Exploring correlations between traffic flows and the build environment in five European cities

Koen Vierling
Bas Recter
Fabian Driessen
Hanna Denekamp

Supervisors: Trivik Verma & Lotte Lourens

11 November 2020

Abstract

In European cities upto 91% of greenhouse gas emissions are caused by car traffic (Rivas et al., 2020). These emissions could be lowered by offering alternative modes of transport, such as walking or public transport.
A study in China found a significant correlation between traffic flow and the built environment based on detector data and the presence of alternative travel modes expressed as Points of Interest. However the results could not be generalised to European cities.
Therefore this study focused on researching this correlation in five European cities. The following research question is formulated: *What is the correlation between measured traffic flows and the amount of POI's related to alternative forms of transport within 250m?* This question was answered through a LISA and regression analysis.

For this analysis traffic flows measured by detectors in Frankfurt, Hamburg, Munich, Rotterdam & Zurich are used. The number of POI's, representing alternative transport, within 250 meters of these detectors are counted. The POI's are defined as: stations, footways, pedestrian area's and cycleways. With these POI's the flow was predicted using an OLS regression model. The model found that only pedestrian areas had a significant negative effect on traffic flow. The other POI's had no significant relation with the traffic flows. Furthermore, a strong spatial correlation was found, indicating that roads around the detectors are a good predictor of local flow.

The model correlation indicates that more walk-able areas, which exist of many pedestrian areas, have lower traffic flows. The correlation for cycleways and public transport stations were not significant. The findings are that in order to reduce traffic flows and emissions, the walk-ability of an area, by creating pedestrian areas, has to be improved.

Contents

1 Introduction

- Faster forms of personal transport brought societies much economic growth (BANISTER2001209). However, this also came at a cost. In Europe, 40% of all CO_2 emissions and 70% of all other pollutants, are the direct cause of road transport (EC, 2020).
- A recent study found that these values were even higher within city centres in Europe, with traffic emissions contributing up to 91% of total emissions (Rivas et al., 2020). There is consensus that the earth is warming-up due to greenhouse gas emissions (Cook et al., 2016). Next to that, air quality is currently one of the biggest health risks in Europe (Marinov et al., 2016). Sustainable mobility is therefore increasingly important in cities.
- The presence of more high quality bicycle lanes, footpaths and public transport access points, could help to make these forms of transport more attractive (Parker et al., 2013;Murray et al., 1998). This could also potentially lead to distance reduction, by constructing more direct travel routes.
- Wang et al., 2018 researched the correlation between traffic flow and the built environment based on detector data and Points Of Interest (POI's) for a big city in China. Among the chosen POI's were factors related to public transport access, but not to the walking or cycling infrastructure. Correlations were found, but they stressed that those findings were perhaps only valid in this specific case. They concluded that it would be interesting to access how universal their findings were, by conducting a similar analysis for European cities.
- Based on the earlier mentioned research(Parker et al., 2013;Murray et al., 1998;Banister, 2011) four relevant POI's related to walking, cycling and public transport infrastructures were selected. Those POI's are: **public transport stations, foot ways, pedestrian area's and cycle ways**. POI's within a radius of 250 meters around a measured flow are assumed to be of influence, based on research mentioned in 2.
- To address the knowledge gap (Wang et al., 2018) formulated, this study will try to determine how POI's the traffic flow on roads in European cities. Insight in this situation could help to identify the most efficient investments to increase the amount of sustainable transport. To asses this, following research question is formulated: *What is the correlation*

between measured traffic flows and the amount of POI's related to alternative forms of transport within 250m?

- The data-analyses will be performed using python, the POI's will be taken from the PYROSM package and street networks from the OSMNX package. The traffic flows were obtained from the Multi-city Traffic dataset as constructed by (Loder et al., 2019). More over, environmental data about the cities was obtained form the Urban Typologies Project (Oke et al., 2019).

- Five cities of which both data sets contained information were selected to be included in this research. These are the following cities: **Frankfurt, Hamburg, Munich, Rotterdam and Zurich.**

2 Related work

- The World Business Council for Sustainability defined ”sustainable mobility” (WBCSD, 2001) as “the ability to meet the needs of society to move freely, gain access, communicate, trade, and establish relationships without sacrificing other essential human or ecological values today or in the future.”

- A study conducted by Banister, 2005 concluded using empirical methods, that one of the key parameters of the sustainable city are that it should have a population over 25000 (preferably over 50000), with medium densities (over 40 persons per hectare). In those cities it should be possible to keep the average trip lengths beneath the threshold required for people to walk or cycle. (Borrego et al., 2006) found that compared to disperse and network cities, compact cities with mixed land use provide better urban air quality.

- Banister, 2011 identified four ways for cities to become more sustainable, of which two are directly linked to the built environment. The first of those focuses on reducing the levels of car use through the promotion of cycling and walking and developing a new transport hierarchy. To achieve this, policy aimed at slowing down urban traffic, parking controls, road pricing, making it easier to use public transport and by reallocating space to public transport is suggested. The second is centered at land use planning measures to achieve distance reduction. This can be achieved by increasing densities and concentrations. This is especially useful for newly built neighbourhoods, whereas the turnover of the building stock is relatively low.

- Walk-ability in the context of sustainable transportation in cities, can be defined as the degree to which walking and cycling are promoted by the community design as an alternative for car usage to reach to reach shopping, schools, and other common destinations (Forsyth, 2015). (Ewing and Handy, 2009) identified specific characteristics that can improve the walk-ability of a city. Among those characteristics were physical features like the proportion of historic buildings, the amount of parks and the noise level.

- Another important factor, is the the distance people are prepared to travel by foot. American studies found that a distance of 400 meters or less, is a distance that Americans rather walk than drive (Yang and Diez-Roux, 2012). (Garcıa-Palomares et al., 2013) found that passengers in Madrid walked around 400 meters on average to reach public transport as well. The walking distance will always be higher than the distance as the crow flies. Therefore POI's within a range of 250 meters are considered to have a potential effect on the traffic flow in this study.

3 Exploratory Data-analysis

- For this research four data sets/sources are used.

  1. UTD19 Data set

- This a large data set that is published by the, ETH Zürich. The data spans 40 cities worldwide.

- In this research we'll analyse the data from the following cities, Frankfurt, Hamburg, Munich, Rotterdam and Zurich. An exploratory data analysis is carried out.

- From this data set we gathered the information about the detectors. This is, to be specific, the detector ID, the detector coordinates and the flow per detector. The flow is measured in vehicles per lane per hour. Other variables are the day of of the measurement, the time from midnight from the measurement, the latitude and longitude of the detector and the city the detector is located in.

- The UTD19 data set was used because it is a free open data source that is scientifically valid.

  2.Urban Typologies project

- The second dataset is the Urban typologies project. From this dataset the sustainability score and CO2 emissions per city were retrieved.

  3.OSMnx & OpenStreetMap

- OSMnx is a python package that allows the user to download spatial geometries and visualize and analyze real world open street networks. These spatial geometries are retrieved from OpenStreetMap. The street networks and the roads within are coupled with the UTD19 dataset.

- With OpenStreetMap the user can select an area of choice and export this as as a data set with spatial geometries. Then, using OSMnx, the data can be analyzed and visualized in python.

- The area to import and analyse were selected by creating a bounding box between the four values of the minimum and maximum of the latitude & longtitude points of the detectors.

  4.PYROSM & OpenStreetMap

- PYROSM is a free open source python package that allows for importing openstreet map data. In this case it is used to import the relevant POI's and retrieve their geometries.
- PYROSM was used because of it's good performance in retrieving POI's
- OSMOSIS was used to get the city maps for specific bounding boxes for PYROSM.

3.1 Inspecting UTD19

- When looking at figure 1, a boxplot with flow per city can be seen. There seem to be a high amount of outliers for high flows. This is likely because of rush hours or other peak times. This means that the mean is skewed upwards. However, it is decided that these values are kept in the dataset, to also account for peak times. And because alternative transport options are perhaps during rushhour even more important, as the road is congested.

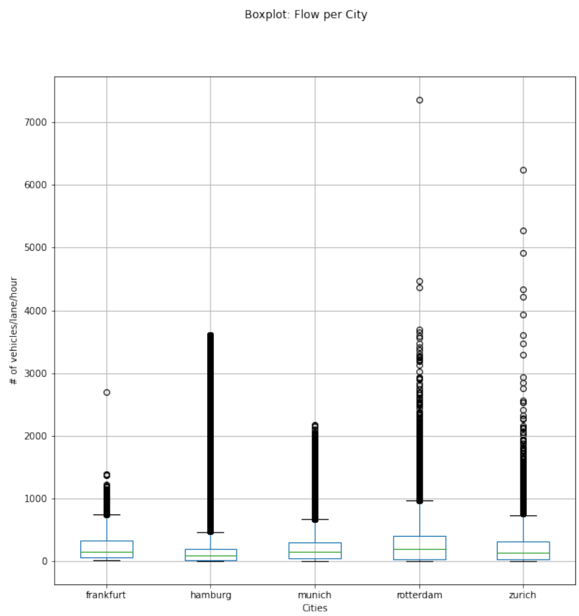

Figure 1: Boxplot: Flow per city

- When looking at figure 2, we can see the total counts of POI's and how these POI's are distributed for each of the cities. Also the types of POI's are reflected.
- It is remarkable that while Munich is bigger than Zurich, Zurich has more POI's registered.
- Rotterdam seems to have more cycle ways relative to the size of the city. The cycle way count fall into the same magnitude of size as Munich, however in total Munich has about three times as much POI's.

	# of days	*# of detectors*	*Mean flow (# of vehicles per hour per lane)*
Frankfurt	1	112	233
Hamburg	108	418	143
Munich	1	520	225
Rotterdam	6	259	260
Zurich	7	1020	207

Table 1: Table with overview of raw data per city

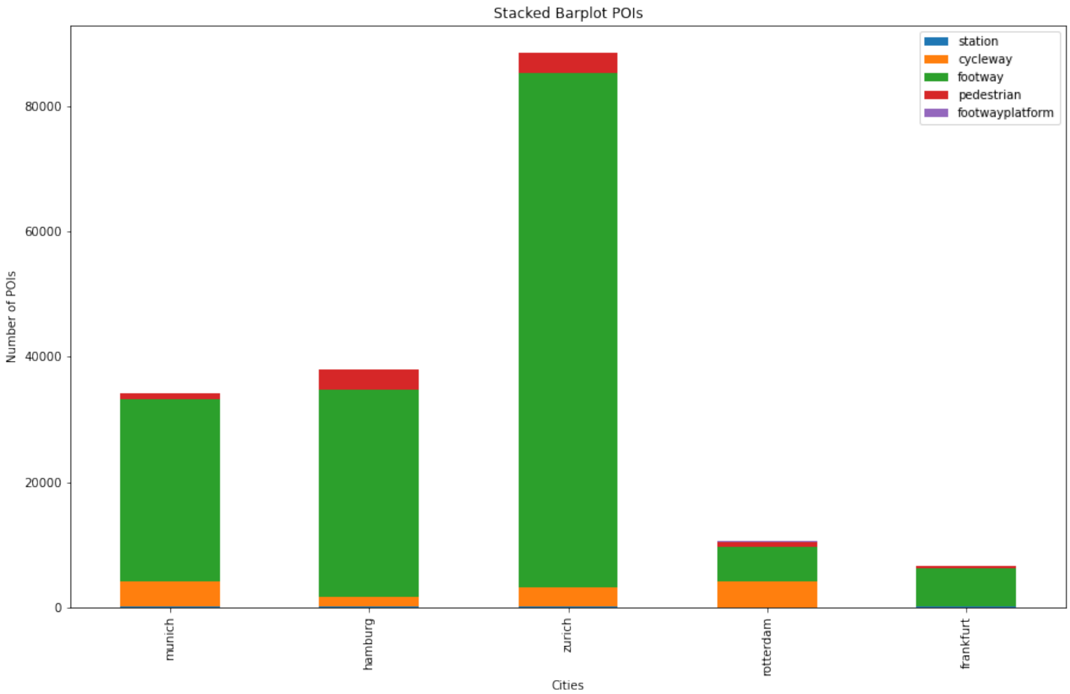

Figure 2: Stacked bar plot with POIs

3.2 Data Interrogation

3.2.1 UTD19

- UTD19 is a dataset with measurements of traffic flows in a variety of cities. Two elements of the UTD-dataseed needed to be combined, one dataset has detector data (such as id, longitude, latitude, roadnames, etc.). The other dataset has measured flow and ids, to link it. Only data for the selected cities were used.
- Some cities have sparser data than others, for example Munich has a single day of detector data while other cities have years. Therefore for each individual detector the mean of the measured flows have been calculated. This number is expressed in cars/hour.

3.2.2 OSMnx & OpenStreetMap

- From OpenStreetMap we gained the streetnetworks of the cities. These were read in and converted to machine readable graphs using OSMnx.
- An important step is to link detector data with roads in the street network. In each city the coordinates of the detectors were associated with the closets edge (or road) in the street network. A visualisation is given in figure 3. This algorithm (associating points with edges) worked better than matching on street names, as now multiple detectors could be associated with one road.
- When roads in the network are associated with detectors it is possible to assign flows to these roads. These flow measurements are derived from the UTD19 data set.

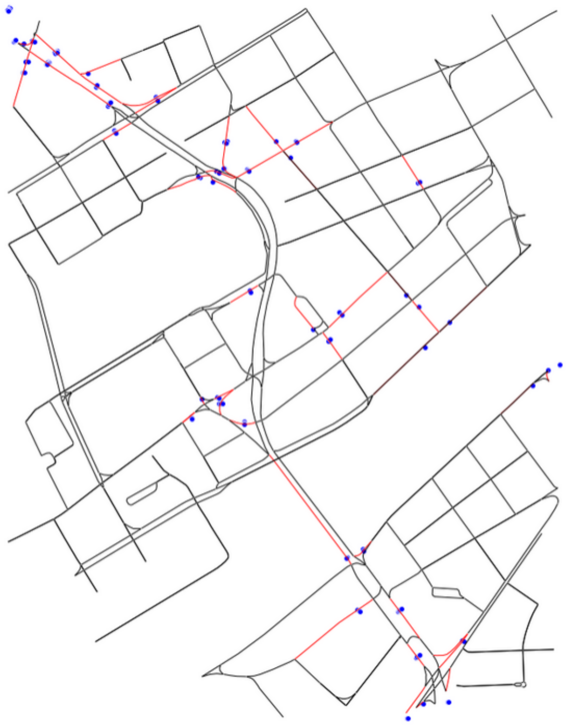

Figure 3: Example of matching detectors in Frankfurt (points in blue) with the assocated edge, also colored red

3.2.3 PYROSM & OpenStreetMap

- from OpenStreetMap the POI's were imported for each of the cities. These were read in and prepared for analysis using the PYROSM package.
- Again in the PYROSM datastructure each POI has a geometry. To express POI's in points the centroids of their geometries were taken. These points were combined with the earlier mentioned detectors and street networks into a single dataframe containing all information necessary for the analysis. A visualisation example of Frankfurt is given in figure 4.

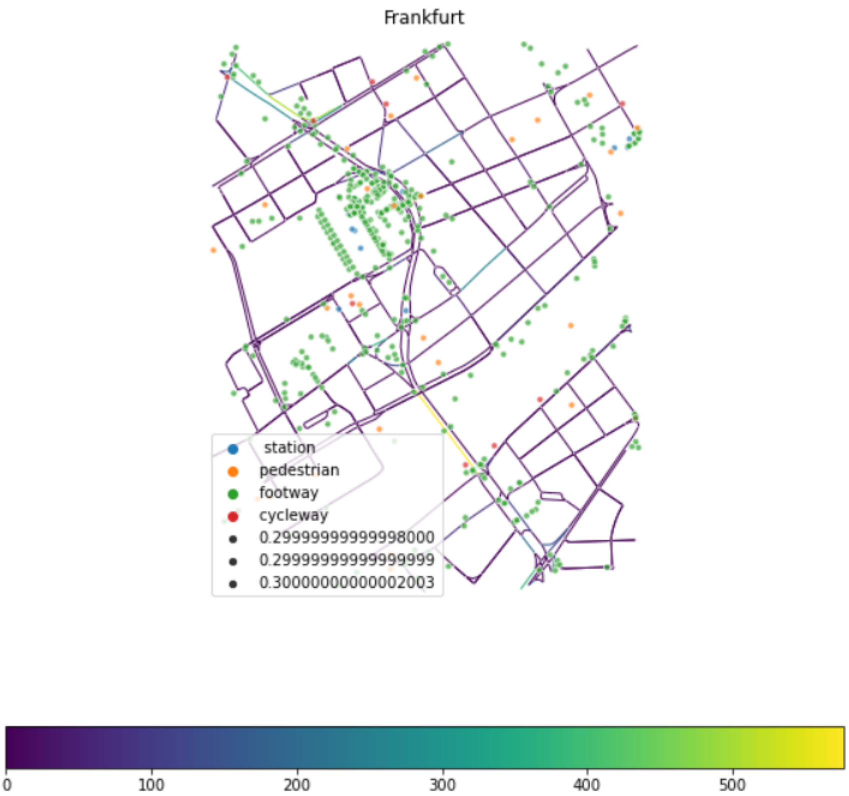

Figure 4: Frankfurt's street network, with the relevant POI's plotted as points, the flows in the street network are also visulised.

3.3 Analysis of data

3.3.1 Setting up the analysis

- As mentioned in sections 3.2.1 & 3.2.2. The detector data has coordinates, the measured flows in detectors were coupled with a street in the OSMnx street network. This creates a dataframe with measured flows and roads.

- To include spatial effects, see section 4.1 for the spatial analysis, that the detectors have on each other the distance bound spatial weights were calculated. The range threshold is set at 250 meters. Thus, if two roads lie within this distance of each other they are neighbours.

- In the same dataframe the POI's retrieved from PYROSM are appended. Again the distance bound spatial weights are calculated with a threshold of 250 meters. Now if looking at a single road the neighbours is a list of both POI's and other detectors. The total count of POI neighbours is calculated for each road, and the count per type of POI.

- Aggregating these results a dataset with information about all of the detectors is retrieved. An example of the dataframe is shown in figure 5.

	detid	flow	City	geometry	NeighboursTotal	station	footway	cycleway	pedestrian	footwayplatform
0	10004022	254.610011	munich	POINT (80765.660 -318226.068)	102.0	1	73	22	6	0
1	10004021	139.225395	munich	POINT (80761.979 -318223.897)	103.0	1	74	22	6	0
2	4000011	243.825830	munich	POINT (83493.409 -315018.647)	18.0	0	16	2	0	0
3	4000014	187.267434	munich	POINT (83485.299 -315013.987)	18.0	0	16	2	0	0
4	4000012	217.739685	munich	POINT (83490.026 -315016.473)	18.0	0	16	2	0	0
...	...	...	...	...	...	...	...	...	...	...

Figure 5: Example of the cleaned and aggregated dataset

3.3.2 Exploring the variables

- To test the hypothesis the flow will be the dependent variable. Initially the predicting variables are the counts of POI neighbours per type. These counts are reflected in the following variables : stations, cycleways, footways, pedestrians and foot-way platforms.

- Below, in figure 6 a pair-plot for all the variables considered in the analysis is presented. The hue, or color of the points represent the city out for that specific point. For each individual city a linear regression fit is represented with a line (in the color of the city).

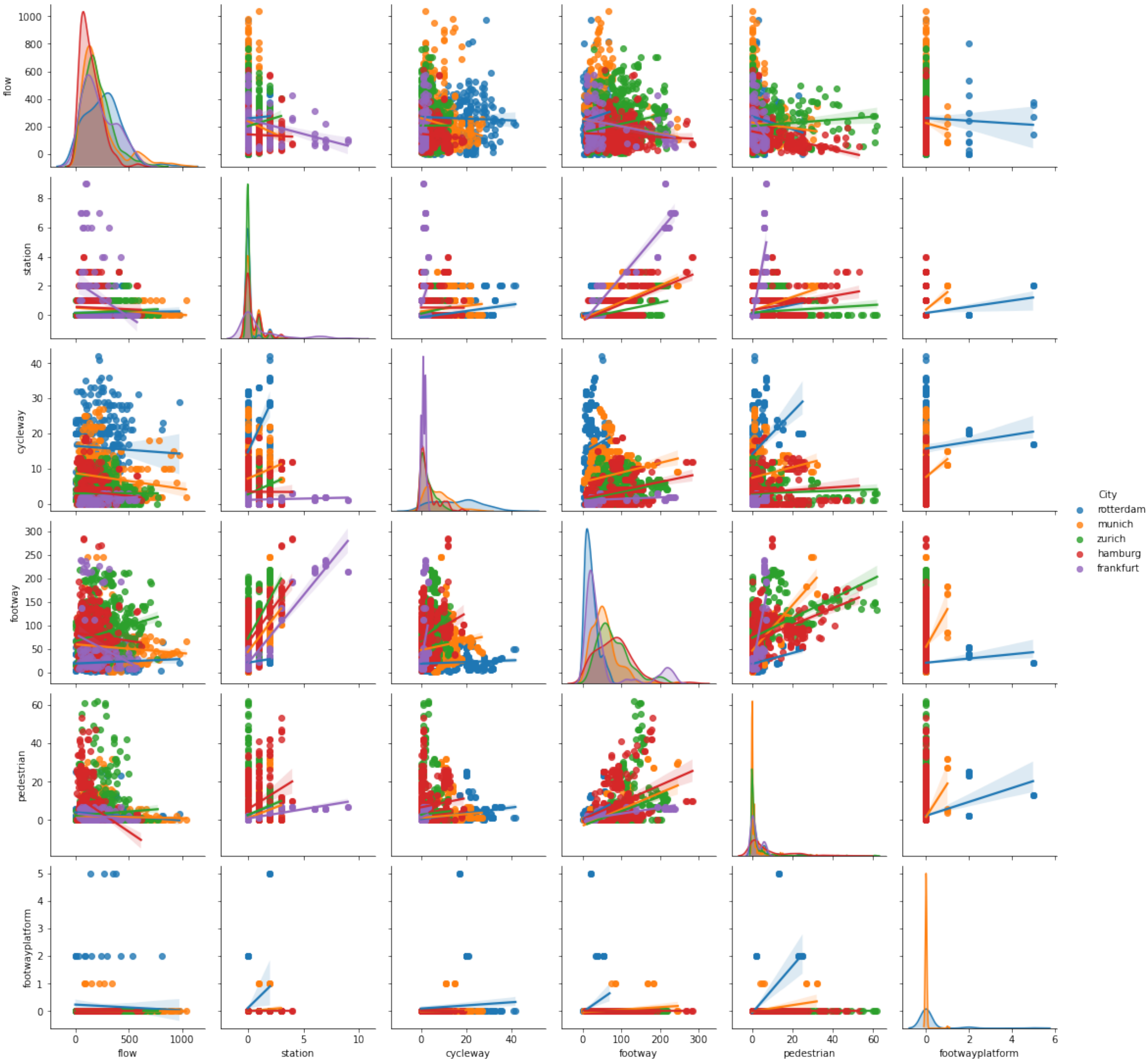

Figure 6: Pairplot, shwoing distributions and relations of the inupt variables for regression

- In figure 6 the diagonal axis shows a kernel density estimation plot for each of the city.

This gives insight regarding the distribution of these variables across cities. For example, Hamburg seems to have the highest flows for the the detectors. Similarily Rotterdam has a high density of foot-ways.

- In the top row of figure 6 the y-axis represents the traffic flows and in each plot the x-axis is a different variable. Most trends seem to be going down, which is inline with the hypothesis.

- In the rows below, correlation with flow and the multicollinearity of the predicting variables are tested. All variables seem to be correlated with the traffic flow, of which some are positive and others negative. There are also, mostly positive, correlations between the different POI's. Too strong correlations could lead to multicollinearity, which could result in unreliable coefficients of a regression model.

- Because most variables seem to indicate some trend with the flow multivariate regression is a good method. However, it is important to take multicollinearity into account when creating the model.

3.4 Visual inspection

Following the pairplot and the aggregated dataframe setup in the sections below, it is possible to visually inspect if there is a relation between POI neighbour count and the local measured road flow. The figures out of which conclusions are drawn are presented in appendix A. An example of a plot is given in figure 7.

- In the figures 21, 20, 18, 7 and 19 (the two digit numbers are given in appendix A) on the left plot the measured flow of detectors is matched with specific streets in the street network.

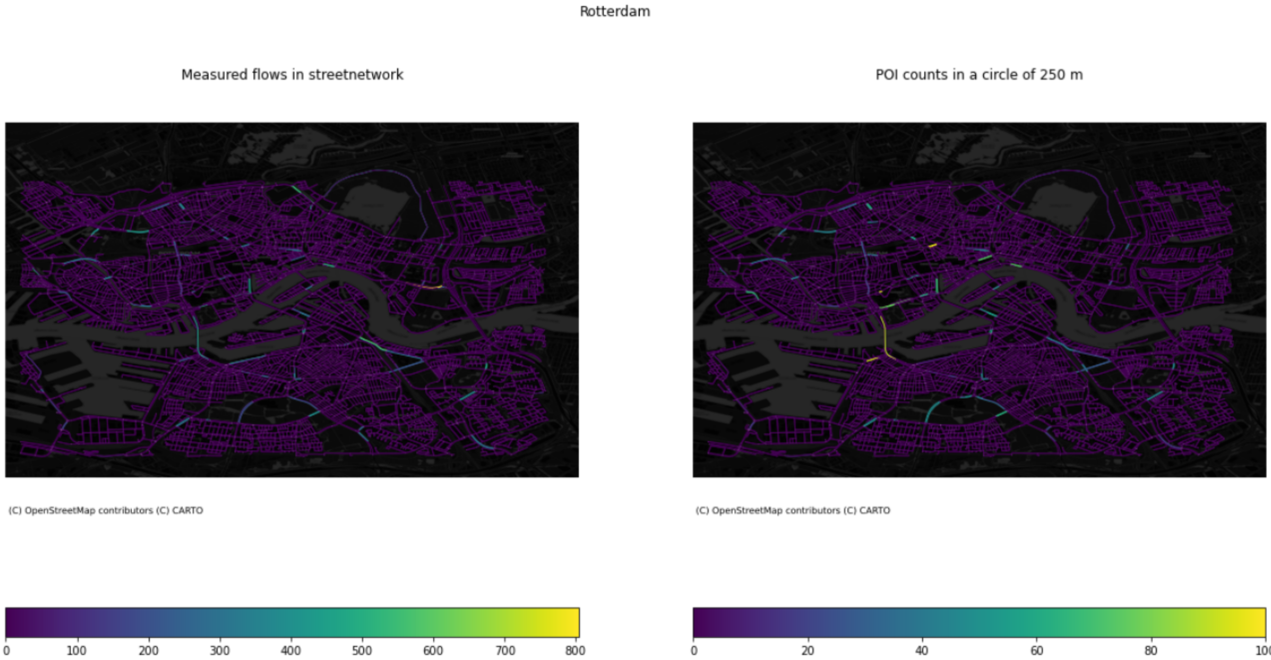

Figure 7: Flows, and amount of POI's per edge in Rotterdam

- Similarly, the count of associated POI's (that are within a range of 250 meters) are plotted on the right.
- Following the hypothesis, roads with a high flow should have a lower amount of POI's in the area.
- In Rotterdam (figure 7) the correlation seems to be less evident. Most high flow roads, also seem to have a high amount of POI's associated with it.
- As seen in figure 21 of Frankfurt the correlation seems to be evident. the roads that show a high traffic flow indeed seem to have less POI's around them.
- As seen in figure 20 of Hamburg the (visual) correlation seems to be less strong. Most roads that show up as high traffic flow in the left figure also show up with a high POI-count. Indicating that perhaps the correlation in Hamburg is not as strong. In the centre of the map there seems to be some correlation though, with high POI-counts and low traffic flows.

- In Munich, figure 18 the correlation seems to be less evident. High flow roads will also have a higher amount of POI's assosciated with the the road.
- In Zurich (figure 19, the correlation seems to play effects in the city centre, however moving to the outer bounds of the city the correlation is less likely.

3.4.1 Visual inspection: Conclusions

- In all of the cities there seems to be some sort of (visual) correlation between the flows in a street and the POI-counts. Of course, this correlation will vary between cities. However, in most cities especially the city centre will have more POI's associated with it, and evidently these roads also have lower flows.
  On the outskirts of the cities, there are roads that have low POI-counts and high flow.
- This makes sense in terms of the hypothesis, as these roads do not have alternative transport options.

- A likely cause is that within a city centre walking or public transport is more feasible and thus that the local traffic flow is reduced. Part of this could be explained by traffic that is going into the city centre, increasing the network flow on the outskirts as well.

3.5 Limitations of data and study

3.5.1 Measurement period

- As can be seen in the table 1 in section 3.1, the measurement period varies. In Frankfurt and Munich they measured just 1 day, where they measured 108 days in Hamburg.
- Depending on the day of the week, there will be more or less flow measured, this will influence the correlation between the POIs and the measured flow.

3.5.2 Measurement interval

- A Limitation from the study is the measurement interval. For the sake of simplicity we chose to compute the detector flow average per day. In the original table the intervals are between 3 and 5 minutes.
- If the original intervals were used the relation could be stronger since in rush hour a higher flow would be measured.

3.5.3 Limited number of detectors

- The limited number of detectors is also a limitation. As can be seen in table 1, there's a difference in number of detectors. With a limited number of detectors it harder to demonstrate the relation between flow and POIs.

3.5.4 limited number of POIs

- The number of POIs in the UTD 19 data set is limited.
- If this data would be more extensive the hypothesis could be rejected or confirmed with more certainty

3.5.5 determination of neighbours

- The neighbours from the streets are determined from a centroid point. The centroid point in this study is the location of the detector and not the whole road itself.
- This is done for simplicity. But it might influence the relation between the flow on streets and the POIs in the area.

3.5.6 Determining neighbours on distance isn't perfect

- Determining the neighbours based on their distance is too, done for simplicity. When someone is traveling near a POI they don't necessarily have that POI as destination or starting point.

- This also influences the relation between the flow on a street and the POIs in the area.

4 Analysis

- Our hypothesis was that when there are many alternative transport options for travelling by car in the built environment, the flow of cars will be less. This would result in a higher sustainability score and less CO2-emssions per capita. We argue that this is true for some alternative transport options, but not for all.

4.1 LISA

- The hypothesis is that the measured flow (average amount of cars per hour) will be lower for streets that have a lot of alternative transport options close by.
- It is possible that the dependent variable (local measured flow) is spatially correlated. If the flow is spatially correlated it would mean that some of the locally measured flow could be explained by the average flow of the neighbours (which are within a radius of 250 meters).
- For network flow spatial correlation would make sense, as if one road is very busy probably the roads around it will also be very busy (due to the high flow of the initial busy road).
- To analyse this effect and provide a statistical basis, the Moran's I for each of the cities was calculated for the variable flow. And as discussed with a distance threshold of 250 meters.
- The analysed roads in Frankfurt are within a relatively small area. Thus with a threshold of 250 meters there are no islands. Thus there are no detectors that have zero detectors around in a radius of 250 meters.
- The figure 8 shows $I = 0.24$, with a significance level of $p = 0.01$. Therefore there is spatial correlation.
- The I should be interpreted as follows, if flow is 1 standard deviation higher than average, the neighbours will be 0.24 standard deviations of flow busier.

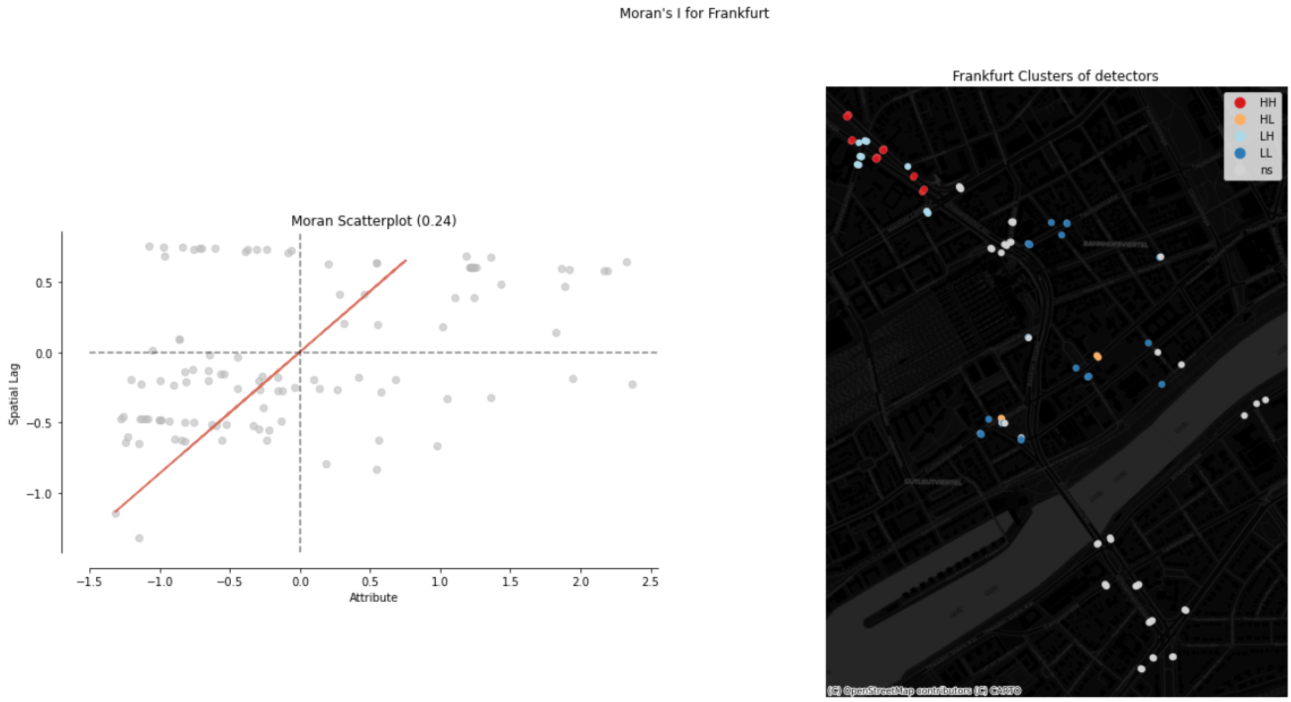

Figure 8: Moran's I for Frankfurt, with $p = 0.01$

- On the right side of the figure in Frankfurt the detectors and their quadrant are plotted on the map. This should be interpreted as follows, red points are in the HH quadrant, these are High High, the flow of this detector is high and the flows of the neighbours are high, these roads can be found on the top left of the figure. Some roads do not have significant spatial correlation, their variance can't be explained by the flows of their neighbours, these are marked gray in the figure.

- Similarly, in Hamburg Moran's I is calculated. In Hamburg, with a range of 250 meters there are three islands. These islands were removed from the analysis for now, and the spatial weights recalculated.

- In Hamburg the $I = 0.34$ with a $p = 0.001$. The I is slightly higher than in Frankfurt, which means that the spatial correlation is stronger in Hamburg, a higher amount of variance is explained by spatial correlation.

- As possible to see in the map on the right, this is likely because detectors are often placed along the same road. These roads often fall in the same quadrant (for example the line of blue detectors that belong to Low-Low). Points seem to be mostly in the high-high or low-low quadrant.

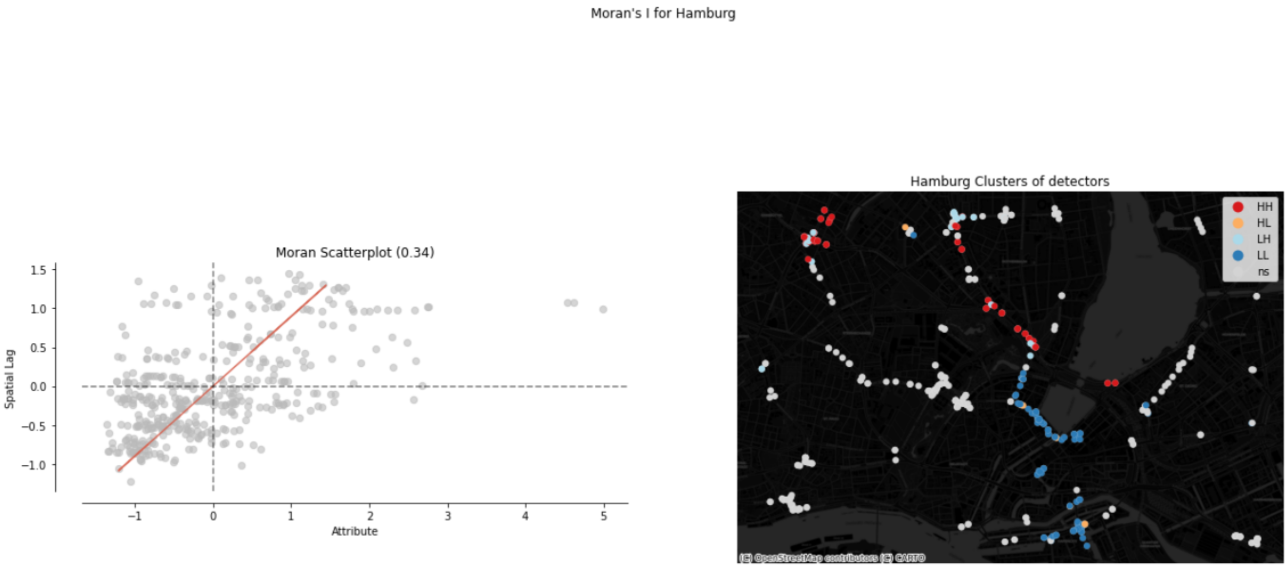

Figure 9: Moran's I for Hamburg with $p = 0.001$

- As seen in the figure 10 the Moran's I is higher in Munich with an $I = 0.7$ and a quite significant values. In Munich there are four islands if the range is set to 250 meters, these were removed from analysis. Looking at the Moran's plot it seems that the spatial correlation is strong in Munich. Similarly as was the conclusion for Hamburg, most points seem to be placed in either the high-high or low-low groups.

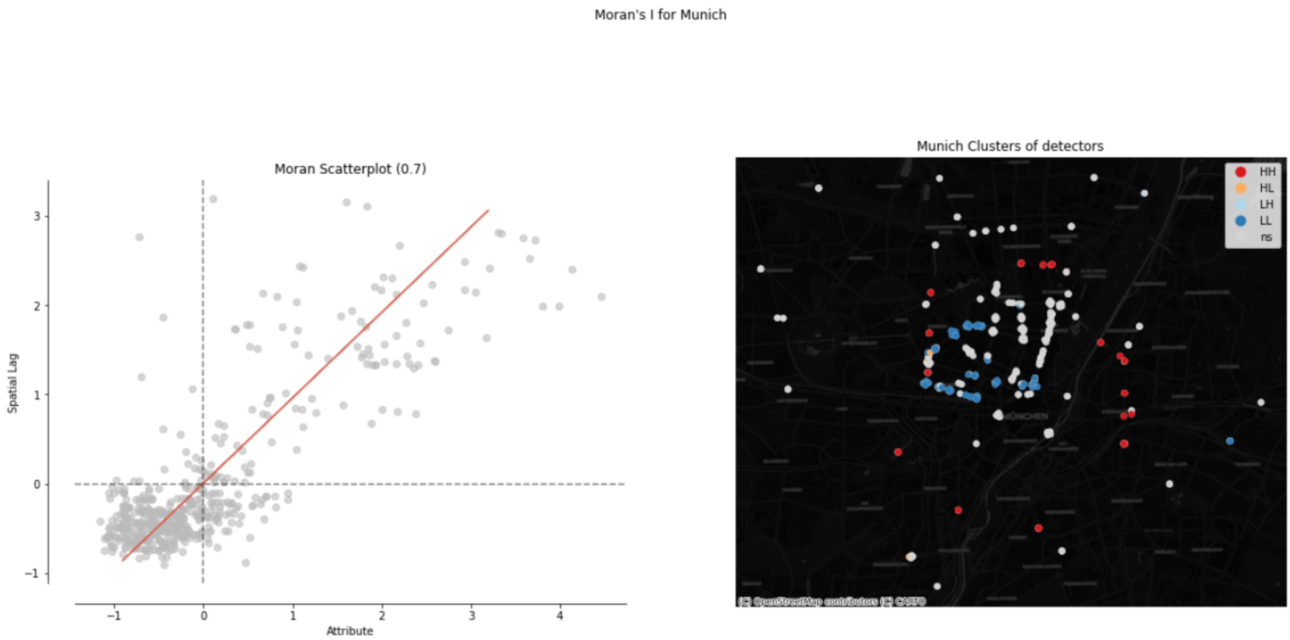

Figure 10: Moran's I for Munich with $p = 0.001$

- In Rotterdam the situation is a little different, with a threshold of 250 meters, there are two islands removed from the analysis. The $I = 0.0956$, which is quite low. This can all be discerned from the Moran plot, the pattern seems less correlated (more random). However the I is in fact signifcant with $p = 0.047$. Because of the low I, the amount of detectors that fall into quadrants is smaller as seen in the sub-figure on the right.

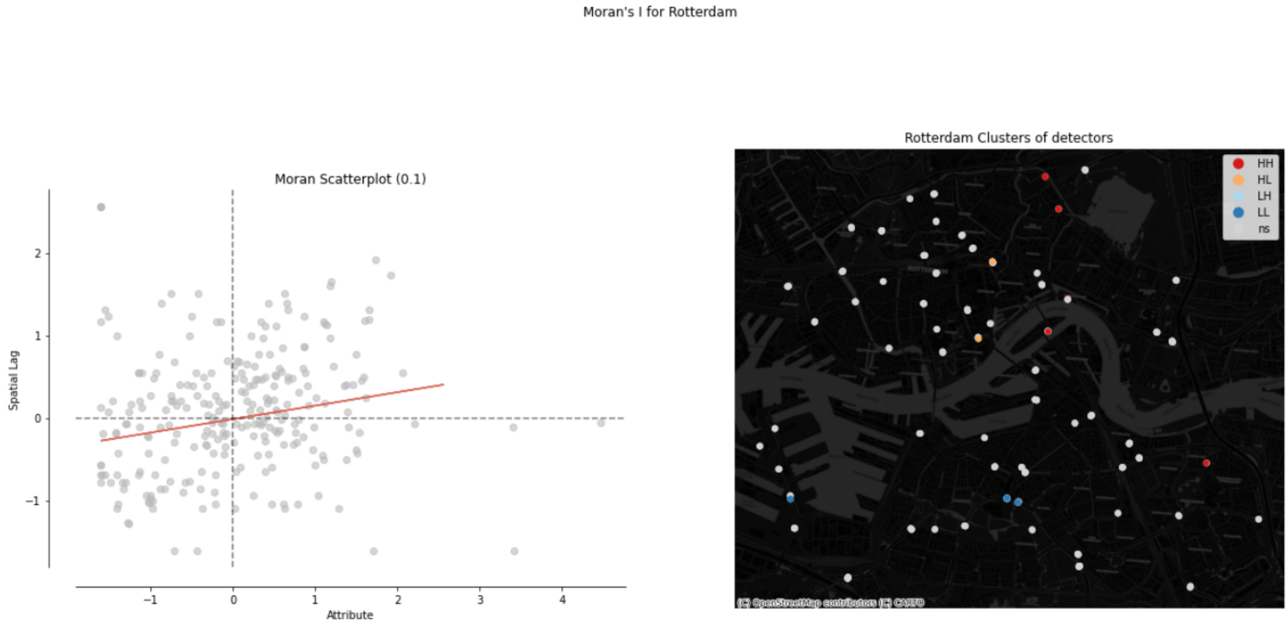

Figure 11: Moran's I for Rotterdam with $p = 0.047$

- In Zurich as seen in the figure 12 especially for the higher flow roads the neighbours also seem to have a higher flow (from the scatter plot). Again, the p value indicates that the spatial correlation is significant.

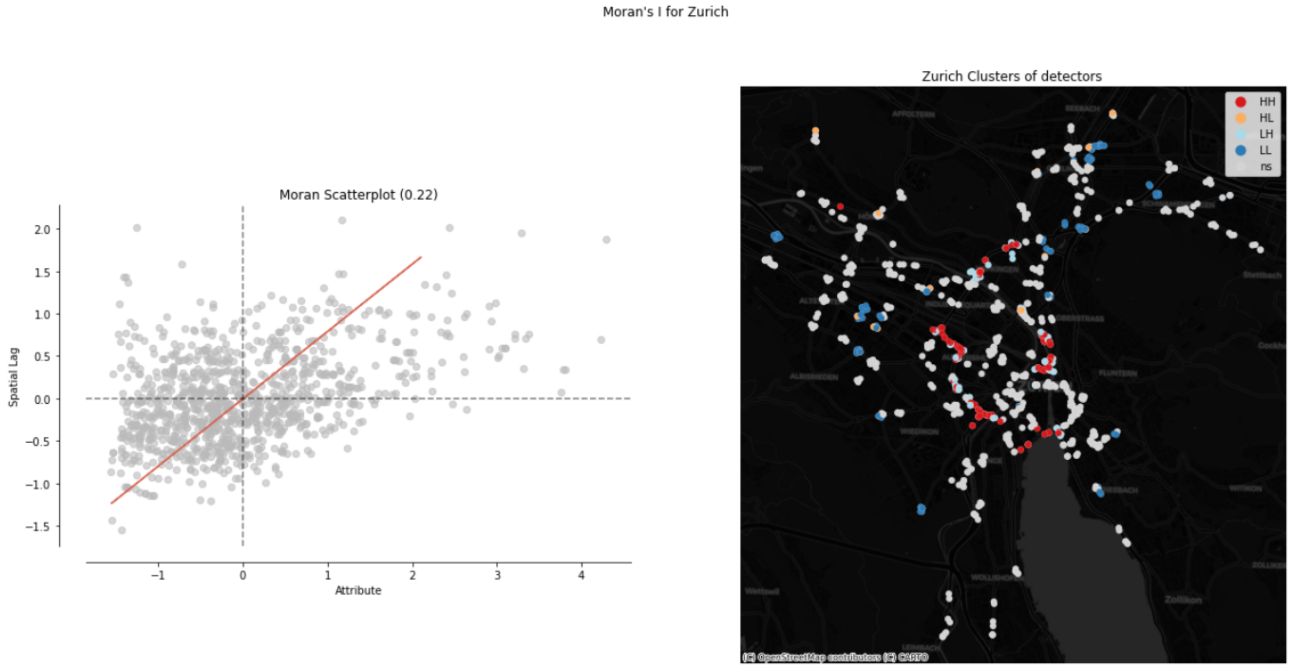

Figure 12: Moran's I for Rotterdam with $p = 0.001$

4.1.1 Conclusion Spatial Correlation

Concluding, the spatial correlation varies between the five cities. However, spatial effects are significant in all of the cities. Therefore some off the variance of the flows is explained by the flows of neighbours.

4.2 Spatial linear regression

To understand the relation between the flow and the alternative transport options a (spatial) regression analysis was conducted. A regression model can help quantify the relationship between the two. Statistics of the model can help us interpret how correct the relation is that we're seeing. The model was build in a iterative way.

4.2.1 Iterations model

- First iteration of the model: All variables were included, with the city as a dummy variable. This model was significant, but the residuals were tested on spatial auto-correlation as argued in section 4.1. The full model results can be found in appendix A.2.1. the Moran's I. The $I = 0.24$ with a $p = 0.01$, thus spatial effects play a significant role and will be tested in the next iteration

- Second iteration: Spatial lag of the flow was included, full model results are given in appendix A.2.2.

- A new variable is included: the spatial lag of flow for each of the detectors. This variable represents the weighted average of neighbouring detector flow.

- Including spatial lag made almost all other predictors non-significant, except for the variable pedestrian. Therefore, the variance of the flows in the streets is quite accurately predicted by the weighted average flow of the neighbours around it.

- Third iteration: All non-significant predictors were removed and the log of pedestrian was taken, since the variable was very skewed.

4.2.2 Results final model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   flow   R-squared:                       0.325
Model:                            OLS   Adj. R-squared:                  0.324
Method:                 Least Squares   F-statistic:                     447.7
Date:                Mon, 09 Nov 2020   Prob (F-statistic):          1.84e-159
Time:                        20:06:35   Log-Likelihood:                -11599.
No. Observations:                1863   AIC:                         2.320e+04
Df Residuals:                    1860   BIC:                         2.322e+04
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             57.7444      6.451      8.951      0.000      45.092      70.397
Lag_Flow           0.7625      0.026     29.270      0.000       0.711       0.814
pedestrian_log    -7.9835      2.792     -2.860      0.004     -13.459      -2.508
==============================================================================
Omnibus:                      334.262   Durbin-Watson:                   2.050
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1280.740
Skew:                           0.837   Prob(JB):                    7.78e-279
Kurtosis:                       6.701   Cond. No.                         533.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 13: OLS results

- Interpretation of coefficients:

· Lag flow: This means that if the average flow of the neighbors is 1 unit higher, then the own flow measured in this detector is 0.76 cars per hour higher.

· Pedestrian: The interpretation of this variable is different, since we took the log of this variable. For every 1% increase in pedestrian, the flow decreases by about 9 cars per hour. For x percent increase, the coefficient has to be multiplied by log(1.x). The general effect is that the flow is lower when it's close to pedestrian areas, such as shopping streets or squares.

- Error analysis:

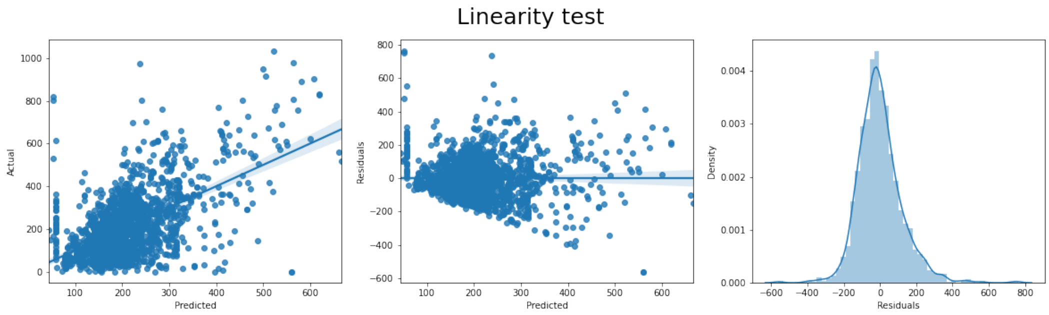

Figure 14: Linearity test

- The residuals have a widening pattern when plotted against the predicted flows. This means that there are some predicting variables still missing.

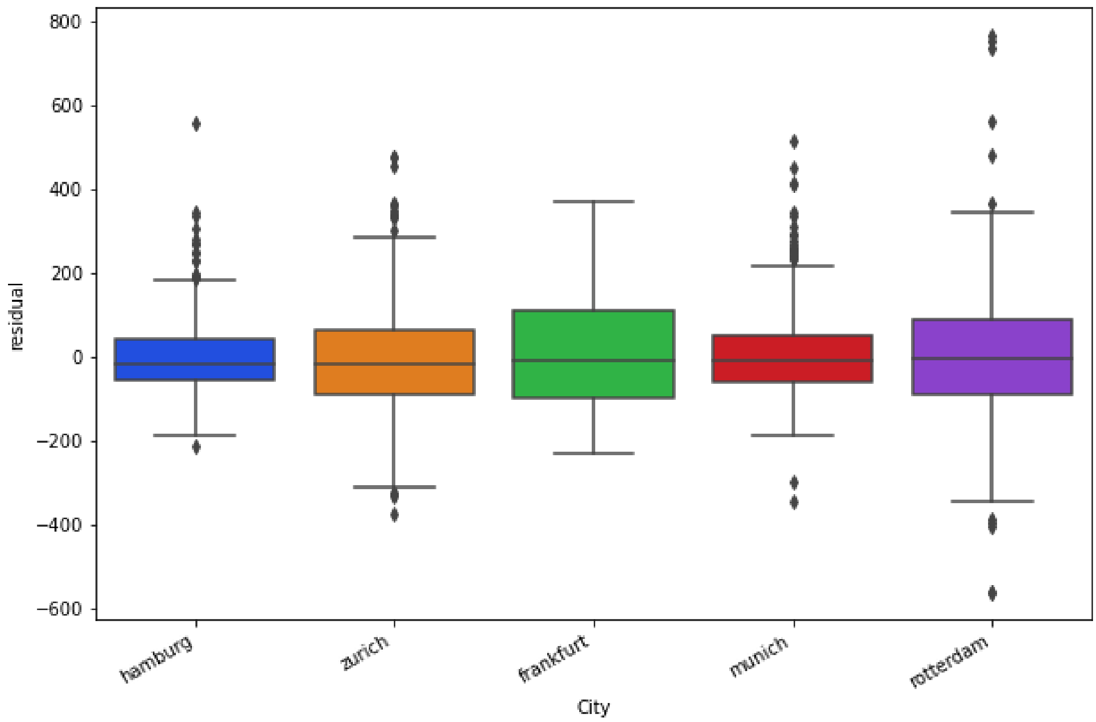

Figure 15: Residuals per city results

On the test set the following scores were found

· R2-score 3: 0.21776849111321606

· MAE 3:91.28649138599641

· rmse 3: 123.28969991409619

The R2-score is quite low and the Mean Absolute Error and Root Mean Square Error are quite high, especially since the predicted flow is sometimes more than two times the actual flow with these errors.

4.2.3 Conclusion regression model

- Based on these results there doesn't seem to be a strong relation between the alternative transport options and the mobility flows, since most of them are not significant.
- Only pedestrian was significant. Pedestrian areas are often used for example for shopping and thus not really offering an alternative travel option. On the other hand it might suggest that built environments that are made "walkable" with pedestrian areas and are less accessible for cars, make people walk, use public transport or go by bike in those parts of the city.
- The flows are spatially correlated, so there is some spatial relation to why some flows are higher than others.
- The model doesn't have a lot of prediction power, the errors are quite high and there seem to be a lot of variance unexplained.

4.3 PCA

To create a single score for the built environment for each city a principal component analysis was conducted on the standardized and averaged counts of POI's per city. This single score was correlated with the sustainability factor of each city to get an indication of the relation between the built environment, mobility and sustainability of cities.

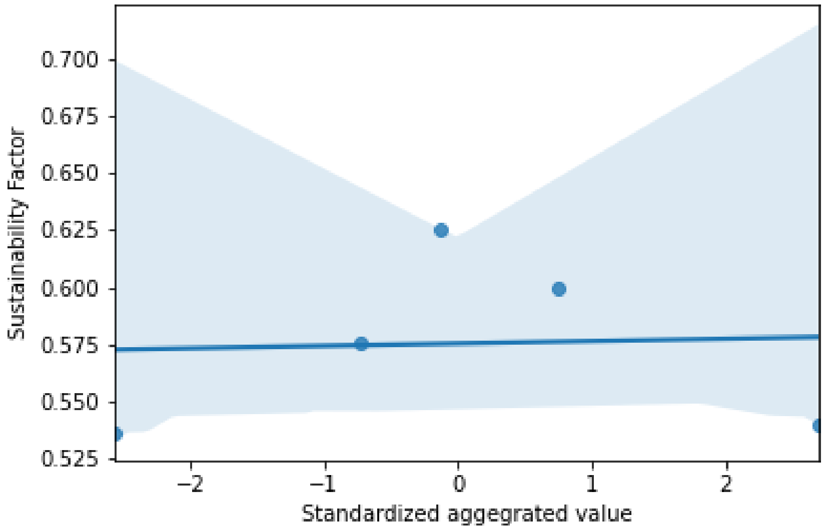

Figure 16: Residuals per city results

- The plot of the correlation between the sustainability factor and the standardized pca score of the poi counts don't show a clear correlation.
- The confidence interval is very wide, since there are only five points, the five cities.
- This supports the point that there is not a strong relation between the alternative transport options and car flows, because if there was then there would be a higher sustainability factor for a city with a higher standardized and averaged poi count.

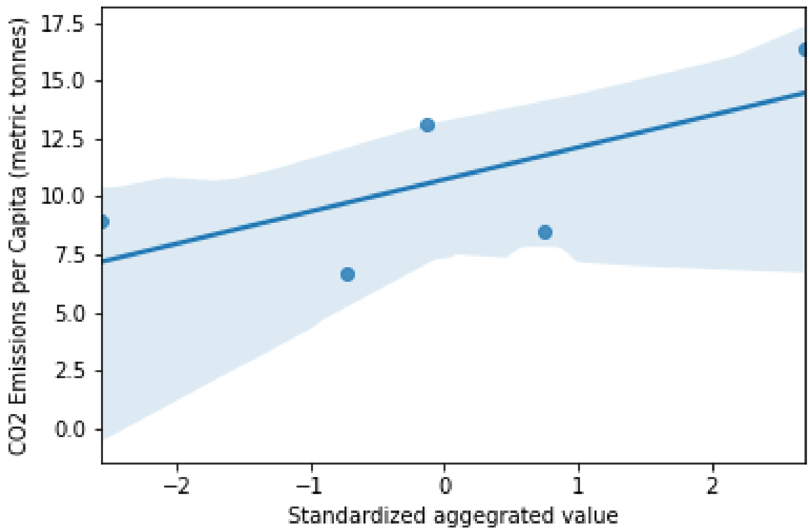

Figure 17: Residuals per city results

- The plot of the correlation between the CO2-emission per capita and the standardized pca score of the poi counts shows a possible positive correlation.
- This indicates that there might still be a relation between the chosen points of interest and being a sustainable city that has less CO2-emission.
- The confidence interval is very wide, since there are only five points, the five cities. This is why it can't be said for certain that the stated correlation is correct.

5 Conclusion

5.1 Findings

- The knowledge gap addressed was whether the alternative transport options, measured as points of interest, would result in lower car traffic flows in European cities, measured with detectors on roads. And whether making changes in this built environment of alternative transport options could result in a more sustainable city.
- The following research question was defined: What is the correlation between measured traffic flows and the amount of poi's related to alternative forms of transport within 250m?
- The regression analysis showed that most counts of points of interest were not significant, except for pedestrian. The network flow is however spatially correlated with flows from other detectors in a radius of 250 meters.
- The regression analysis showed that there was little connection between the built environment in a distance of 250 meters, measured as points of alternative travel options, and the car flows.
- There was a correlation between the pedestrian areas and the flows. In areas with more pedestrian areas, the traffic flows are lower.
- There is also some spatial correlation, between the detectors. This is shown in the LISA analysis (section 4.1) and the spatial lag of flow is significant in the regression. There were areas with a lot of high values clustered and others with low values clustered. This could indicate that there is some relation between the location and built environment to the flows, however it is not well explained by the part of the built environment that we researched.
- The lack of correlation between the aggregated score of the point of interest per city and the sustainability factor confirmed the lack of a relation between the alternative transport options and the car flows.
- The correlation with the CO2-emissions, however, did show a positive correlation. This would mean that cities with relatively more alternative transport options will result in less CO2-emissions.

5.2 Implication and explanation of findings

- The results fill the gap of Wang et al. (2018) that suggested a correlation between the flow and built environment based on detector data en the close by point of interest for a big city in China, but couldn't say whether this correlation would exist in European cities.

  · This correlation does exist in European cities, but not all significant factors of Wang et al. (2018) were included in this analysis. For example hospitals were not taken into account in our analysis. We only focused on the alternative transport as points of interest.

  · Adding hospitals and other points of interest could help with the unexplained variance in our regression model.

- Research suggests that to make a city walkable, it needs to have a good pedestrian network, but also it needs to be safe and have plentiful destinations(R. Rafiemanzelat, M. Emadi, A. Kamali, 2017). According to this research, a more walkable city results in a more sustainable city.

  · The correlation that is found between the pedestrian areas and the flows on streets, seem to suggest that people choose to walk when there are many pedestrian areas.

  · These areas seem to be located in centres of cities, where it is often safe to walk and which has plentiful destinations.

  · An alternative explanation could be that these parts of the city are made less accessible to cars. Cars are diverted around the center instead of going through the center to reach their destination. This could indicate that even though the flows are lower around pedestrian areas, the total flows of the city do not decrease. Maybe even increase, since cars cannot take the most direct route. This could be an interesting aspect for future research.

- Other research suggested that the presence of more high quality bicycle lanes, footpaths and public transport access points, make these forms of transport more attractive (Parker et al.,2013;Murray et al., 1998), which could reduce the car traffic flows.

· Our analysis suggests that there is no predictive relation between the amount of bicycle lanes and public transport stations and the traffic flows, even though this was suggested in literature.

· This could be because we only looked at detectors and the points of interest within 250 meters distance and not at the routes people were taking. There still might be a relation between the alternative travel options, such as public transport and by bike, for routes and the traffic flows.

· That way you look at point of destination instead of nearby points.

- As mentioned earlier, in the past there has been a lot of research on the relation between sustainable transport and the built environment (Borrego et al., 2006;Banister, 2005;Banister, 2011;Ewing and Handy, 2009). Those studies focused primarily on determining the specific characteristics of a city or measures that make alternative forms of transport more attractive choices. Most of them did not study the correlation between measured traffic flows and the existing infrastructures for alternative forms transport nearby. Only (Yang and Diez-Roux, 2012) used a similar method for a city in China.
- In this study, weak but significant, correlations were found between traffic flows and the presence of POI's. Those findings could help in the decision making process to select promising options to reduce traffic flows in the city by expanding the infrastructures for alternative transport. However, additional research in which more European cities are included is necessary to do any detailed recommendations.

5.3 Final conclusion

- Creating sustainable cities is important to tackle the global warming. To make cities more sustainable, people have to make more sustainable travel choices, since traffic emissions contribute up to 91% of total emissions in European city centres (Rivas et al., 2020).
- One of the causes of choice of travel mode, might be in the built environment (Ewing and Handy, 2009). Stimulating people to use more sustainable travel modes, than by car could be done through changing the built environment.

- Understanding how offering alternative travel options relate to the flow of cars helps to guide policy on changing the built environment to reduce traffic flows.
- From the analysis, it can be concluded that creating more pedestrian areas in a city, that have many destinations at a walking distance could result in a reduce in traffic flow.
- However, these relations are not very strong. There are probably other aspects of the built environment that could explain this relation for areas with high flows and areas with low flows better.
- In the current analysis only the counts of alternative transport options are used as a predicting variable for the network flow. Perhaps the relation between amount of alternative transport options and the choice of the mode of transportation would reflect the reduction in total traffic in the city better. This would be an interesting approach for a continuing research.
- This analysis contributes only a small part of creating a better understanding of the built environment and the impact thereof on traffic flows. A lot of research still remains to be done.

References

Banister, D. (2005). *Unsustainable transport: City transport in the new century*. Taylor & Francis.

Banister, D. (2011). Cities, mobility and climate change [Special section on Alternative Travel futures]. *Journal of Transport Geography*, *19*(6), 1538–1546. https://doi.org/https://doi.org/10.1016/j.jtrangeo.2011.03.009

Borrego, C., Martins, H., Tchepel, O., Salmim, L., Monteiro, A. & Miranda, A. (2006). How urban structure can affect city sustainability from an air quality perspective [Urban Air Quality Modelling]. *Environmental Modelling Software*, *21*(4), 461–467. https://doi.org/https://doi.org/10.1016/j.envsoft.2004.07.009

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R. & Rice, K. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, *11*(4), 048002. https://doi.org/10.1088/1748-9326/11/4/048002

EC. (2020). *Urban mobility*. Retrieved November 3, 2020, from https://ec.europa.eu/transport/themes/urban/urban_mobility_en#:~:text=Urban%20mobility%20accounts%20for%2040,of%20other%20pollutants%20from%20transport.

Ewing, R. & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, *14*(1), 65–84.

Forsyth, A. (2015). What is a walkable place? the walkability debate in urban design. *Urban design international*, *20*(4), 274–292.

Garcıa-Palomares, J. C., Gutiérrez, J. & Cardozo, O. D. (2013). Walking accessibility to public transport: An analysis based on microdata and gis. *Environment and Planning B: Planning and Design*, *40*(6), 1087–1102.

Loder, A., Ambühl, L., Menendez, M. & Axhausen, K. W. (2019). Understanding traffic capacity of urban networks. *Scientific reports*, *9*(1), 1–10.

Marinov, M. B., Topalov, I., Gieva, E. & Nikolov, G. (2016). Air quality monitoring in urban environments. *2016 39th International Spring Seminar on Electronics Technology (ISSE)*, 443–448.

Murray, A. T., Davis, R., Stimson, R. J. & Ferreira, L. (1998). Public transportation access. *Transportation Research Part D: Transport and Environment*, *3*(5), 319–328. https://doi.org/https://doi.org/10.1016/S1361-9209(98)00010-8

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

Parker, K. M., Rice, J., Gustat, J., Ruley, J., Spriggs, A. & Johnson, C. (2013). Effect of Bike Lane Infrastructure Improvements on Ridership in One New Orleans Neighborhood. *Annals of Behavioral Medicine*, *45*(suppl$_1$), S101–S107. https://doi.org/10.1007/s12160-012-9440-z

R. Rafiemanzelat, M. Emadi, A. Kamali. (2017). City sustainability: the influence of walkability on built environments. *Transport Research Procedia*, *24*, 97–104. https://doi.org/https://doi.org/10.1016/j.trpro.2017.05.074

Rivas, I., Beddows, D. C., Amato, F., Green, D. C., Järvi, L., Hueglin, C., Reche, C., Timonen, H., Fuller, G. W., Niemi, J. V., Pérez, N., Aurela, M., Hopke, P. K., Alastuey, A., Kulmala, M., Harrison, R. M., Querol, X. & Kelly, F. J. (2020). Source apportionment of particle number size distribution in urban background and traffic stations in four european cities. *Environment International*, *135*, 105345. https://doi.org/https://doi.org/10.1016/j.envint.2019.105345

Wang, S., Yu, D., Ma, X. & Xing, X. (2018). Analyzing urban traffic demand distribution and the correlation between traffic flow and the built environment based on detector data and pois. *European Transport Research Review*, *10*(2), 50.

WBCSD. (2001). *World mobility at the end of the twentieth century and its sustainability.* http://web.mit.edu/aeroastro/sites/waitz/publications/WBCSD.report.pdf

Yang, Y. & Diez-Roux, A. V. (2012). Walking distance by trip purpose and population subgroups. *American Journal of Preventive Medicine*, *43*(1), 11–19. https://doi.org/https://doi.org/10.1016/j.amepre.2012.03.015

Appendices

A Appendix A

A.1 Visual inspection of correlation between flow and POI neighbours

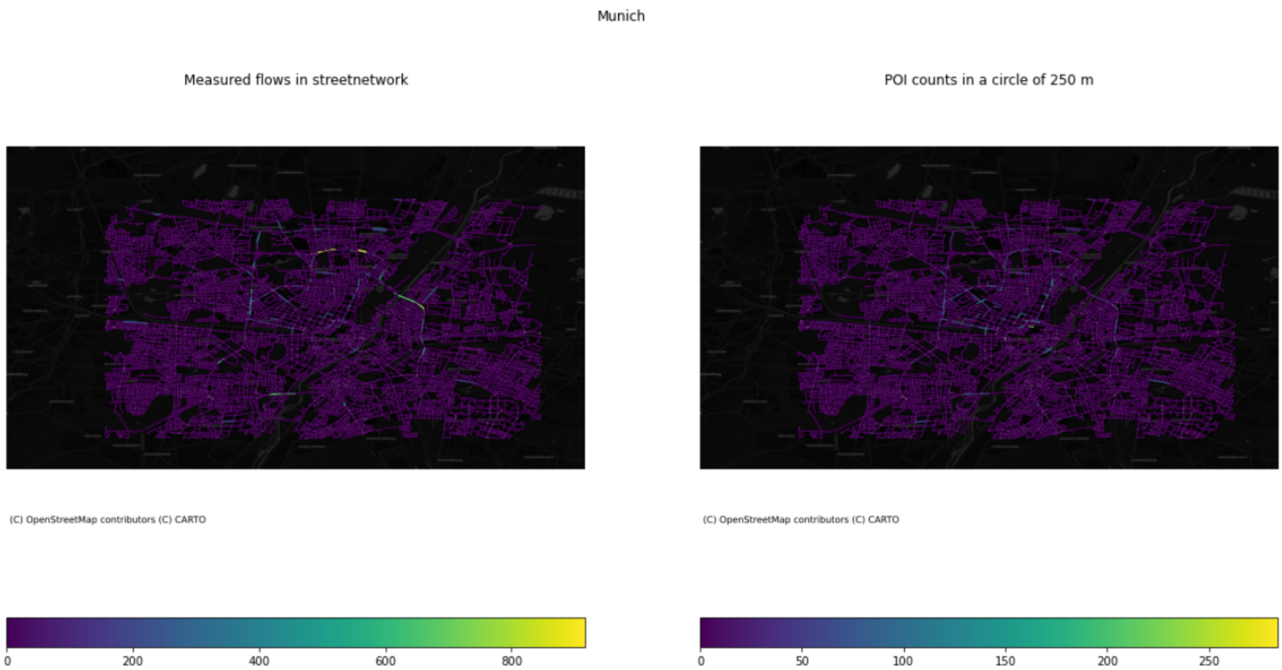

Figure 18: Flows, and amount of POI's per edge in Munich

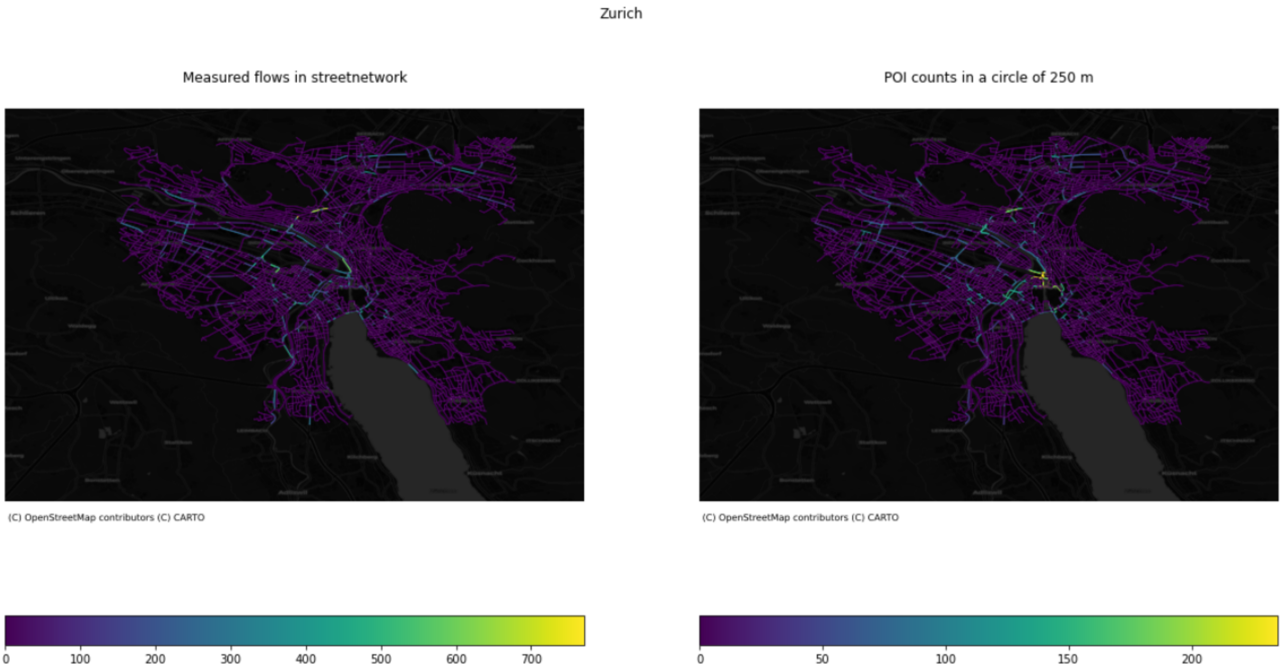

Figure 19: Flows, and amount of POI's per edge in Zurich

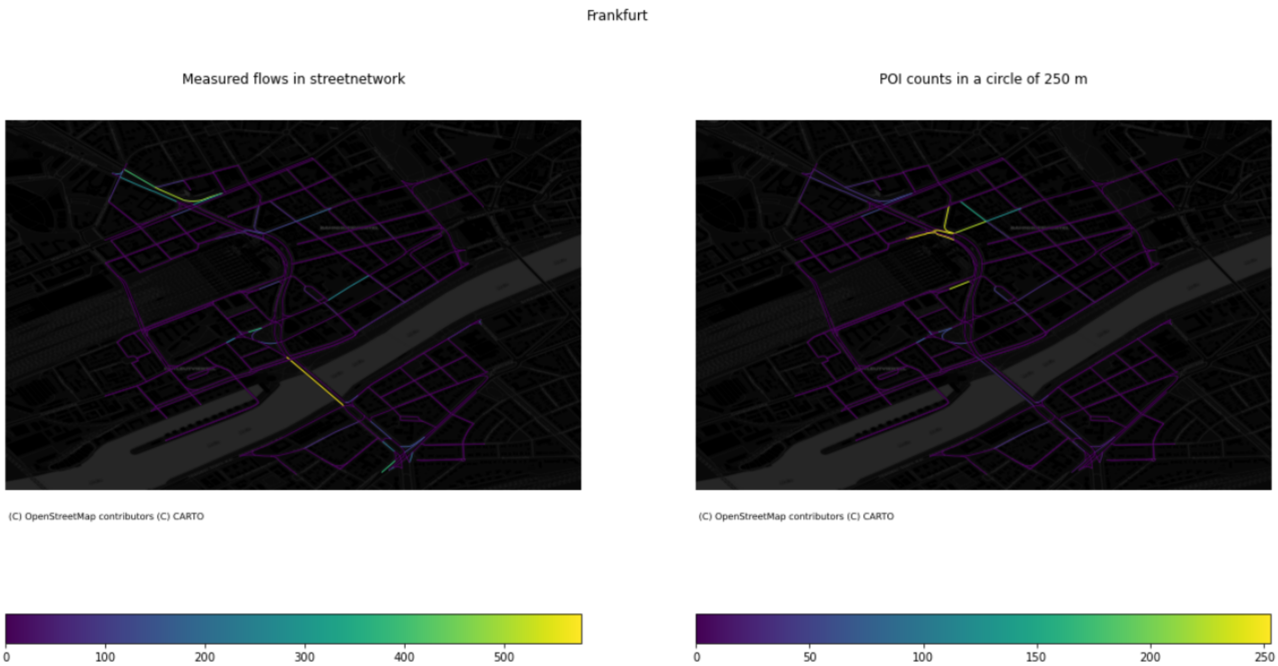

Figure 21: Flows, and amount of POI's per edge in Frankfurt

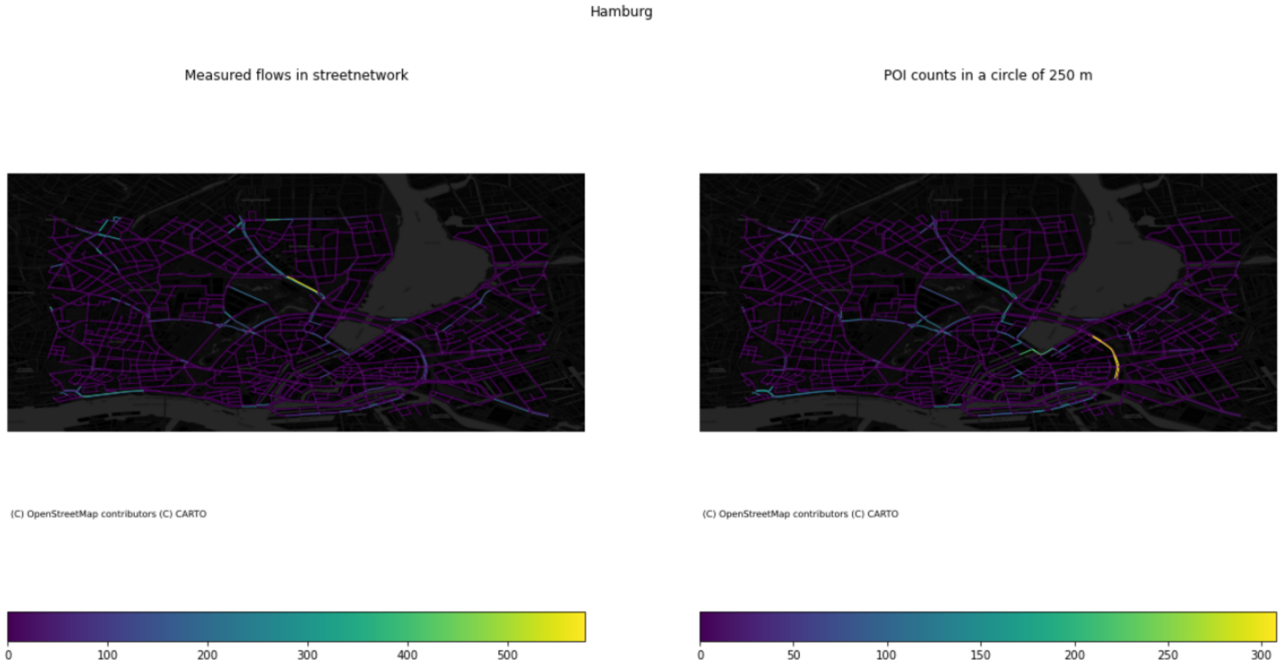

Figure 20: Flows, and amount of POI's per edge in Hamburg

A.2 Appendix B

A.2.1 First model iteration

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   flow   R-squared:                       0.079
Model:                            OLS   Adj. R-squared:                  0.075
Method:                 Least Squares   F-statistic:                     17.69
Date:                Tue, 10 Nov 2020   Prob (F-statistic):           2.08e-28
Time:                        10:36:37   Log-Likelihood:                -11888.
No. Observations:                1863   AIC:                         2.380e+04
Df Residuals:                    1853   BIC:                         2.385e+04
Df Model:                           9
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const             177.1601      6.051     29.277      0.000     165.292     189.028
station           -22.6534      5.318     -4.260      0.000     -33.083     -12.223
cycleway           -1.6305      0.683     -2.386      0.017      -2.971      -0.290
footway             0.5303      0.108      4.918      0.000       0.319       0.742
pedestrian         -1.7916      0.491     -3.651      0.000      -2.754      -0.829
footwayplatform    10.6612     13.658      0.781      0.435     -16.126      37.449
City_frankfurt     63.4750     13.824      4.592      0.000      36.363      90.587
City_hamburg      -45.8559      8.008     -5.726      0.000     -61.562     -30.150
City_munich        50.5475      7.135      7.084      0.000      36.554      64.541
City_rotterdam    108.5972     11.811      9.195      0.000      85.433     131.762
City_zurich         0.3963      7.044      0.056      0.955     -13.418      14.210
==============================================================================
Omnibus:                      520.395   Durbin-Watson:                   2.081
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1532.523
Skew:                           1.424   Prob(JB):                         0.00
Kurtosis:                       6.410   Cond. No.                     4.17e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 7.03e-29. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

A.2.2 Second Model iteration

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   flow   R-squared:                       0.328
Model:                            OLS   Adj. R-squared:                  0.325
Method:                 Least Squares   F-statistic:                     100.5
Date:                Wed, 11 Nov 2020   Prob (F-statistic):          5.28e-153
Time:                        11:56:04   Log-Likelihood:                -11595.
No. Observations:                1863   AIC:                         2.321e+04
Df Residuals:                    1853   BIC:                         2.327e+04
Df Model:                           9
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const             49.7918      7.097      7.016      0.000      35.873      63.711
station           -8.0628      4.563     -1.767      0.077     -17.012       0.886
cycleway          -0.5765      0.584     -0.987      0.324      -1.722       0.569
footway            0.1764      0.093      1.900      0.058      -0.006       0.359
pedestrian        -1.0923      0.417     -2.622      0.009      -1.909      -0.275
Lag_Flow           0.7471      0.029     26.208      0.000       0.691       0.803
City_frankfurt    17.1049     11.932      1.433      0.152      -6.297      40.507
City_hamburg       1.6114      7.070      0.228      0.820     -12.254      15.477
City_munich       15.5660      6.235      2.497      0.013       3.338      27.794
City_rotterdam    22.7655     10.552      2.158      0.031       2.071      43.460
City_zurich       -7.2561      6.024     -1.205      0.229     -19.071       4.558
==============================================================================
Omnibus:                      322.802   Durbin-Watson:                   2.055
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1200.352
Skew:                           0.816   Prob(JB):                    2.22e-261
Kurtosis:                       6.578   Cond. No.                     1.31e+18
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 6.15e-29. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```