EPA1316A
INTRODUCTION TO DATA SCIENCE

Managing Congestion and Traffic Volume for a Cleaner Urban Transport System

PROJECT-01

November 10th, 2022

Authors:
Paul Ludwig Branzk
Alexander Dietz
Ryan Tan Yi Wei
Canan Tezgec

TU Delft

Table of Contents

List of Figures

List of Tables

Abstract

With the ongoing urbanization and globalization we see a sharp increase in the population of cities and need for mobility. These phenomena increase the demand of transportation and emphasize the need for policy-makers to adapt for a sustainable mobility transition.
Since individual means of transport , such as light-duty vehicles, are main drivers of the overall transport emission [Enzmann and Ringel, 2020], cities should centre their policies on managing their environmental impact. We approach this from two specific areas - congestion and volume which were identify are key contributors to emission problem. As such, our research questions are:

A For what commuting purposes do citizens use personal vehicles for, and how can we reduce this dependency with public transport?

B How can we control the surge in street occupancy during peak hours?

C Do typologies of cities hint to us the emerging difficulties in reducing the dependencies on cars?

We use the UTD19 dataset of ETH Zurich, city typology classification, and distances to public transportation and amenity data which we collected to address these questions.
By doing an elaborate analysis on a city and inter-city level we conclude that to address the surge in occupancy, it is necessary to "flatten the curve" and promote flexible and off-peak time travel. On the question of car-dependency, we show that the distance to public transportation is not the only factor influencing the decision on the mode of transportation. In this context, we provide a theoretical approach on the patterns of certain car-dependent behaviour and further suggest how policy-makers can influence their decisions with tailored policy recommendations.
We also compare these patterns between multiple cities to draw deeper insights into how city typology could have played a part in inducing such commuting behaviour.
These findings illustrate the wickedness of a sustainable mobility transformation and should spark the interest of policy-makers and other researchers to investigate these topics.

1 Introduction

With 4.4 billion inhabitants, urban areas have become hubs of growth, increased productivity and innovation [United Nations Habitat, 2022]. During the next decade speed and scale of urbanization will intensify so that by 2050 around 68% of the world's population will live in cities [United Nations Habitat, 2022, WBG, 2021]. In addition, it is expected that driven passenger kilometres will double to 100 billion by 2050 while at the same time motor vehicle ownership will grow to 2.5 billion [Oke et al., 2019].

As a result, it is estimated that without mitigation efforts, CO2 emissions increase by 60% within the next 30 years [ITF, 2017]. As urban systems already use two-thirds of the world's energy and cause 70% of today's CO2 emissions, most of which come from industrial and motorized transport systems [ITF, 2017, WBG, 2021], massive decarbonization of cities is recognized as a key driver to tackle climate change [WBG, 2021] and hence, a focus of policy-making.

In order to meet rapid urbanization, increasing demand for mobility, and decarbonization goals, cities have to focus on promoting a sustainable mobility transition. Since personal means of transport such as light-duty vehicles are main drivers of the overall transport emission [Enzmann and Ringel, 2020], cities should centre their policies on reducing it.

However, the approaches each cities should undertake will be different since they each are constrained by their respective physical forms and transportation culture. When promoting policies of sustainable mobility transitions, one has to be cautious of generalizing solutions. [Oke et al., 2019] attempts to address this issue of non-generalizability by classifying cities into 12 city typologies based on urban dimensions and indicators perceived to be core to the mobility system of the city. However, they have only used this classification in comparison to city-level environmental outcomes which is largely descriptive but is limited in prescribing a call to action.

In this context, we aim to contribute to the debate on three aspects by addressing certain knowledge gaps:

Firstly, in transitioning to a cleaner city, public transportation is the key alternative mode of transport to ensure the efficiency and scalability of the city's mobility system. Yet, a large percentage of individuals still is in need of personal vehicular transportation or actively decides against public transportation. Hence, we ask:

For what commuting purposes do citizens use personal vehicles for, and how can we reduce this dependency with public transport?

Secondly, we investigate the reasons for congestion for the urban environment. Congestion and traffic volume are not necessarily dependent on each other, where overall traffic volume can be lower but rate of congestion remains. It largely boils down to how the city manages peak hour demands where congestion is the most prevalent. Hence we ask:

How can we control the surge in street occupancy during peak hours?

Finally, with much urgency in working towards a zero-emission future, any means to compare and generalize findings across relevant cities can help in knowledge sharing of solutions. We extend the work done by [Oke et al., 2019] and seek ways to use these typologies to understand complex traffic patterns occurring on the street level. Hence, we ask:

Do typologies of cities hint to us the emerging difficulties in reducing the dependencies on cars?

We explore these three research questions on five cities that represent different city typologies and hence, different characteristics. Los Angeles and Toronto exemplify the typology of Auto Innovative cities. Zurich, as a reference city accounts for the MassTransit Moderate and Paris and Hamburg represent the Typology of MassTransit Heavyweight.

The analysis is based on traffic flow and occupancy data collected from detectors across the selected cities. We explore this data numerically and spatially and provide an unsupervised learning approach to gain new insights and patterns in the data.

2 Related Work

Various studies have been carried out to reduce emissions in the urban transport system. The following section touches on the most important ones that are relevant to our project.

2.1 Key Factors affecting Traffic Emissions

Two key factors contributing to traffic emissions have been identified and will be introduced in the following section.

Traffic Congestion

Every vehicle commute involves the following movement - idling, accelerating, cruising and decelerating. The level of congestion plays a significant role in determining the proportion of the trip spent on the various stages- which in turn describes the total efficiency of the overall trip [Barth and Boriboonsomsin, 2008]. If the average traffic speed is below a certain threshold, typically around 25mph, the CO2 emissions will increase exponentially. There is a need to prevent traffic from reaching the far left of the emission-speed curve.

The dangers of congestion is two fold - 1) the increased emissions, 2) the localized pollutants on streets like particulate matter, which poses a major health risk.

Cities' failure to adapt to rapid urbanization and the increasing demand for mobility has resulted in urban traffic systems being congested [Ng et al., 2013]. It would be necessary for cities to ensure that this surge in traffic demand is well spread to relief pressure on the transport infrastructure [Choi and Toh, 2010]. Traffic demand are extremely time-sensitive due to the nature of certain trips. For instance, morning peak hours are characterized by the rush of students and office workers commuting to schools and workplaces. As such, studying the time of travel will be important in understanding commuting behaviour and its direct impact to transport emissions.

Traffic Volume

Even with perfect road traffic efficiency, the emissions is still largely reliant on the effective demand or the dependence of vehicular transport modes. According to a report from the European Environment Agency [EEA, 2022], road transportation is responsible for 71.7 percentage of the EU's total transport greenhouse gas emissions. Cars are at the top of the road transport emission, which constitutes 60.6 percentage of the emission. Globally, transport accounts for approximately 23 percentage of overall energy-related carbon dioxide emissions, and road transportation is responsible for 75 percentage of the transportation emission [IEA, 2017]. The EU and global data both demonstrate that road transportation, primarily cars for personal commutes, is mainly responsible for emissions.

2.2 Traffic Modelling and Emissions Analyses

Existing models have been created to study emission from traffic flow pattern: [Mensink et al., 2000] combined the urban traffic flow model with more detailed metrics relating to degree of emissions such as vehicle type, fuel type, vehicle age, trip length and function of road type - to predict the hourly emissions of CO, NOx, VOC, PM, SO2 and Pb for each street in Antwerp.

[Qin and Chan, 1993] studied the extent in which street-level air pollution was caused by traffic sources in Guangzhou, China. It has considered diurnal traffic volumes from different vehicle types, vehicle speeds and idle conditions of motor vehicle emissions as key variables to understand how it contributes to the different types of emissions. Findings show that NOx play a significant percentage in pollution and more attention should be paid to controlling these emission sources.

2.3 Traffic Flow Analysis

[Tsuboi, 2021] provides a theoretical framework to understand traffic data typically recorded by cameras or detectors. This aids us to fully utilize the data set and interpret our findings better.

- **Flow** is defined by the vehicles per hours over the detector. This metric can be useful in gauging traffic volume, where it provides an objective quantity of vehicles used.
- **Occupancy** is defined by the percentage of time that a vehicle over the detector. This metric is highly representative of congestion where it tells you the extent in which a road is used.
- **Average Speed** is defined by aggregating the speeds of vehicles within the the recording interval. This metric can also be used to identify congestion when coupled with traffic flow. Practically, this metric will be useful for planners to design speed limits.

With our research question focusing on traffic emission, traffic volume and congestion would be important behaviours to analyze - which traffic flow and occupancy are effective for. Average speed would be overlooked as occupancy is a more representative metric of congestion.

2.4 Commuting Behaviours and Mode Share

Commuting behaviours are rather distance sensitive and they are useful in explaining why certain modes of transport are preferred over others for different trips.

Walking and Pedestrian behaviours.

In transport planning, it is a common rule of thumb that a 5 minute walk or 400m is considered to be the distance people are willing to walk before opting to drive [mor, 2019].

Willing distances depend on purpose of trips.

The willing duration for walking trips also depends on their purpose. Trips for shopping, errands and reaching transportation are shorter, while recreational walks tend to be longer [NHTSA, 2008]. On average, the commute to school and work are naturally much longer because of the lack of flexibility of where these institutions may be located.

Importance of access to public transport.

This behaviour influences the willingness to use of public transport as residents seek a reasonable travel distance to transit facilities. The willing distance to access a public transport also depends on the length of the entire trip - as such bus stops are estimated to be 400m while train stations are 800m [El-Geneidy et al., 2014]. This implies the importance of local transit accessibility or ”last-mile” connections which is key to shift commuting behaviours [Choi and Toh, 2010].

Spatial Configurations in influencing commuting behaviour.

The spatial configuration of neighbourhoods can be influential in convincing residents to walk, cycle and use public transport options. The 20-minute neighbourhood [DELWP, 2019] is an example of such efforts, where planners focus on improving local accessibility to amenities and services to reduce the need to depend on vehicle usage.

Table 1: This table summarizes the OSM Classification based on [Zhang et al., 2018, Guth et al., 2021] and comprises five street levels (A1-A5). For each level, a concise description is provided.

Level	OSM Classification	Description
A1	motorway, trunk	free-ways, such as motorways and trunk ways, usually with limited access
A2	primary	important roads that often link large towns or main road within cities
A3	secondary	forming a link in the national route network
A4	tertiary	connect smaller settlements and minor streets to major roads
A5	residential, living streets	access roads to housing, without function of connecting settlements

2.5 Functional Classification System of Roads

Roads in cities provide various functions in the hierarchy of movements [Forbes, 2000]. The two main functions are access and mobility which are conflicting in nature.

- Roads with high mobility but low accessibility (such as motorways) are characterized with few entries and exits as they primarily support efficient long distance vehicular travel.
- Roads with high accessibility but low mobility (such as local roads) have entries and exits due to its purpose for connecting to residences and amenities.
- Roads in the intersection between accessibility and mobility (such as secondary roads) provide a smooth transitions between the above types.

Exploring traffic in relation to these functions can be useful in identifying commuting behaviours. For instance, high motorway traffic during peak hours could imply reliance on cars for work commutes. We seek a simpler classification system; the Traffic Flow data set (UTD) uses OSM's functional road class classification [OpenStreetMap contributors, 2017], which have too many labels to compare road functions and scale of mobility effectively. [Zhang et al., 2018, Guth et al., 2021] faced similar issues in their data analyses on OSM road networks and we draw inspiration from their classification methods:

3 Methodology

Our approach can be divided into five sub-steps illustrated in Figure 1. First, we scoped our problem with a literature review and identified important subtopics such as factors influencing pollution, traffic flow analysis, and general commuting behaviour. Doing so, helped us to formulate a research question and work on a holistic approach. Afterwards, we collected data from multiple sources that we used to address the different topics we identified in our literature review. We cleaned the data to enhance the quality of our analysis. We conducted an exploratory data analysis to get a better overview and understanding of the available data sources. This numerical and visual analysis helped us to further scope the problem, provide a first check of the feasibility of our assumptions, and prepare our final analysis. For our analysis, we first had a deeper detailed view of one specific city and analyzed its characteristics by using clustering, an unsupervised learning method. We used clustering to identify previously unknown patterns in the data.

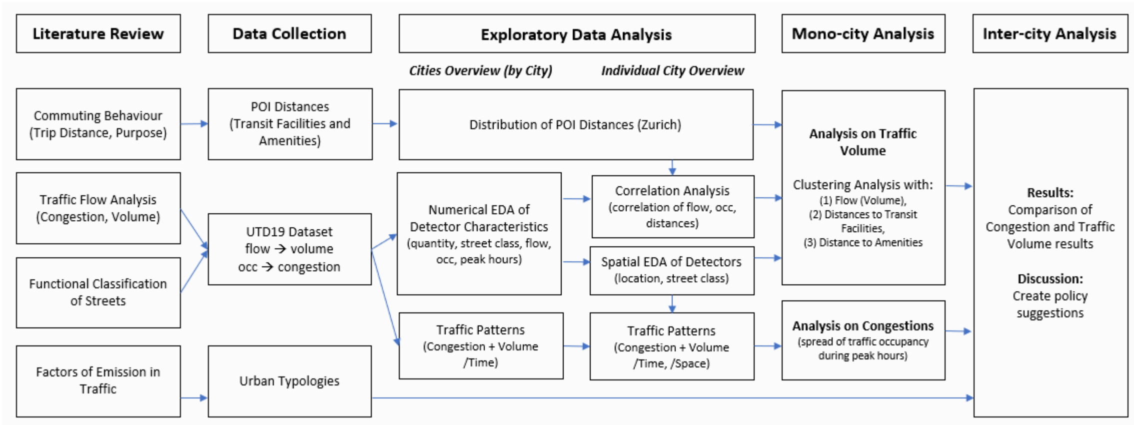

Figure 1: The diagram displays the methodological work-flow of this research project. It is divided into five different levels - Literature Review, Data Collection, EDA, Mono-city Analysis and Multi-city Analysis. Every level is based on the previous one.

4 Data Collection and Processing

In the following section, the applied data sets are introduced (see Table 2) which are utilized in this project.

4.1 UTD19

UTD19 is a large-scale traffic data set providing information from over 23541 stationary loop detectors covering 40 cities worldwide. It provides insights into vehicle flow, street occupancy, and speed, which are detected within relatively small time intervals of three to five minutes. Furthermore, it includes the speed limit, the road type, and the number of lanes for each measurement. Moreover, all detectors and the associated roads are geo-tagged in WGS84 coordinates.

To enhance the analysis of the dataset, we applied multiple data cleaning steps. First, we assigned every detector its respective functional classification level from one to five. Moreover, we dropped all measurements with errors. These can be identified by having a number in the column ”errors”. We checked the feasibility of values for the remaining rows by checking the location and day of remarkable measurements.

As UTD19 is a ”consolidated” data set, the location of detectors can be inconsistent between cities where they vary in proportions of road types. Furthermore, we saw that some streets have multiple detectors close to each other while other streets only have one or a few distributed across the length of the street. Moreover, we see that the timeframe and interval for each city is different which leads to a difference in the amount of available data. Also not all cities reported on all columns. Hamburg, for instance, is missing values for speed while Paris does not has data for the speed-limit. Hence, we could not use these columns for our analysis.

4.2 OSMnx

OSMnx [Boeing, 2017], a Python package, is another source of data where it queries geospatial data from OpenStreetMap APIs such as buildings, street networks, and boundaries for visu-

Table 2: The table summarizes the three data sets that were used within this project. Additionally, information about the source, included features, spatial level, number of cities and the year are provided.

Data Sets	Sources	Indicators	Spatial Level	Cities	Year
UTD19	ETH Zurich, (open access)	Vehicle flow, occupancy, speed, etc.	Street Level	40	2017 - 2019
City Typology	MIT, (open access)	CO2 emission, road deaths, etc.	City Level	331	2013 - 2017
OSMnx	OpenStreetMap API	Street Network, points of interests, etc	City and Street	All	Present

alization and analysis. We have choosen this package due to the fact that it is open source and provides a high reliability because of the community support. Moreover, it is highly flexible and thus, we were able to specify it for our visualization and analysis needs. In addition, it provides the opportunity to calculate distance between nodes in the network.

4.3 Distances to Points of Interest

We seek to understand from the traffic data of detectors the purpose of the trips and the public transport alternatives available to it - this could potentially be implied by distance to amenities and transit facilities respectively.

Proximity to amenities by Manhattan distance can reflect the probability of the street being use for the start and end of a journey to access that amenity. Convenience to transit facilities implies the potential of using public transport alternatives especially if it is within reasonable distances [NHTSA, 2008]. We have chosen bus, trams and trains stations as transit facilities; supermarket, school, work, shops and food as amenities to reflect 5 common yet distinct trip purposes [Choi and Toh, 2010].

Routing Algorithm

Distances from detectors to the above points of interests (POI) were generated separately by our algorithm which is largely made up of OSMnx's query and routing functions. Firstly, we queried the walking street network and POIs location data. Then, we establish the target and origin nodes on the street by finding the least euclidean distance from the detector and POI to the network respectively. Finally, we computed the shortest route through OSMnx's routing function and extracted the distances. Essentially, our POI Distance Metric would be the sum of the distance from POI/detector to street, and distance of shortest path- which is a more representative measurement of the urban experience. Our complete algorithm and its details can be found in the appendix - see appendix A. There are possible errors in this method. When establishing origin and target nodes, it was assumed that the direct path of least euclidean distance from a POI to the street to be the most ”natural way” to access the street network - however, in extreme cases, this leads to an unfavourable node like a motorway where it causes awkward routes which heavily overestimates the distance. This issue cannot be solved that easily algorithmically, so we will address these errors by filtering them out from our data set.

4.4 Urban Typologies

The Urban Typologies data set [Oke et al., 2019] provides an assortment of transport and social-economic indicators across cities. Altogether, 64 indicators from seven urban dimensions are included. We use this data to guide parts of our discussion and provide a more diverse range of potential points for analysis. Furthermore, we can see if our results align with the proposed typologies of the cities.

5 Exploratory Data Analysis

The exploratory data analysis is divided into two parts. First, a general overview on the selected cities and their measurements is provided. Later, Zurich - as our references city - is explored in more detail.

5.1 Cities Overview

This section provides an overview about the general characteristics of all selected cities.

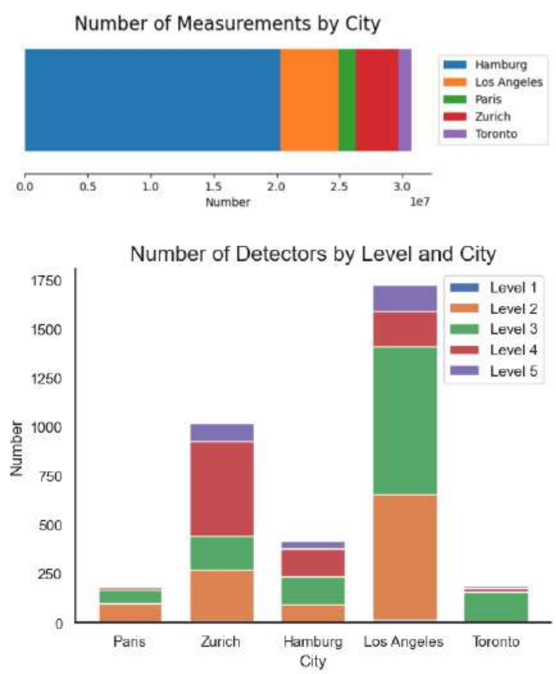

Figure 2: The Number of Measurements by City displays the distribution of data between the selected cities. The Number of Detectors by Level and City figure presents the distribution of detectors on street-level for each city.

5.1.1 Detector Characteristics

The number of detectors in relation and their respective levels, the amount of data and time-frames of measurement are discussed.

Analysis on Data Structure

The pre-processed data consists of 30.729.014 rows and 20 columns. Hamburg, with 20.280.386 rows contains the most data. Los Angeles includes 4.611.100 rows , while Paris has 1.375.866 rows and Zurich comprises 3.419.332 rows. Toronto, with 1.042.330 rows, has the lowest amount of data. Each row refers to one measurement in time, while columns contain information about the traffic such as flow, occupancy, intervals, street-level or city-code.

Distribution of Detectors

The total number of detectors varies between 188 and 1722 detectors. Toronto has the lowest number of detectors, directly by Paris. Los Angeles, which is the upper limit, and Zurich have more detectors compared to the other three cities. While Zurich and Los Angeles have detectors at all levels, Toronto's detector level is limited to level three and Paris is represented mostly by level two and three. In general, the street-level distribution shows that level three, with a total of 1294 detectors, comprises most detectors. In contrast, level one has the lowest with 21 detectors. In context to urban street networks that makes sense since free-ways are less likely to occur due to high structural density.

Measurement Time-frame

The time range for the data collection is dependent on the data collectors' approach. The selected five cities have different data collection periods- see Table 3. In Paris, the data was collected for approximately one year, while in Zurich and Los Angeles, the range dropped to seven and six days, respectively. Besides, the frequency of data collection within the day changes across the cities. The data for Paris is collected each hour, while Hamburg and Los Angeles get the data every 3 minutes. Even though the number of detectors and the collection periods change considerably among the selected cities, all of them have sufficient data to conduct a further analysis.

Table 3: The table summarises the periods of data collection for each city. It includes the number of days, the specific time of the year and the intervals at which the data are measured.

City	Number of Days	First Day	Last Day	Interval
Paris	336	2016-01-01	2016-12-01	60 mins
Zurich	7	2015-10-26	2015-11-01	5 mins
Hamburg	105	2016-08-27	2016-12-09	3 mins
Los Angeles	6	2017-10-02	2017-10-10	3 mins
Toronto	61	2016-09-01	2017-01-31	15 mins

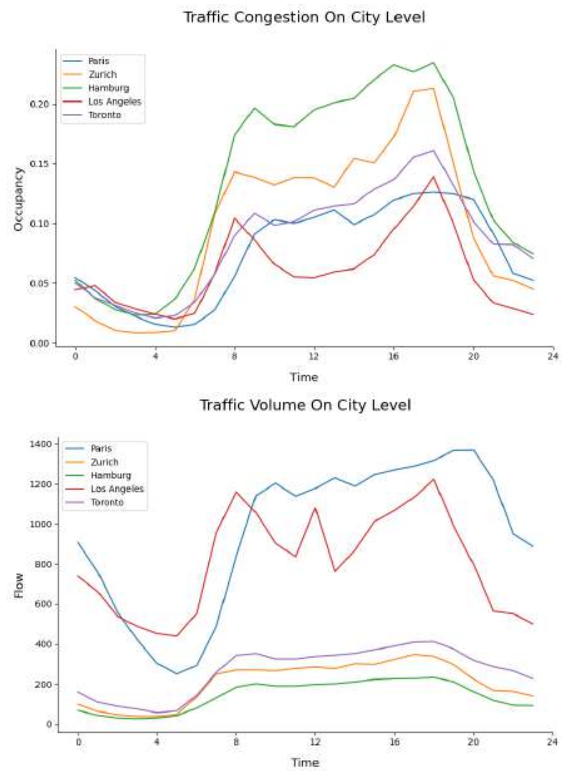

Figure 3: Traffic Congestion On City Level displays the average street occupation in relation to the time. Traffic Volume On City Level illustrates the average traffic flow for each city. Each city is marked by a different colour.

5.1.2 Traffic Patterns of Cities by Time

Time has an all-encompassing influence on our societies. To ensure good compatibility between cities, interval boundaries were synchronized to one-hour intervals.

Traffic Congestion

We see that all cities show a similar rapid increase at 5am and a beginning decrease at 8pm in congestion. Thus, a clear distinction between day and night is observable within all cities. During the day, Los Angeles has a very distinct curve with reoccurring morning and evening peaks, indicating commute travels. In contrast, the other cities do not have a clear morning peak. Instead the occupancy steadily increases during the day until it peaks at the evening. Hamburg, constantly displaying occupancy over 0.15 during day-time possess the highest traffic congestion. While Zurichs' evening peak differs from the daily average, Paris curve is more flat.

Traffic Volume

We see that Los Angeles and Paris have generally speaking a higher traffic volume than the other three cities. It is noticeable that the traffic flow falls for each city between 12pm and 4am. The peak hours are between 5 and 8pm. Los Angeles also clearly shows the morning rush hour at 8am and a lunch rush hour at 12am.

Congestion and Volume

By comparing both graphs, the following relations are discovered. Los Angeles and Paris with the highest traffic flow display the lowest occupancy rates. This is reasonable since flow and occupancy are related to each other in the way that high occupancy indicates congested traffic and hence low flow. The two diagrams - apart from Los Angeles - visualise this relationship as they have a similar shape and inverted hierarchy. The difference within Los Angeles indicates that during the morning and evening peak the road network is functioning at its capacity limit. In contrast, the increased traffic volume due to the lunch rush hour is handled without any increased congestion.

Congestion Response During Peak Hour

The way in which all cities respond to the peak hour is also different. As seen from Figure 4, the rate in which traffic demand falls after peak hour is the same, but the rate in which traffic demand rises before the peak hour is different. Cities like Paris and Hamburg has allowed for a slower and flatter escalation of traffic demand as compared to Toronto or Zurich. This is ideal as it implies traffic has been more effectively distributed to prevent congestion.

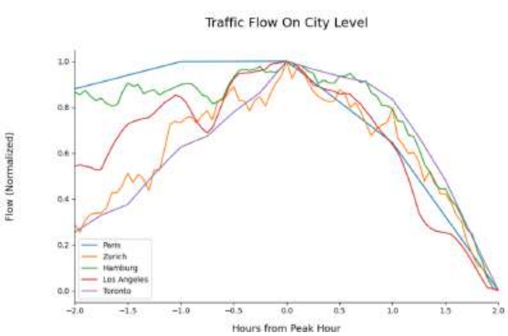

Figure 4: The Graph displays the normalizes occupancy data, centered in respect to each cities peak +- two hours. Time intervals are not synchronized

5.2 In-depth Review of Zurich

Within this section, Zurich as our reference city is analyzed. First, the current urban environment and road architecture is put into context. Afterwards, we analyze the spatial distribution of detectors, the correlation between distances to amenities and traffic, the distribution of POIs and traffic patterns in relation to time and space.

5.2.1 Spatial EDA - Detectors in Zurich

Spatial Context

Geographically, Zurich is located in north-central Switzerland. From an infrastructure perspective, two motorways are bypassing Zurich. In the north, it's the west-east connecting A1 and in the west the A3 going to the south. Between 1967 to 1974 planners intended to connect those motorways in the centre of Zurich. Therefore, the A3-west and A1-east leading traffic into Zurich were built. Although the project was never implemented, Zurich's infrastructure network of today is influenced heavily by the A1E from the north and the A3W from the South. Nowadays, a level two motorway is connecting both losing ends, bypassing the city centre through the Industriequartier via the Hardbruecke.

Due to the low scale, information about the exact location of every detector is limited. Hence, the map provides a more general overview of how data was measured within the environmental context of Zurich.

Figure 5 displays a high detector density within the city centre of Zurich. Thus, indicating that most data is aggregated within the centre of Zurich. In addition, possible points of interest within the infrastructure system, such as feeder roads, bridges or traffic junctions, appear to have been selected intentionally rather than randomly. For more specific insights, each level is inspected separately.

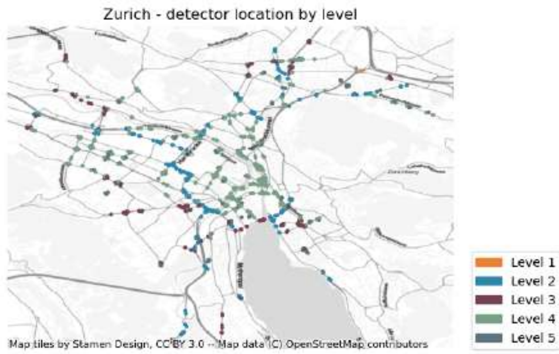

Figure 5: The map displays the exact location of all 1022 detectors that were left after the UTD19 data processing within Zurich. Colours indicate the classified street level.

Level 1 contains four detectors that measure traffic of the A1E. The motorway connects the centre of Zurich with the A1. Within Zurich, traffic is directed through the Milchbucktunnel. Detectors were located at the entry of Milchbucktunnel and at the junction of A1E to A1. At all possible exits, level two and level four detectors registered vehicles leaving the motorway.

Level 2 comprises a total of 264 level two detectors, mainly monitoring the motorway A3W and its connecting bypassing street A3/A4 towards A1E. The number of detectors is much higher than level one since more junctions exist.

Level 3 has a total of 171 level-three detectors. Detectors are fairly well distributed on several different streets. Nevertheless, three points of interest are located. Kreuzplatz in the southeast, Goldbrunnenplatz in the west and the ZSC-Lion Platz in the north. With higher street levels more streets have to be considered. To keep resources reasonably low, data collection was probably scoped to these three areas of interest.

Level 4 has the most detectors with a total number or 484. Instead of focusing on specific points of interest, the number of detectors was increased and the focus shifted to the city centre of Zurich.

Level 5 detectors measured traffic at 97 different locations which are sprawled over the whole of Zurich. Mostly detectors were placed within the vicinity of level four detectors.

The placement of detectors is influenced by the environmental properties as they determine the traffic infrastructure such as street networks. Therefore, affects the total number, as well as the density of detectors. As a result, the data collection and hence the data itself is limited. When comparing cities, this has to be acknowledged.

5.2.2 Correlation Analysis

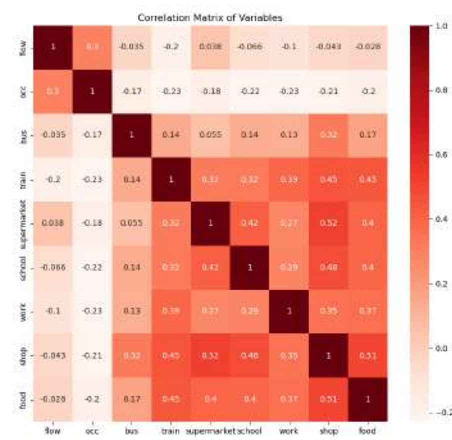

Figure 6: The graph displays the average street occupation of Zurich in relation to the time. Each street level is marked by a different colour.

The correlation between the flow, occupancy, transit facilities, and amenities can be seen in Figure 6. There is a high correlation between the transit facilities and the amenities. We can interpret this high correlation in two ways: either amenities choose their locations according to the public transportation, or municipalities locate the public transit close to the amenities. Besides, there is a high correlation between the amenities. For instance, the correlation between the distance to the shops and the distance to the supermarket is 0.52. On the other hand, occupancy and flow have a low correlation with all the other variables. Neither occupancy nor flow can be explained by these variables linearly. Hence, the linear regression would not work in explaining flow and occupancy due to the weakness of the relationship.

5.2.3 Distribution of Point of Interests in Zurich

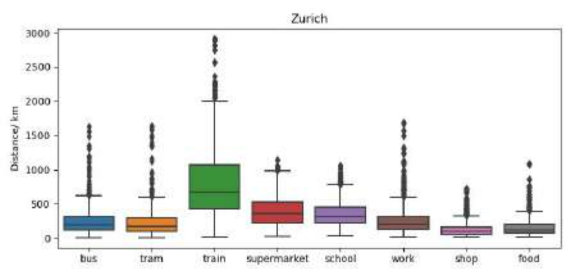

Figure 7: The boxplot displays how the distances are distributed for every POI within Zurich. Each boxplot accounts for one POI. The distance is displayed in km

The distances to POIs provides suggest certain patterns of how the city may be spatially configurated [Hickman et al., 2013]. Overall, we see rather low median values to both transit facilities and amenties, however, there is a significantly large number of outliers - see figure 14. This could be due to Zurich being a fairly large city with a dense city centre, as such, have an amount isolated areas outside the city centre.

Distances to transit facilities are low, with 75% of detectors within 600m of a bus or tram. However, trains are on average much higher, with a fair number of streets longer than 2km. This could show that transportation within the city is more demanded as compared to inter-city connections, since general accessibility to bus and trams are much higher than trains.

Distances to amenities are fair, where most are within 1km radius of any detector. It is lower than auto-heavy cities like Los Angeles, but is higher than mass-transit heavyweight cities like Hamburg.

5.2.4 Traffic Patterns of Zurich by Time and Location

The following section explores traffic patterns on different street-levels in relation to time and space.

Traffic Congestion

In section 5.1.2 the following general traffic patterns have been identified.

- Low traffic congestion during the night
- Rapid increase of traffic congestion within morning hours (5am - 8am)
- Steady increase of traffic congestion during the day
- Strong congestion peak during evening rush hour

Scoping traffic patterns down onto street-level, similar patterns can be observed. Level four has the highest degree of congestion, while level one has the lowest. During the night all levels show a low degree of traffic congestion. Additionally, level two, three and four display an identical increase during the morning hours. In contrast, the increase in congestion is lower at levels one and five. It is worth noting that level four congestion never decreases during the day, opposed to level one or three. As level four detectors were focused on traffic within the city centre, we conclude that the city centre becomes increasingly congested throughout the day.

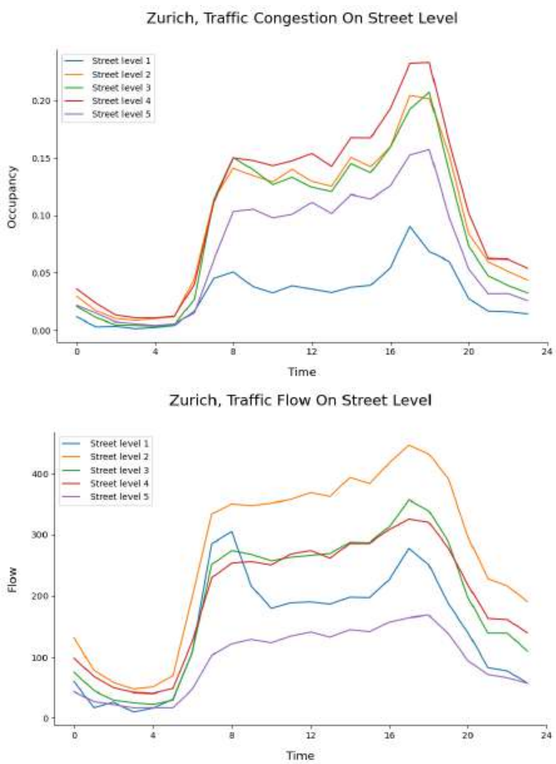

Figure 8: Zurich, Traffic Congestion On Street Level displays the average street occupation in relation to the time for each street level. Zurich, Traffic Volume On City Level illustrates the average traffic flow in relation to the time for each street-level. Each street level is marked by a different colour.

Traffic Flow

We see that level five has the lowest traffic volume, while level two has the highest. All curves, except level one, display a similar character. In the morning (5am - 7am) a rapid increase is identified. During the day, traffic volume stays consistently high. Towards the evening, traffic volume increases until it peaks between 5pm and 6pm. Afterwards it drops. In contrast, level one shows a clear morning and evening peak, with a relaxation in between. These peaks indicate commute travel. Hence, we conclude that level one is used as a feeder road guiding commute travelers in and out of the Zurich.

Volume and Congestion

The general relations from section 5.1.2 between traffic congestion and volume differ within the context of Zurich. For instance, we see that level one, while having a strong traffic volume increase in the morning, displays a comparable low occupancy peak. We assume that it is due to Zurich's' well developed highway system.

We see that the detectors with the highest occupancy during peak hour are found in the city centre, especially around the central station. The lowest occupancy during peak hour can be seen in the western parts of the city. This is due to the fact that these are mostly residential areas. Furthermore, it is possible to identify streets such as highways. For example, we see a hotspot in the north which is the intersection of multiple highways and intercity roads. Moreover, it is also the way to the Zurich International Airport and home to the Zurich Oerlikon train station. Oerlikon is the seventh largest train station of Switzerland and can also be seen as a bottleneck of the Swiss train system as almost all trains towards the north must pass it. Hence, it is an area with a high number of commuters and transit trips leading to a high occupancy during rush hour in the late afternoon.

6 Results

We have divided our analysis into two parts to first gain insights on one individual city, Zurich, and then interpolate the gained assumptions to multiple cities and make a comparison between all five cities. For that we used cleaned datasets and applied the k-means clustering method. K-means is useful for a large number of observations with medium number of clusters. In general, clustering provides the benefit of simplifying large datasets which will help our analysis as we have measurements in a high frequency from a large number of detectors and timeframes of multiple days. With clusters we can make more general statements about streets or even neighborhoods of cities. Furthermore, we can detect previously unknown pattern in the street flow of a city.

6.1 Mono-City Analysis

To gain more insights, we first analyze a single city and illustrate the relationships between flow and distances to public transportation and amenities. In order to run a k-means analysis, we need to obtain the optimal k. We will use the elbow technique to do so.

We can identify certain spatial conglomeration of detectors to clusters. We see also how clusters partly align with the street levels as cluster five follows the motorway connecting the highways A1E and A3W. Cluster zero tends to show areas with a high availability of amenities and public transportation. This also shows that in Zurich, amenities and public transportation are close to each other. Furthermore, we see spatial conglomerations of clusters. In particular interests are clusters that have cluster hotspots as these are areas that might be spatially separated but share common characteristics. This would be relevant when implementing policies as one policy that works well in one hotspot should see an increased chance of success in another hotspot of the same cluster. In our analysis, we see this for example for Cluster one (orange) which can be found in the middle left and top right of Figure 12.

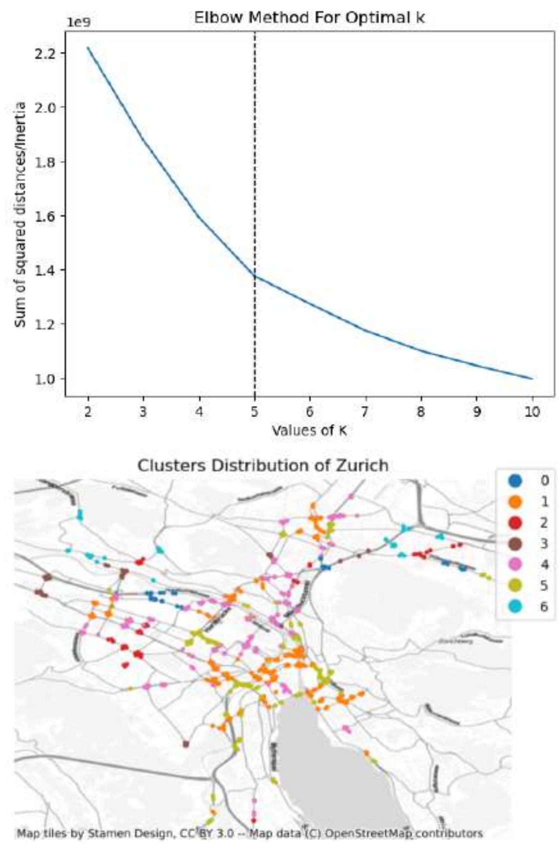

Figure 12: Elbow Method Zurich displays the optimal k for k-means clustering. It shows the explained variation as a function of the number of clusters.
The cluster distribution map of Zurich displays the generated clusters using k-means clustering (k=7). Every color indicates a different cluster.

We expand these insights with a Kernel Density Estimation (KDE) graph of the identified clusters. This step also helps to evaluate the quality of our clustering model. In an ideal scenario, each cluster would be separated from each

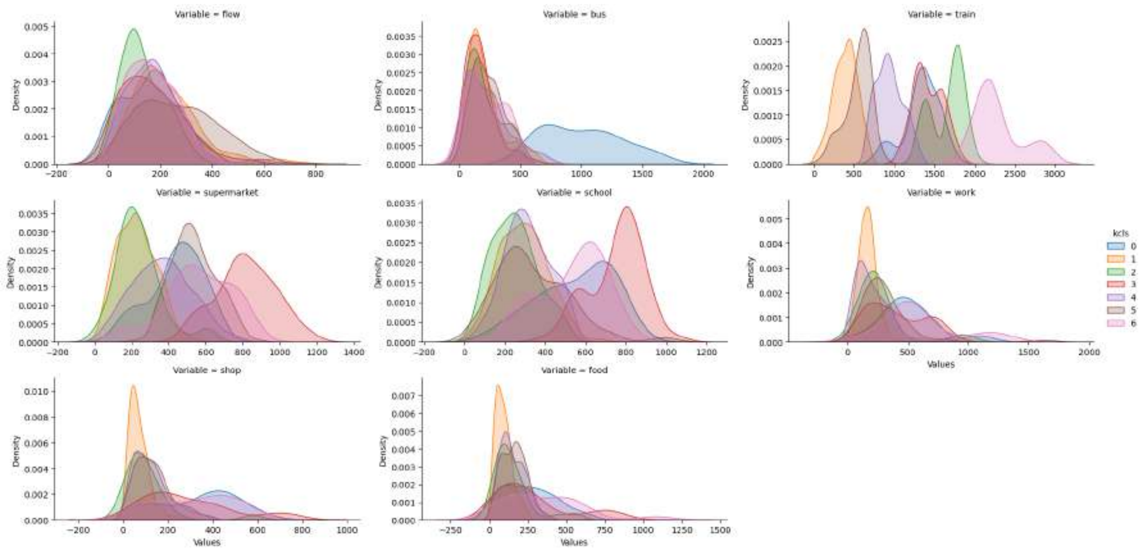

Figure 11: KDE-Plot for Zurich clusters.

other and there would be only a minimal overlap between clusters.

We see that the overlap between the clusters is different for each variable. This means that the values are not largely separated from each other and hence, difficult to cluster. We see the best cluster results for train stations.

Cluster 2 has the lowest traffic flow in the KDE plot, which have low distances to bus stations and amenities. This could imply effective planning where buses have serve that area rather well. Upon inspection of their location on the map, these areas are well connected neighbourhoods on the fringes of the city centre.

Cluster 5 has the highest traffic flow in the KDE plot. It is further observed that the detectors are quite close to transit facilities as well as amenities, with the exception of supermarkets but is still average as compared to the other clusters. This illustrates an interesting traffic problem - some areas that functions to access amenities have good public transport alternatives but are likely underused and is car heavy. Examining the detector locations on the map, these areas are streets into the city centre. This could allude that there could be a proportion of residents who prefer to enter the city centre through cars.

6.2 Inter-City Analysis

In the following, we want to compare our findings of Zurich with other cities. In particular, we analyze how different typologies affect the relationship between flow and distance to public transportation, and distance to amenities. To provide a holistic picture we will analyze two *Auto Innovation* (Los Angeles, Toronto), two *Masstransit Heavyweight* cities (Paris, Hamburg) and one *Masstransit Medium* city(Zurich).

It would be difficult to compare the exact numerical quantities within the clusters, hence, 13 will summarize the findings from our analysis. It focuses two or three (depending on total number of clusters) clusters with the highest traffic volume and display the corresponding distance metrics with labels of low/mid/high for succinctness. It presents a quick overview of the potential pub-

City	Cluster	Bus	Train	Spmkt	School	Work	Shop	Food	Type
Hamburg	5	High	Mid	Mid	Mid	High	Mid(v)	High	1
	2	Low	Low	High	Mid	Mid	Mid(v)	Mid	1
	4	Low	High	Low	High	Low	Low	Low	2
LA	0	Low	Mid	Low	Low	Low	Low	Low	3
	3	Low	High(v)	High	Low	Low	Mid(v)	Mid(v)	2
Paris	3	High	High	High	High	High	High	High(v)	1
	1	Low	Low	Low	Low	Low	Low	Low	3
Toronto	5	Low	Low	Mid	Low	High	High	High	3
	2	High	Mid	High	High	High	High	Mid	1
Zurich	5	Low	Low	Mid	Low	Low	Low	Low	3
	1	Low	Low	Low	Low	Low	Low	Low	3
	4	Low	Mid	Mid	Low	Low	Low	Low	2

Figure 13: Table displays the clusters of detectors of each city by the highest traffic volume, and its corresponding distance metrics. Low/Mid/High was qualitative assessed, (v) was placed if metric had high variance

lic transport alternatives and general commute purposes for each cluster.

Analyzing patterns between commuting purpose and accessibility of public transport could allow to understand why these car-dependent trips exist. We generalize three possible theories based on the possible combinations of yes/no alternatives and short/long commute purposes:

1. **"High quantities of long distance travel"**. The high distances to all amenities and public transport implies the road serves the purpose of long distance travel/commutes e.g.logistics or inter-city work commutes, where public transport is unlikely to be an easy alternative. An example would be Paris Cluster 5.

2. **"Public transport type does not fit purpose of commute"**. The proximity to an amenity imply a specific commuting purpose, however, the typical "distance" of the commute is unlikely to be served by the public transport available or accessible. An example would be Los Angeles Cluster 2, where it is likely used for school and work commutes, but trains, which are quite apt for such distance of commutes, are not within reasonable distances.

3. **"Public transport is available but not preferred"**. Proximity to amenities and public transport is good; all commute purposes have the potential to be served by public transport infrastructure, but residents do not utilize them for reasons including lifestyle preferences or inadequacies of public transport. An example would be Zurich Clusters 5.

This classification of problem theories is also labelled on Figure 13.

How does city typologies influence this theory?

We draw similarities in problem theories between cities with similar typologies (Auto-Innovative, Mass Transit) and reflect on the problems they face in managing traffic volume.

- **Auto-Innovative Cities** i.e. Toronto, Los Angeles. These cities face the theory of "(3) Public transport is available but not preferred". This is expected of American and Canadian cities which have a history of car-dependencies; and this likewise reflected in the spatial configuration of their cities. Even with great government efforts to reduce car use [Peters, 2019], the cultural lock-in inhibits progress into a car-lite city. Public transport in both cities are widely accessible, but residents still choose to take cars; policies might be ready, but people are slow to make this shift. This points to "softer" measures that help residents adapt better.

- **Mass Transit Heavyweight** i.e. Hamburg, Paris. These cities face the theory of "(1) High quantities of long distance travel". A possible reason for this is the geographical characteristics of Hamburg and Paris where they are generally big and dense cities where it is common for people to stay outside the city and enter the city every day for work and school. This presents a potential paradox that even though bus/train systems are intensive and sufficient, people still use vehicles for long distance trips. This points to possible reasons such as the lack of robustness of existing public transport infrastructure or complexities in resident's commuting patterns, which might not been adequately account for that have not been adequately accounted for.

- **Mass Transit Moderate** i.e. Zurich. At present, it is hard to identify problem theories that are unique to Mass Transit Moderates as there is no other city to make that comparison to. However, when comparing with Mass Transit Heavyweights, Mass Transit Moderates do not share the problem theory (1), which could conservatively hint that structural measures might not be a solution to deal with long distance travels.

7 Conclusion and Discussion

With this research we have addressed three research questions.

Firstly, we deep dived into what commuting purposes do citizens use personal vehicles for, and we can reduce them. We show that three theories of commuting exist that put the purpose and the use of personal transportation in relation. There are long distance travelers, who cannot use public transportation as an alternative. Moreover, we identified a group that has public transportation available but they require a different kind of public transportation. Lastly, there is the group that have good access to public transportation but actively decides to not use it.

Secondly, we addressed studied the patterns of street occupancy during peak hours by cities. We compare each city and observed similarities and differences in ways each city's traffic demand escalates and dissolves. We also saw this congestion problem to be independent of the above issue of traffic volume and believe that distributing demands around the peak hour can be a low-hanging fruit that any city can employ.

Thirdly, we questioned how typologies of cities hint the emerging difficulties in reducing the dependencies on cars. We show that Auto Innovative cities tend to struggle with theory (3), the group that has access to public transportation but actively does not decide to use it. For Mass Transit Heavyweight we see that the biggest issue is theory (1) the high quantity of long distance travel. For Mass Transit Moderate, we conclude that even as it is still Mass Transit-oriented, they do not share theory (1) with the heavyweights which imply how intensifying public transport further for public transport-intensive cities might not be useful.

7.1 Traffic Congestion

Through our exhaustive exploratory analysis we have shown that all cities - except Paris - are somewhat sensitive to high traffic volume. Therefore, they fall victim to reoccurring traffic congestion. Although the data is limited in regard to provide an answer for its reason we assume that morning and evening peaks considerably increased due to commute travel. Therefore, peak hours are unavoidable due to time sensitive activities of daily commutes. However, the way in which each city experience this surge in travel demand varies and is reflected by how fast occupancy escalates and falls in our peak hour analysis. This difference could be influenced by factors such as work culture or even policies, that allows for a better "spread".

As traffic congestion describes the state in which the transportation system is pushed to its capacity limits, cities should aim to distribute traffic demands especially for individual transportation around peak hours. This way, stress for the system is reduced and thus the level of congestion. Due to the relation between congestion and emissions, cities would be able to decrease emissions without reducing the the actual number of rides. Additionally, no changes within the actually infrastructure system has to be made. This makes this approach especially sustainable.

Hence, we recommend policy-makers to promote a more flexible use of the transport system. For instance, they could introduce discounted tickets for off-peak hours as it is already done in some Dutch cities, like The Hague. Furthermore, we see an increased number of people working from home who thus, do not have to commute to their work place. Policy-makers should support this phenomena and contribute to more flexible work laws that open more opportunities for home office.

7.2 Traffic Volume

We have shown that there are three theories that describe the flow of commuters across a city and can explain the behavior we have found in the data.

Firstly, there are long distance travellers that use the city's infrastructure primarily to pass through or for a long distance work commute. Another part of this group are trucks and supply chain vehicles that deliver goods to the city. This traffic volume is hardly replaceable with public transportation as it rather serves a more unique purpose and hence, requires a special setting that is currently only possible with the flexibility of individual transport. Still, policy-makers should tackle this group of traffic flow to lower emissions overall. For instance, policy-makers can decrease the emission of traffic volume by incentivizing more emission-efficient vehicles, such as cars powered by electricity or hydrogen. Furthermore, the potential of off-peak time delivery could be investigated. The idea is that supply chain vehicles, which tend to have higher emissions, should partly only be allowed during off-peak times as this will decrease the chance for them to be in a congestion. Vehicles in a congestion are a main contributor to emissions.

Secondly, we argue for a group of commuters that has a mode of public transportation nearby but it does not necessarily fit the purpose of travel. For instance, there might be bus stops, yet, the commuters need to access a distant business district of the city which is not or hardly accessible with solely the bus network. Here, regional public transportation modes such as trains are necessary. This implies that policy-makers need to be aware of the connection between different modes of public transportation and also the individual journey route of the commuter. We recommend to identify key routes between neighborhoods to optimize the potential of a bus and train network. Doing so, will minimize the chance of taking the car for this journey and thus, lower the traffic volume and consequently the emissions.

Thirdly, we described a group of commuters that has all modes of public transportation closeby, yet, still decides not to use them. Hence, we argue that for this group the distance to public transportation is not the only factor influencing its use and hence, also not the only factor to minimize street traffic volume. Cities like Los Angeles and Toronto seem to have a well-established network of public transportation close to the investigated streets, yet, we see high traffic flow. Thus, it is necessary to gain a wider overview on the factors influencing the use of public transportation. For one, it is a question of cost-efficiency. Especially in the United States of America and Canada, fuel is relatively cheap lowering the cost of a trip by car. In addition, it is necessary to incorporate the cost of public transportation and its structure. European cities have a long history of public transportation while North American tend to be more car-focused, which can also be seen in the different typologies of the respective cities.

The use of public transportation is however, not just the result of a cost analysis. For one, it is essential to account for the accessibility of the public transportation network. Typical questions are what is the interval of e.g., buses at a station, how many different lines stop at a station, how long does it take to reach central points of interests. This highly affects the feasibility of a potential trip using public transportation. There might be social stigmas linked to it. For example, some might see public transportation as the transport mode of lower income groups. Furthermore, they emphasize the individual freedom of individual car transport. This point is also linked to the fact that a significant group of commuters does not feel safe in public transportation, especially during the evening and night.

Even if a large percentage of the population is willing to commute by public transportation the systems in place need to be able to handle peak demand. Policy-makers have vast influence on this demand. For example, in summer 2022 the German government passed a legislation that enabled everyone to buy a monthly ticket for just 9€ enabling each citizen to take any local or regional mode of public transportation. This lead to an extreme increase in travellers, especially as it was also paired with the German summer vacation season. The public transportation systems were not prepared for such demand and hence, the consequences were full trains, and high delay and waiting times. Therefore, the measure was seen as contra-productive by some.

Overall, policy-makers should be aware that public transportation is a valid way of decreasing the traffic volume in cities. However, it requires a holistic approach that goes beyond city planning that only accounts for the distance to the next bus station. Public transportation needs to be seen as an essential part of the city and hence, planning should also include socio-economical factors. Policy-makers must raise awareness and work on the public image of public transportation. As this is a process that needs to develop over a longer period of time, it is recommend to start campaigns in schools and ensure the accessibility of schools by public transportation. The effect could be multiplied by ensuring the affordability of public transportation for children with a discounted ticket and by ensuring the safety of a ride with a stronger presence of law enforcement at public transportation stations.

7.3 Limitations

Nevertheless, this work comes with some limitations and biases. Cities are unique in their urban environment as they have different urban regulations and history. This results in the fact, that even cities in the same typology can show vastly different numbers in flow and occupancy. We analyzed five cities and made the assumptions that these would be representative of their respective typology. We had a large amount of data points but they were not equally distributed between the cities. This limits the explanatory power of underrepresented cities in our comparison. Moreover, it also limited the individual clustering as a smaller number of data points might not have equally showed the character of a detector and its respective street.

We have seen that for the clustering the elbow method did not provide a clear result for most cities. Hence, we had to manually decide on the optimal number of clusters for k-means. We have used k-means because of the our data size and the amount of clusters, yet, it provides sub optimal results for noisy data and outliers.

7.4 Future Research

We propose that future research should follow a similar approach for a larger number of cities. Moreover, it would be beneficial to have a more diverse set of cities across continents, urban, and socio-economical characteristics. More diversity can also be achieved in the comparison of multiple typologies. in order to identify common themes and differences.

In addition, we noticed that detectors were not equally distributed across all cities. Hence, it would help to establish common standards for detector locations concerning location in relation to the city, functional street classification, and the total amount of detectors for a city. An additional study incorporating these current limitations could provide a more insightful comparison between cities.

References

[mor, 2019] (2019). The 5-minute walk.

[Barth and Boriboonsomsin, 2008] Barth, M. and Boriboonsomsin, K. (2008). Real-world carbon dioxide impacts of traffic congestion. *Transportation research record*, 2058(1):163–171.

[Boeing, 2017] Boeing, G. (2017). Osmnx: A python package to work with graph-theoretic openstreetmap street networks. *Journal of Open Source Software*, 2(12):215.

[Choi and Toh, 2010] Choi, C. C. and Toh, R. (2010). Household interview survey from 1997 to 2008 – a decade of changing travel behaviours. In Ebrahim, N., Tan, A., Sun, G., Ely, M., and Ai, F. J., editors, *JOURNEYS. Sharing Urban Transport Solutions*, page 52. LTA Academy, LTA Academy, Land Transport Authority, Singapore.

[DELWP, 2019] DELWP (2019). Plan melbourne 2017 - 2050 addendum 2019 - planning.

[EEA, 2022] EEA (2022). Eea 2021: The year in brief.

[El-Geneidy et al., 2014] El-Geneidy, A., Grimsrud, M., Rania, W., Tétreault, P., and Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, 41.

[Enzmann and Ringel, 2020] Enzmann, J. and Ringel, M. (2020). Reducing road transport emissions in europe: Investigating a demand side driven approach. *Sustainability*, 12(18):7594.

[Forbes, 2000] Forbes, G. (2000). Urban roadway classification before the design begins. *Urban Street Symposium E-C019*, pages 1–8.

[Guth et al., 2021] Guth, J., Keller, S., Hinz, S., and Winter, S. (2021). Towards detecting, characterizing, and rating of road class errors in crowd-sourced road network databases. *Journal of Spatial Information Science*, 2021(22):1–31.

[Hickman et al., 2013] Hickman, R., Hall, P., and Banister, D. (2013). Planning more for sustainable mobility. *Journal of Transport Geography*, 33:210–219.

[IEA, 2017] IEA, I. E. A. (2017). Co2 emissions from fuel combustion 2017. https://www.oecd-ilibrary.org/content/publication/co2_fuel-2017-en.

[ITF, 2017] ITF, I. T. F. (2017). Itf transport outlook 2017 (summary in english). https://www.oecd-ilibrary.org/content/component/e979b24d-en.

[Mensink et al., 2000] Mensink, C., De Vlieger, I., and Nys, J. (2000). An urban transport emission model for the antwerp area. *Atmospheric Environment*, 34(27):4595–4602.

[Ng et al., 2013] Ng, K. M., Reaz, M. B. I., and Ali, M. A. M. (2013). A review on the applications of petri nets in modeling, analysis, and control of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):858–870.

[NHTSA, 2008] NHTSA (2008). National survey of bicyclist and pedestrian attitudes and behaviornbsp;.

[Oke et al., 2019] Oke, J. B., Aboutaleb, Y. M., Akkinepally, A., Azevedo, C. L., Han, Y., Zegras, P. C., Ferreira, J., and Ben-Akiva, M. E. (2019). A novel global urban typology framework for sustainable mobility futures. *Environmental Research Letters*, 14(9):095006.

[OpenStreetMap contributors, 2017] OpenStreetMap contributors (2017). Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org.

[Peters, 2019] Peters, A. (2019). How los angeles plans to get people out of cars by 2028 - fast company.

[Qin and Chan, 1993] Qin, Y. and Chan, L. (1993). Traffic source emission and street level air pollution in urban areas of guangzhou, south china (p.r.c.). *Atmospheric Environment. Part B. Urban Atmosphere*, 27(3):275–282.

[Tsuboi, 2021] Tsuboi, T. (2021). Traffic flow analysis and management. In Sepasgozar, S., Shirowzhan, S., Sargolzae, S., and Bienvenido-Huertas, J. D., editors, *Design of Cities and Buildings*, chapter 8. IntechOpen, Rijeka.

[United Nations Habitat, 2022] United Nations Habitat, U. (2022). World cities report 2022. https://unhabitat.org/wcr/.

[WBG, 2021] WBG, W. B. G. (2021). World bank group climate change action plan 2021–2025 : Supporting green, resilient, and inclusive development. https://openknowledge.worldbank.org/handle/10986/35799.

[Zhang et al., 2018] Zhang, Z., Wang, J., Hart, J. E., Laden, F., Zhao, C., Li, T., Zheng, P., Li, D., Ye, Z., and Chen, K. (2018). National scale spatiotemporal land-use regression model for pm2.5, pm10 and no2 concentration in china. *Atmospheric Environment*, 192:48–54.

Appendix

A Distance to Point of Interest Algorithm

```python
def shortestpath(det_data, city, POI_data, POI_name, k = 5):
    """
    Arguments
    ------------
        det_data: dataframe; detector point data
        city: string; Name of City
        POI_data: geopandas; amenity point data
        POI_name: string; name of amenity
        k: number of points to consider for street routing analysis
    Returns
    ------------
        det_data: updated dataframe; with distance to POI as columns
    """

    #Query walking street network
    G = ox.graph_from_place(city, network_type="walk")
    G = ox.utils_graph.get_largest_component(G, strongly=True)

    det_data[POI_name] = 9999 #dummy values

    #Loop through all detector points
    for i in range(len(det_data)):
        ori_x = det_data.iloc[i].long
        ori_y = det_data.iloc[i].lat

        #Generate euclidean distance of all points
        dist = [] #distance values
        g = [] #point data

        #Loop through all POI points
        for index, row in POI_data.iterrows():
            pt = row.geometry.centroid
            des_x = pt.x
            des_y = pt.y
            #update point data
            g.append(pt)
            #update distance
            dist.append(ox.distance.euclidean_dist_vec(ori_y, ori_x, des_y, des_x))

        #Return K closest points by Euclidean Distance
        top = [g[i] for i in np.argpartition(np.array(dist), k)[0:k].tolist()]

        #Compute Street-Level Distance
        #get distance to nearest node
        orig_node, o_dist = ox.distance.nearest_nodes(G, ori_x, ori_y, return_dist=True)
        street = [] #street distance values

        #Loop K closest Points
        for t in top:
            #get distance to nearest node
            target_node, t_dist = ox.distance.nearest_nodes(G, t.x, t.y, return_dist=True)
            #total distance = distances to origin & target node + distance of shortest path
            street.append(o_dist + t_dist + nx.shortest_path_length(G=G, source=
```

```
    orig_node, target=target_node, weight='length'))

        #Find closest points by Street-Level Distance
        det_data[POI_name].iloc[i] =  min(street, default = 999)

    return det_data
```

Listing 1: Algorithm to find shortest path to Point of interest given a set of detectors.

B Distance to Point of Interest (Cities)

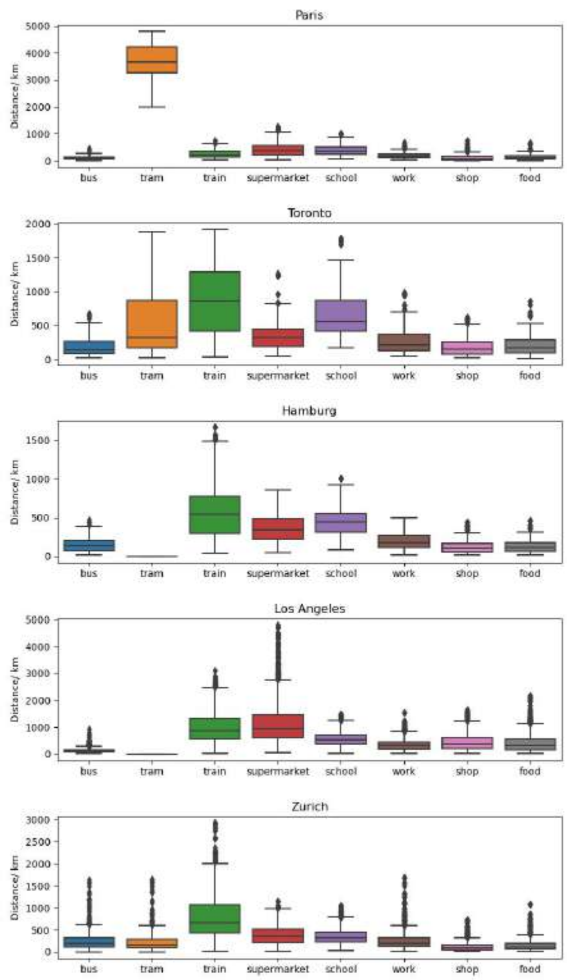

Figure 14: This displays the distance to POIs(Public Transportation, Amenities) for each of the selected cities.

C Traffic Congestion (Cities)

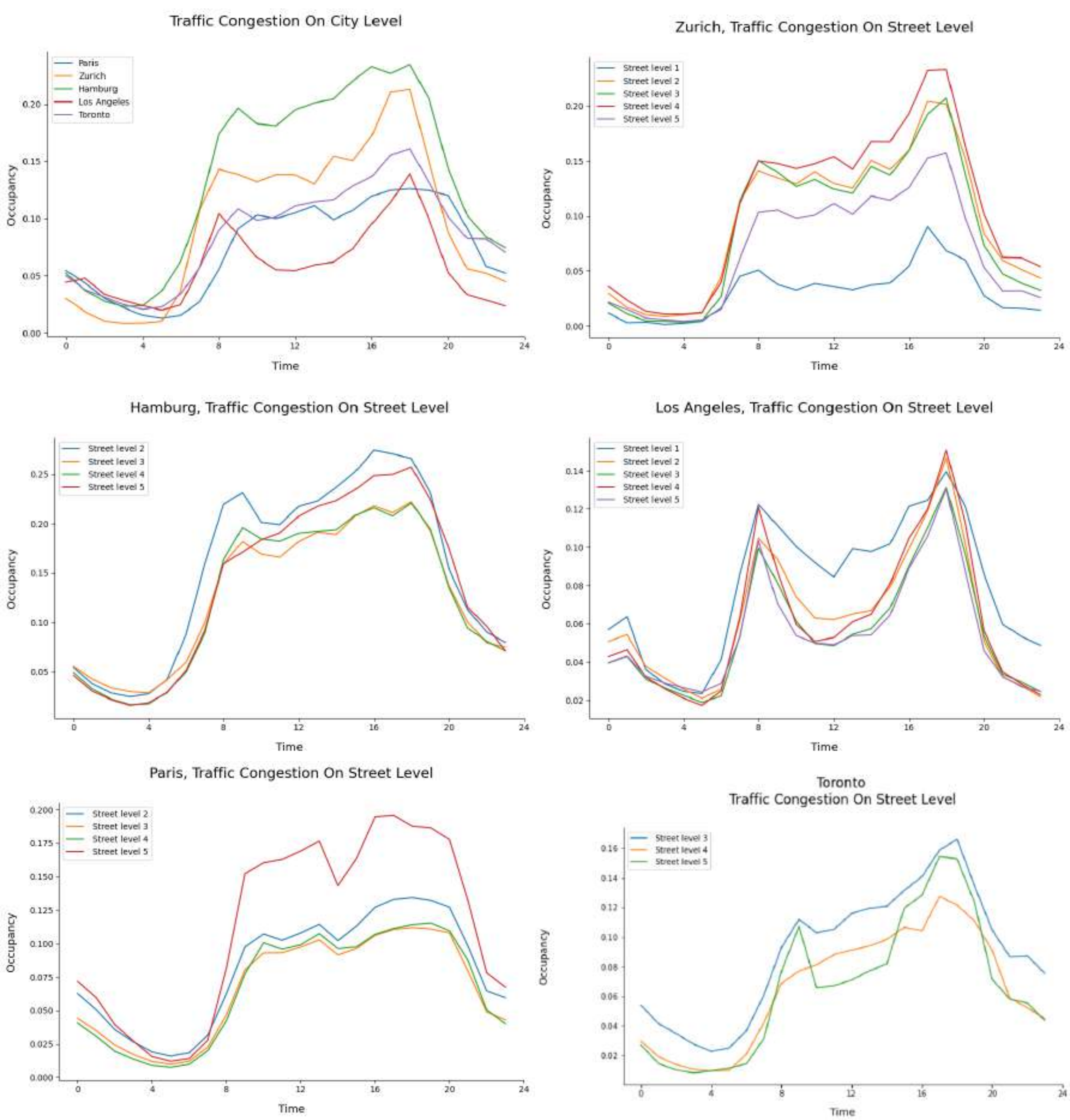

Figure 15: Figures display the average congestion for different cities in relation to time.

D Traffic Flow (Cities)

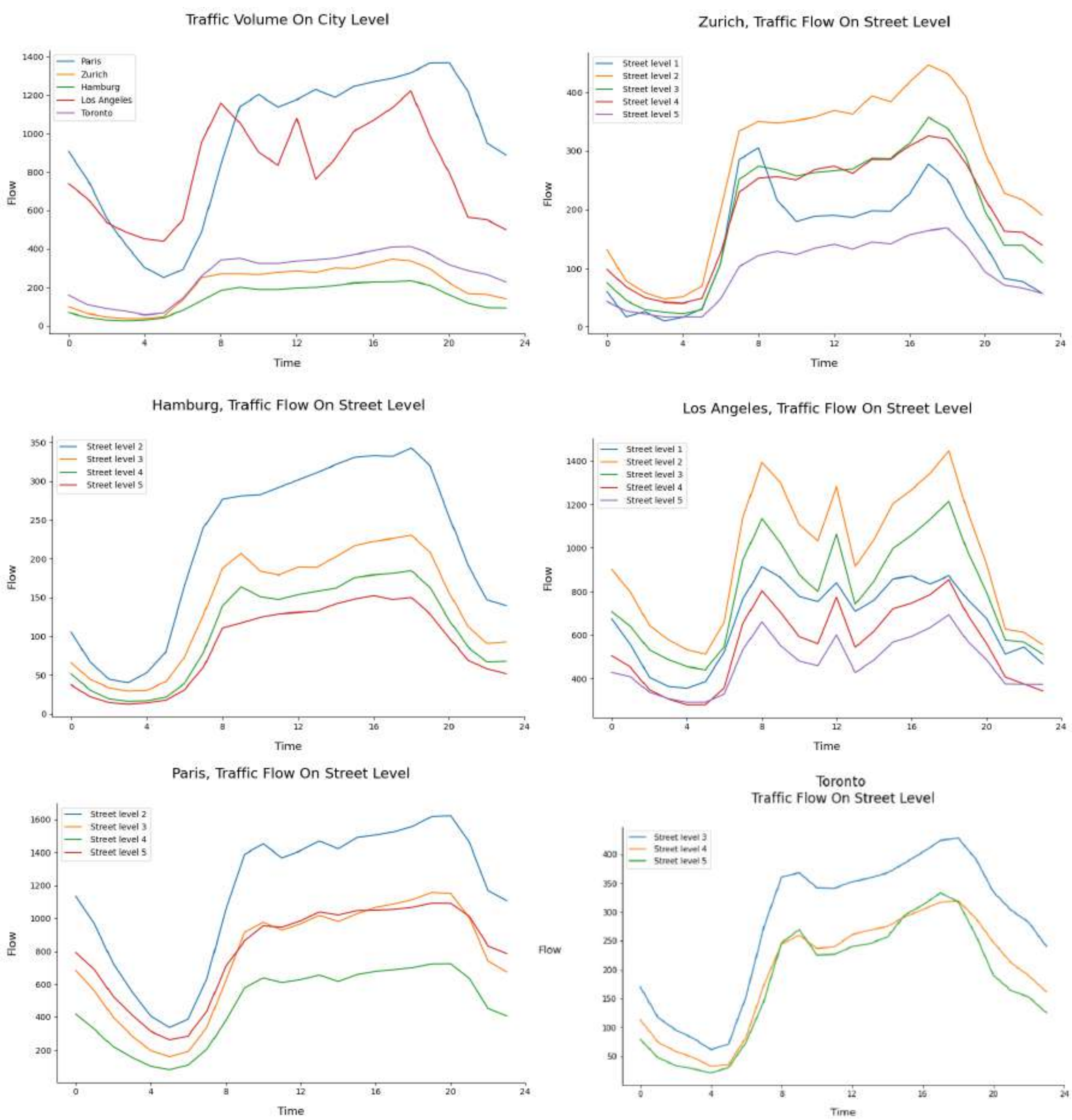

Figure 16: Figures display the average traffic volume for different cities in relation to time.

E Clustering Analysis - Optimal K for k-means

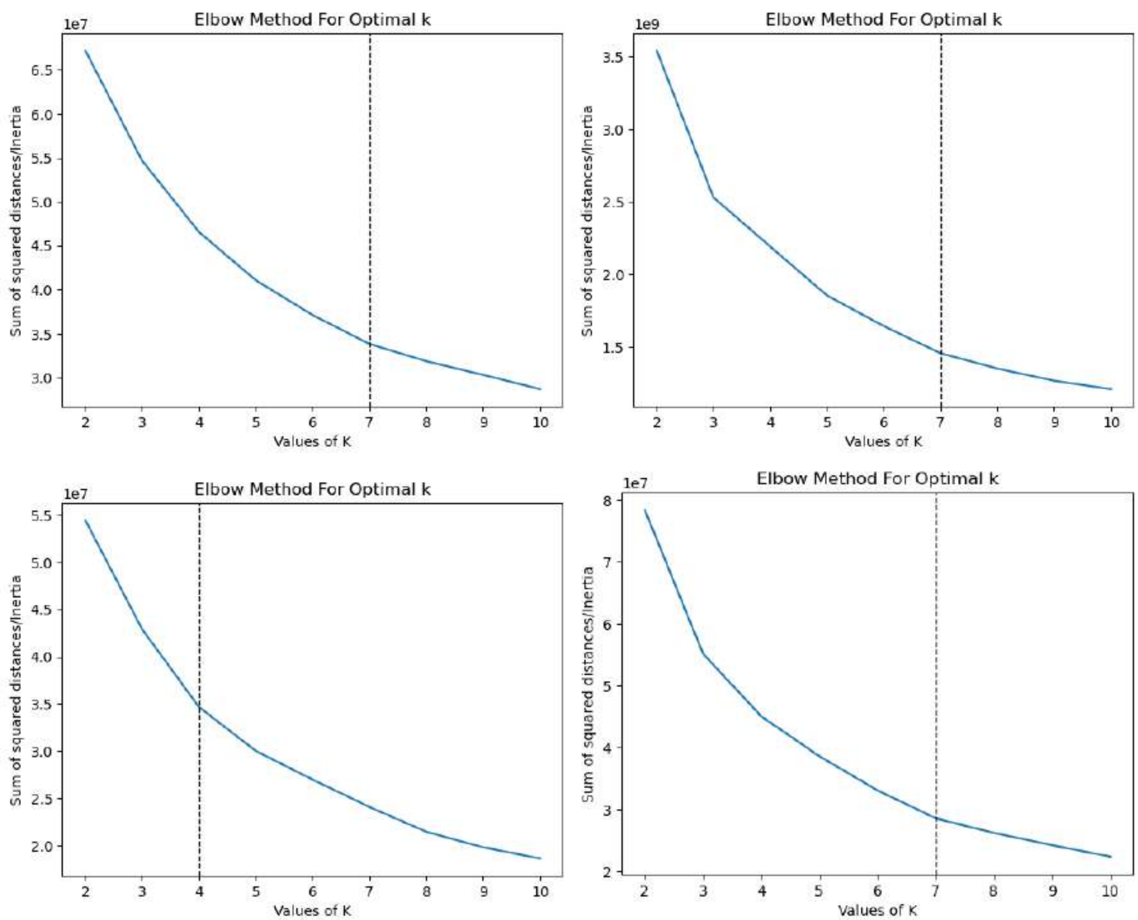

Figure 17: Figures display the optimal k for k-mean.

F Correlation Analysis - Pairplot

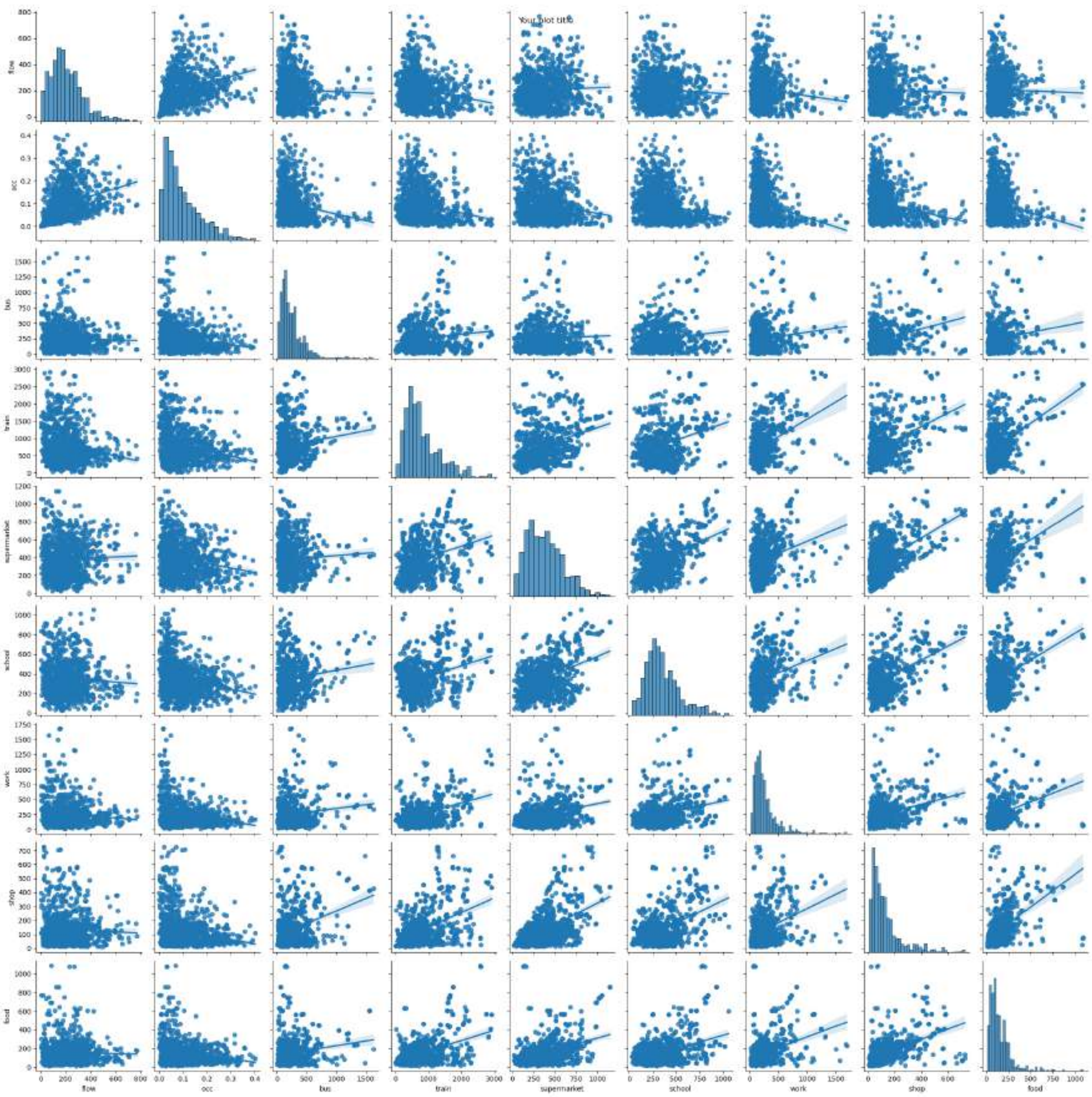

Figure 18: The figure displays the correlation between all features.

G Map of Occupancy and Flow for all cities

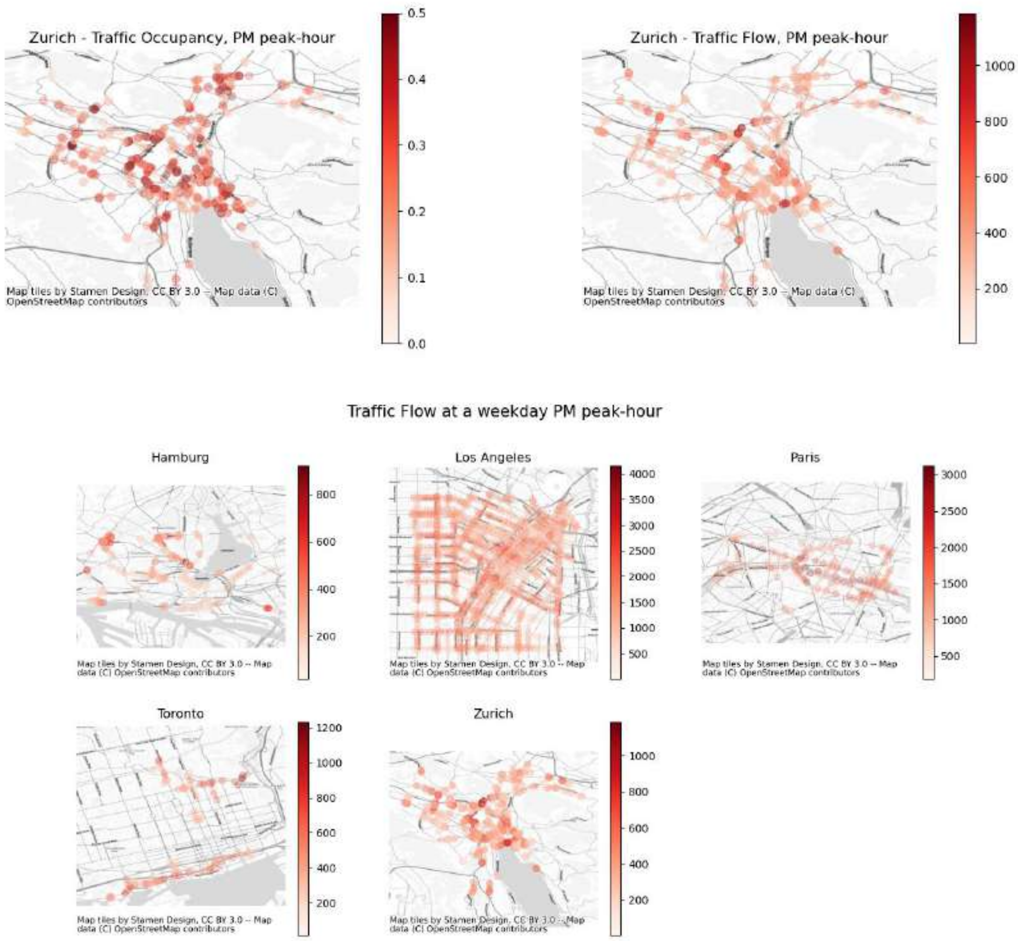

Figure 19: Figures display the average traffic occupancy and traffic volume for Zurich in relation to space

H Detector Clustering - KDE plots, Maps

H.1 Hamburg

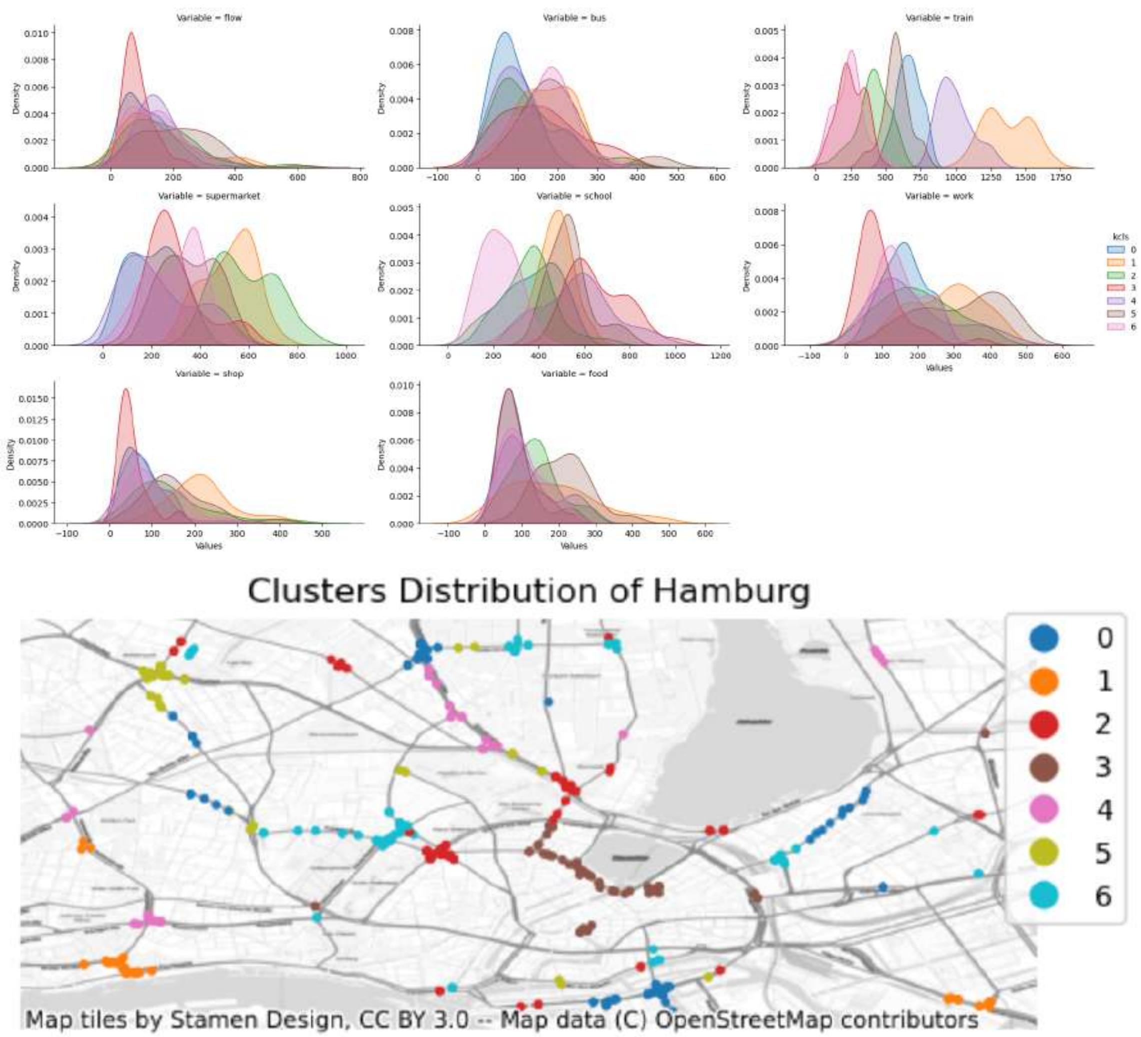

Figure 20: Figures display the KDE and cluster map for Hamburg.

H.2 Los Angeles

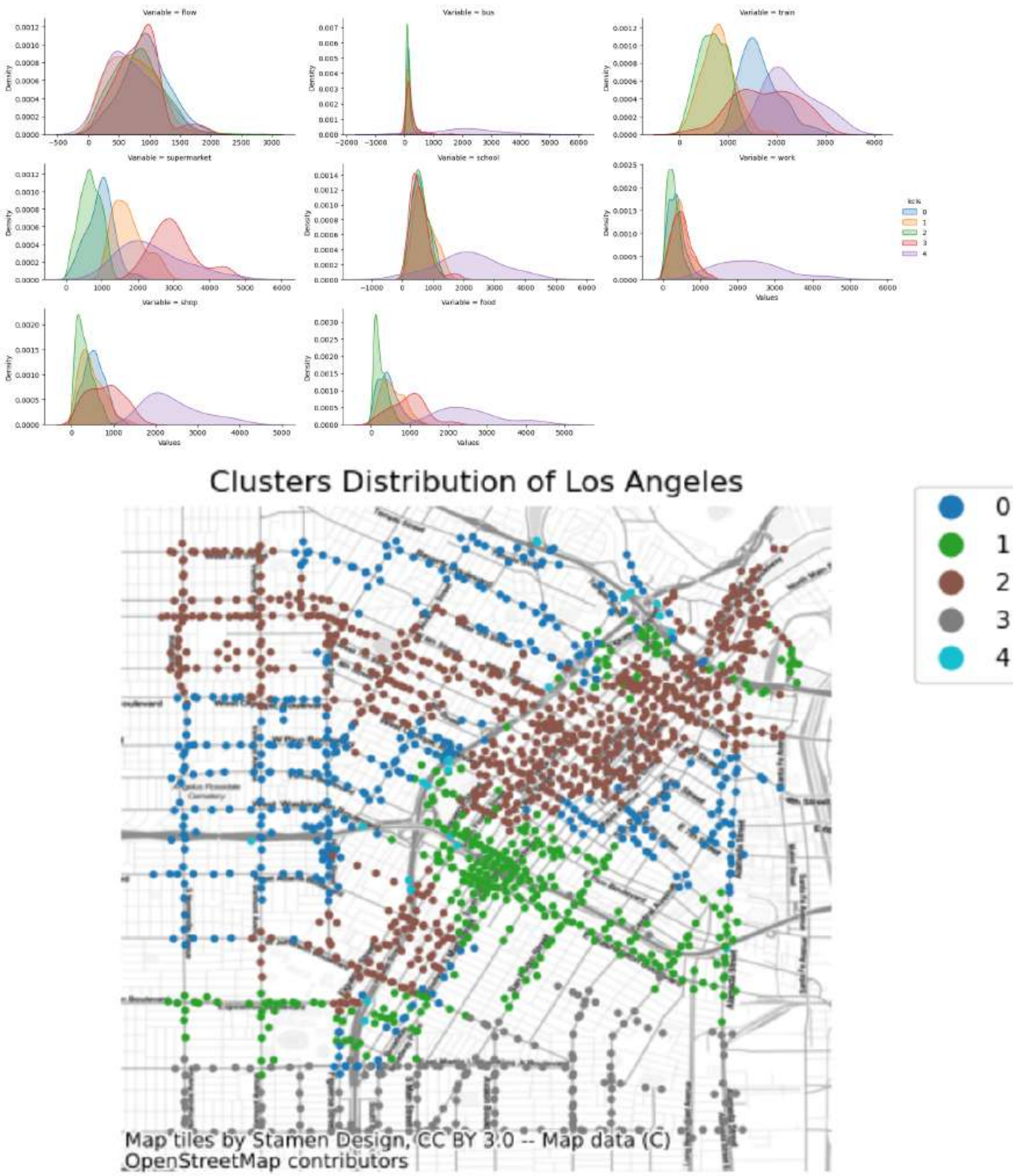

Figure 21: Figures display the KDE and cluster map for Los Angeles.

H.3 Paris

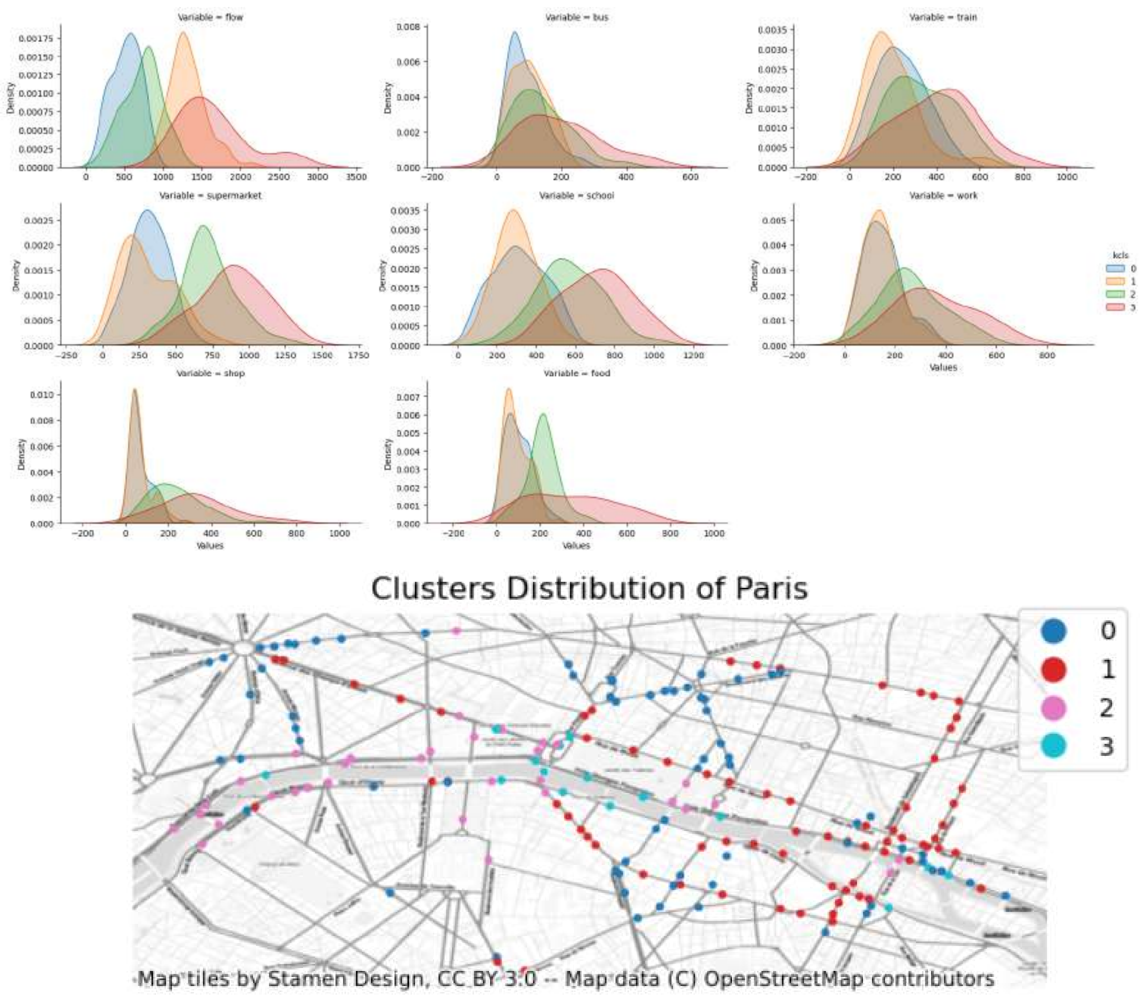

Figure 22: Figures display the KDE and cluster map for Paris.

H.4 Toronto

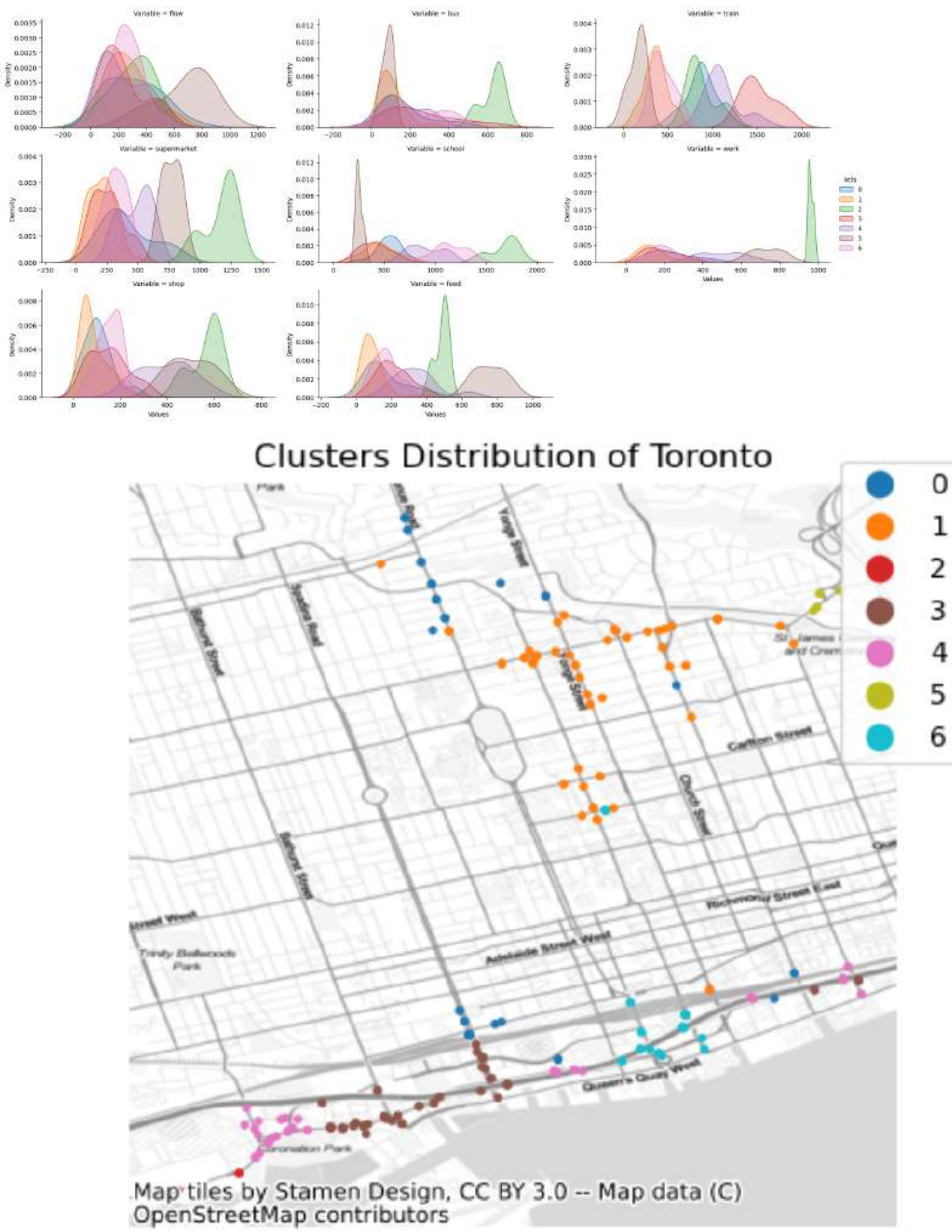

Figure 23: Figures display the KDE and cluster map for Toronto.