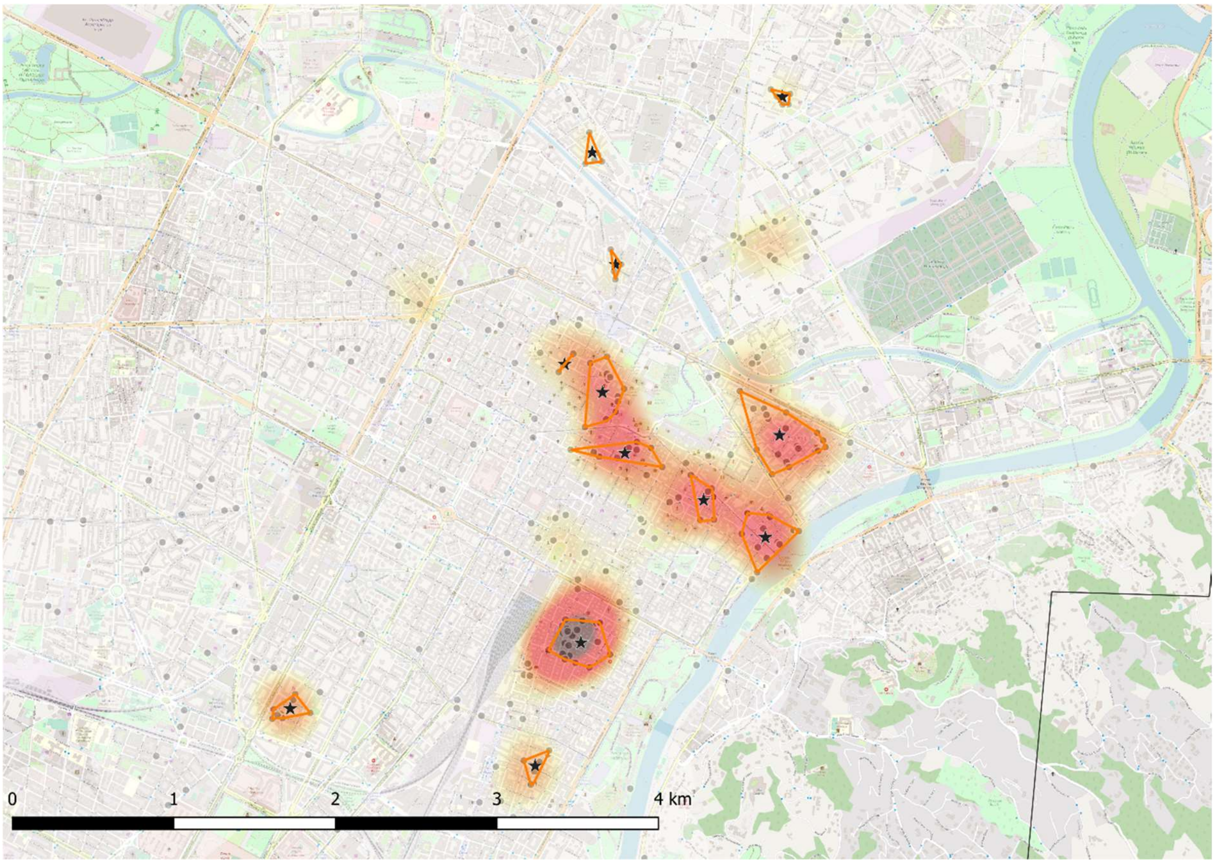
d) Praha



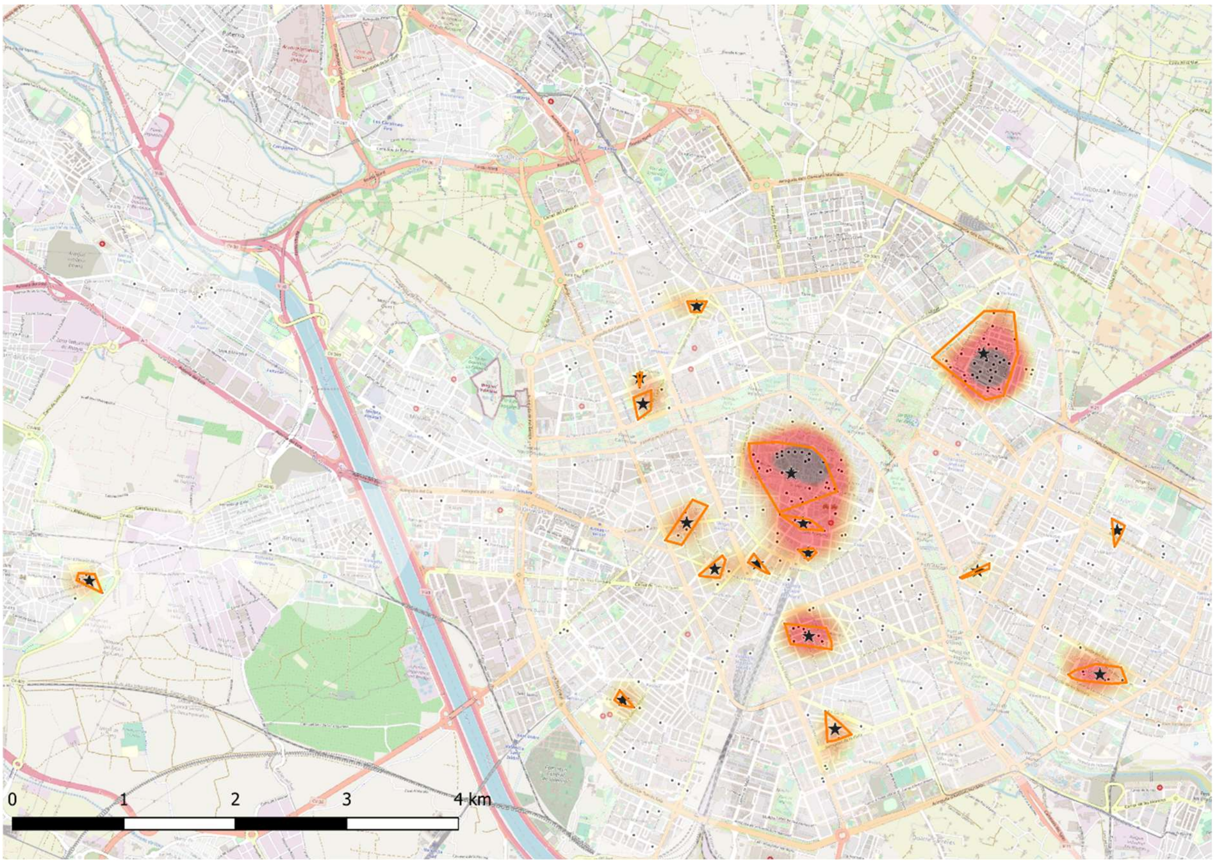
e) Roma



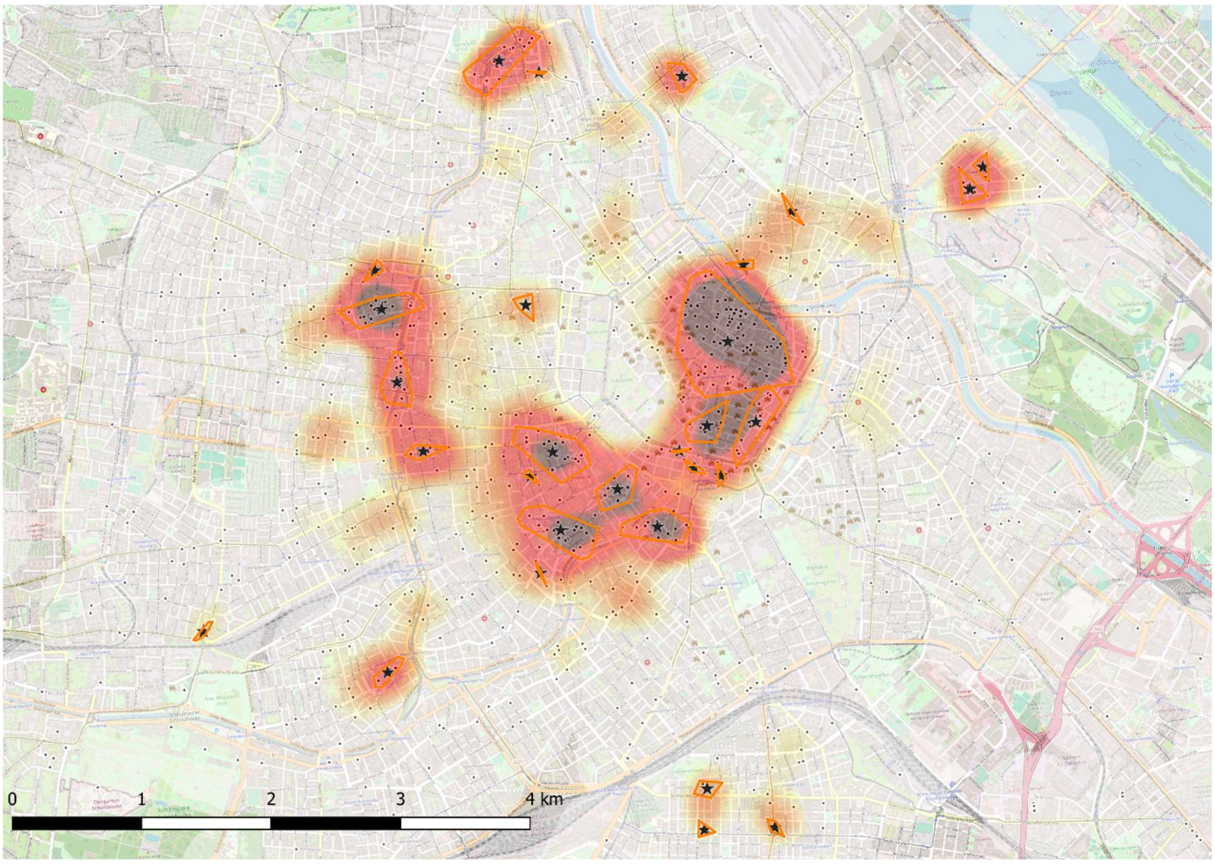
e2) Roma detail



f) Torino



g) Valencia



h) *Wien*

Figure 46 - Visualization of NLAs distribution through clustering and heatmap: a) Budapest, b) Milano, b2) detail from Milano, c) München, d) Praha, e) Torino, f) Roma, f2) detail from Roma, g) Valencia, h) Wien

In Table 15 the number of relevant amenities considered for each city is reported, also indicating how many of them have been used in the cluster process (not all the amenities were situated in a position which allowed them to create or be part of a cluster) and the number of created clusters. In the last columns an attempt to show the correlation between the number of clusters in a city and the number of its inhabitants is shown: specifically, the values in the last column are the results of a ratio between the total population of each urban centre and the number of night life related amenities in it, resulting in the average number of inhabitants for each amenity. Milano came out to be by far the studied city with the highest number of night life related amenities, followed by Praha and Wien: the high value encountered in Milano case could be due to the extension of the urban centre and bigger amount of population living inside of it; indeed the city with more amenities for inhabitants, and so with less inhabitants for each amenity, is Praha, which is both one of the city with the higher amount of amenities and the city with the lowest population.

Table 15 – Number of night-time related amenities, the ones being part of a cluster, found clusters and virtual number of inhabitants for each amenity for each city.

City	Amenities within the Urban Centres	Amenities falling inside a cluster	Number of created clusters	Inhabitants for each amenity
Budapest	917	446	14	1879
Milano	1680	370	35	1879
München	985	347	30	1607
Praha	1156	608	26	1033
Roma	773	165	23	3225
Torino	484	120	12	2494
Valencia	474	207	18	3195
Wien	1143	434	29	1736

4.3. Public transport coverage of the different Night-life areas

After running the reachability computation with the Traveltime plugin, as explained in Chapter 3.5, we got to know the number of inhabitants reached by public transport within a certain time threshold, at a certain departure hour, from each NLA. Comparing this value with the number of inhabitants enclosed in the corresponding circular buffer area, we deducted the coverage percentages for each NLA. Thanks to averaging processes we could aggregate these results at a city level. The whole procedure is explained with an example on the city of Budapest. As first step, in Table 16, we collected information on the number of amenities composing each NLA, so that it was possible to assign a weight to each NLA: the weight is taken as the quota of amenities inside a specific NLA related to the total clustered amenities. Then we collected the number of inhabitants falling inside the buffer area around each NLA: in the example we consider a time threshold of 30 minutes, so the buffer area has a 7.5 km radius.

Table 16 – Collection of data about the weights of each NLA and the number of inhabitants inside the buffer area. Example for Budapest with a 30 minutes time threshold.

NLA	Number of amenities within the NLA	NLA'sWeight	Pop in a7.5 km radius
1	19	4.26%	1013656
2	8	1.79%	1021537
3	245	54.93%	1048070
4	44	9.87%	1075425
5	7	1.57%	1103274
6	17	3.81%	1091852
7	34	7.62%	983869
8	9	2.02%	924971
9	24	5.38%	859978
10	9	2.02%	932369
11	10	2.24%	1043766
12	9	2.02%	1004003
13	6	1.35%	989419
14	5	1.12%	1030314

Then we collected the results of the computation, which are namely the numbers of inhabitants reached by every NLA, and we computed the coverage percentage for each NLA to obtain a weighted average when jointly considering all NLAs. For the example showed in Table 17, regarding the city of Budapest, a time threshold of 30 minutes was used; for visual reasons, only four departure times are shown. For each Departure time, which in this example vary from 01:00 to 01:45 the number of inhabitants inside the buffer area reached from each NLA is reported and the related coverage obtained as a percentage of inhabitants reached over the ones enclosed in the buffer area. For each departure time, the values referred to each NLA are then averaged two times, the first time through a standard average function and the second one through a weighted average, where the weights are the ones computed in Table 16 deducted from the number of amenities in each NLA.

Table 17 - Example of data treatment with the results of the Traveltime computation. Example for Budapest, with 30 minutes time threshold.

	Time:	01:00	Time:	01:15	Time:	01:30	Time:	01:45
NLA	Reached 1	Coverage	Reached 2	Coverage	Reached 3	Coverage	Reached 4	Coverage
1	283425	27.96%	367517	36.26%	262872	25.93%	398217	39.29%
2	349851	34.25%	386091	37.80%	349695	34.23%	409600	40.10%
3	374694	35.75%	352363	33.62%	337629	32.21%	346001	33.01%
4	319107	29.67%	331150	30.79%	315561	29.34%	308098	28.65%
5	485011	43.96%	409911	37.15%	447302	40.54%	350035	31.73%
6	272949	25.00%	303612	27.81%	266904	24.45%	348316	31.90%
7	375242	38.14%	356620	36.25%	341643	34.72%	309897	31.50%
8	348790	37.71%	313081	33.85%	297670	32.18%	283861	30.69%
9	328753	38.23%	315060	36.64%	295898	34.41%	301937	35.11%
10	87433	9.38%	101766	10.91%	90599	9.72%	104439	11.20%
11	349851	33.52%	374692	35.90%	349831	33.52%	383122	36.71%
12	313920	31.27%	316113	31.49%	266501	26.54%	308350	30.71%
13	415280	41.97%	322612	32.61%	366204	37.01%	362069	36.59%
14	326254	31.67%	355597	34.51%	337712	32.78%	394768	38.32%
Average	330754	32.75%	329013	32.54%	309002	30.54%	329194	32.54%
Weighted	352199	34.23%	342439	33.28%	323002	31.37%	335362	32.57%

Finally, in Table 18 the two aggregated results obtained for each departure time, are further averaged over the nine different times to obtain two global values aggregated at city level (but not at time threshold level, indeed we kept the results referred to different time thresholds separated).

Table 18 – Coverage aggregated at city level. Example for Budapest, with 30 minutes time threshold.

	Coverage_30 min
Average	30.95%
Weighted	31.57%

Then, we considered the inhabitants served at least once, following the two definitions introduced in chapter 3.5.3. For the first one, whose data are collected in Table 19, all the numbers of inhabitants connected to at least a NLA for each departure time are shown. An aggregation at city and time threshold level was made here too, averaging the values. In this case it didn't make sense to consider the weights, since the considered values are not referred to a specific NLA but to the whole set. The second definition, which express the inhabitants served at least once by any NLA at any departure time, is reported as a single value, collected in Table 20.

Table 19 – Inhabitants served at least once for each departure time and percentage over the total population. Example for Budapest, with 30 minutes time threshold.

Time	POP TOT	Onceserved	Perc

200	1,723,314	736799	42.75%
215		674223	39.12%
230		633376	36.75%
245		682648	39.61%
300		710195	41.21%
315		642444	37.28%
330		697690	40.49%
345		671972	38.99%
400		774960	44.97%
TOT		691590	40.13%

Table 20 – Inhabitants served at least once at any departure time and percentage over the total population. Example for Budapest, with 30 minutes time threshold.

total_onceserved	**Perc**
870610	50.52%

Generalising the previous results that were only related to Budapest, all the results have been collected for each city, as shown in Table 21. It is also possible to see them on graphs, as in Figure 47 and Figure 48, in order to have a visual comprehension of the data. In the visual representations the values are shown in decreasing order considering the coverage ratio. We can see how this order would have been different if we changed the considered parameter. The reasons behind these differences will be explored in chapter **Errore. L'origine riferimento non è stata trovata.**.

Table 21 - Results of the traveltime computation for each city

City	**Inhabitants in Urban Centre**	**Coverage_30 min**	**Weighted Coverage_30min**	**Coverage_45min**	**Weighted Coverage_45min**	**Onceserved_30min**	**total_onceserved_30min**	**Onceserved_45min**	**total_onceserved_45min**
Budapest	1,723,314	30.95%	31.57%	49.76%	50.85%	40.13%	50.52%	65.03%	80.83%
Milano	3,157,550	7.13%	7.79%	16.65%	19.64%	24.22%	31.83%	36.39%	43.07%
München	1,582,533	16.34%	18.19%	38.66%	42.36%	47.01%	65.62%	67.55%	85.45%
Praha	1,194,604	15.13%	15.19%	36.23%	37.38%	46.14%	60.40%	75.37%	93.06%
Roma	2,493,124	10.48%	10.31%	20.88%	20.67%	38.46%	56.62%	56.85%	80.31%
Torino	1,207,163	5.94%	5.35%	8.89%	9.01%	18.93%	44.46%	25.74%	57.52%
Valencia	1,514,210	18.52%	18.40%	32.81%	33.43%	48.16%	55.19%	54.79%	59.62%
Wien	1,984,266	21.94%	24.16%	52.98%	57.96%	61.55%	69.00%	81.82%	88.27%

.

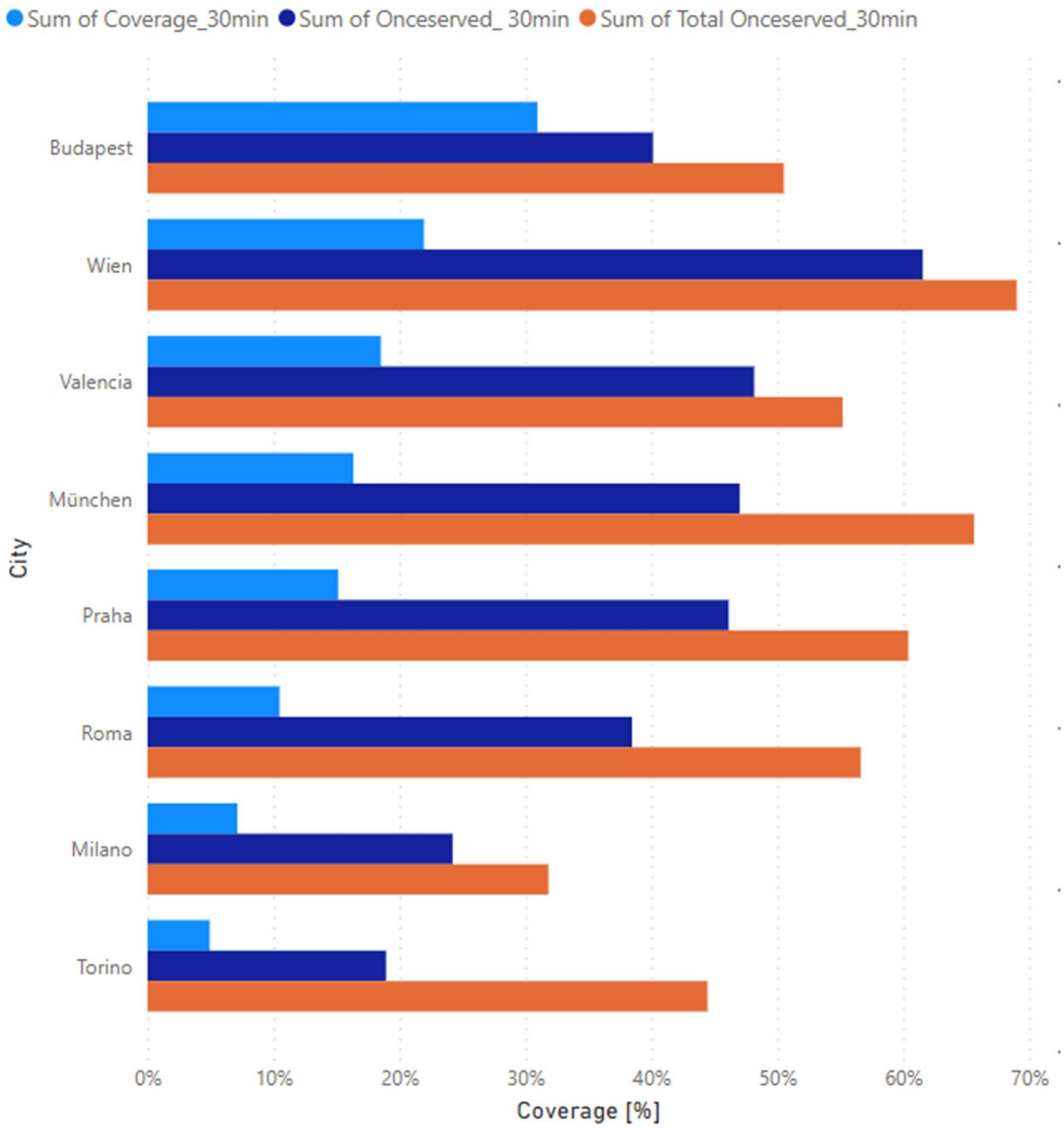


Figure 47 – Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 30 minutes.

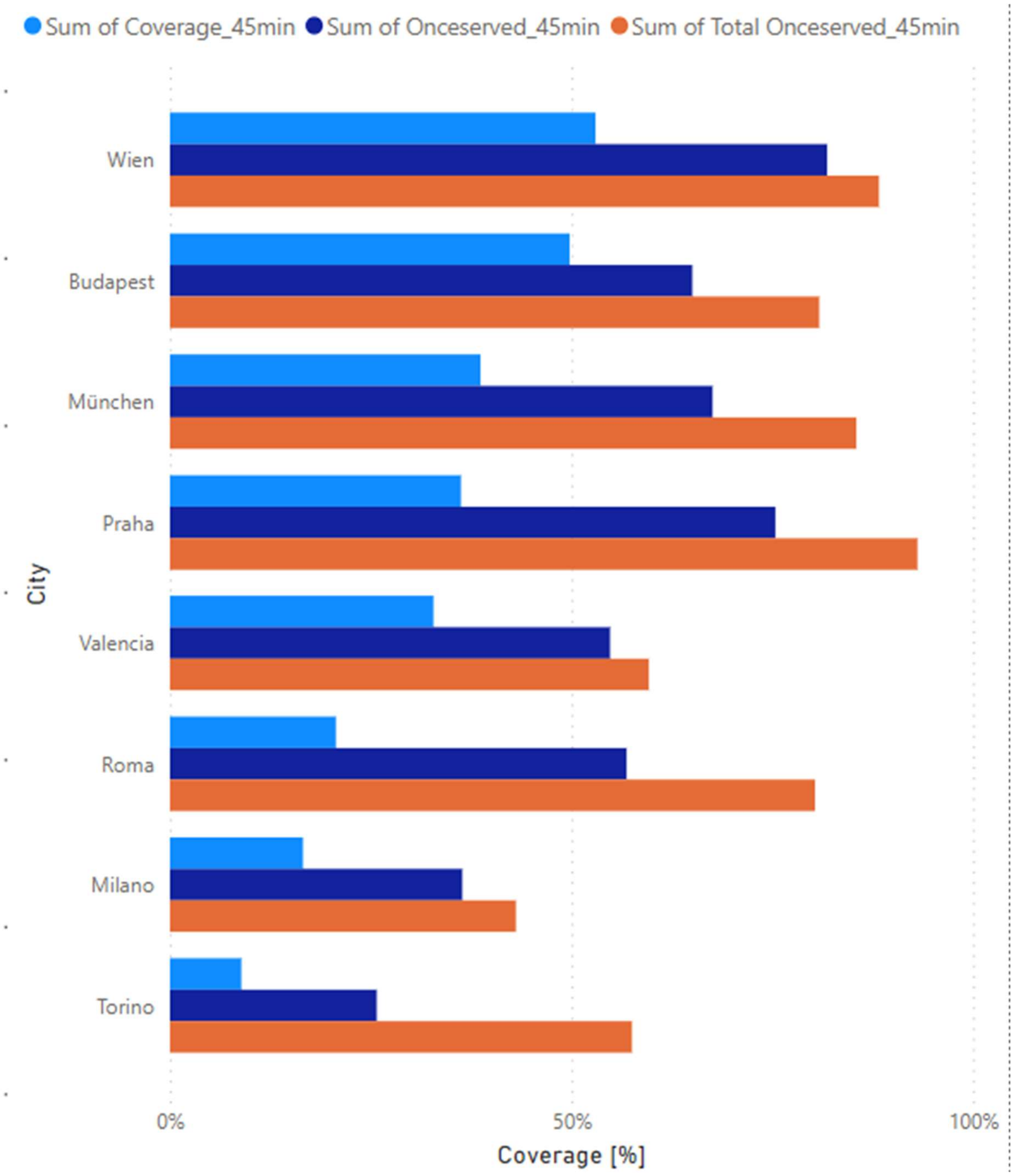


Figure 48 - Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 45 minutes.

We can see the changing order of cities, when we consider the coverage, for different time thresholds from the numerical data in the table and the visual stats in the previous two pictures. So, for example. In the 30 minutes time threshold Budapest is the city with the highest coverage, while in the 45 minutes time threshold, it is Wien. This happens because a more compact peripheric belt in Wien, means a larger number of inhabitants close to the public transport stops than in Budapest, where the peripheral areas are more sprawled. We can also see as Valencia loose two positions in favour of München and Praha: this is due to the lack of night lines covering the suburbs in Valencia, which on the other side, are present in Praha and München. Concerning the *Once served* parameter, Wien appears to have the best results in both time thresholds configurations, confirming the networks' high grade of penetration in the suburbs. When considering the parameter *Total once served*, which indicates a total value of inhabitants served during the study time not being averaged, we can see how Wien has the highest value for the 30 minutes time thresholds but not for the 45 minutes one, which belongs to Praha: this means that the Praha network covers an higher percentage of inhabitants in the suburbs, but less coherently in time compared to Wien, since the averaged value is lower.

All the results found and explained in this chapter, are collected separately for each city in Appendix B.

4.4. Benchmark analysis

4.4.1. Number of inhabitants versus service offer

We made a benchmarking exercise through numerical and visual correlations between the service coverage, the service offer and the number of inhabitants in the study area. In Figure 49 and Figure 50 we can see the correlation between the public transport offer, measured as vehicle per km in the two hours range, and the total population for each city (the dimension of the points is given by the computed service coverage related to the 30 minutes time threshold in

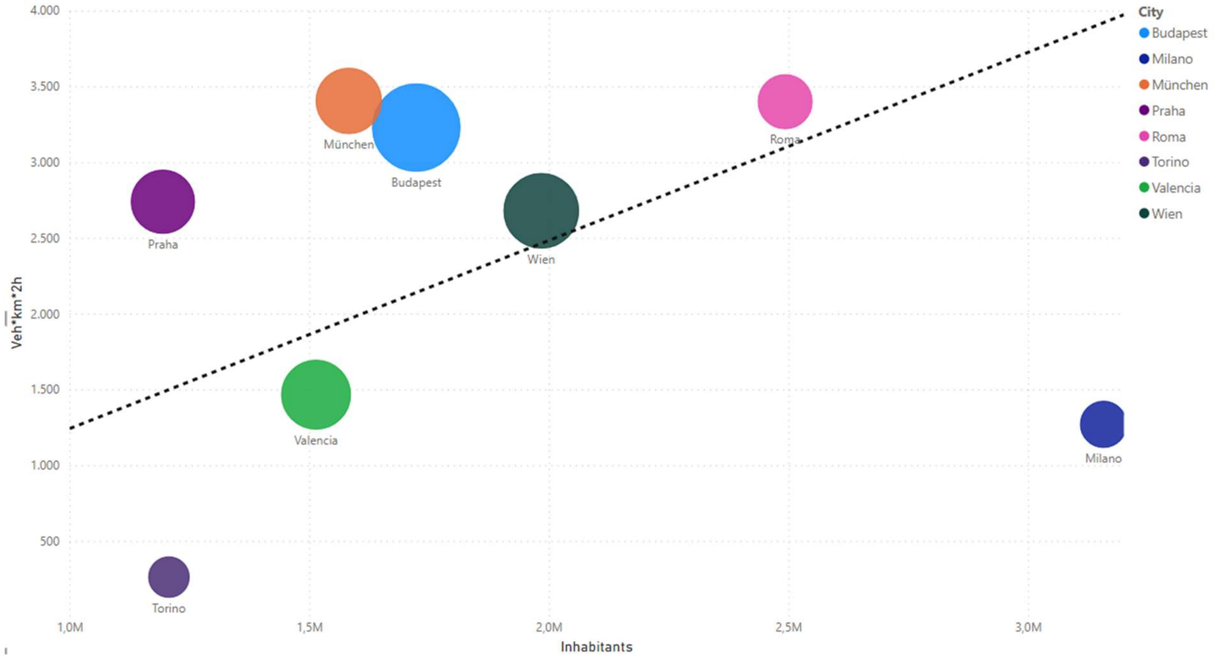


Figure 49 and the 45 minutes time threshold in Figure 50). From this correlation we expected cities to be around an interpolated diagonal which could identify an empirical length of public transport lines pro capite (this value is based on the chosen cities in the sample, so it is not to be considered a global value). The position of the points above or under the diagonal will suggest some information about the cities themselves: cities falling above it are either highly sprawled or present an over-development of the service, on the other side, when falling beneath, it could be because of a compact structure of the urban agglomerate or an under-developed network.

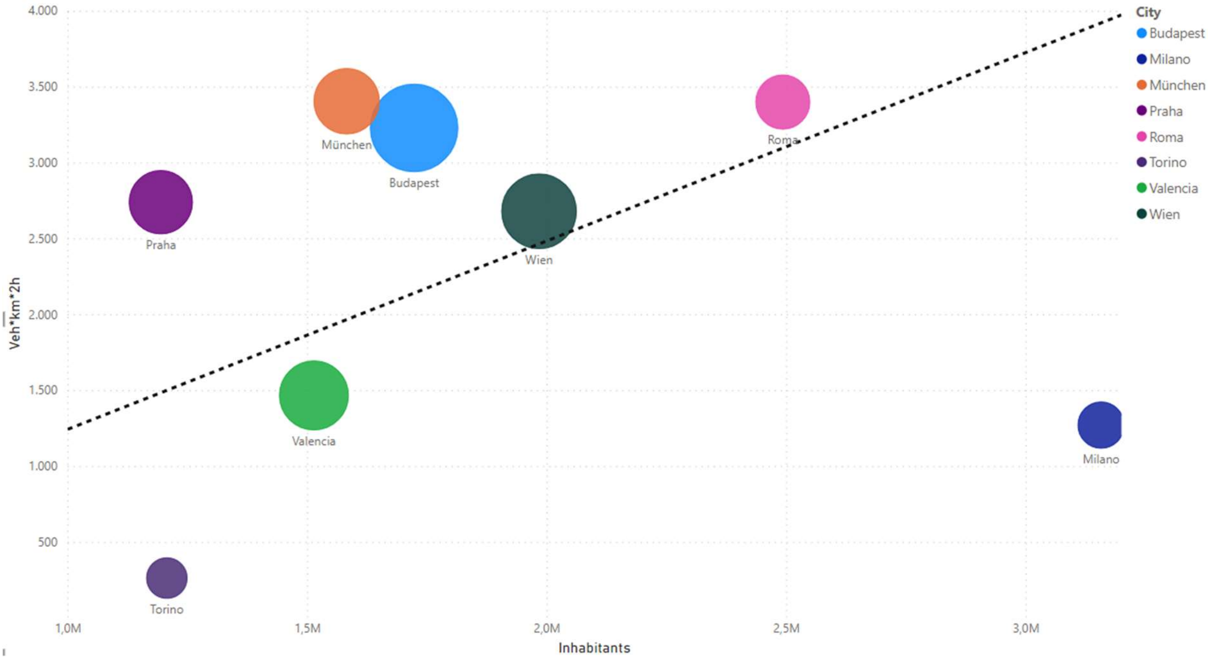


Figure 49- Correlation between length of the networks and number of inhabitants in the study area. The dimension of the points is given by the value of service coverage for the 30 minutes time threshold.

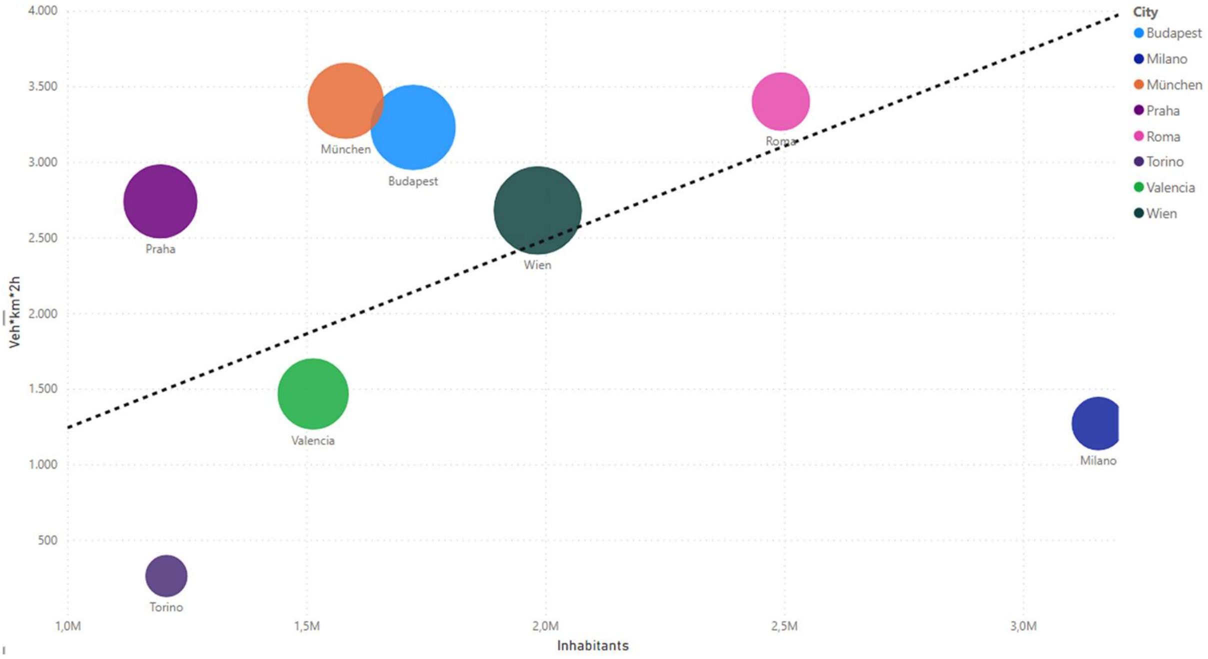


Figure 50- Correlation between length of the networks and number of inhabitants in the study. The dimension of the points is given by the value of service coverage for the 45 minutes time threshold.

The cities of Wien, Roma and Valencia are the ones closer to the interpolated average of km of lines per inhabitant. For Wien and Valencia, since they are also two of the most performing cities of the sample in terms of service coverage, we can state that, always considering this specific sample, the offered service of those two cities is adequate. For Roma, since the service coverage is one of the lowest in the sample, this collocation indicates how the city conformation, which is highly sprawled, is not beneficial in terms of service coverage and/or the network design don't fit the land use. So even with an offer superior to what expected for the related total population, the service is still not performing well as others.

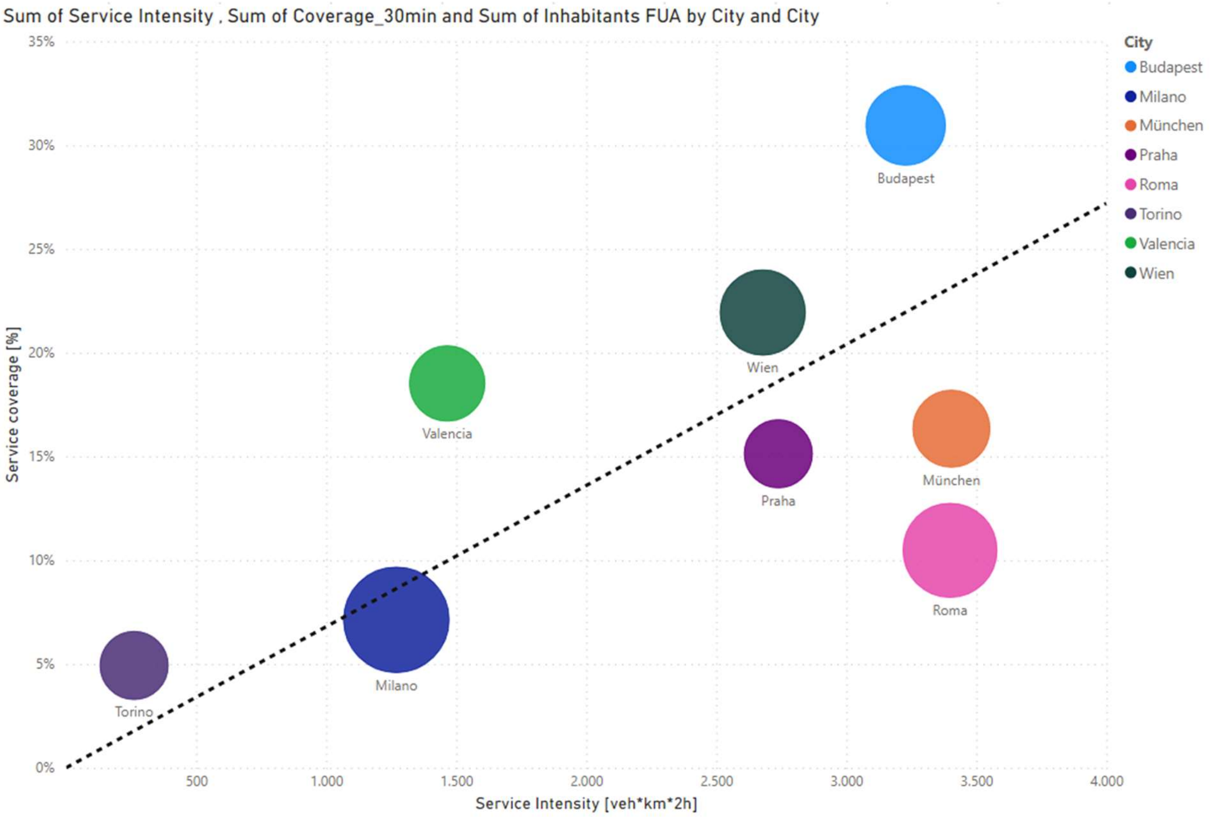
The most performant city in terms of service coverage, came out to be Budapest, which in this figure lays far above the line, meaning a greater length for each inhabitant than the sample average; this can be due to three different factors (and one doesn't exclude the others):

- The city of Budapest is more sprawled than the others, so it needs a longer public network to reach a higher number of inhabitants.
- The system design is not coherent with the NLAs position.
- The NLAs are sprawled and so from each of them a smaller portion of the city is reachable. When this is the case, we should meanwhile observe an increase of the Once_served parameter value.

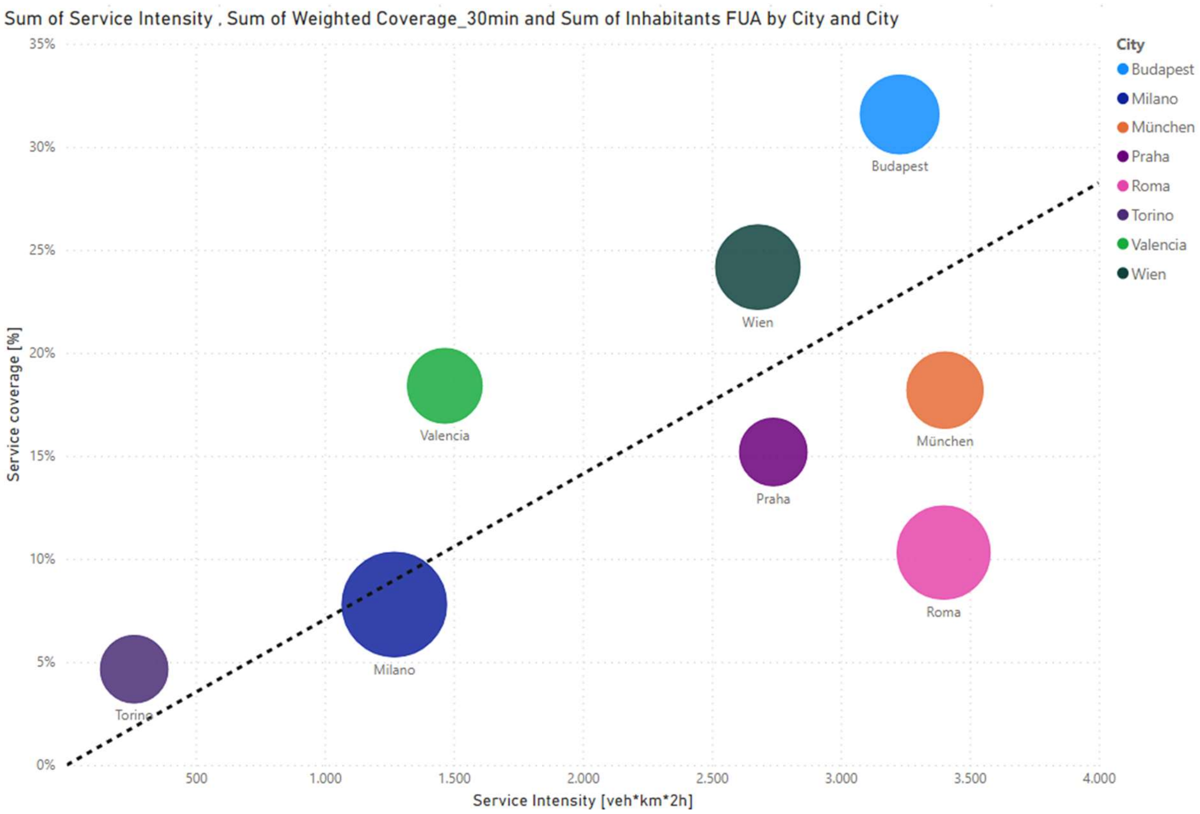
The other two cases of cities laying above the diagonal are Praha and München. Both cities present an evident sprawl of population and that is the reason of an over-extension of the network, justified by two similar service coverages, which place themselves in the middle of the sample results. For Torino and Milano, which were the cities with the lowest service coverages, the position under the diagonal signifies an under-development of the offer, which should be higher in order to offer an adequate coverage.

4.4.2. Service offer versus service coverage, average over all Night-life areas and departure times

In Figure 51 and Figure 52, we can see the relation between the public transport offers and the service coverage percentages relatively for time thresholds of 30 minutes and 45 minutes. Both the results considering weighted and not weighted means are shown. We expect points to lay on a diagonal, where the ones falling beneath are sign of a sprawled city and/or a layout of the lines not coherent with our NLAs, while the ones falling above are signs of an urban compactness and/or effective lines design. The dimension of the points represents the number of inhabitants in the study area. In this case, the dimension of the points is the same in the two graphs but their disposition can change within the two different time thresholds.

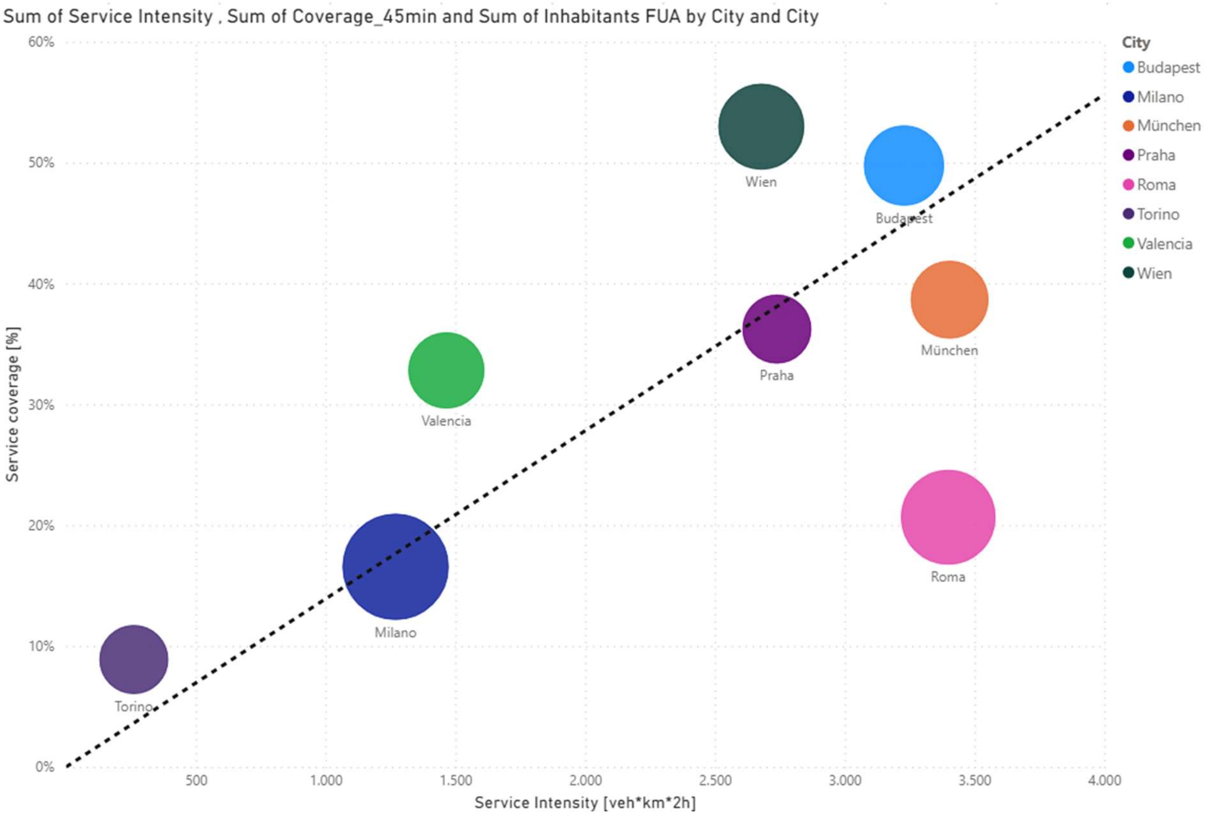


a)

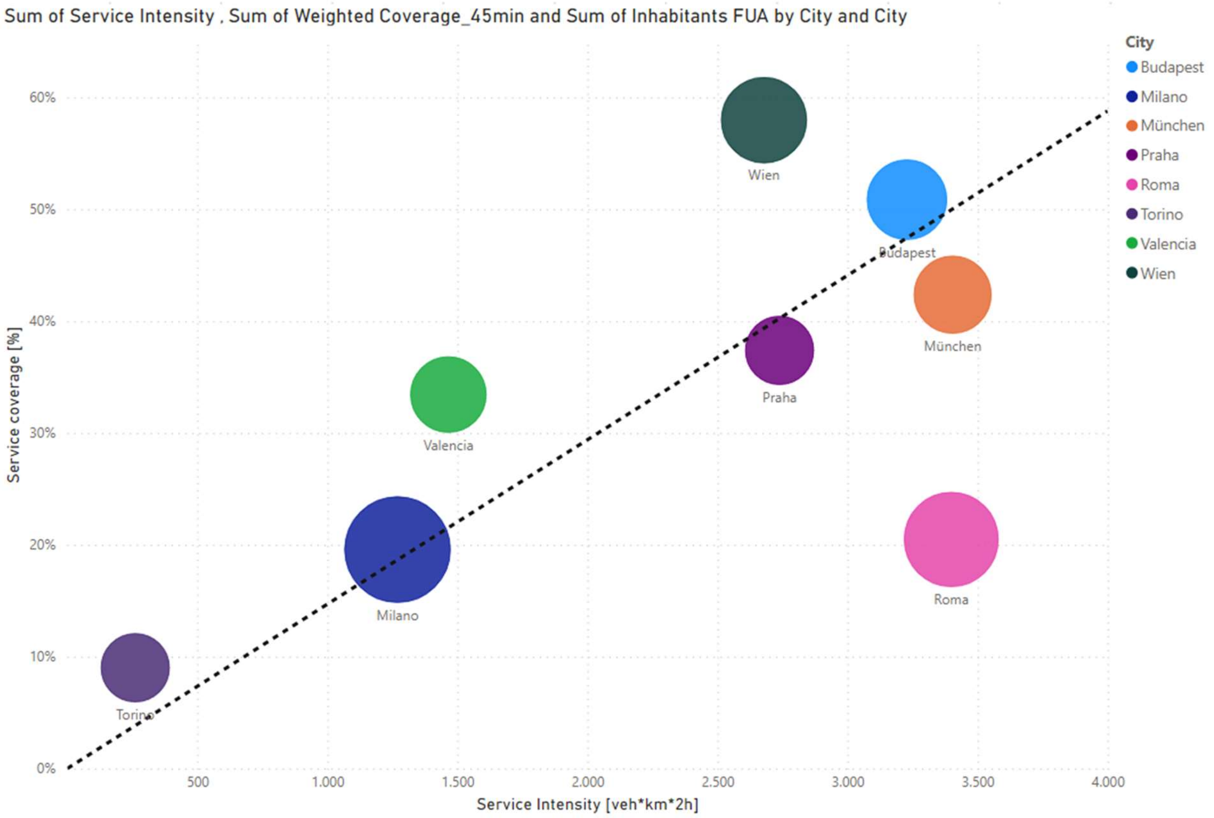


b)

Figure 51- Correlation between service Intensity and service coverage for the 30 minutes time threshold. a) service coverage computed as the mean of all the coverage percentages of each NLA at each departure time considered without weight; b) service coverage is weighted over the number of amenities inside each cluster as explained in 4.3.



a)



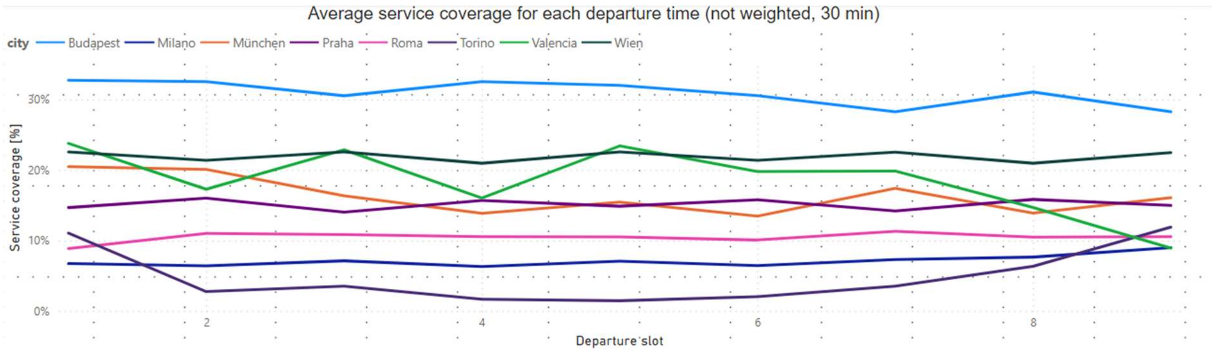
b)

Figure 52- Correlation between service intensity and service coverage for the 45 minutes time threshold. a) service coverage computed as the mean of all the coverage percentages of each NLA at each departure time considered without weight; b) service coverage is weighted over the number of amenities inside each cluster as explained in 4.3

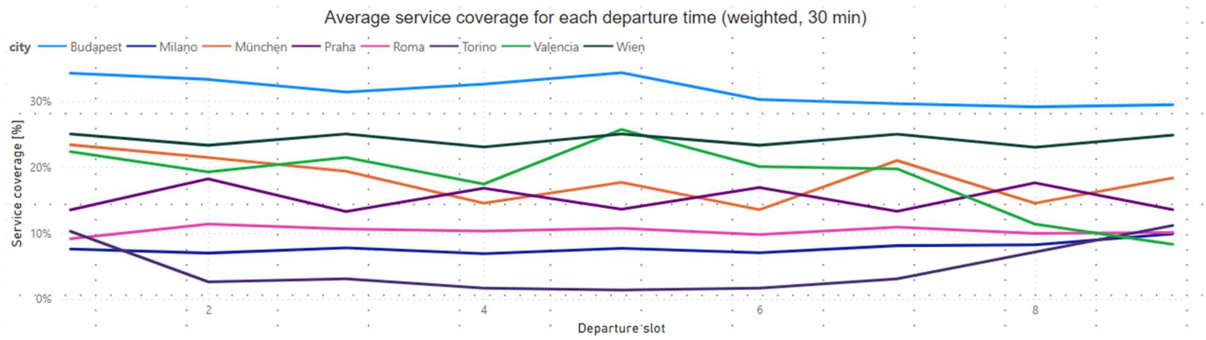
We can see how the cities laying on the diagonal are Praha, Milano and partially Torino; these cities have, always considering the sample, a service coverage that is on average in line with the quantity of service offered. Valencia, Wien and Budapest positions over the diagonal shows how the offer of the two networks is lower than expected considering the good efficiency percentage reached, indicating a convenient disposal of the network and/or a compact population distribution: specifically the compact population distribution is coherent with Valencia and Wien cases but not with Budapest, which appear as a sprawled city and whose density of inhabitants per km^2 is the lowest of all the sample. So the position over the diagonal in this graph can only be due to an optimal network disposition. The positions of the points related to Budapest, Valencia and Wien are also the only ones to change from one time threshold to the other. Namely, Valencia's values of service coverage become lower than the München and Praha ones, while Wien and Budapest switch position with each other. This can be due for Valencia to the fact that the night service finishes to run earlier and so the last portion of the two hours study time remains uncovered: this brings, more on the long time (45 minutes) than on the short times (30 minutes), to a penalization in terms of service coverage: this can be seen later in Figure 65 where all the partial service coverage percentages for each departure time are shown. Another reason can be that Valencia's network doesn't penetrate deep in the suburbs, and so most peripheral areas are not served, and this has a stronger impact when we consider the 45 minutes time threshold. The switch between Wien and Budapest is easily explainable saying that Wien urban centre is smaller and more populated than that of Budapest and so more compact, so a longer travel time allows a sharper increase in the number of reached inhabitants compared to Budapest. Roma and München are below the diagonal, with a similar offer measure but a great distance between the two values of service coverage. The point representing Roma is generally farther from the diagonal than the one representing München, which in the 45 minutes time threshold tend to get closer to the diagonal. Roma, even being the second most dense urban centre, has an evident sprawled urban pattern, which makes it difficult to reach a large number of inhabitants and so a high value of service coverage. The same can apply, downscaled, for München, which is anyway closer to the diagonal than Roma. This also due to the number of inhabitants: at equal offer, the city with less inhabitants is naturally better served, and this can be seen in Figure 45, where the offer taken for each 1000 inhabitants is shown.

4.4.3. Average service coverage, average over all Night-life areas

We can finally see graphs as in Figure 53 and Figure 54 where a line reporting the coverage values for each of the nine departure times is shown for each city and for each time threshold. Both the not weighted and weighted (on NLAs basis) results are shown. Since we used different time spans for different cities, we considered here the departure times as time slots, so that we could have a comparison: that means the departure slot number one will be at 01.00 for Budapest, as 03.00 for Milano, as 02.00 for Torino etc.

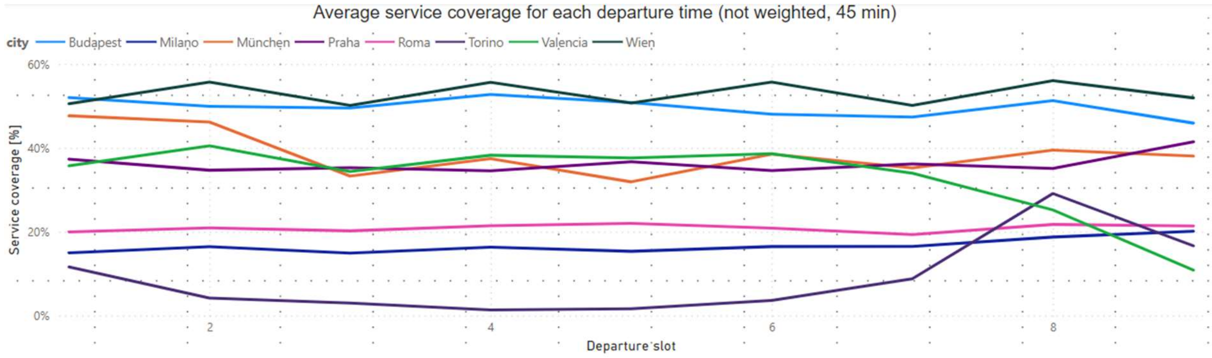


a)

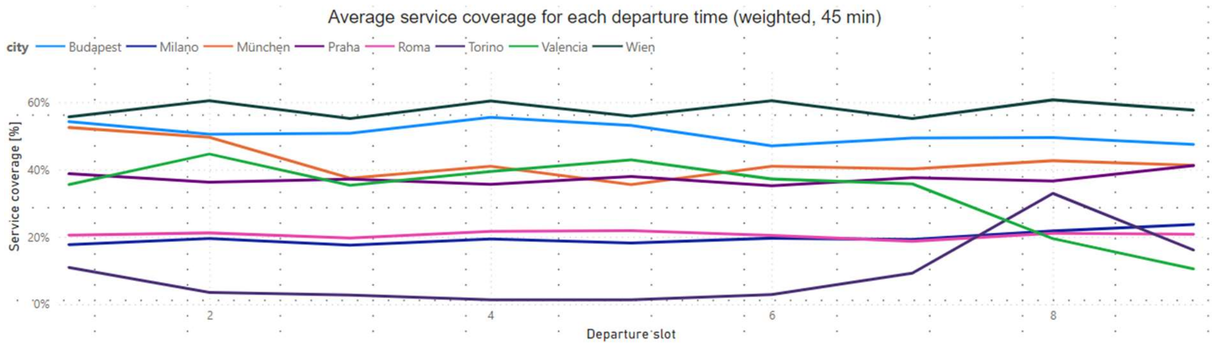


b)

Figure 53 - Average service coverage for each city for each departure slot for a time threshold of 30minutes; a) with not weighted values, with NLA weighted values.



a)



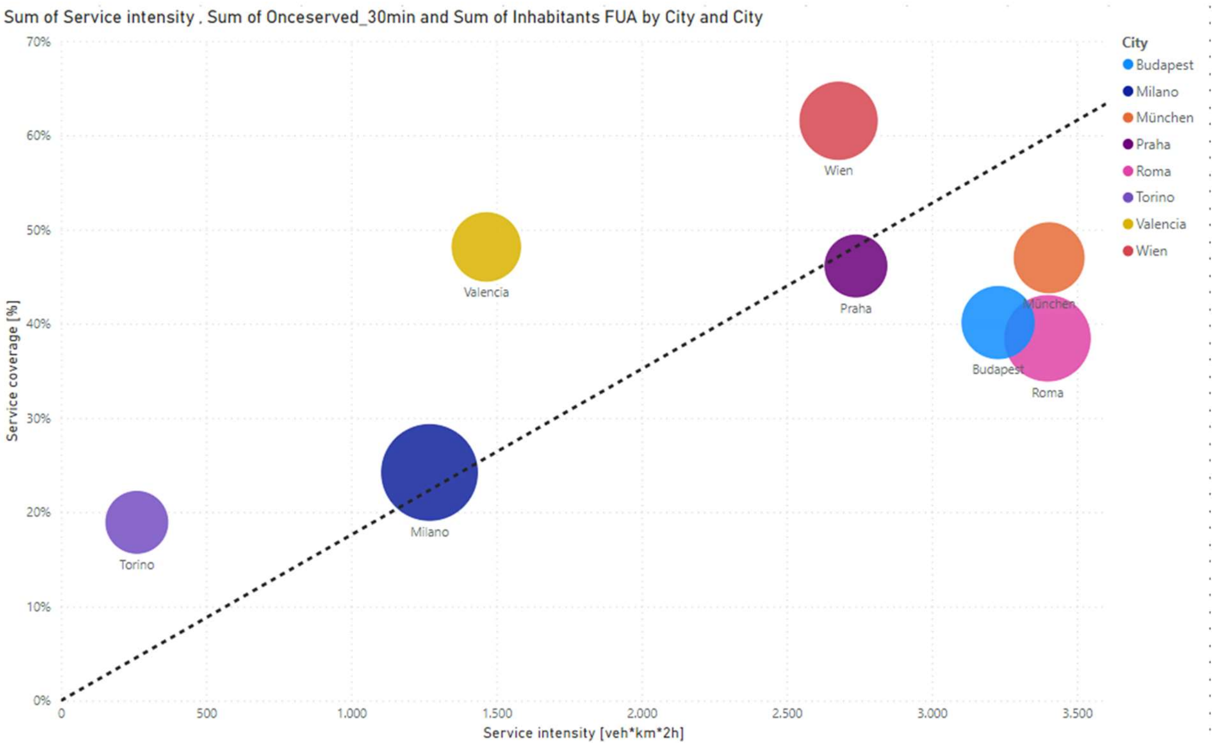
b)

Figure 54 - Average service coverage for each city for each departure slot for a time threshold of 30minutes; a) with not weighted values, with NLA weighted values.

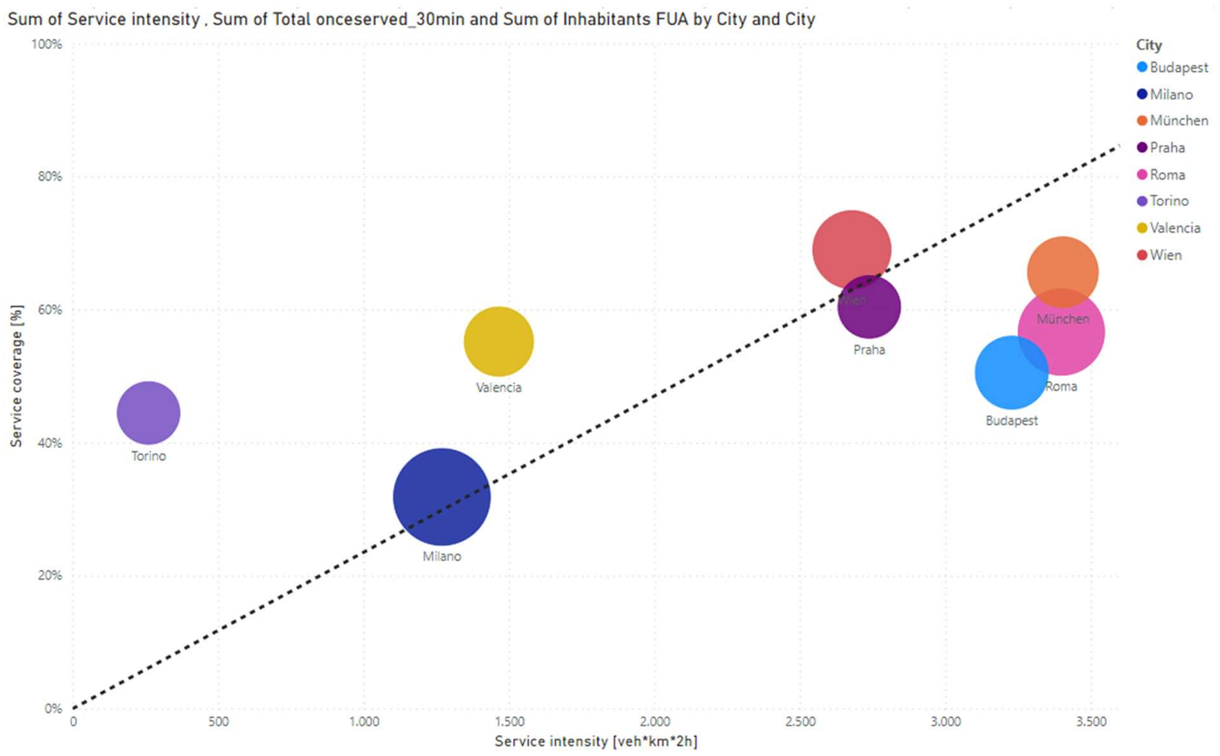
We can see how the service coverages don't change much when considering the weight of the NLAs. As we have seen in 4.4.2 the city with the highest service coverage ratio is Budapest when considering a 30 minute time threshold and Wien when considering a 45 minutes one. We can also see how some lines have a more constant period than others, in reflection to the different kind of time schedules and relative frequency (explored in paragraph 4.4.6): for example Wien presents a constant pattern with constant period due to a regular time schedule, while Valencia shows a dramatic fall during the last departure slots because of the end of the service. Torino's peak at departure slot number 8 shows how the service coverage for that specific time slot is actually higher than the ones of other 3 cities (namely Milano, Roma and Valencia), showing how the frequency influence the time-aggregated service coverage values: indeed, in the time aggregated results Torino has the lowest value by far.

4.4.4. Service offer versus service coverage, all Night-life areas jointly considered

We can further examine both urban sprawl and the dispersion of night-life areas computing the relation between public transport offer in the study area (service intensity) and the number of inhabitants served at least once from any NLA (Onceserved). In this case we expect the cities falling beneath the interpolated diagonal to be more sprawled. In Figure 55 we can see the correlation between service intensity and the percentage of inhabitants served at least by one NLA for the 30 minutes time threshold: in figure a) we see the percentage of inhabitants served at least once computed as an average of the percentages at each departure time, in b) we see data about the totality of the two hours. The same but for the 45 minutes time threshold is shown in Figure 56 . In both Figure 55 and Figure 56 the size of the dot is proportional to the number of inhabitants in the Urban Centres of each city.

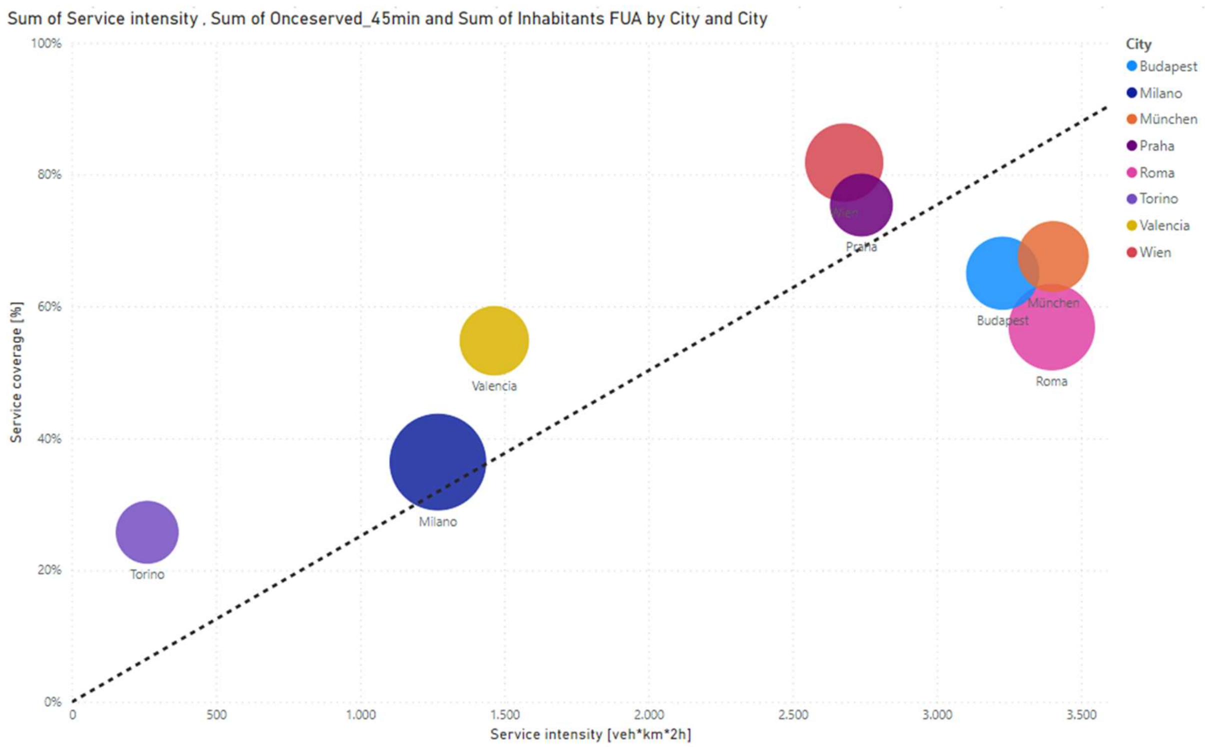


a)

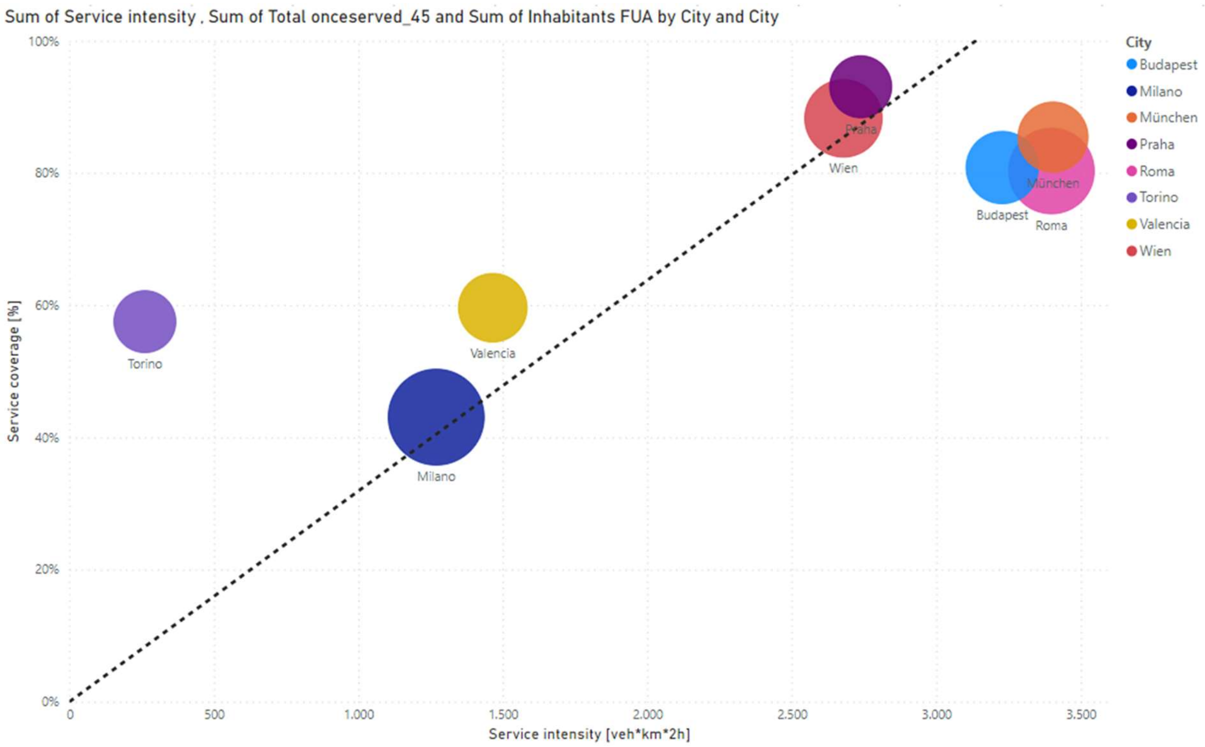


b)

Figure 55 - Correlation between offer and percentage of inhabitants served by at least one NLA for the 30 minutes time threshold. a) percentage of inhabitants served at least once computed as an average of the percentages at each departure time (Average Once Served); b) percentage of inhabitants served at least once in the two hours study time (Total Once Served).



a)



b)

Figure 56 - Correlation between offer and percentage of inhabitants served by at least one NLA for the 45 minutes time threshold. a) percentage of inhabitants served at least once computed as an average of the percentages at each departure time (Average Once Served); b) percentage of inhabitants served at least once in the two hours study time (Total Once Served).

We can see how the position of some points indicating cities change from one parameter definition to the other, for each of the two-time thresholds. The values, taken from Table 21,and their variation are showed in Table 22. We can see as for both the time thresholds, the highest variations between values answering to two different definitions in the same time threshold (columns of Table 22 named *Variation – 30* and *Variation – 45*) happen for Torino, Roma, München and Praha, whose values increase from the parameter considering the averages to the one considering the whole study time. Indeed, the first parameter considers an average of all the nine values obtained at the nine departure hours, meaning that a network with a constantly high offer, so with a frequency closer to 15 minutes, performed better than a network with a low frequency. The second parameter, instead, considered the inhabitants served at least once in the two hours composing the study time, so it doesn't consider the frequency of the lines but only their extension and design. For these reasons we can see as Torino, presenting indeed a low frequency network, shows the higher variation of value from one parameter to another at both the time thresholds. We can also highlight how the values change between the two-time thresholds following the same definition, as shown in Table 22 in the columns called *Variation – Av* and *Variation - Tot*. Praha and Budapest turned out to be the cities with the highest variation between the two-time thresholds. Praha and Budapest are indeed the cities with the two longest networks (Table 14), which help to understand why a longer travel time induces a larger variation. On the other side, Torino, Milano and Valencia are the cities with the shortest networks, which lead to a minor variation of the percentage of served inhabitants between the two-time thresholds.

Table 22 – Values of the parameters indicating the percentage of inhabitants served at least once for both the two given definitions and the two considered time thresholds. Followed by the variation between one definition and the other for both time thresholds (Variation - 30 and Variation – 45) and the variation between the two time thresholds following the same definition(Variation – Av and Variation –

Tot). The conditional formatting shows with the lightest green the highest variation, while the deepest red the lower ones.

City	Onceserved_30min	total_onceserved_30min	Onceserved_45min	total_onceserved_45min	Variation - 30	Variation - 45	Variation - Av	Variation - Tot
Budapest	40.13%	50.52%	65.03%	80.83%	10.39%	15.79%	24.90%	30.31%
Milano	24.22%	31.83%	36.39%	43.07%	7.60%	6.68%	12.17%	11.25%
München	47.01%	65.62%	67.55%	85.45%	18.62%	17.90%	20.54%	19.82%
Praha	46.14%	60.40%	75.37%	93.06%	14.27%	17.69%	29.23%	32.66%
Roma	38.46%	56.62%	56.85%	80.31%	18.15%	23.46%	18.39%	23.69%
Torino	18.93%	44.46%	25.74%	57.52%	25.53%	31.79%	6.81%	13.06%
Valencia	48.16%	55.19%	54.79%	59.62%	7.03%	4.83%	6.63%	4.43%
Wien	61.55%	69.00%	81.82%	88.27%	7.45%	6.45%	20.27%	19.27%

4.4.5. Average service coverage, all Night-life areas jointly considered

We can see in Figure 57 and Figure 58 how the resulting service coverage considering jointly all the NLAs change over time (Onceserved), using again the notion of departure slot, so that even having different time ranges for different cities we can make a comparison.

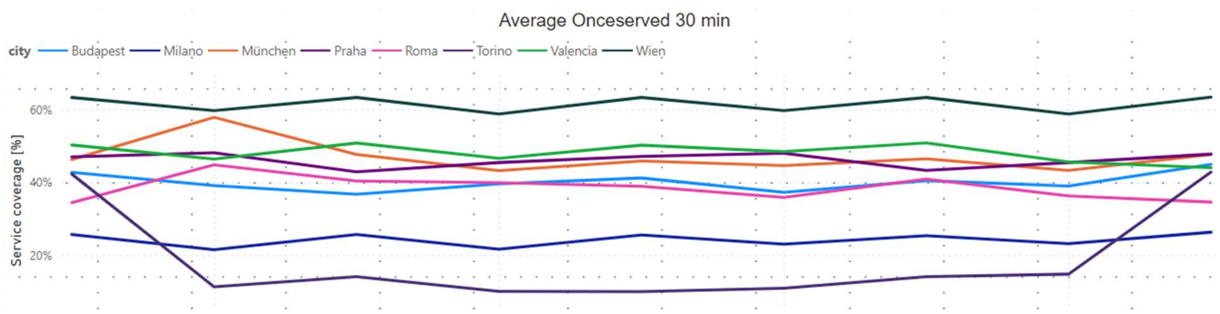


Figure 57 – Average service coverage for each city considering all the NLAs jointly. Time threshold of 30 minutes.

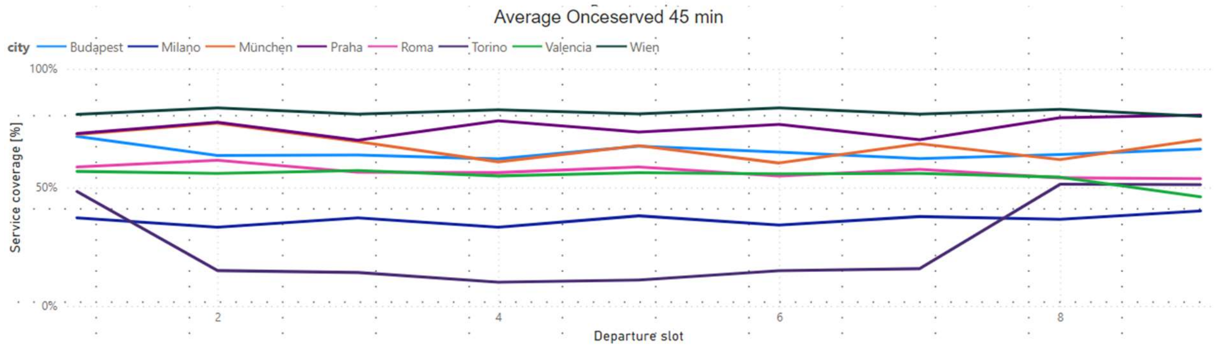


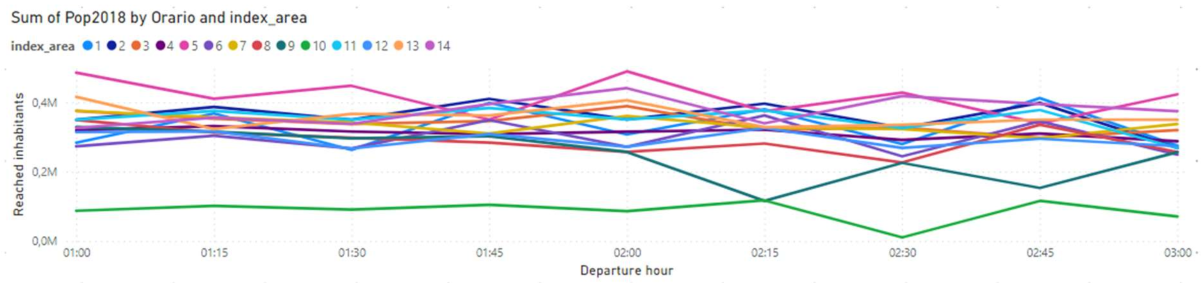
Figure 58 - Average service coverage for each city considering all the NLAs jointly. Time threshold of 45minutes.

As in paragraph 4.4.3, where similar graphs related to service coverage were shown, we can state how evident is the impact of different time schedules. Indeed, we can again see a constant pattern for the city of Wien and the presence of peaks for Torino, whose results in terms of service coverages are highly impacted by the scarce frequency of the service. Contrary to what observed in 4.4.3, where Budapest was leading in the 30 minutes threshold and Wien in the 45 minutes one, Wien is the city with the highest values of service coverage in both 30 and 45 minutes thresholds. This means that a larger portion of the city (and so, of its inhabitants) is reached by the service in Wien compared to Budapest.

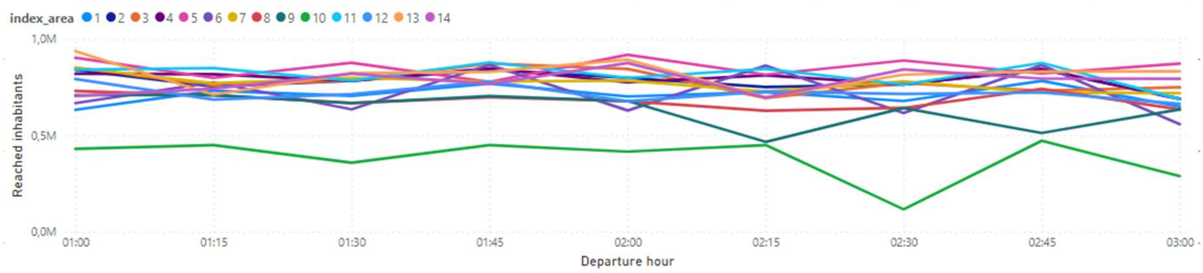
4.4.6. Time variations due to the impact of service frequencies

The computation in Traveltime plug-in was made considering nine different departure times with a 15 minutes frequency within the two night hours considered to be the night time service peaks. Considering different departure times in two hours allowed us to consider the influence of frequency over the service coverage and once served results. Showing in a graph to the amounts of inhabitants reached by each line at each departure time, we can visually state the presence of a pattern indicating the frequency values. Indeed, the presence of regular frequency lines causes a periodical succession of peaks (where the period is indeed the inverse of frequency).

Budapest: In Budapest's network lines have a non-regularly time-spaced schedule, which means that the frequency is not regular but changes from time to time for each of those lines. Nonetheless, lines with fixed frequency exist as well, and their presence is clear in the two graphs of Figure 59, where peaks present themselves with a 30 minute period, signifying that for most lines the frequency per hour is two. The green line on the bottom, is the one responding to NLA number 10, whose little amount of covered inhabitants is due to the de centralised position. The case of this NLA is also interesting because it shows an especially dramatic drop at departure hour 02:00: it is due to a missing passage of a certain bus for a longer time than during the other hours, which is related to the previous consideration about the irregular schedules of the network.



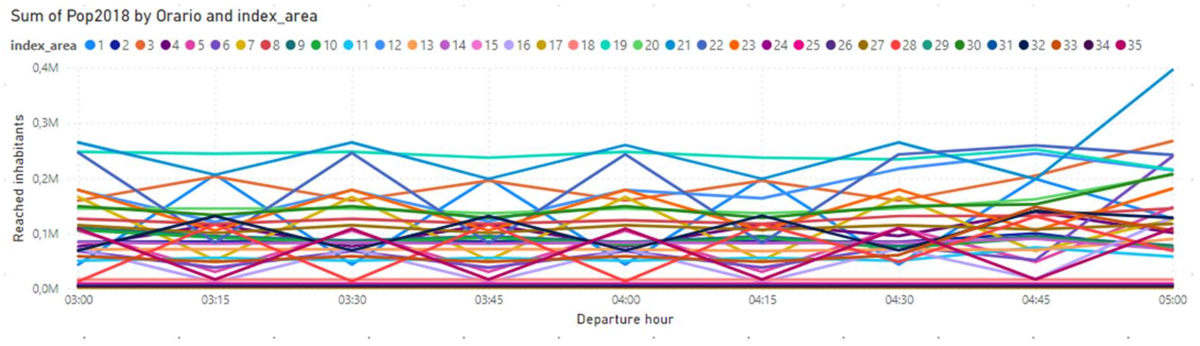
a)



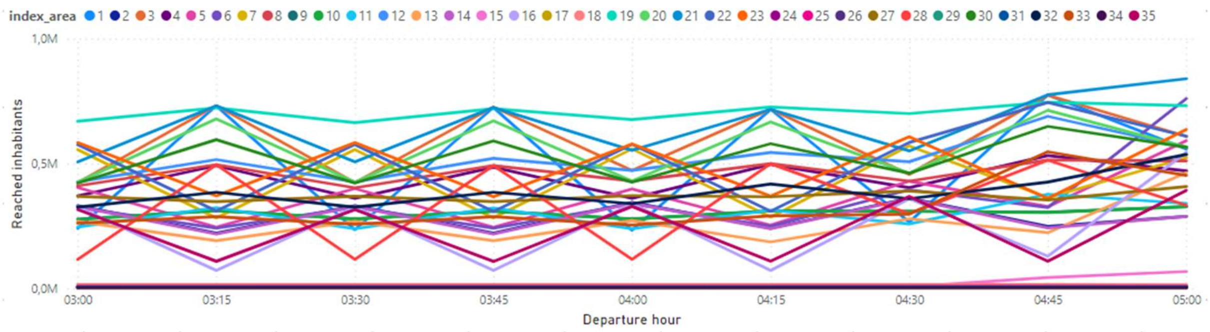
b)

Figure 59 - Evolution of covered inhabitants for each NLA in Budapest during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Milano: In Milano's network, the frequency is fixed at 2 vehicles per hour for each line: indeed, we can see in Figure 60 peaks with 30 minute period. The graphs are disturbed at the end of the study time by the entering in service of early day – time lines, which caused an increase of the coverage. Some NLA in Milano are not connected to the night time network, so from them only the inhabitants reachable walking in ten minutes are covered: the lines indicating these NLAs are the ones laying on the bottom.



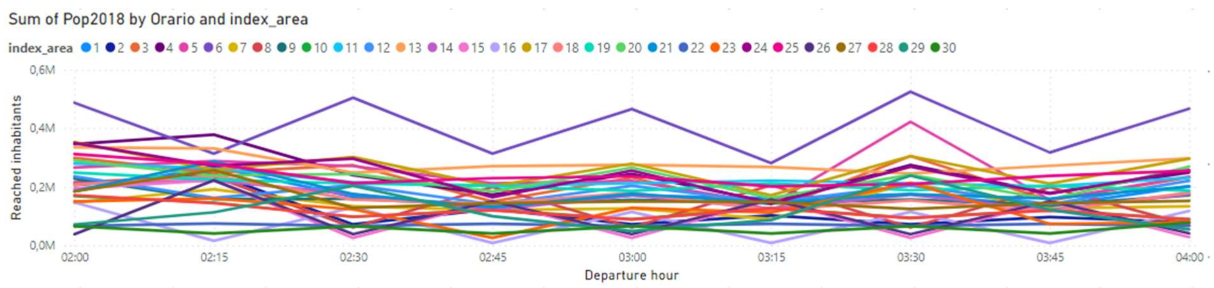
a)



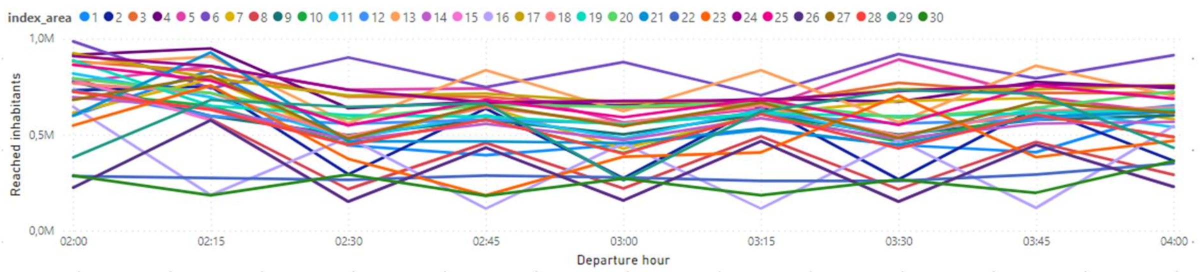
b)

Figure 60 - Evolution of covered inhabitants for each NLA in Milano during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

München: In München network different lines have different schedules, which most of the time remain the same for the whole night time service, but in some times change from hour to hour. These frequencies vary from 2 vehicles per hour, to 4. In Figure 61 is possible to see the influence of the lines having a passage every 30 minutes (which are the majority) and how there is a general inflection in terms of coverage which converge at departure time 03:00.



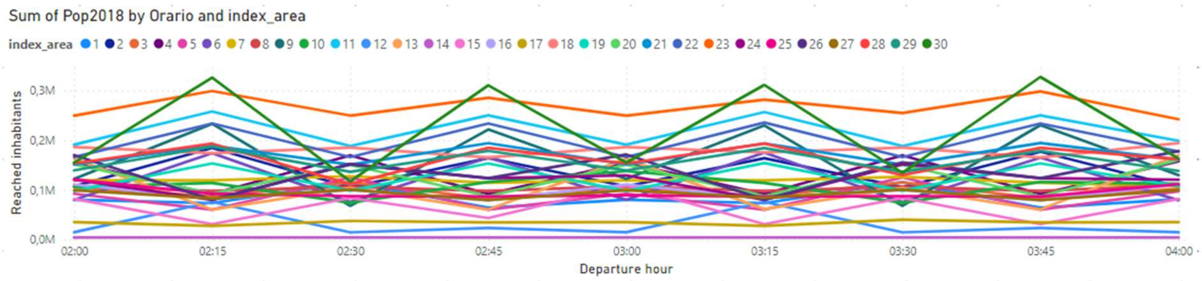
a)



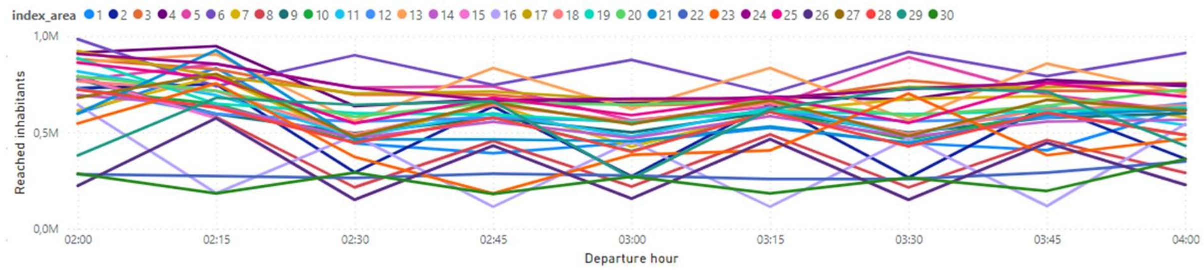
b)

Figure 61 - Evolution of covered inhabitants for each NLA in München during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Praha: Praha's lines have different frequencies: for half of the lines runs one vehicle per hour, for the other half two vehicles and for a single line, only one vehicle runs in the two study hours. Since the NLA are mostly situated close to the lines with the highest frequency (2 vehicles per hours), the graphs in Figure 62 show peaks with a 30 minutes periods. An example of an NLA served by a low frequency line can be seen following the blue line referred to NLA number 1 for the 30 minutes time threshold graph: the peaks have an hourly period and not a 30 minutes one: this evidence can't be spotted in the graph related to the 45 minutes time threshold, probably because with a longer travel time it is possible to catch lines with higher frequencies, which are dominant in the visual representation.



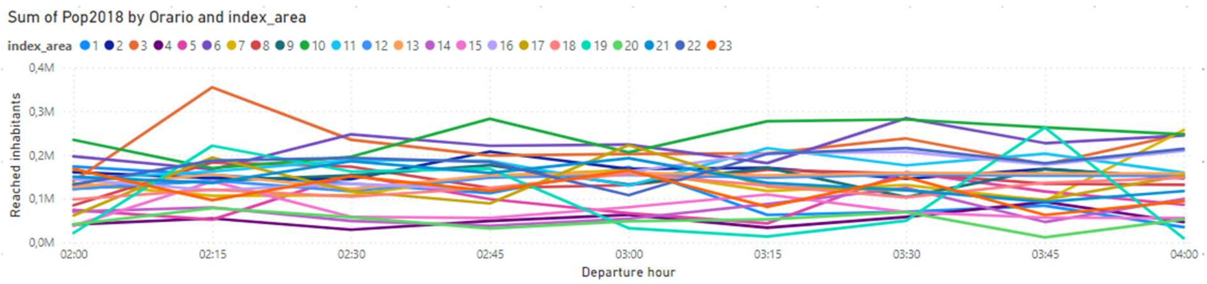
a)



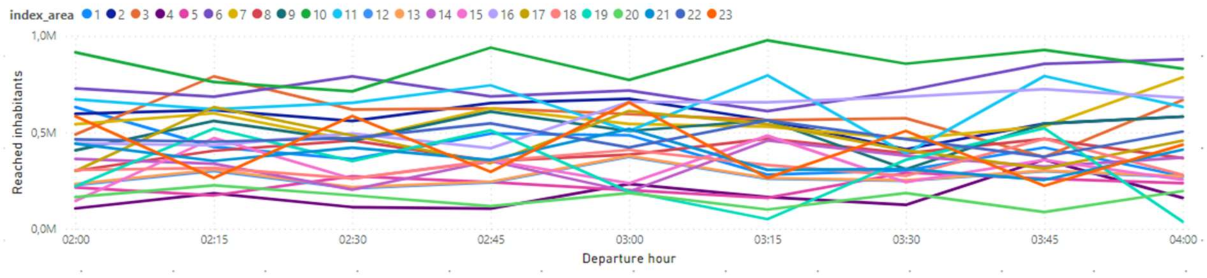
b)

Figure 62 - Evolution of covered inhabitants for each NLA in Praha during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Roma: In Roma's network each line has different frequency for different hours: generally two vehicles passes every hour for each line, but the amount of time between one and the other always change. Indeed, is not easy for Roma's case to identify a pattern of peaks with a precise period, as visible in Figure 63 .



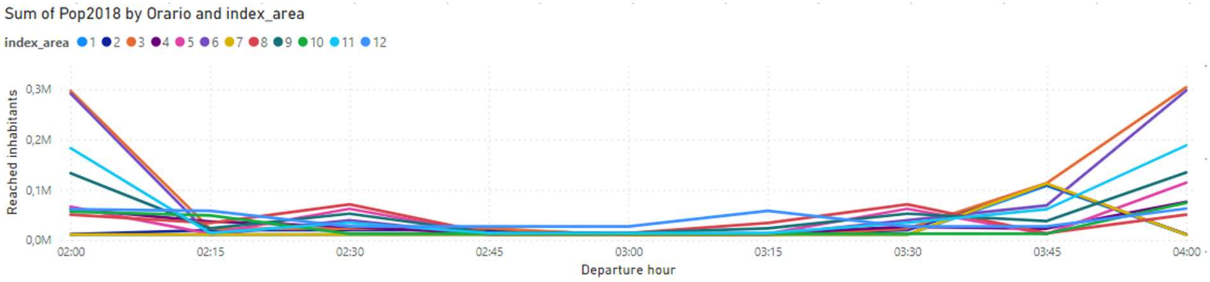
a)



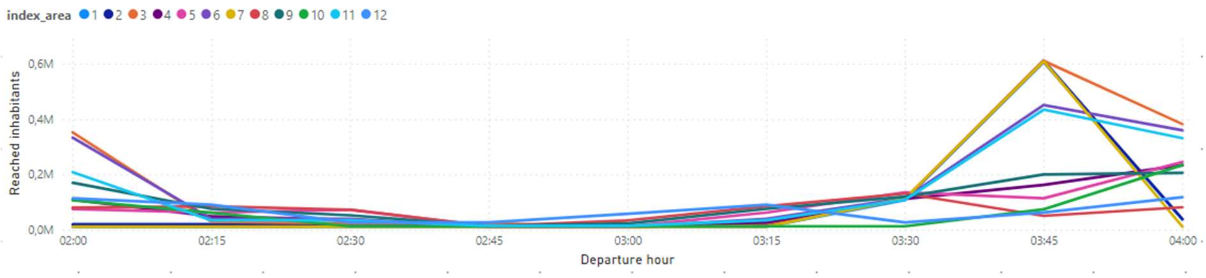
b)

Figure 63 - Evolution of covered inhabitants for each NLA in Roma during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Torino: In Torino's network all the lines have one vehicle departing at 02:00 and one at 04:00 and all the lines depart from the same place at the same time. That explains the two peaks and the plateau in the middle in Figure 64. In the two time thresholds the peak is reached at a different hour. That can be related to the interchange of vehicles made possible by an incremented travel time and the particularly centralized conformation of the network.



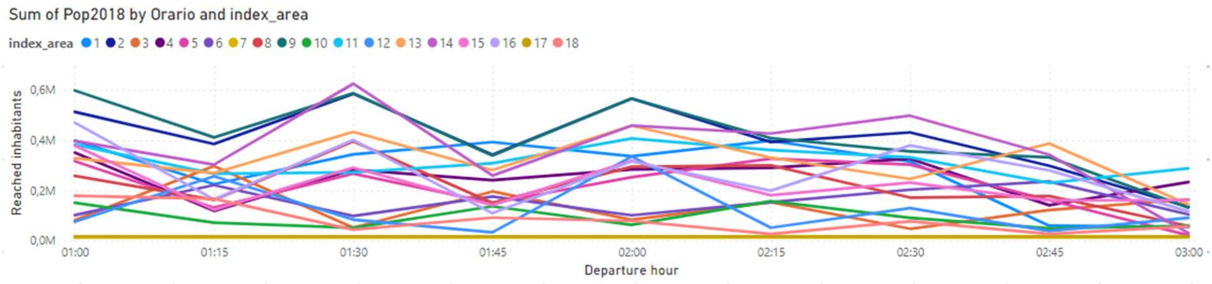
a)



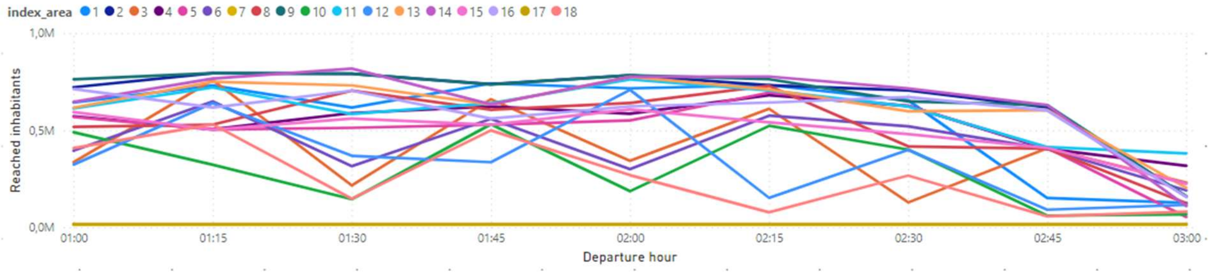
b)

Figure 64 - Evolution of covered inhabitants for each NLA in Torino during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Valencia: In Valencia's network every line has its own schedule with variable frequencies. Nonetheless there are usually two vehicles for hour, which allow to individuate a pattern of peaks with 30 minutes period, even if not extremely clear (Figure 65). The night time service stop at 03:00 and it is clear from the two graphs how the coverage decrease abruptly when that hour is reached.



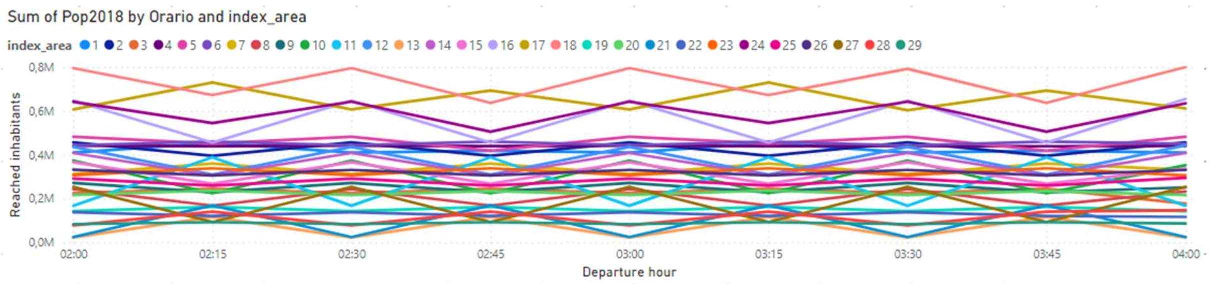
a)



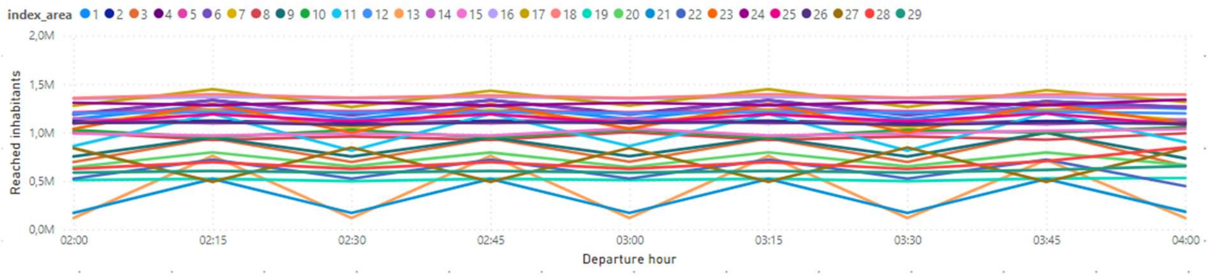
b)

Figure 65 - Evolution of covered inhabitants for each NLA in Valencia during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

Wien: In Wien's network every bus line presents a vehicle every 30 minutes, while the underground lines run every 15 minutes. This lead, in Figure 66 to a clear pattern of peaks with 30 minutes periods and a series of horizontal lines, indicating the NLAs served by the underground only. In the graph referred to the 45 minutes threshold, most lines are horizontal: the highest travel time allows more service users to change mean and catch an underground, whose effect in term of frequency is dominant on the graph.



a)



b)

Figure 66 - Evolution of covered inhabitants for each NLA in Wien during the two hours study time a) 30 minutes time thresholds b) 45 minutes time threshold.

5. Conclusions

The goal of the thesis was to tackle the matter of public transport night-time service, a matter so crucial for social inclusion and road safety, especially for the younger generations. To do so we developed a benchmarking exercise over a sample of eight European cities. For each city of the sample we obtained, through open-source data, information regarding the public transport night-time offer (public transport agencies), the population distribution (Urban Atlas), and the night-time related amenities distribution (Overpassturbo). With public transport offer data we computed an offer parameter based on the length and the frequency of the service, with the data about night-time related amenities we created a series of clusters indicating the so called Night Life Areas (NLA), which are to be intended as areas with an high density of night time related amenities (such as bar, pub, etc.) and were considered as the departure points of the trips, while the population distributions, disaggregated at blocks level, worked as the set of possible destination points, With this information we could compute, through the QGIS plug in Traveltime, the service coverage of each network, intended as the percentage of inhabitants served in a specific time threshold by each NLA, considering 9 different departure time with a 1 minutes frequency in a two hour time range.. We also computed for each city and each time threshold the numbers of inhabitants served at least once during the study time. At the end of the computations, we shown a series of correlation between the considered eight cities considering different parameters, such as the total population, the service coverage results, the inhabitants served once time and the public transport offer. We have seen how the resulting parameters concerning the service coverage change depending not only the length of the network but also on the frequency of the same, especially when we aggregate over the two studied hours. We have also highlighted the influence of different shapes and city sprawl patterns. A synthetic resume of each city's performance is given:

Budapest: the urban centre is the third largest one in the sample, which combined to a not huge number of inhabitants, leads to the lowest population density value in the sample. On the other side, Budapest has the longest public transport network of the sample (both considering the overall and the portion inside the urban centre only), which leads to one of the highest values of service intensity. In terms of service coverage, Budapest has the highest result by far for the 30 minutes time threshold and the second one for the 45 minutes one. The values related to inhabitants served at least once (both the one considering an average from all the departure times, both the one considering the overall) are not the best ones, which is justified by the dimension of the urban centre.

Milano: the urban centre is by far the biggest of the sample, such as the population: anyway, the pattern of it is not uniform but the result of urban expansions of Milano (the main centre) and multiple smaller centres, mostly ruled by different municipalities and served by different public transport agencies, which don't offer a night time service. That led to an oversized urban centre, where important portions are not served at all. Milano presents the second shortest network and a following second worst value of service intensity, The results is that Milano had the second worst service coverage in both time thresholds, resulting to be the city where the smallest percentage of inhabitants are served at least once during the whole study time for the 45 minutes time threshold.

München: both the urban centre dimension, the number of inhabitants and the density are slightly under the average of the sample. The network length is also around the average, but the high frequency of the lines allows München to have the highest value of service intensity, and the second one of service intensity per 1000 inhabitants. The values of service coverage are above the average in both time threshold, such as the ones related to the inhabitants served at least once, which are in the top three values in both time thresholds.

Praha: the urban centre is the second smallest of the sample, and the population the smallest one, while the density is the second lowest one; on the other side the network is the second longest, which bring to the highest value of service intensity per 1000 inhabitants. This brings to high values of inhabitants served at least once, reaching the highest value for the parameter regarding the overall and the second one for the one regarding the average (both for the 45 minutes time threshold).

Roma: the urban centre is the second biggest one, as the population is the second largest. The density is the second highest of the sample. The network is the third longest and the service intensity value is the second highest: this changes when considering the service intensity per 1000 inhabitants, which value is only the fifth highest. Even with a long network, the high number of inhabitants and their dispersion, allows Roma to have the third lowest value of service coverage (for both time thresholds) and general under average values for the parameters related to inhabitants served at least once.

Torino: the urban centre is the smallest and the number of inhabitants slightly the second lowest: the density is by far the highest of the sample. The network is by far the shortest, and accompanied by the lowest frequency of the lines, leads to the worst values of service intensity. The results in terms of service coverage and inhabitants served at least once are always the worst, except for the inhabitants served at least once considering the overall of the departure time, where the values are the second worst (because of Milano issues on this parameter).

Valencia: is the third smallest city of the sample in term of urban centre surface and inhabitant's amount: the density is slightly under average. The network length, such as the two service intensity values are the third lowest ones. In terms of service coverage, is the city changing more position from one time threshold to the other, being the third best in 30 minutes threshold and the fifth one in the 45 minutes one. The other parameters are on the average, with just the second highest value of inhabitants served at least once considering an average of all departure times for the 30 minutes time threshold.

Wien: the urban centre dimension is under samples' average, while the population is over the average: the density is the third highest one. The network length, such as the values related to service intensity are under the average. Nevertheless, the value of service coverage is the send highest for the 30 minutes time thresholds, and the best one in the 45 minutes one. The values related to inhabitants served at least once are the highest except for the one considering the overall of the departure time, which is slightly lower than the one referred to Praha.

Criticalities of this job mostly lied on the clustering process which led to the creation of the Night Life Areas: both concerning the input data and the parameters of the process itself. Indeed, we created clusters considering only the density of points representing amenities, without considering their relative importance. It would rather have been interesting to weight these points if we had quantitative information such as the number of served customers, the size of the premises, the working personnel or the revenues of each amenity. Regarding the parameters used for the clustering process, it would have been interesting to define an objective set of parameters to study such phenomenon (as existing for the criminalities phenomena), but there is no literature about it, and since it was not the purpose of this work, subjective parameters were used. A possibility studied at the beginning of the project and then abandoned as excessively computing expensive, was to consider the factors amenities and population distribution invertedly, defining for each cadastral block the number of reachable amenities Another criticality lied in the difficulty to find reliable, uniform and easily implementable information about the public transport offer. Indeed, the available files containing this information, in GTFS format, were most of the time not updated and incomplete. That signified a difficulty to gain information on service frequencies, and a discrepancy with the results obtained with the plugin Traveltime, which, on the other hand, uses updated and complete files. Nonetheless, since we relied on Traveltime and their data for the computations, the criticalities referred to public transport offer were only spotted in the Correlation chapter, where offer related parameters based on these data were shown. Another criticality was the choice of a standard procedure to outline the neighbour area of each NLA in order to compute the service coverage: indeed, we choose to consider a measure based on vehicles average linear speed giving two different time thresholds, but neither the speed, nor the assumption of an isotropic linear viability (and less of all the completely subjective choice of 30 and 45 minutes as time thresholds) are reflecting real scenarios. It would be interesting to investigate the relationship between each NLA and the whole Urban Centre and not only specific neighbour areas, to make a comparison in term of coverage between different modal choices or their combination. Another possibility would be to invert the factors of the problem, so computing a needed travel time in order to assess a defined service coverage.

Conducing a benchmark exercise, we were only able to gain an overview of the matter and a general prospective of how these eight different cities have different services, that related to different conformations and dimension offer a different grade of coverage. It would be interesting, as a future development, to study a framework able to study hypothetical scenarios of the public transport networks, such us the introduction of new lines (or the transposition in night time of already existing day time lines) and/or an increase of their frequencies: a platform where this proposal could be developed is the scientific study proposed by Indaco Biazzo called City Isochrone[8] .

The methodology employed here is not exclusively applicable to the night time service and to the night life areas, but it could also be translated to other service times, such as the daytime service or the festivity one, or related to the location of some important events; on the other hand it could be translated for other type of areas then the ones referred to night life, such as shopping, sport, medical facilities.

References

A.Tennøy, Knapskog, M., & Wolday, F. (2022). Walking distances to public transport in smaller and larger Norwegian cities. *Transportation Research Part D: Transport and Environment*. doi:https://doi.org/10.1016/j.trd.2022.103169

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 359-364. doi:https://doi.org/10.1016/j.aap.2008.12.014.

Aud Tennøy, M. K. (2022). Walking distances to public transport in smaller and larger Norwegian cities. *Transportation Research Part D: Transport and Environment, 103*. doi:https://doi.org/10.1016/j.trd.2022.103169.

Bäcklund, H., Hedblom, A., & Neijman, N. (2011). *DBSCAN A Density-Based Spatial Clustering of Application with Noise*. Linköping: Linköpings Universitet.

Balas, E. (1989). The prize collecting traveling salesman problem. *Networks*, 621-636.

Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., & West, a. G. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, 104*. doi:doi: 10.1073/pnas.0610172104

Carrasco, M. &.-L. (2016). The impact of neighborhood amenities on residential property values: A hedonic pricing approach. *Journal of Property Research*, 33(2), 121-137. doi:10.1080/09599916.2016.1165169

Catalogue, The Data Visualization. (2023). *Heat Maps: Visualizing Data Density*. Retrieved from https://datavizcatalogue.com/methods/heat_map.html

Chen, C., Zhang, D., Zhou, Z.-H., Li, N., Atmaca, T., & Li, S. (2013). B-Planner: Night bus route planning using large-scale taxi GPS traces. *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 225-233. doi:10.1109/PerCom.2013.6526736

Chia, J., Lee, J., & Kamruzzaman, M. (2016). Walking to public transit: exploring variations by. *International Journal of Sustainable Transportation, 10*(9), pp. 805-814. doi:10.1080/15568318.2016.1156792.

Costa, P., Neto, G. M., & Bertolde, A. (2017). Urban Mobility Indexes: A Brief Review of the Literature. *Transportation Research Procedia Vol.25*, 3645-3655. Retrieved from https://doi.org/10.1016/j.trpro.2017.05.330

Curtis, A., Droste, N., Coomber, K., Guadagno, B., Mayshak, R., Hyder, S., . . . Miller, P. (2019). Off the rails—Evaluating the nightlife impact of Melbourne, Australia's 24-h public transport trial. *International Journal of Drug Policy, Vol.63*, 39-46. Retrieved from https://doi.org/10.1016/j.drugpo.2018.10.006

Davis, M. (2011, 09 27). *Data model diagrams for GTFS*. Retrieved from Lin.ear th.inking: http://lin-ear-th-inking.blogspot.com/2011/09/data-model-diagrams-for-gtfs.html

[8] http://citychrone.org/world

Dijkstra, L., Poelman, H., & Veneri, P. (2019). *The EU-OECD definition of a functional urban area.* OECD Regional Development Working Papers. Paris: OECD Publishing. doi:https://doi.org/10.1787/d58cb34d-en

Domènech, A., & Gutiérrez, A. (2017). A GIS-Based Evaluation of the Effectiveness and Spatial Coverage of Public Transport Networks in Tourist Destinations. *ISPRS Int. J. Geo-Inf.*, 6(3(, 83. doi:https://doi.org/10.3390/ijgi6030083

Edrisi, A., Barzegari, V., & Nourinejad, M. (2021). Serial formation and parallel competition in public transportation. *Transportmetrica A: Transport Science, 17*, 1193-1216.

El-Geneidy, A., & Levinson, D. (2022). Making accessibility work in practice. *Transport Reviews*, 129-133. doi:10.1080/01441647.2021.1975954

El-Geneidy, A., & Levinson, D. (2022). Making accessibility work in practice. *Transport Reviews*, pp. 42:2, 129-133. doi://doi.org/10.1080/01441647.2021.1975954

esri. (n.d.). *How Geographically Weighted Regression (GWR) works*. Retrieved from pro.arcgis.com: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-geographicallyweightedregression-works.htm

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.* Munich: University of Munich.

European Environment Agency . (2017). *Copernicus Land Monitoring Service* . Copenaghen.

European Road Safety Observatory. (2021). *Mobility & Transport - Road Safety*. Retrieved from European Comission: https://road-safety.transport.ec.europa.eu/statistics-and-analysis/data-and-analysis/facts-and-figures_en

Florczyk, A., Melchiorri, M., Corban, C., Schiavina, M., Maffenini, L., Pesaresi, M., . . . Zanchetta, L. (2019). *Description of the GHS Urban Centre Database 2015.* Luxembourg: Publications Office of the European Union. doi:10.2760/037310

Frate, L. (2021). *Tutorial N.6. Calcolo dell'Home range (Kernel Density Estimation).* Retrieved from ECOVIEW: https://www.ecoview.it/2019/01/04/tutorial-n-6-calcolo-dellhome-range-kernel-density-estimation/

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS,. *Applied Geography vol. 63*, 408-417. Retrieved from https://doi.org/10.1016/j.apgeog.2015.08.002

GHSL - Global Human Settlement Layer. (2023). *GHSL - Global Human Settlement Layer*. Retrieved from ghsl.jrc.ec.europa: https://ghsl.jrc.ec.europa.eu/about.php

Google. (2022). *Static transit*. Retrieved from Google transit : https://developers.google.com/transit/gtfs/reference/#field_types

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. . *Environmental Planning B: Planning and Design*, 37(4), 682-703. doi:10.1068/b36080

Hidalgo, C. A., Castañer, E., & Sevtsuk, A. (2020). The amenity mix of urban neighborhoods. *Habitat International, 106*, 102205. doi:https://doi.org/10.1016/j.habitatint.2020.102205

IBM. (n.d.). *What is liner regression* . Retrieved from IBM.com: https://www.ibm.com/topics/linear-regression#:~:text=Resources-,What%20is%20linear%20regression%3F,is%20called%20the%20independent%20variable.

Jackson, C., & Owens, E. (2011). One for the road: Public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics 95*, 106–121. Retrieved from https://doi.org/10.1016/j.jpubeco.2010.09.010

Jasim, I. A., Al-Jaberi, A. A., Al-Mamoori, L. A.-M., Al-Ansari, N., & K., S. (2022). Do the population density and coverage rate of transit affect the public transport contribution? *Cogent Engineering*. doi:10.1080/23311916.2022.2143059

Joly, I. (2004, October 27-29). Travel Time Budget – Decomposition of the Worldwide Mean. *ISTAT. Italian National Statistical Institute. Time Use: What's New in Methodology and Application Fields?*, p. 23.

Joly, I. (2007). The role of travel time budgets – Representation of a demand derived from activity. *11th World Conference on Transport Research* -. Berkeley, USA: WCTRS, University of California.

JRC - ISPRA EC. (2019). *GHSL Data Packages Instructions for data access. V1.0.*

Le, J., & Ye, K. (2022). Measuring City-Level Transit Accessibility Based on the Weight of Residential Land Area: A Case of Nanning City, China. *Land*, 11(9), 1468. Retrieved from https://doi.org/10.3390/land11091468

Levine, N. (2007). *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations.* Houston, TX: Ned Levine & Associates.

Lichtman-Sadot, S. (2019). Can public transportation reduce accidents? Evidence from the introduction of late-night buses in Israeli cities,. *Regional Science and Urban Economics Vol. 74*, 99-117. Retrieved from https://doi.org/10.1016/j.regsciurbeco.2018.11.009

Liu, M., & Jiang, Y. (2021). Measuring accessibility of urban scales: A trip-based interaction potential model. *Advanced Engineering Informatics*, 48. Retrieved from https://doi.org/10.1016/j.aei.2021.101293

Ma, Li, & Han. (2016). A planning tool for maximising transit services. *Transportmetrica B: Transport Dynamics, 4*(1), 1-22. doi:http://dx.doi.org/10.1080/21680566.2015.1017023
Miranda, H. (2010). *sustainable urban mobility and the case of Curitiba.* São Carlos: School of Engineering of São Carlos, University of São Paulo, São Carlos.
Murray, A. T. (2001). Strategic analysis of public transport coverage. *Socio-Economic Planning Sciences*, 175-188. Retrieved from https://www.sciencedirect.com/science/article/pii/S0038012101000040
Nagarajan, V. (2017, 03 29). Lecture Notes: Prize-Collecting Steiner Tree (Primal-Dual).
Neis, P., & Zielstra, D. (2015). OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. *ISPRS International Journal of Geo-Information*, 4(1), 92-115. doi:10.3390/ijgi4010092
Palacios, S. M., & El-geneidy, A. (2022, February). Cumulative versus Gravity-Based Accessibility Measures: Which One to Use. *Findings*. doi:https://doi.org/10.32866/001c.32444.
Poelman, H., & Dijkstra, L. (2015). *Measuring access to public transport in European cities.* Directorate-General for Regional and Urban Policy. Retrieved from https://ec.europa.eu/regional_policy/sources/work/2015_01_publ_transp.pdf
Poelman, H., Dijkstra, L., & Ackermans, L. (2020). *How many people can you reach by public transport, bicycle or on foot in European cities?Measuring urban accessibility for low-carbon modes.* Directorate - General for Regional and Urban Policy.
Sabouri, S., Park, K., Smith, A., Tian, G., & Ewing, R. (2020). Exploring the influence of built environment on Uber demand. *Transportation Research Part D: Transport and Environment*.
Saif, M. A., Zefreh, M. M., & Torok, A. (2019). "Public Transport Accessibility: A Literature Review". *Periodica Polytechnica Transportation Engineering*, 36-43.
Sakshat virtual lab . (n.d.). Retrieved from https://www.civil.iitb.ac.in/~gpatil/utsp_vlab/tripgeneration.php#:~:text=In%20contemporary%20transportation%20planning%20language,Origin%20and%20Attraction%20or%20Destination.
Sala, L., Wright, S., Cottrill, C., & Flores-Sola, E. (2021). Generating demand responsive bus routes from social network data analysis,. *Transportation Research Part C: Emerging Technologies, vol.128*. Retrieved from https://doi.org/10.1016/j.trc.2021.103194
Salado García, M., Díaz Muñoz, M. A., Bosque Sendra, J., Carvalho Cantergiani, C., & Rojas Quezada, C. (2006). Movilidad sostenible y SIG. Propuesta. *XII Congreso Nacional de Tecnologías de la Información Geográfica* (pp. 1777-1794). Granada, Spain: Universidad de Granada.
Sarker, R., Mailer, M., & Sikder, S. (2020). Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany. *Smart and Sustainable Built Environment, 9*(1), pp. 38-53. doi:https://doi.org/10.1108/SASBE-07-2017-0031
Sarker, R., Mailer, M., & Sikder, S. (2020). Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany. *Smart and Sustainable Built Environment, 9*(1), pp. 38-53. doi:https://doi.org/10.1108/SASBE-07-2017-0031
Scagnolari, S., Walker, J., & Maggi, R. (2015). Young drivers' night-time mobility preferences and attitude toward alcohol consumption: A Hybrid Choice Model,. *Accident Analysis & Prevention, vol. 83*, 74-89. Retrieved from https://doi.org/10.1016/j.aap.2015.07.002
SEDAC. (n.d.). *Gridded Population of the World (GPW), v4*. Retrieved from sedac.ciesin.columbia.edu: https://sedac.ciesin.columbia.edu/data/collection/gpw-v4
si. (n.d.). si. *si*.
Silva, F. B., Poelman, H., & Dijkstra, L. (2021). The JRC-GEOSTAT 2018 a novel small-area population estimation for Europe. *European Forum for Geography and Statistics.* Germany, virtual.
Sun, Y., Fan, H., Li, M., & Zipf, A. (2016). Identifying the city center using human travel flows generated from location-based social networking data. *Environment and Planning B: Planning and Design*, 480-498.
Supernak, J. (1967). TRAVEL-TIME BUDGET: A CRITIQUE. *Transportation Science*, pp. 261 - 265.
Tomasiello, D. B. (2022). A time interval metric for cumulative opportunity accessibility. doi:https://doi.org/10.31235/osf.io/ux5ah
Tomasiello, D. B., Herszenhut, D., Oliveira, J. L., Braga, C. K., & Pereira., a. R. (2022). *A Time Interval Metric for Cumulative Opportunity Accessibility.* doi:10.31235/osf.io/ux5ah.
Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The Location of Emergency Service Facilities. *Operations Research 19 (6)*, 1363-1373. Retrieved from https://doi.org/10.1287/opre.19.6.1363
Transitfeeds. (n.d.). *Transitfeeds*. Retrieved from Transitfeeds: https://transitfeeds.com/
wiki.openstreetmap. (2022, 4 13). *About Open Street Map*. Retrieved from Open Street Map: https://wiki.openstreetmap.org/wiki/About_OpenStreetMap
wiki.openstreetmap. (2022, 10 15). *Map features* . Retrieved from Open Street Map: https://wiki.openstreetmap.org/wiki/Map_features#Amenity
Xiao, L., Fan, X., Mao, H., Xu, C., Lu, P., & Luo, S. (2016). When Taxi Meets Bus: Night Bus Stop Planning over Large-Scale Traffic Data. *7th International Conference on Cloud Computing and Big Data (CCBD)*, 19-24. doi:10.1109/CCBD.2016.015.

Xu, S. Z., Yang, L., Zhou, M., Zhang, L., & Qiu, B. (2015). Mapping the popularity of urban restaurants using social media data,. *Applied Geography vol. 63*, 113-120. Retrieved from https://doi.org/10.1016/j.apgeog.2015.06.006

Yeran, S. (2015). Identifying the city center using human travel flows generated from location-based social networking data. *Environment and Planning B: Planning and Design*. doi:10.1177/0265813515617642

Zhang, Y., Li, H., & Ren, G. (2022). Quantifying the social impacts of the London Night Tube with a double/debiased machine learning based difference-in-differences approach,. *Transportation Research Part A: Policy and Practice, vol. 163*, 288-303. Retrieved from https://doi.org/10.1016/j.tra.2022.07.015

APPENDIX A: Parameters’ set for NLAs clustering.

In order to cluster the leisure – related amenities into Night Life Areas (NLAs) the DBSCAN tool on QGIS has been used.

This algorithm requires two parameters as inputs, namely the maximum distance between clustered points and the minimum size of cluster size:

- Minimum cluster size is the minimum number of features to generate a cluster (minPts).
- Maximum distance between clustered points is the distance beyond which two features cannot belong to the same cluster (epsilon).

By first choosing an epsilon and a minPts we define a neighbourhood around each data point. For each data point, the number of other data points in its epsilon neighbourhood is counted. If this count is less than minPts, the point is marked as a "noise" point, otherwise it is marked as a "core" point. For each core point, a new cluster is created and added to the cluster set. Then, all other core points that are directly or indirectly reachable from the initial core point (i.e., within epsilon distance) are recursively added to the cluster. The algorithm continues until all data points have been assigned to a cluster or marked as noise. The resulting clusters are dense regions in the data space, while the noise points are isolated points or points on the fringes of clusters.

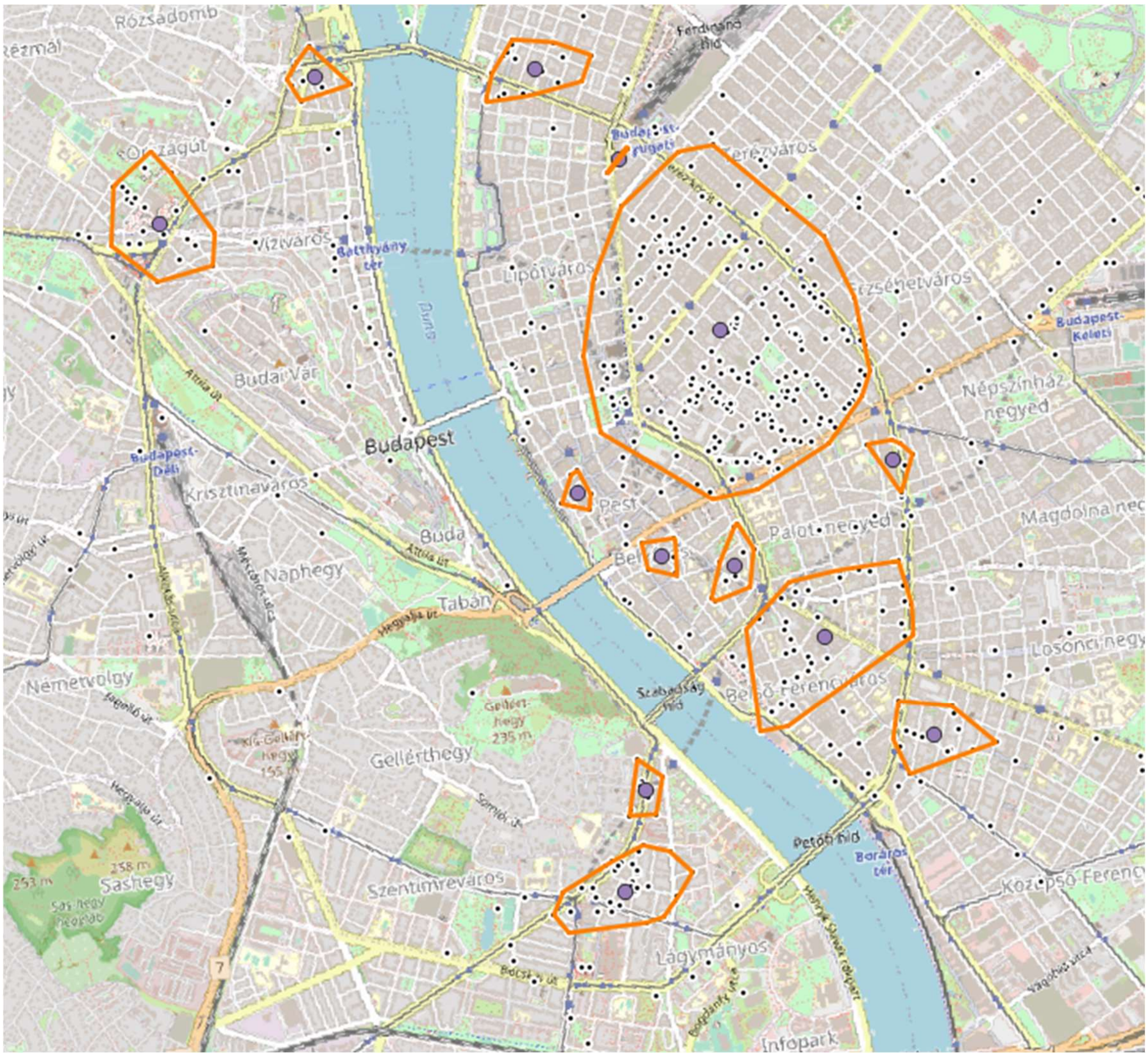
In (A. Figure 1) is a comparation between two cases in which we choose as maximum distance 125 m (in orange) and 150 m (in blue). The clusters created with a 150 m epsilon are bigger and sometimes involving multiple clusters created with 125 m epsilon.

A. Figure 1 - Explanation of how the results change changing the parameters settings. City of Wien.

In the following pages of the appendix, the results of DBSCAN with max_distance set at 150 m will be shown; for the case in which the result didn’t look satisfying, so when the cluster in the city centre looked to be to extended (Budapest, München, Praha, Wien) also the results with max_distance set at 125 m will be shown, so that a visual comparation was possible. The *minimum cluster size* parameter was set at 5 for all the cities.
The goal of the appendi is to show how DBSCAN works and why the used set has been chosen.

1. Budapest

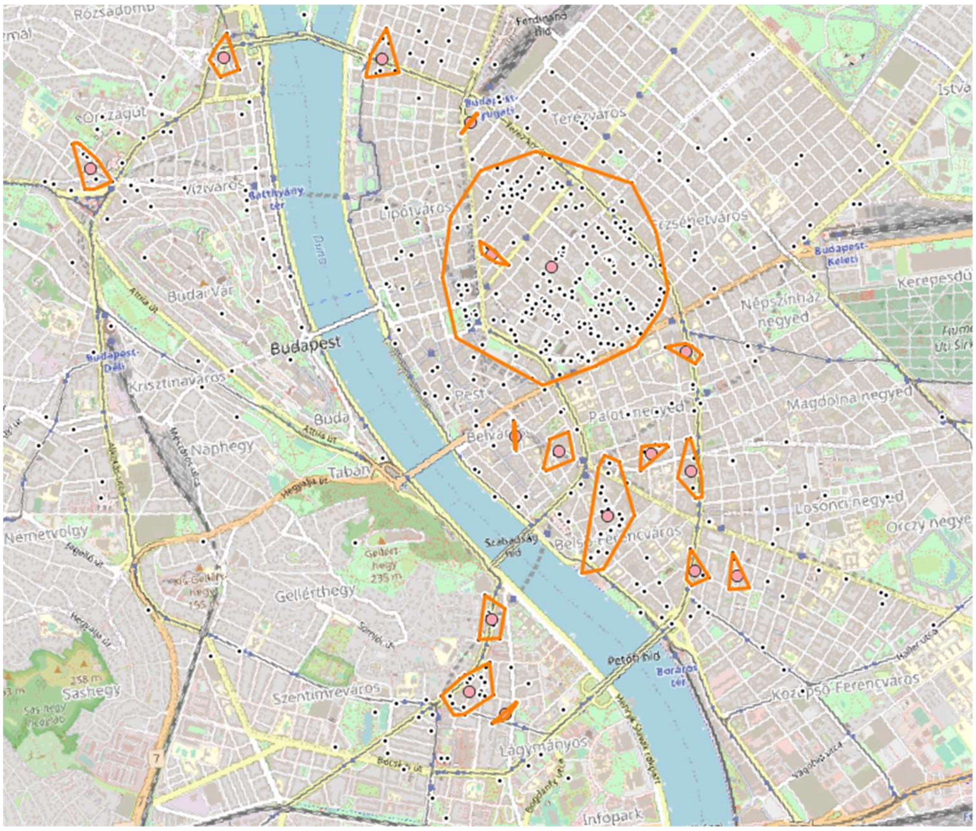
Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	14
Max number of points in a cluster	245



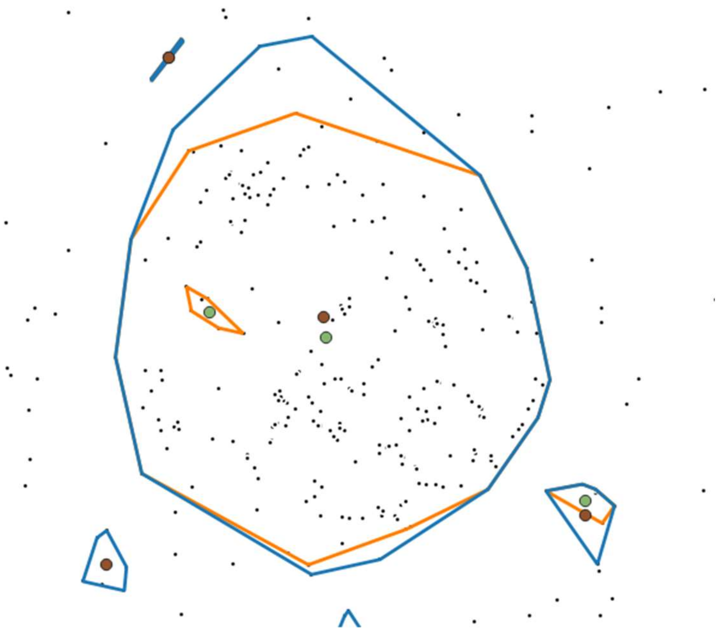
A. Figure 2 - Budapest, max_dist = 150 m

A wide cluster is generated in the city centre. To make it smaller we move the max_distance parameter from 150m to 125m. With the *measure* tool on QGIS (which allow to measure the distance between two points) has been measured that from the centroid to the perimeter points of the cluster, the length never goes over 1 km.

Minimum Cluster Size	5
Max distance between clustered points	125
Resulting Clusters	18
Max number of points in a cluster	228

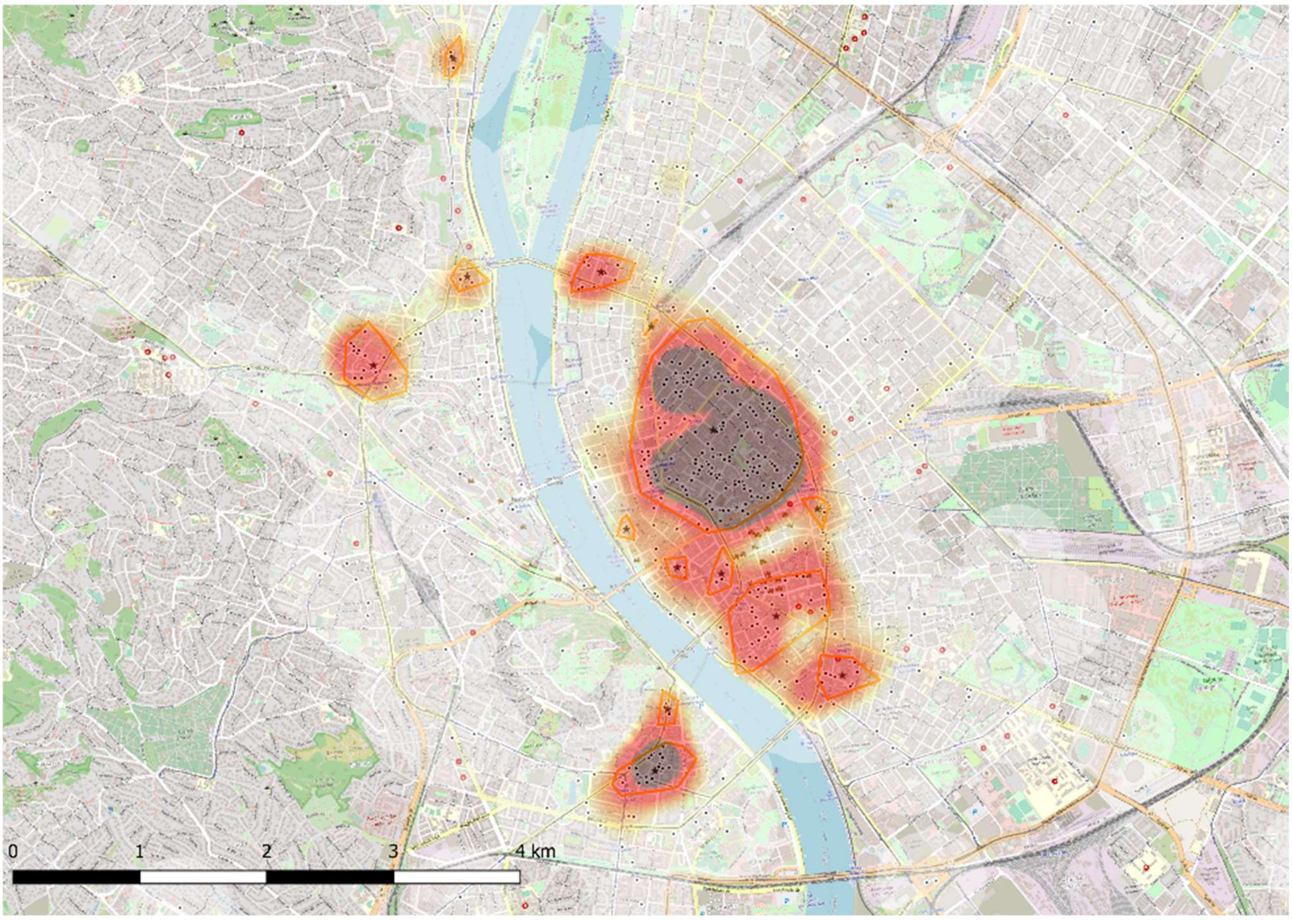


A. Figure 4 - Budapest, max_dist = 125 m



A. Figure 3 - Budapest, comparision between max_dist = 125 m and 150 m

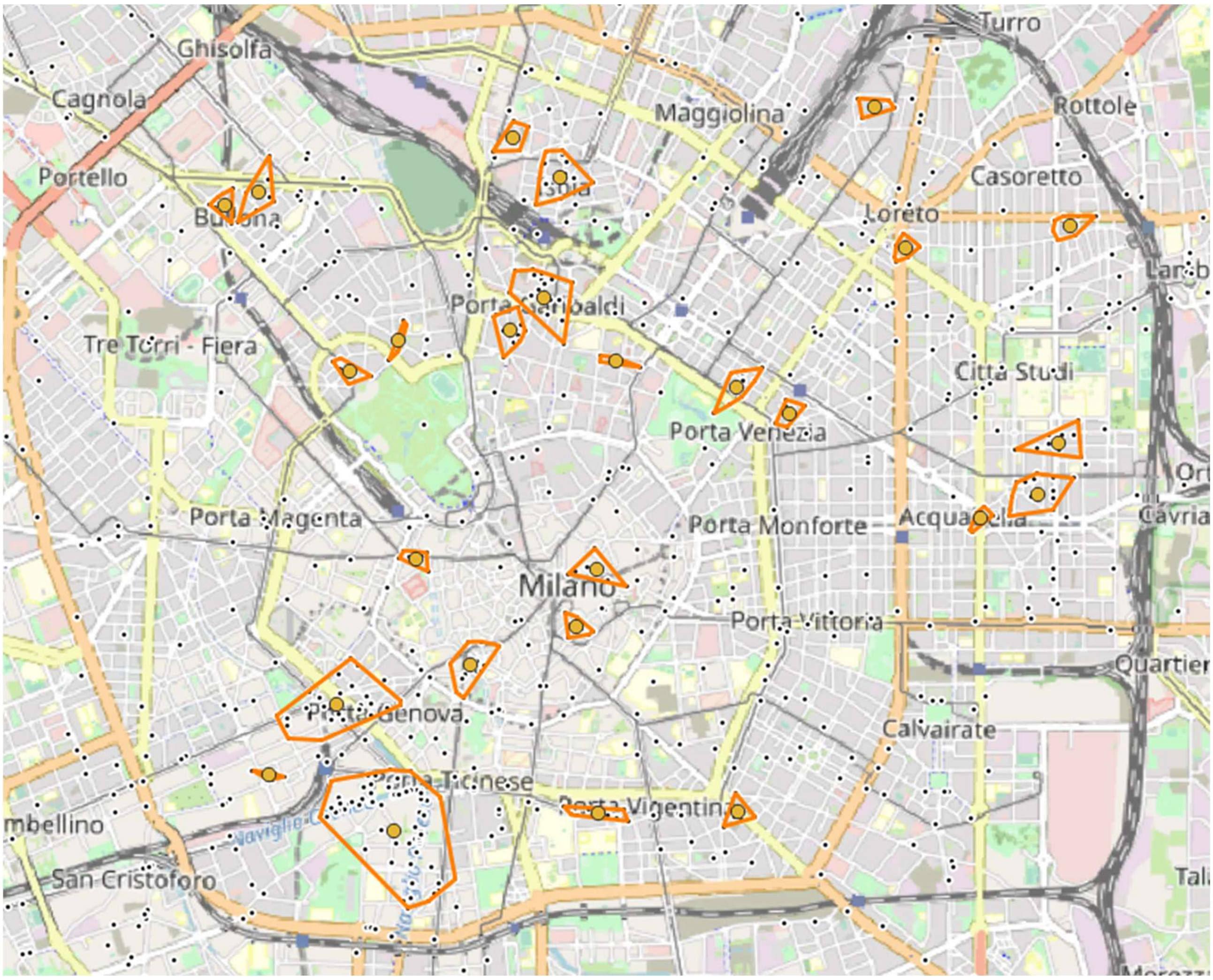
In blue we can see dle cluster creted with max_dist = 150 m and in orange the one created with max_dist = 125 m. The situation reamains stable; the surface changes from 1,43 km^2 to 1,63 km^2 and the number of points inside the area goes from 245 to 228. We decided to keep max_distance = 150 m



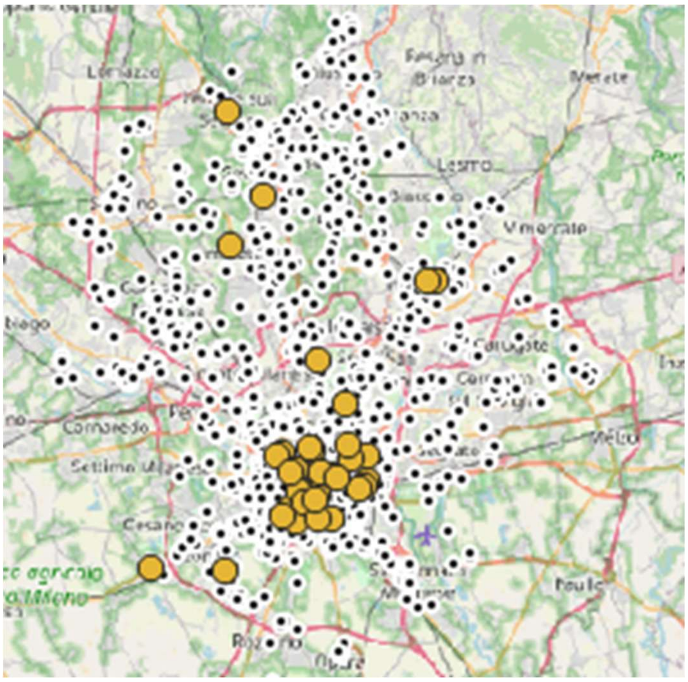
A. Figure 5 - Budapest, Heatmap and NLAs computed with max_dist = 150m.

2. Milano

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	35
Max number of points in a cluster	77

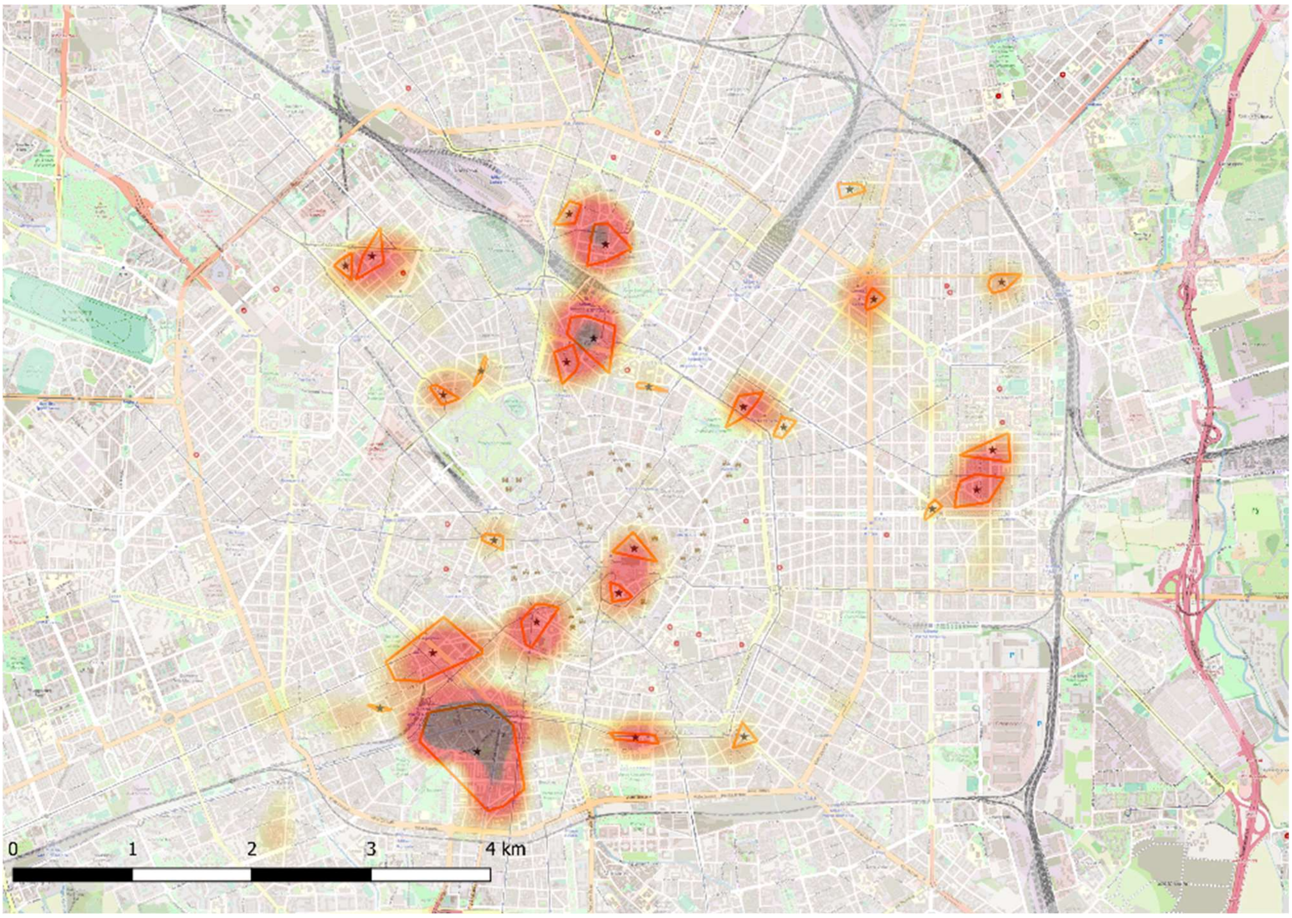


A. Figure 6 - Milano, max_dist = 150 m



A. Figure 7 - Milano, max_dist = 150 m

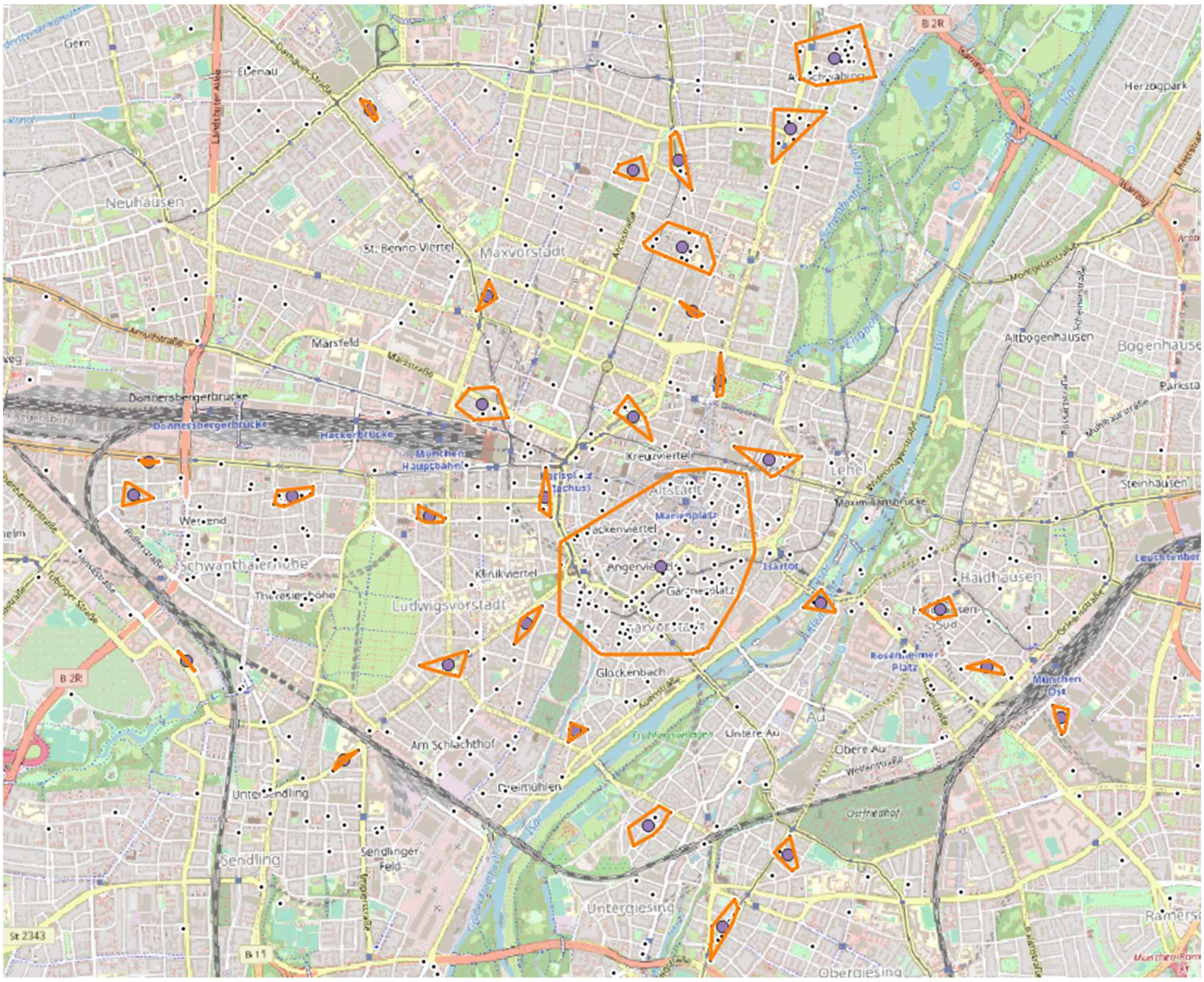
There is no creation of a central wide cluster. 150 m is an appropriate value for the parameter max_distance.



A. Figure 8 - Milano, Heatmap and NLAs computed with max_dist = 150m

3. München

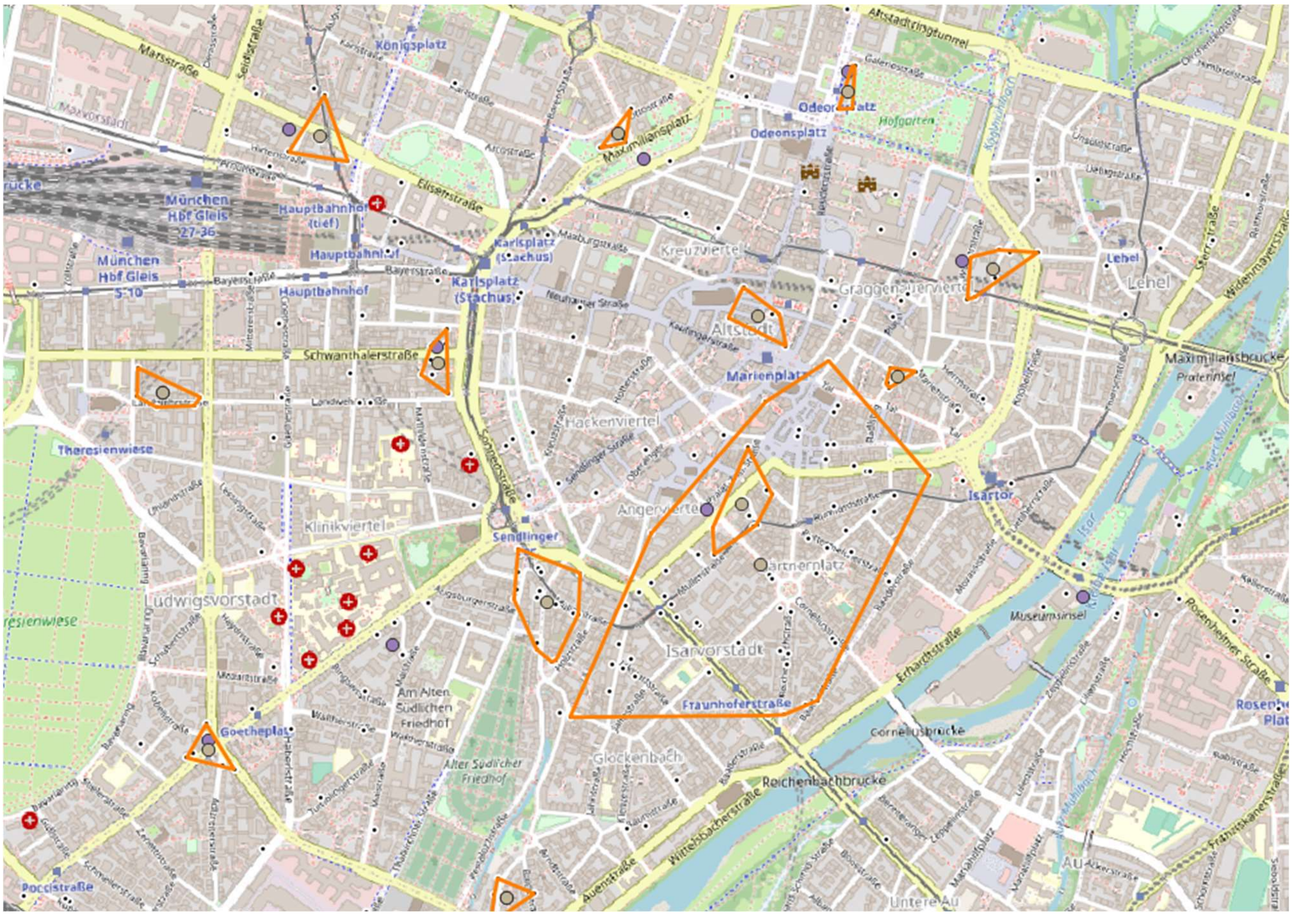
Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	30
Max number of points in a cluster	113



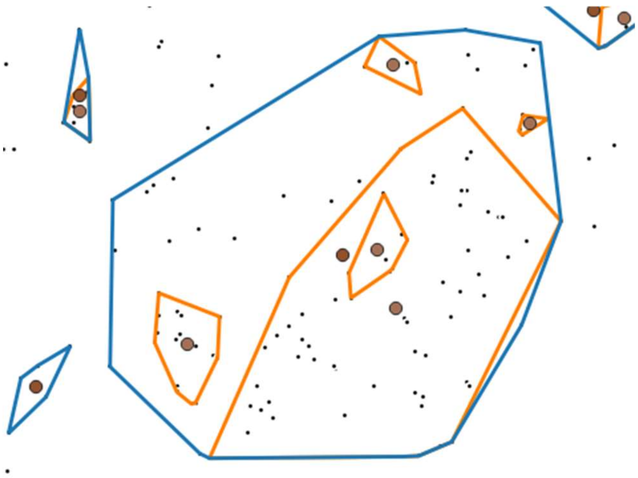
A. Figure 9 - München, max_dist = 150 m

A wide cluster is generated in the city centre. To make it smaller we move the max_distance parameter from 150m to 125m. With the *measure* tool on QGIS (which allow to measure the distance between two points) has been measured that from the centroid to the perimeter points of the cluster, the length never goes over 1 km.

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	26
Max number of points in a cluster	54

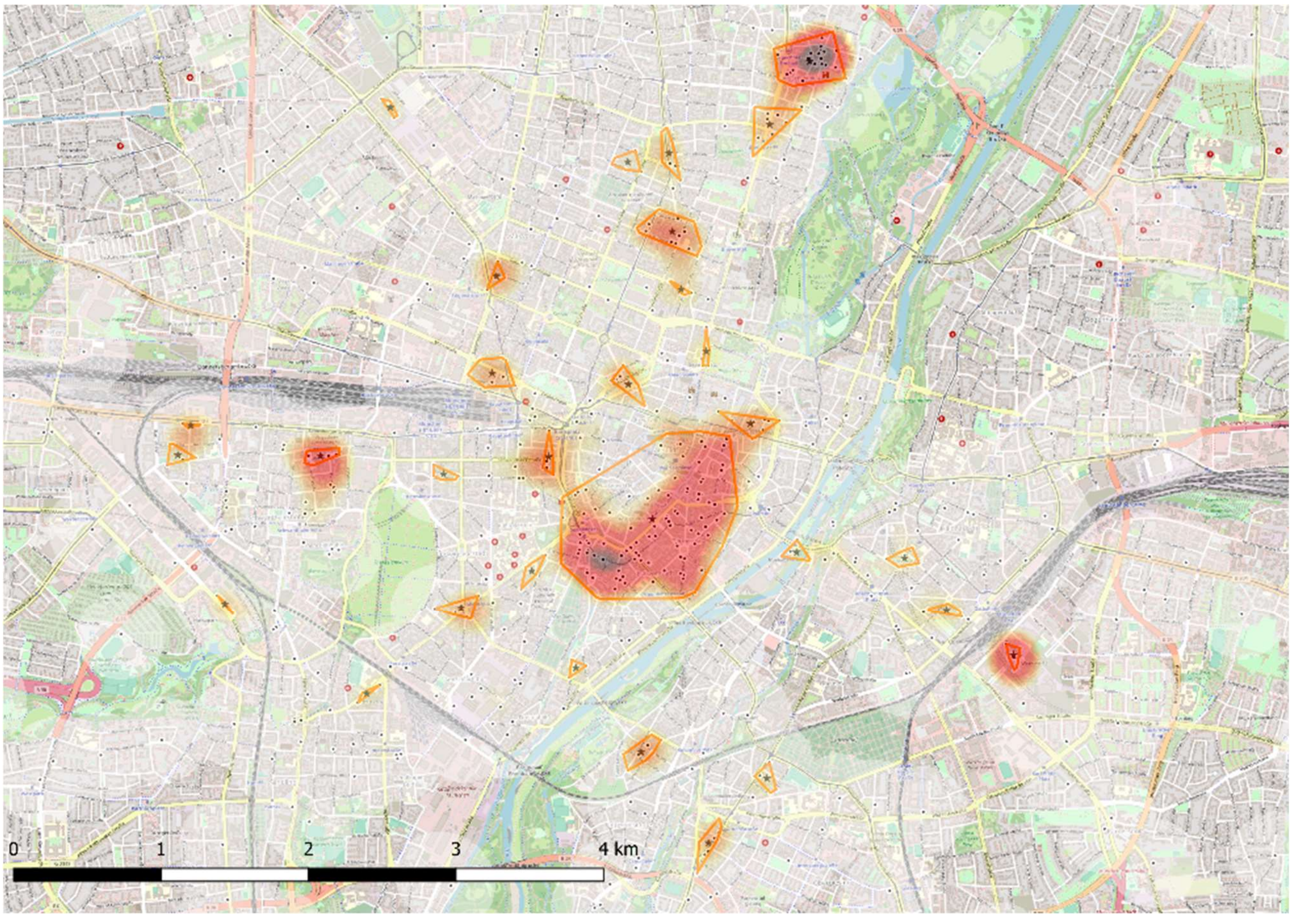


A. Figure 10 - München, max_dist = 125 m



In blue we can see dle cluster creted with max_dist = 150 m and in orange the one created with max_dist = 125 m. The situation change but not significantly; the surface changes from 1,05 km^2 to 0,49 km^2 and the number of points inside the area goes from 113 to 54. We decided to keep max_distance = 150

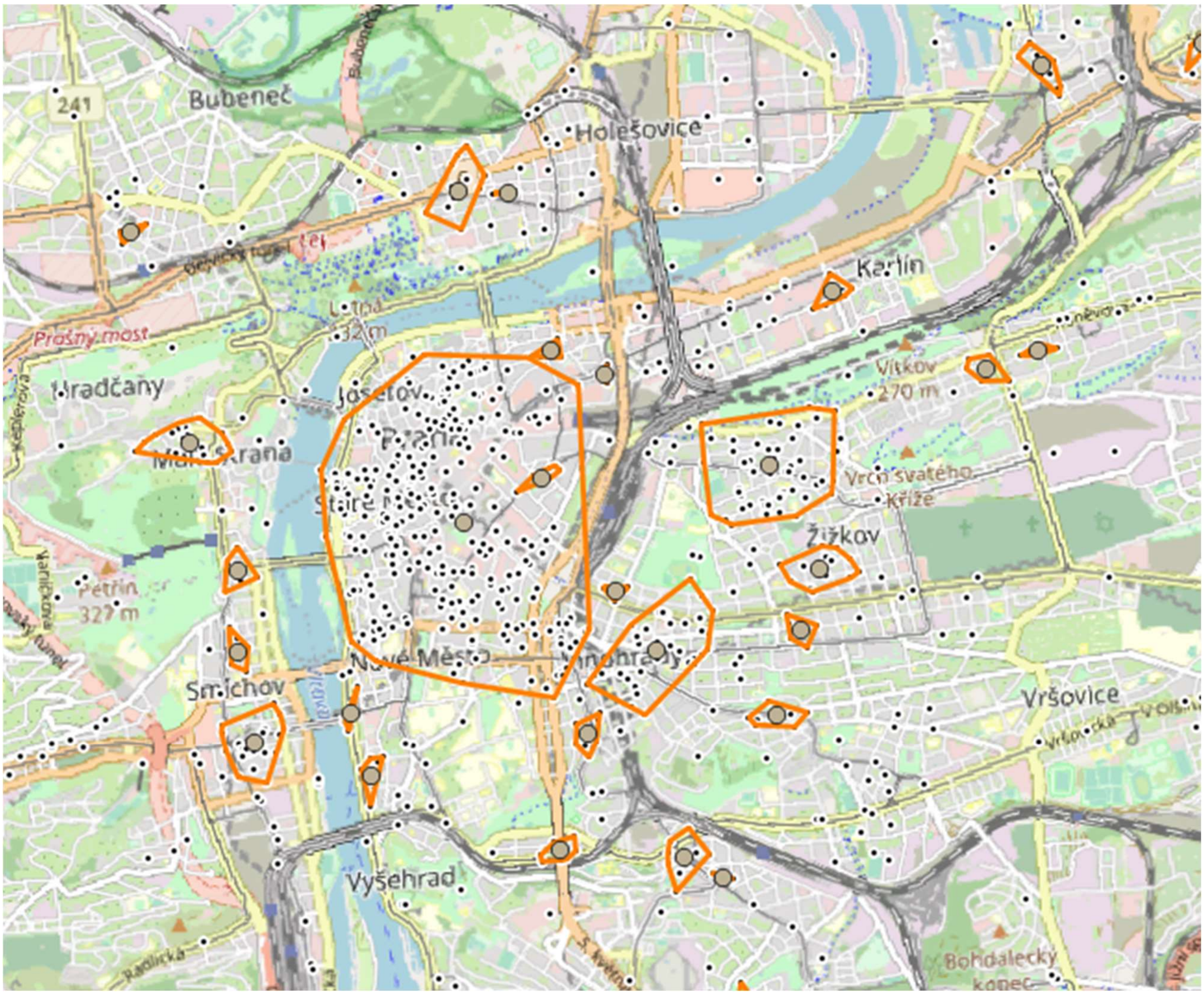
A. Figure 11 - München, comparision between max_dist = 125 m and 150 m



A. Figure 12 - München, Heatmap and NLAs computed with max_dist = 150m

4. Praha

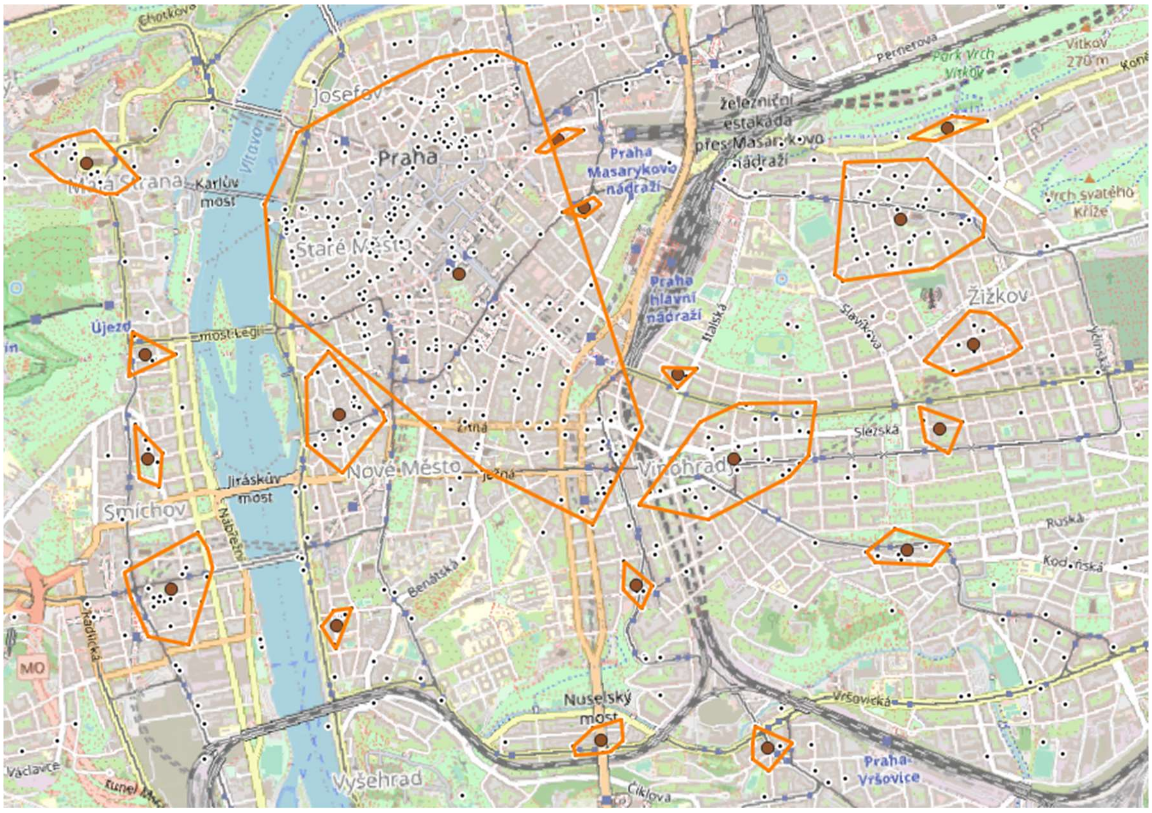
Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	26
Max number of points in a cluster	302



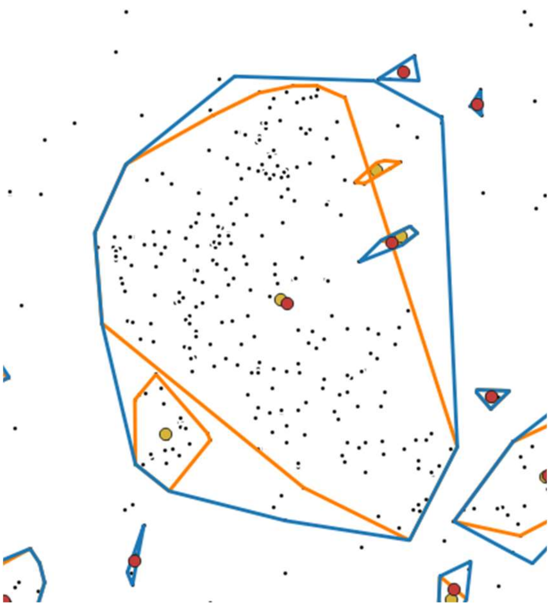
A. Figure 13 - Praha, max_dist = 150 m

A wide cluster is generated in the city centre. To make it smaller we move the max_distance parameter from 150m to 125m. With the *measure* tool on QGIS (which allow to measure the distance between two points) has been measured that from the centroid to the perimeter points of the cluster, the length is around 1 km and eventually more.

Minimum Cluster Size	5
Max distance between clustered points	125
Resulting Clusters	25
Max number of points in a cluster	254

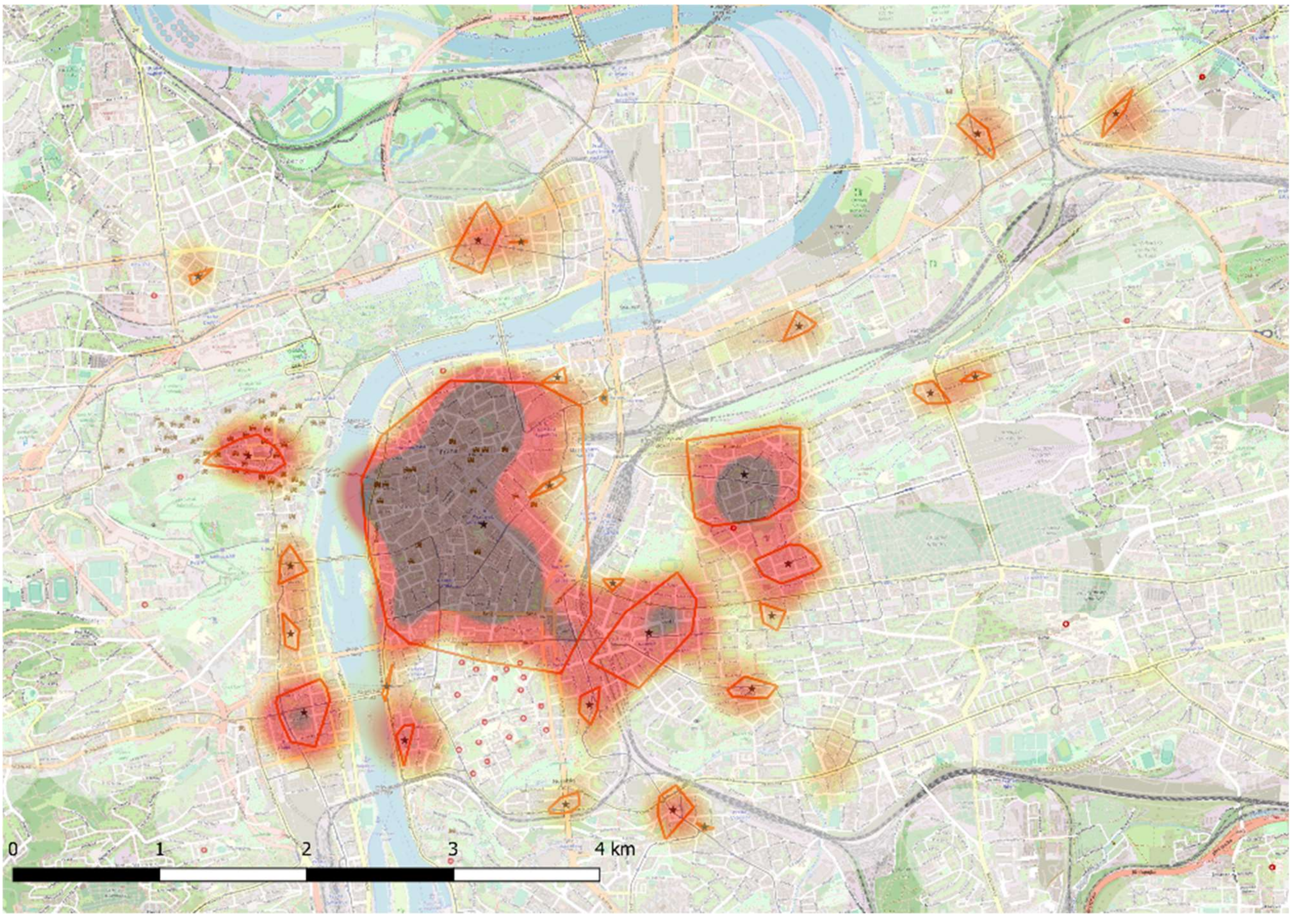


A. Figure 14 - Praha, max_dist = 125 m



A. Figure 15 - Praha, comparision between max_dist = 125 m and 150 m

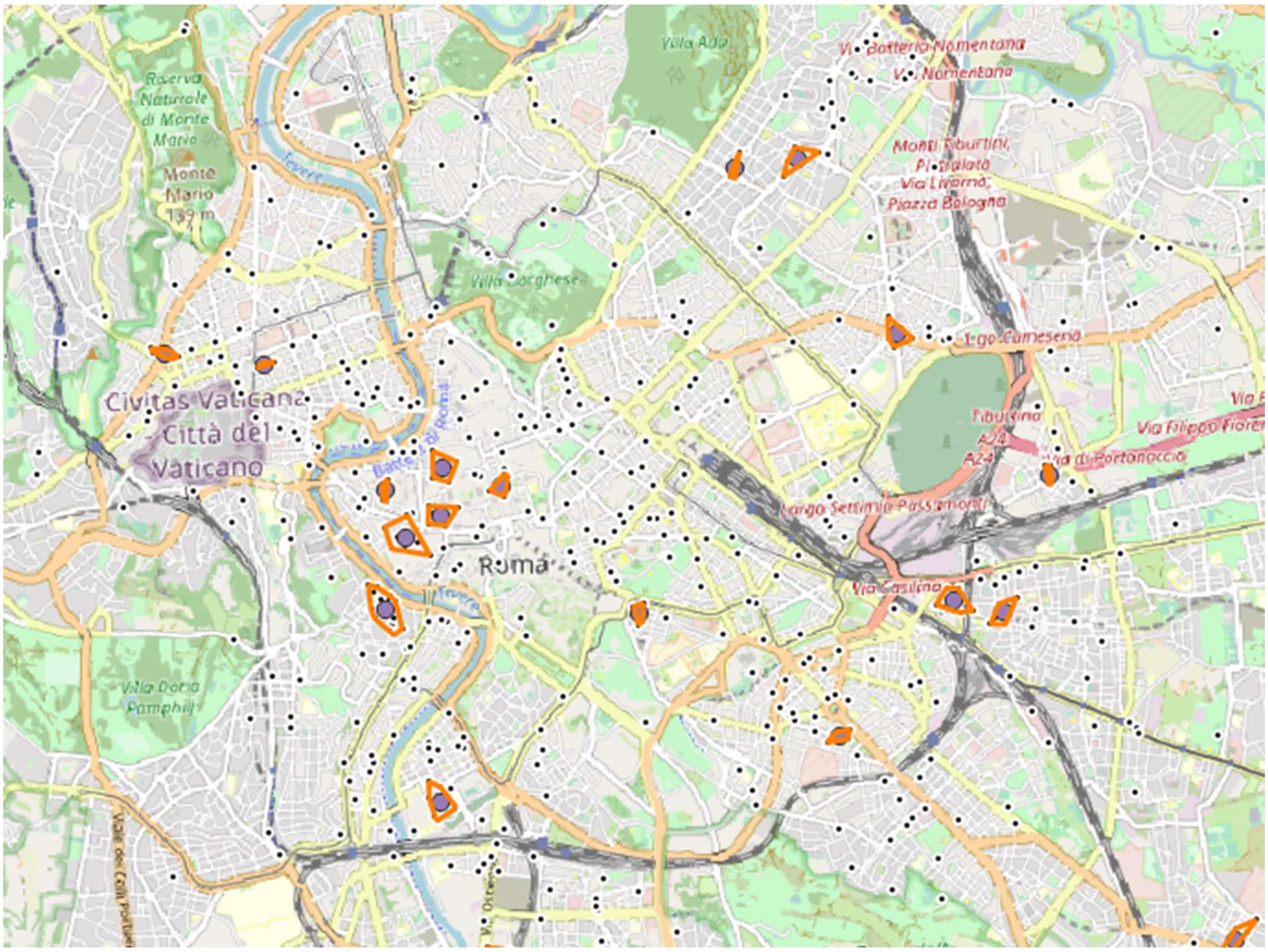
In blue we can see dle cluster creted with max_dist = 150 m and in orange the one created with max_dist = 125 m. The situation changes but not sifnigically; the surface changes from 2,58 km^2 to 1,87 km^2 and the number of points inside the area goes from 113 to 54. We decided to keep max_distance = 150



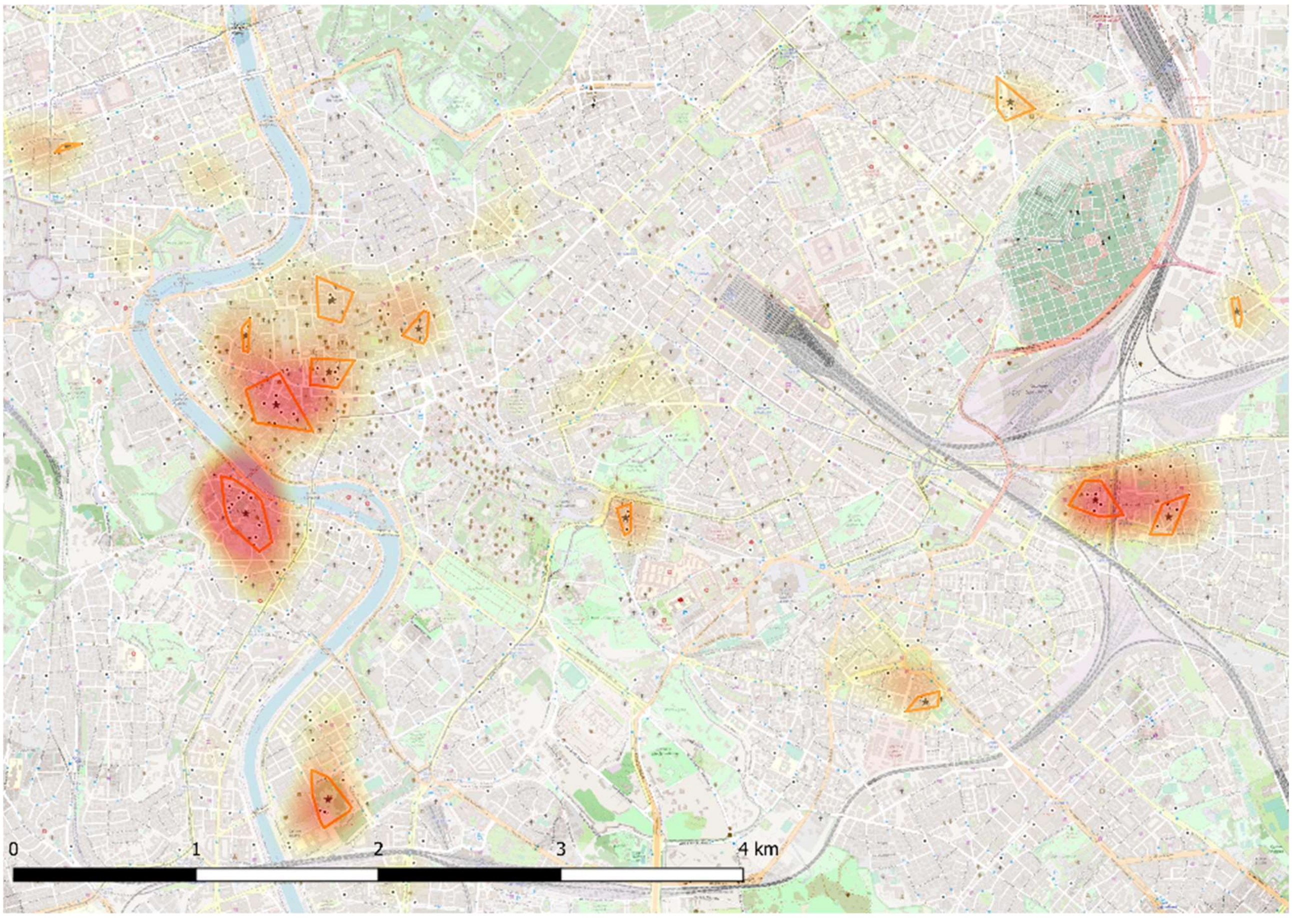
A. Figure 16 - Praha, Heatmap and NLAs computed with max_dist = 150m

5. Roma

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	23
Max number of points in a cluster	20



A. Figure 17 - Roma, max_dist = 150 m

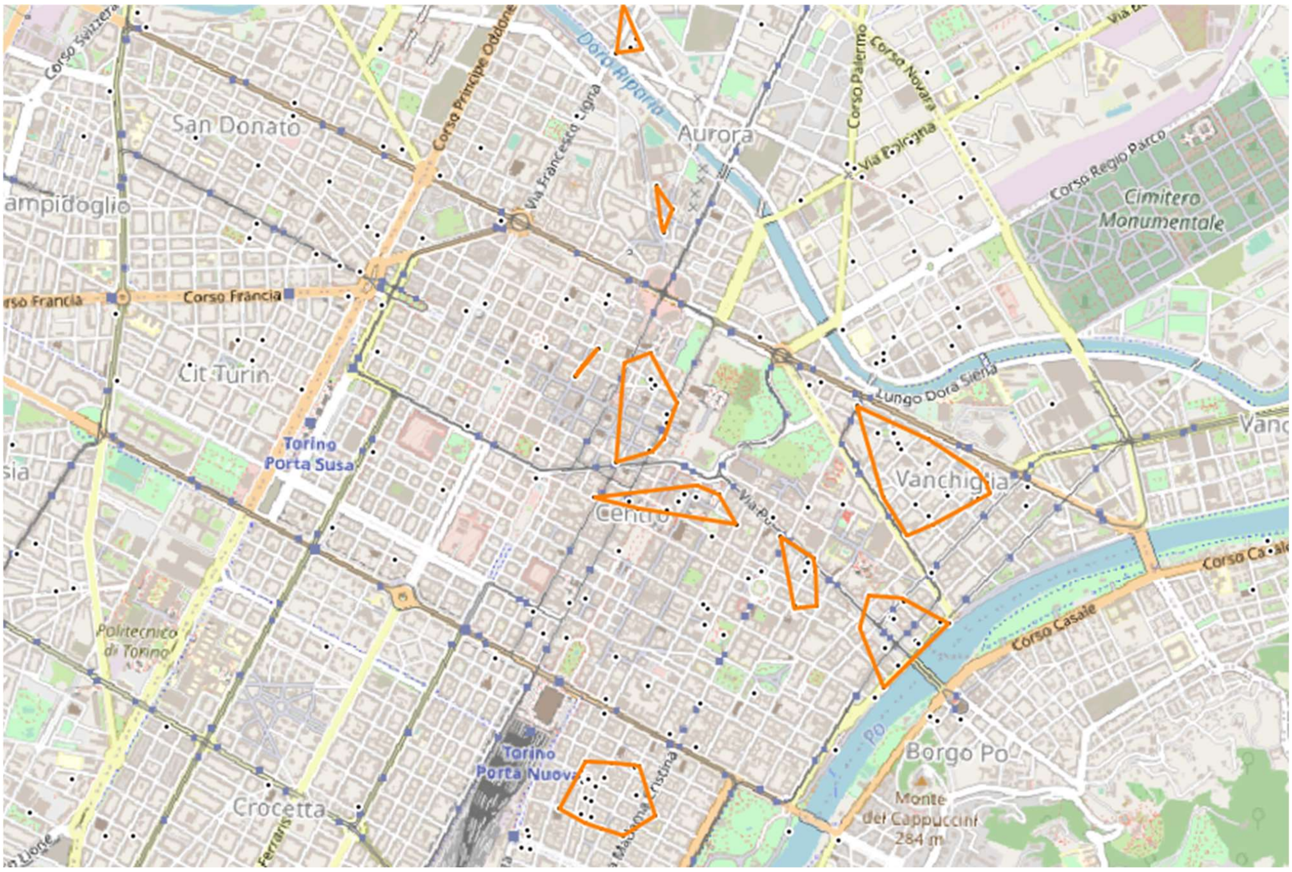


A. Figure 18 - Roma, Heatmap and NLAs computed with max_dist = 150m

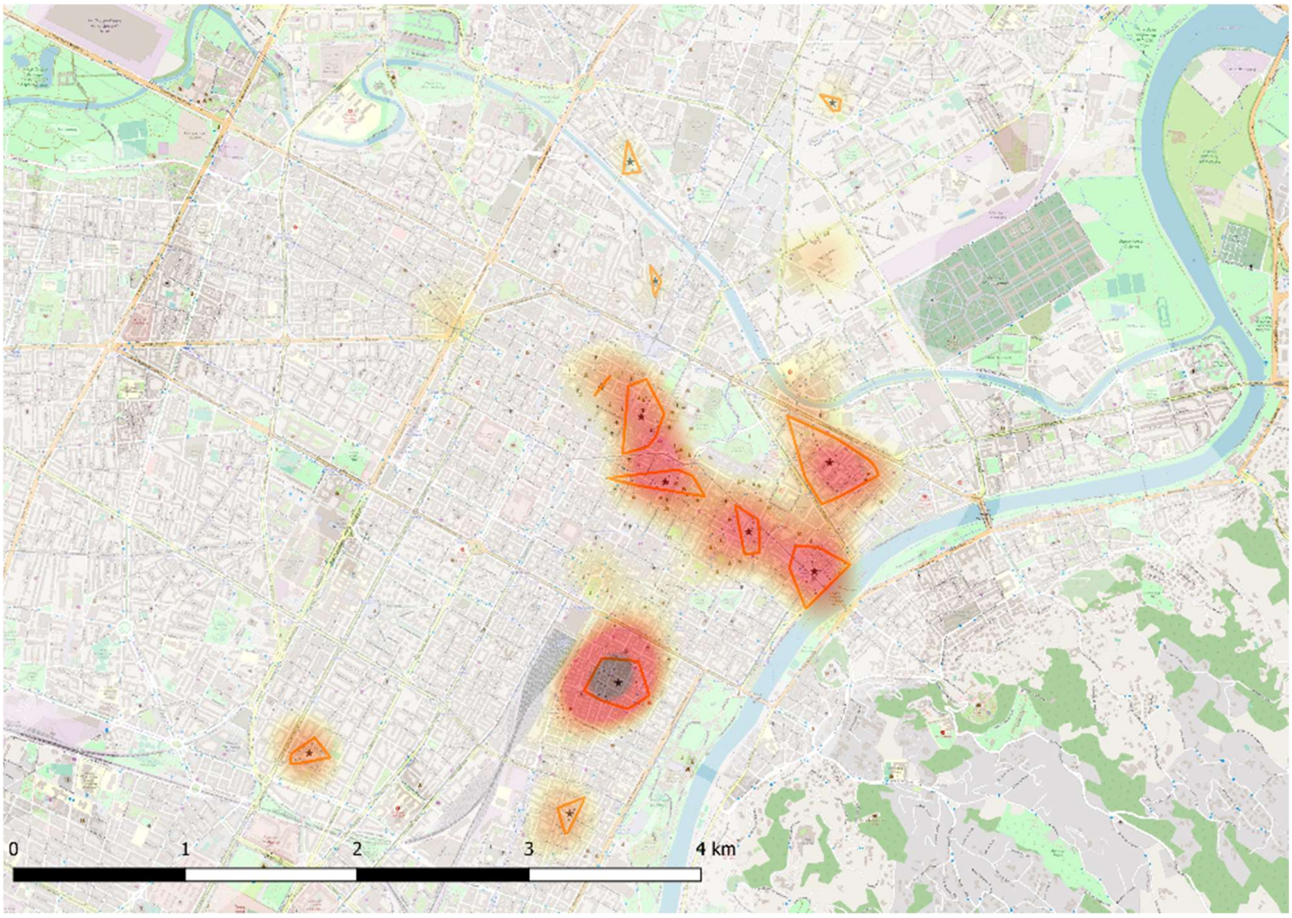
There is no creation of a central wide cluster. 150 m is an appropriate value for the parameter max_distance

6. Torino

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	12
Max number of points in a cluster	23



A. Figure 19 - Torino, max_dist = 150

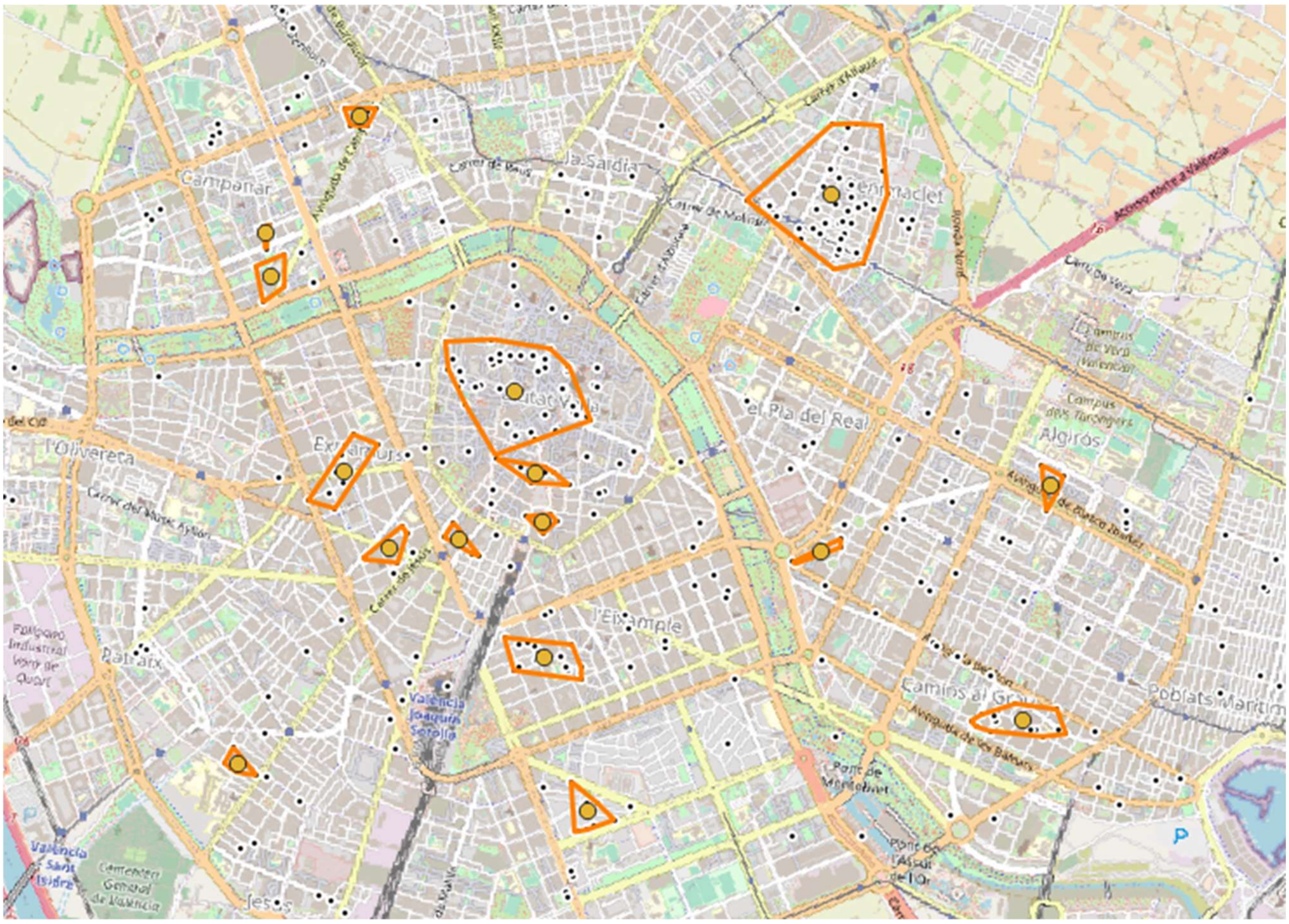


A. Figure 20 - Torino,, Heatmap and NLAs computed with max_dist = 150m

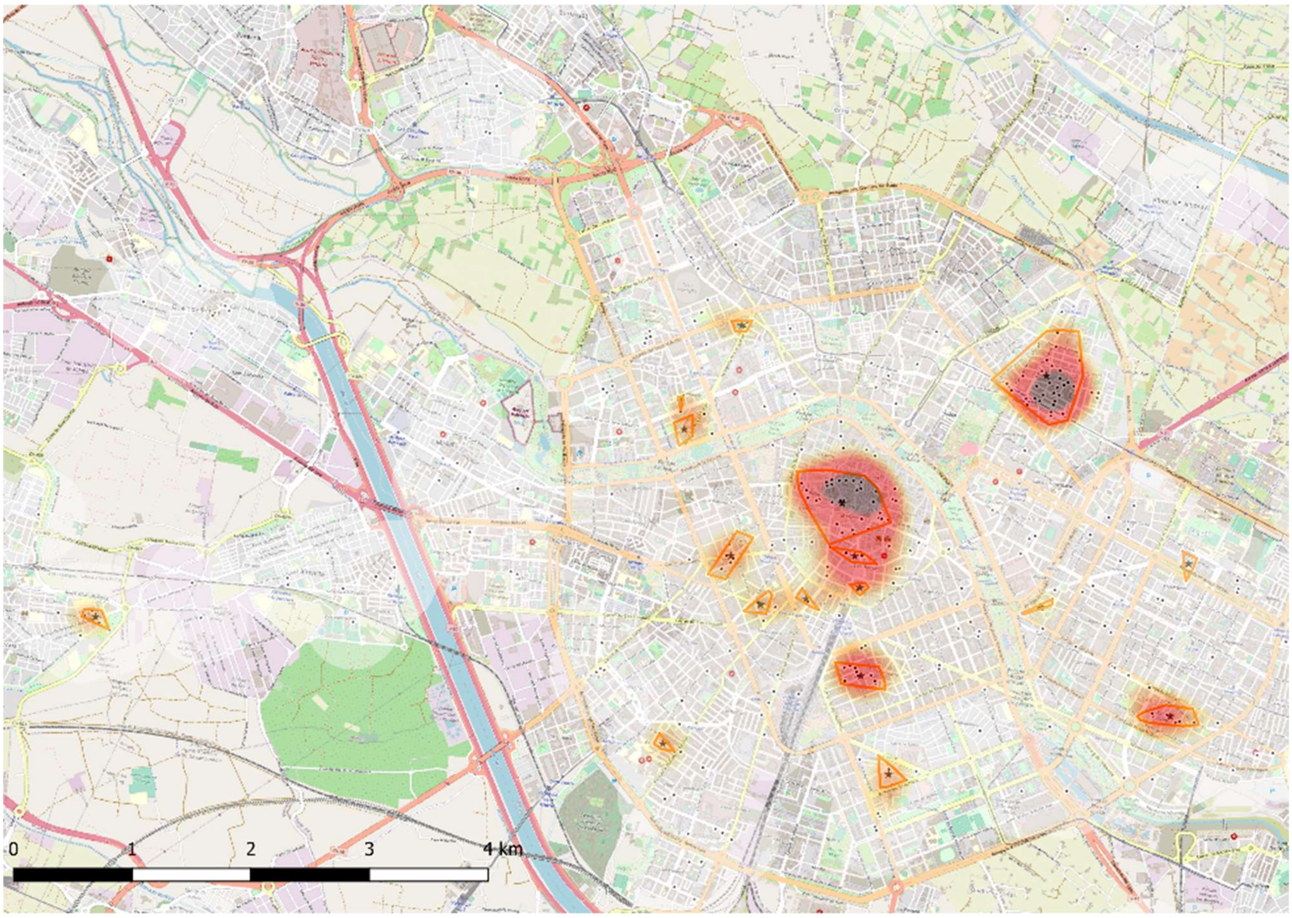
There is no creation of a central wide cluster. 150 m is an appropriate value for the parameter max_distance.

7. Valencia

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	18
Max number of points in a cluster	47



A. Figure 21 - Valencia, max_dist = 150 m

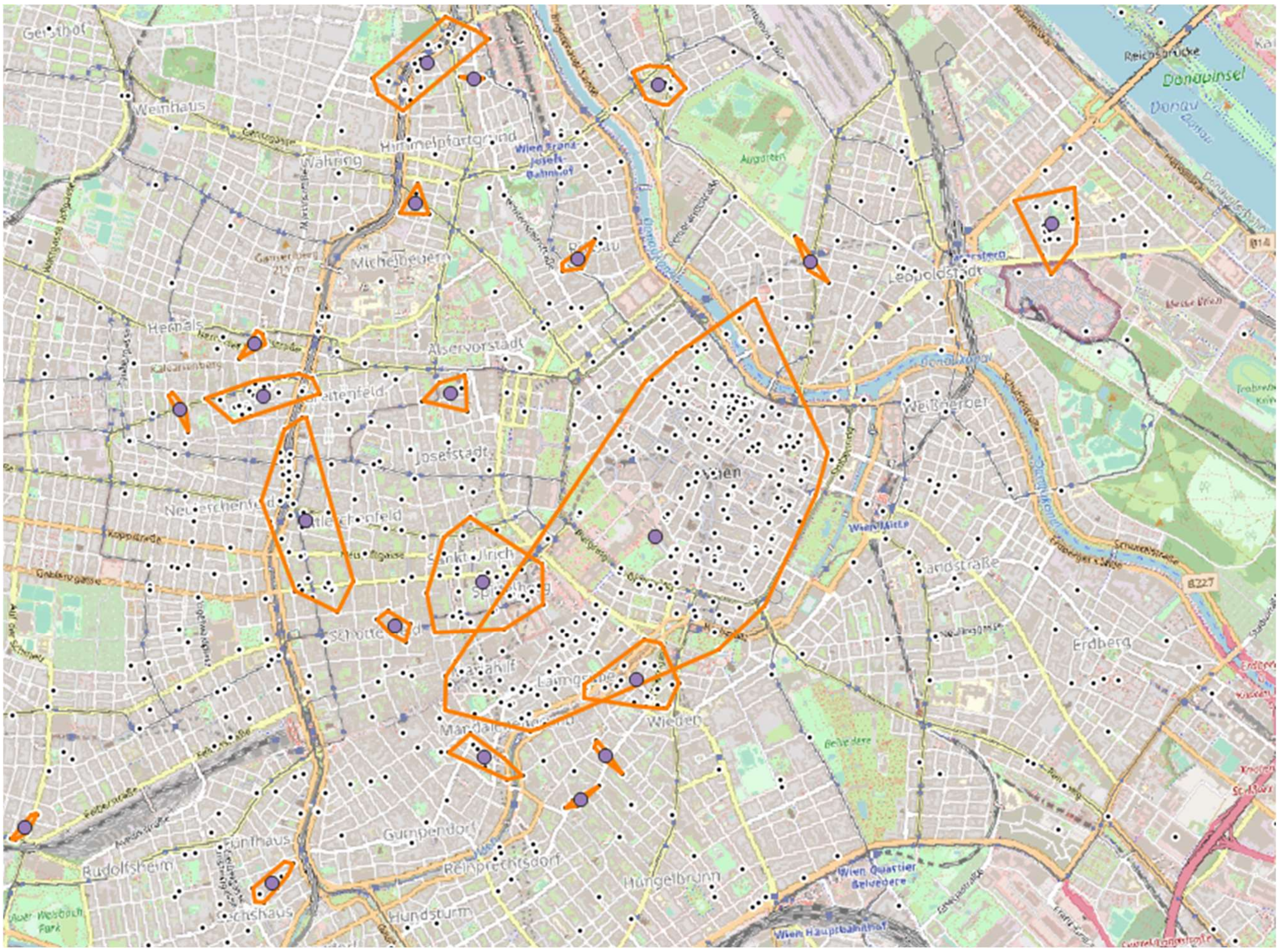


A. Figure 22 - Valencia, Heatmap and NLAs computed with max_dist = 150m

There is no creation of a central wide cluster. 150 m is an appropriate value for the parameter max_distance.

8. Wien

Minimum Cluster Size	5
Max distance between clustered points	150
Resulting Clusters	24
Max number of points in a cluster	223

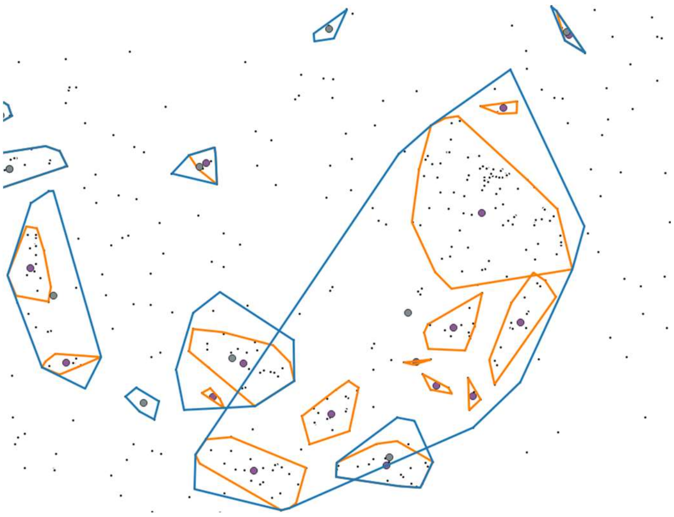


A. Figure 23 - Wien, max_dist = 150 m

A wide cluster is generated in the city centre. To make it smaller we move the max_distance parameter from 150m to 125m. With the *measure* tool on QGIS (which allow to measure the distance between two points) has been measured that from the centroid to the perimeter points of the cluster, the length is eventually more than 1.5 km.

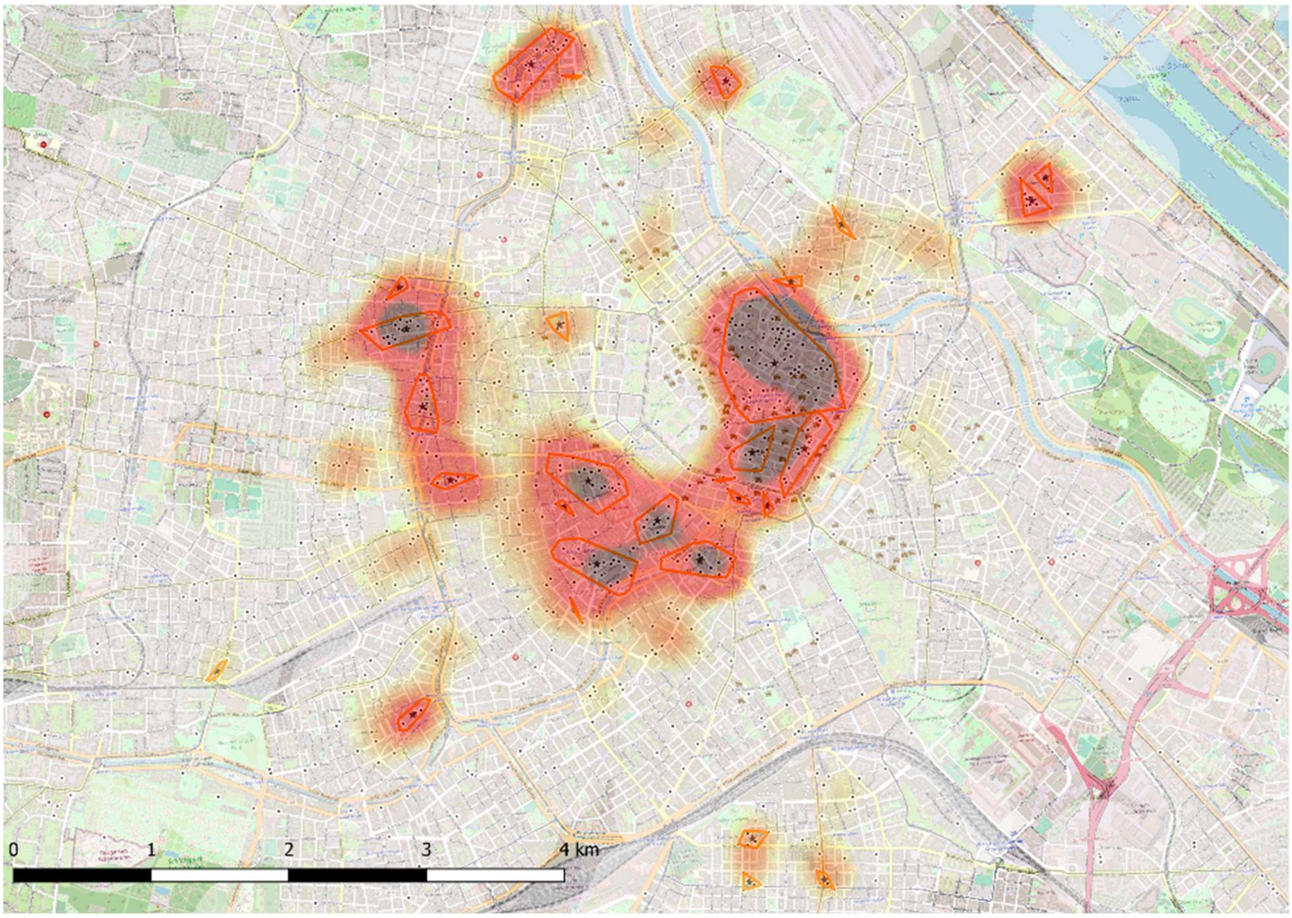
Minimum Cluster Size	5
Max distance between clustered points	125
Resulting Clusters	29
Max number of points in a cluster	104

A. Figure 24 - Wien, max_dist = 125 m



A. Figure 25 - Wien, comparision between max_dist = 125 m and 150 m

In blue we can see dle cluster creted with max_dist = 150 m and in orange the one created with max_dist = 125 m. The situation change significantly; the surface changes from 2,79 km^2 to 0,58 km^2 and the number of points inside the area goes from 223 to 104. We decided to change the parameter value to 125 m.



A. Figure 26 - Wien, Heatmap and NLAs computed with max_dist = 125 m.

APPENDIX B – Results overview

In the following Appendix, for each of the eight studied cities, an overview of the extrapolated data will be exposed. Namely will be introduced information related to:

- Population distribution and Public traffic offer
- Amenities and Night Life Areas
- Results from Traveltime computations (Coverages, Percentages of inhabitants served at least once).

Explanatory pictures of the cities are shown as well.

In order to give a general overview of the data explored for each city, some graphs present in the thesis are here reproposed.

Table 23 - Surface and population density for each Urban Centre; computed with Urban Atlas 2018 data

City	Inhabitants Urban Centre	Surface [km^2]	Density [Inh/km]
Budapest	1,723,314	433	3980
Milano	3,157,550	754	4188
München	1,582,533	349	4534
Praha	1,194,604	295	4050
Roma	2,493,124	482	5172
Torino	1,207,163	207	5832
Valencia	1,514,210	333	4547
Wien	1,984,266	392	5062

Table 24 - Length of night-time public transport lines (km) for the total network and for the study area; offer measures computed as service intensity and service intensity per thousand inhabitants; average frequency for each line operated in the two hours study time; Network intensity.

City	Total length	Length_Urban centre	Service efficiency	Service efficiency per 1000 inh	Average Frequency	Network intensity
	km	Km	Veh*km*2h	Veh*km*2h / (Inh/1000)	minutes	Km/km²
Budapest	1276.2	1078.7	3228.2	1.87	40	2,49
Milano	322.5	317.4	1269.6	0.40	30	0,42
München	773.1	726.0	3404.2	2.15	26	2,08
Praha	1159.2	896.9	2738.5	2.29	39	3,04
Roma	1104.1	849.7	3398.7	1.36	30	1,76
Torino	159.3	130.7	261.4	0.22	60	0,63
Valencia	411.2	366.3	1465.3	0.97	30	1,10
Wien	573.0	570.2	2678.7	1.35	26	1,45

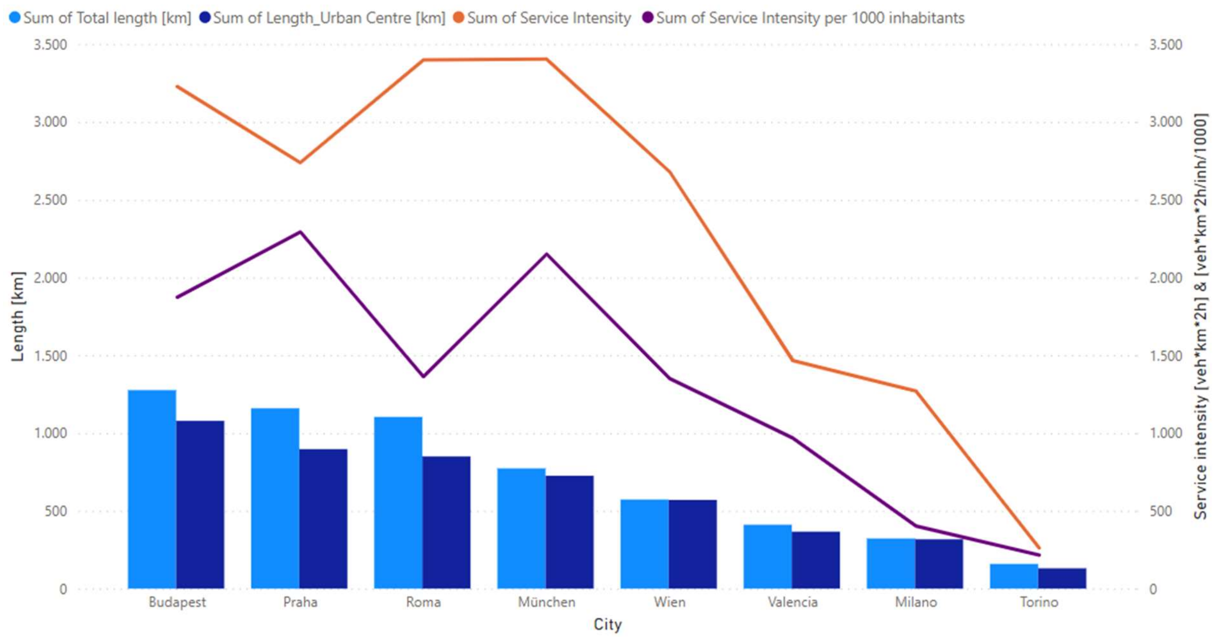


Figure 67 - Comparison between networks lengths (the total ones the ones inside the study area) and offer measures.

Table 25 - Results of the traveltime computation for each city

City	Inhabitants in Urban Centre	Coverage_30 min	Weighted Coverage_30min	Coverage_45min	Weighted Coverage_45min	Onceserved_30min	total_onceserved_30min	Onceserved_45min	total_onceserved_45min
Budapest	1,723,314	30.95 %	31.57 %	49.76%	50.85%	40.13%	50.52%	65.03%	80.83%
Milano	3,157,550	7.13%	7.79%	16.65%	19.64%	24.22%	31.83%	36.39%	43.07%
München	1,582,533	16.34 %	18.19%	38.66%	42.36%	47.01%	65.62%	67.55%	85.45%
Praha	1,194,604	15.13%	15.19%	36.23%	37.38%	46.14%	60.40%	75.37%	93.06%
Roma	2,493,124	10.48 %	10.31%	20.88%	20.67%	38.46%	56.62%	56.85%	80.31%
Torino	1,207,163	5.94%	5.35%	8.89%	9.01%	18.93%	44.46%	25.74%	57.52%
Valencia	1,514,210	18.52 %	18.40 %	32.81%	33.43%	48.16%	55.19%	54.79%	59.62%
Wien	1,984,266	21.94 %	24.16 %	52.98%	57.96%	61.55%	69.00%	81.82%	88.27%

.

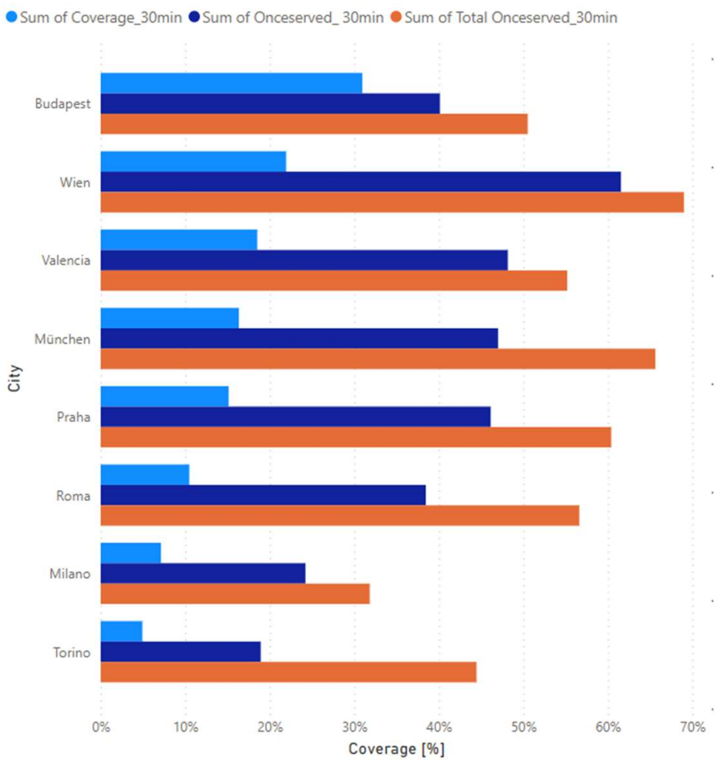


Figure 68 – Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 30 minutes.

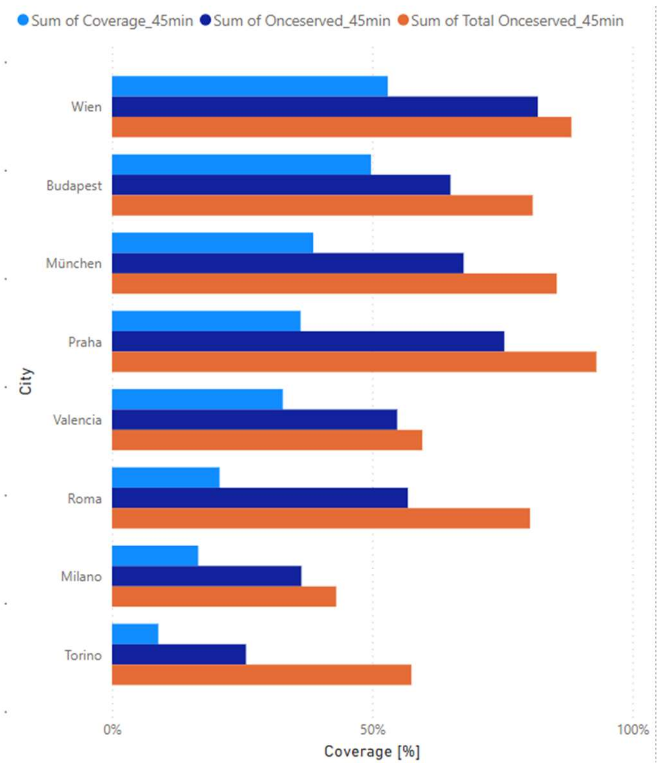
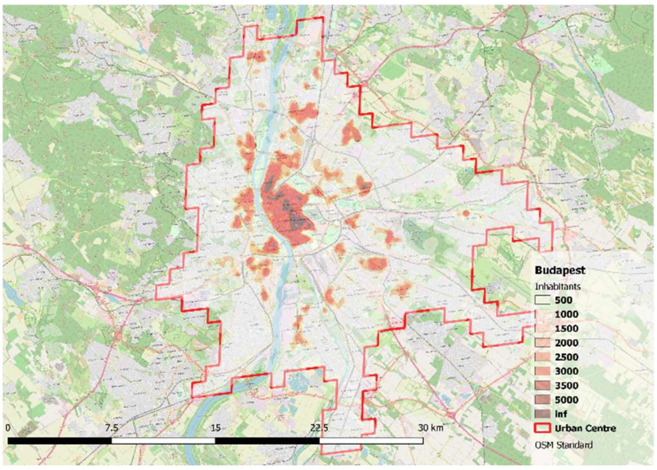


Figure 69 - Visual representation of the results obtained through the reachability computation for each city. The results are exposed following a decreasing order of the coverage. Results for a time threshold of 45 minutes.

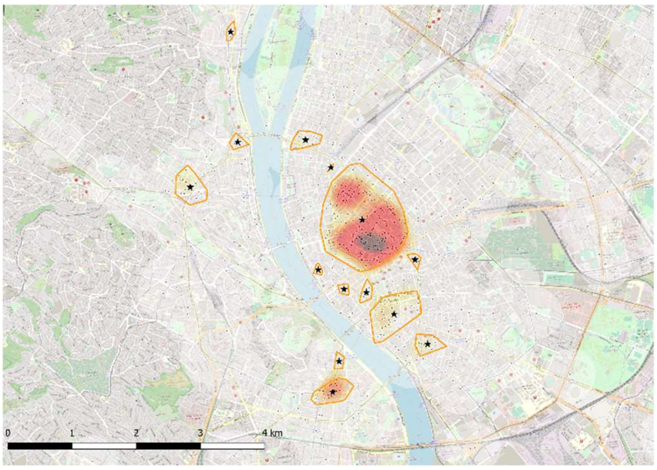
Budapest

Population distribution and Public transport offer



Population	1.723.314	inhabitants
Surface	433	Km^2
Density	3980	Inh/km^2
Total length	1276	km
Length in FUA	1079	km
Offer	3228,25	Veh*km*2h
Offer x 1000 inh.	1,87	Veh*km*2h/inh

Amenities and Night Life Area

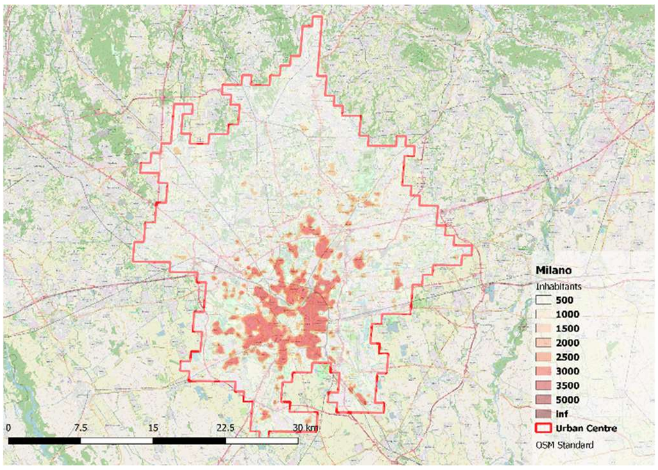


Amenities in FUA	917	-
Clustered Amenities	446	-
Number of cluster	14	-
Inhabitants for each amenities	1879	-

Coverage in 30 min	30,95%	%
Weighted coverage in 30 min	31,57%	%
Coverage in 45 min	49,76%	%
Weighted coverage in 45 min	50,85%	%
Averaged once served 30 min	40,13%	%
Total once served 30 min	50,52%	%
Averaged once served 45 min	65,03%	%
Total once served 45 min	80,83%	%

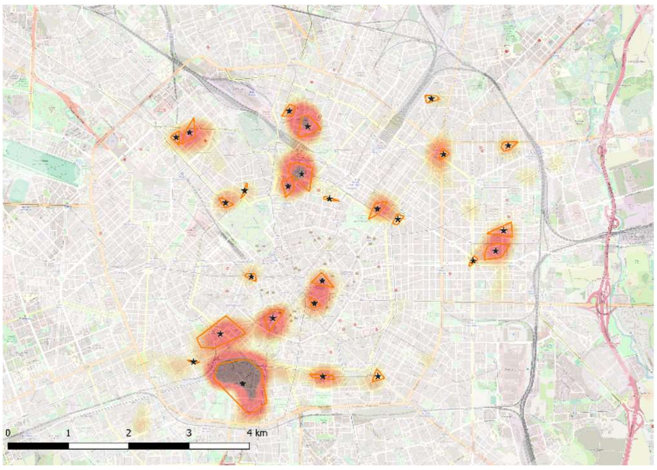
Milano

Population distribution and Public transport offer



Population	3.157.550	inhabitants
Surface	754	Km^2
Density	4188	Inh/km^2
Total length	323	km
Length in FUA	317	km
Offer	1269,6	Veh*km*2h
Offer x 1000 inh.	0,40	Veh*km*2h/inh

Amenities and Night Life Area

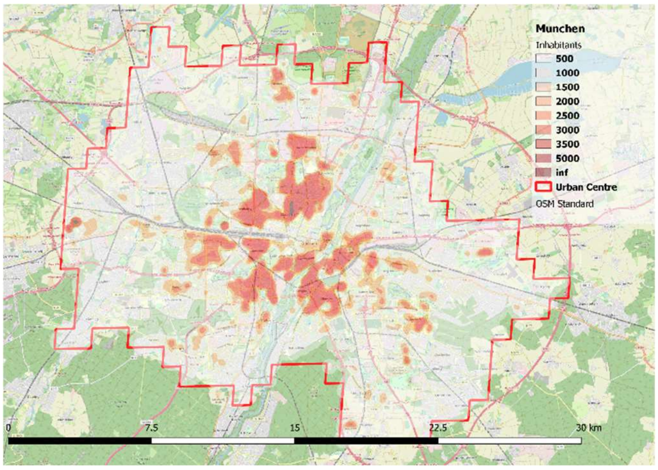


Amenities in FUA	1680	-
Clustered Amenities	370	-
Number of cluster	35	-
Inhabitants for each amenities	1879	-

Coverage in 30 min	7,13%	%
Weighted coverage in 30 min	7,79%	%
Coverage in 45 min	16,65%	%
Weighted coverage in 45 min	19,64%	%
Averaged once served 30 min	24,22%	%
Total once served 30 min	31,83%	%
Averaged once served 45 min	36,39%	%
Total once served 45 min	43,07%	%

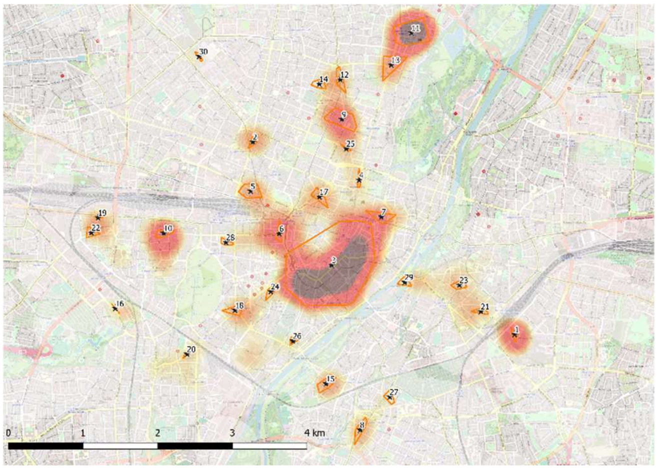
München

Population distribution and Public transport offer



Population	1.582.533	inhabitants
Surface	349	Km^2
Density	4534	Inh/km^2
Total length	773	km
Length in FUA	726	km
Offer	3404	Veh*km*2h
Offer x 1000 inh.	2,15	Veh*km*2h/inh

Amenities and Night Life Area

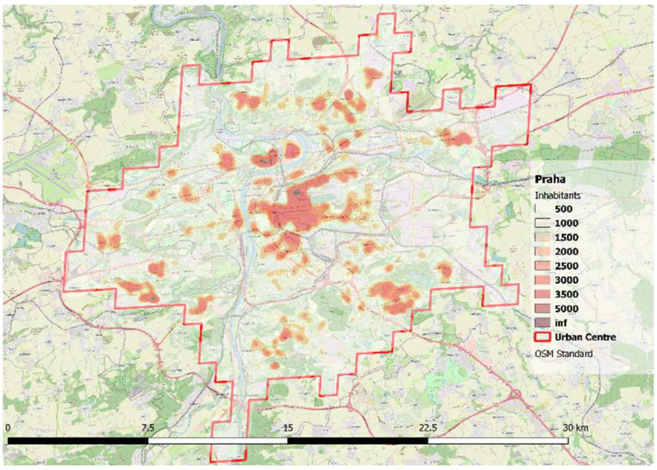


Amenities in FUA	985	-
Clustered Amenities	347	-
Number of cluster	30	-
Inhabitants for each amenities	1607	-

Coverage in 30 min	16,34%	%
Weighted coverage in 30 min	18,19%	%
Coverage in 45 min	38,66%	%
Weighted coverage in 45 min	42,36%	%
Averaged once served 30 min	47,01%	%
Total once served 30 min	65,62%	%
Averaged once served 45 min	67,55%	%
Total once served 45 min	85,45%	%

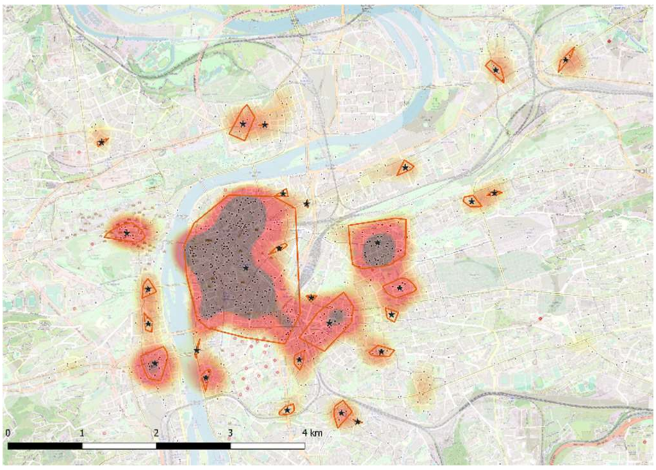
Praha

Population distribution and Public transport offer



Population	1.194.604	inhabitants
Surface	295	Km^2
Density	4050	Inh/km^2
Total length	1159	km
Length in FUA	897	km
Offer	2738,56	Veh*km*2h
Offer x 1000 inh.	2,29	Veh*km*2h/inh

Amenities and Night Life Area

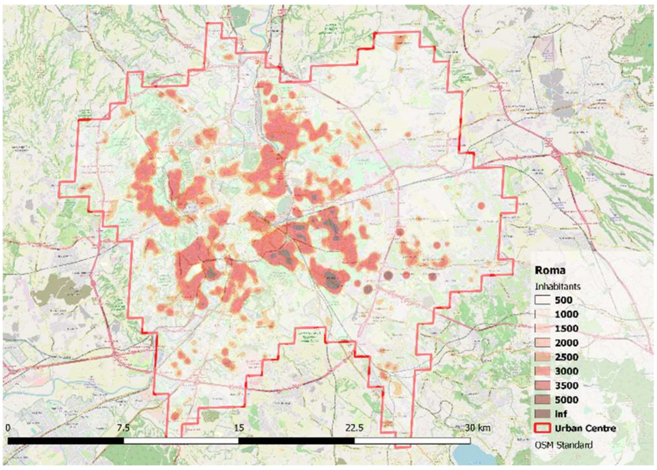


Amenities in FUA	1156	-
Clustered Amenities	608	-
Number of cluster	26	-
Inhabitants for each amenities	1033	-

Coverage in 30 min	15,13%	%
Weighted coverage in 30 min	15,19%	%
Coverage in 45 min	36,23%	%
Weighted coverage in 45 min	37,38%	%
Averaged once served 30 min	46,14%	%
Total once served 30 min	60,40%	%
Averaged once served 45 min	75,37%	%
Total once served 45 min	93,06%	%

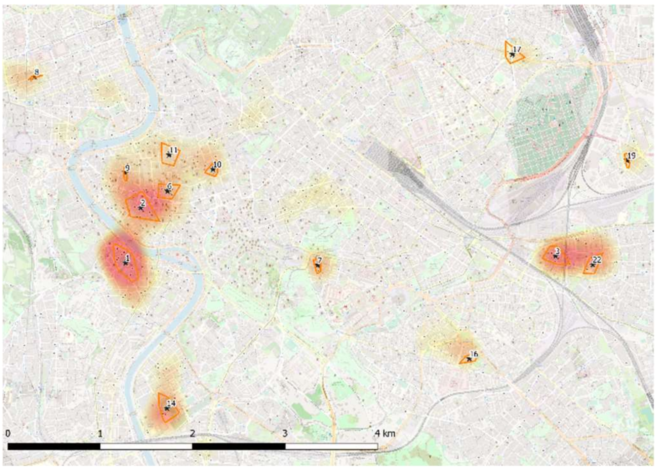
Roma

Population distribution and Public transport offer



Population	2.493.124	inhabitants
Surface	482	Km^2
Density	5172	Inh/km^2
Total length	1104	km
Length in FUA	850	km
Offer	3398,72	Veh*km*2h
Offer x 1000 inh.	1,36	Veh*km*2h/inh

Amenities and Night Life Area

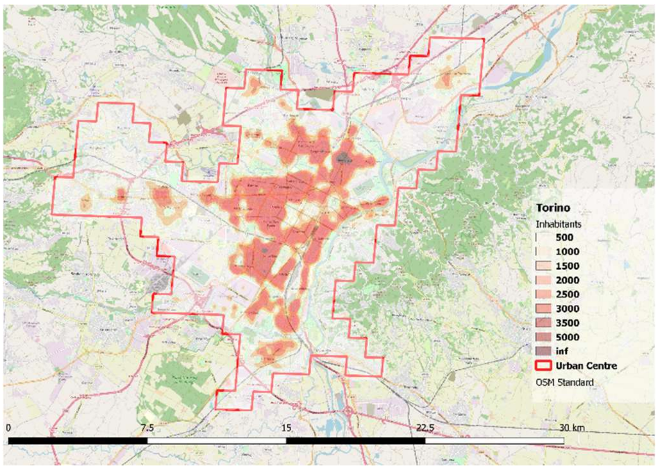


Amenities in FUA	773	-
Clustered Amenities	165	-
Number of cluster	23	-
Inhabitants for each amenities	3225	-

Coverage in 30 min	10,48%	%
Weighted coverage in 30 min	10,31%	%
Coverage in 45 min	20,88%	%
Weighted coverage in 45 min	20,67%	%
Averaged once served 30 min	38,46%	%
Total once served 30 min	56,62%	%
Averaged once served 45 min	56,85%	%
Total once served 45 min	80,31%	%

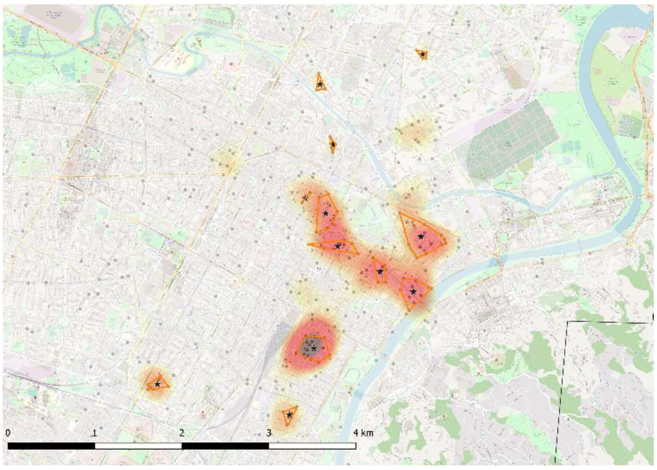
Torino

Population distribution and Public transport offer



Population	1.207.163	inhabitants
Surface	207	Km^2
Density	5832	Inh/km^2
Total length	159	km
Length in FUA	131	km
Offer	261,4	Veh*km*2h
Offer x 1000 inh.	0,22	Veh*km*2h/inh

Amenities and Night Life Area

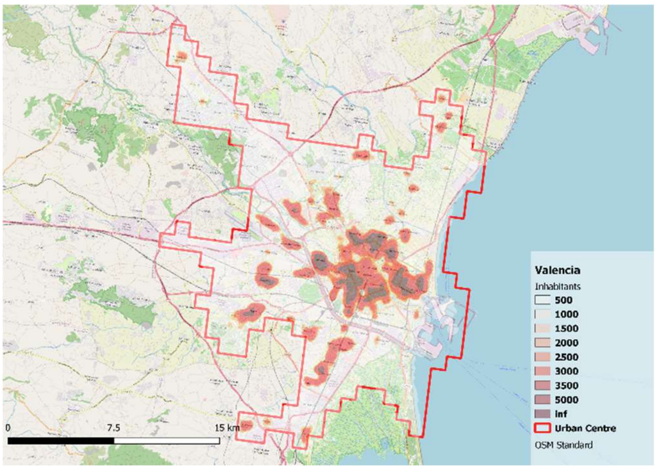


Amenities in FUA	484	-
Clustered Amenities	120	-
Number of cluster	12	-
Inhabitants for each amenities	2494	-

Coverage in 30 min	5,94%	%
Weighted coverage in 30 min	5,35%	%
Coverage in 45 min	8,89%	%
Weighted coverage in 45 min	9,01%	%
Averaged once served 30 min	18,93%	%
Total once served 30 min	44,46%	%
Averaged once served 45 min	25,74%	%
Total once served 45 min	57,52%	%

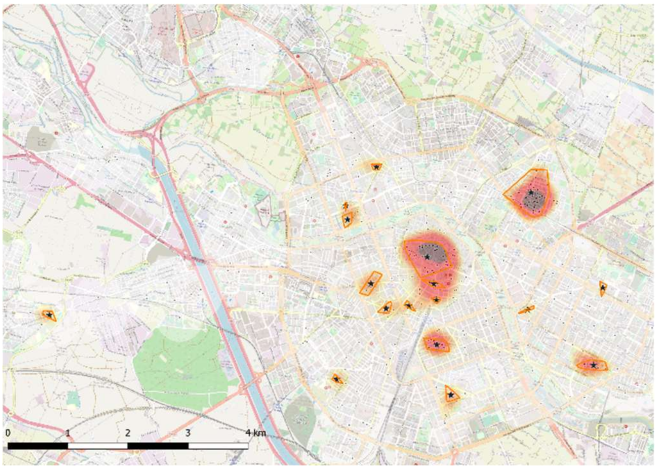
Valencia

Population distribution and Public transport offer



Population	1.514.210	inhabitants
Surface	333	Km^2
Density	4547	Inh/km^2
Total length	411	km
Length in FUA	366	km
Offer	1465,32	Veh*km*2h
Offer x 1000 inh.	0,97	Veh*km*2h/inh

Amenities and Night Life Area

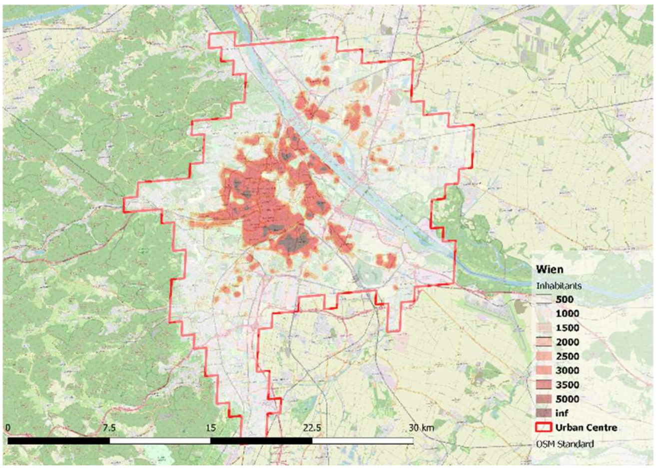


Amenities in FUA	474	-
Clustered Amenities	207	-
Number of cluster	18	-
Inhabitants for each amenities	3195	-

Coverage in 30 min	18,52%	%
Weighted coverage in 30 min	18,40%	%
Coverage in 45 min	32,81%	%
Weighted coverage in 45 min	33,43%	%
Averaged once served 30 min	48,16%	%
Total once served 30 min	55,19%	%
Averaged once served 45 min	54,79%	%
Total once served 45 min	59,62%	%

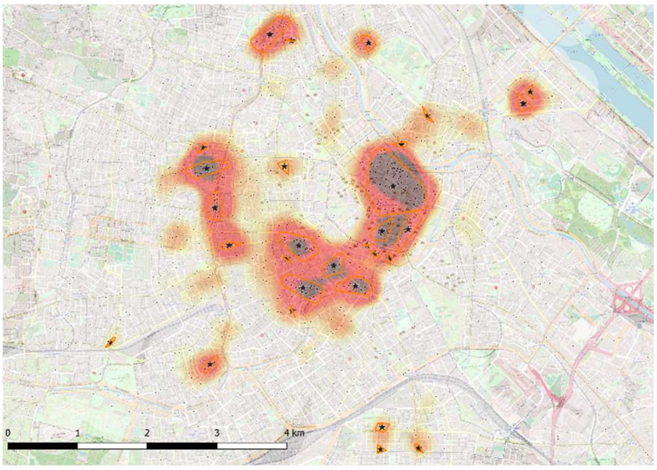
Wien

Population distribution and Public transport offer



Population	1.984.266	inhabitants
Surface	392	Km^2
Density	5062	Inh/km^2
Total length	573	km
Length in FUA	570	km
Offer	2678,76	Veh*km*2h
Offer x 1000 inh.	1,35	Veh*km*2h/inh

Amenities and Night Life Area



Amenities in FUA	1143	-
Clustered Amenities	434	-
Number of cluster	29	-
Inhabitants for each amenities	1736	-

Coverage in 30 min	21,94%	%
Weighted coverage in 30 min	24,16%	%
Coverage in 45 min	52,98%	%
Weighted coverage in 45 min	57,96%	%
Averaged once served 30 min	61,55%	%
Total once served 30 min	69,00%	%
Averaged once served 45 min	81,82%	%
Total once served 45 min	88,27%	%