PREDICTING NUMBER OF EMERGENCY AMBULANCE CALLS THROUGH SOCIO-ECONOMIC INDICATORS FOR THE HAGUE NEIGHBOURHOODS USING MULTI-REGRESSION.

Final assignment

M.Sc. Engineering and Policy Analysis
EPA 1316 Introduction to Urban Data Science

By

Tim Beens

Auriane Tecourt

Alma Liezenga

Anna Noteboom

November 2020

Professor: Dr. Trivik Verma

TU Delft Technology, Policy and Management

Msc Engineering and Policy Analysis
Faculty Technology, Policy, and Management
Technische Universiteit Delft

ABSTRACT

Socio-economic indicators have long been suspected to be a viable predictor for ambulance emergency calls [2]. Recent research even shows valid predictions of emergency medical visits based on these indicators in Belgium [15]. This study aimed to extend this work by predicting the number of ambulance calls for the city of The Hague (The Netherlands) based on a carefully selected set of socio-economic indicators. After a study of related works, a first selection of 20 potential indicators was made, retrieved from open-source data by the Dutch CBS and Den Haag Cijfers, combined with private data on emergency calls. Through exploratory data analysis and modelling the selection was slimmed down to 2 final predictors: number of crimes and population size. The multi-regression model was trained on data from 2017 and tested on data from 2018. This resulted in a Root Mean Square Error of $RMSE = 192$ and an R^2 of 0.61. This research suggests that socio-economic indicators can impact the number of ambulance calls to some extent but that, in the case of the Hague, only very few of them are relevant. Furthermore, the model may be good for being based on so few predictors, but it is by far not accurate enough to base the distribution of ambulance resources on it. These results support further investigation into the prediction of ambulance calls with the help of socio-economic indicators.

CONTENT

1 INTRODUCTION

Relations can be identified between socio-economic indicators, such as income, education level, migration background, and general level of health on one hand and ambulance calls on the other. This relation is currently not employed to allocate ambulance resources and predict the expected ride time in the Netherlands [20] because it is uncertain whether socio-economic indicators can actually be used as a predictor for ambulance calls. In this report a prediction model will be developed for ambulance calls in the neighbourhoods of The Hague based on socio-economic indicators to evaluate their usefulness for predicting ambulance calls.
The corresponding research question was as follows: can socio-economic factors accurately predict ambulance calls for the neighbourhoods of The Hague? Two datasets were used to answer this question:

- A subset of a dataset on emergency calls of The Netherlands collected from January 2017 to September 2020. This data is, unfortunately, not open source yet.
- Open-source data from the CBS and Den Haag in Cijfers on socio-economic indicators in The Hague neighbourhoods.

The final model based on a multi regression employed two indicators to predict the number of ambulance calls: population size and number of crimes. It had a root mean square error $RMSE = 192$ and a coefficient of determination R^2 of 0.61. Therefore, it is suggested that socio-economic indicators can impact and, to some degree, predict the number of ambulance calls. However, this prediction in its current form was not reliable enough to use for redistribution of ambulance resources. These results supported further investigation into considering socio-economic indicators for redistribution of ambulance capacity. A reliable model might be produced through the use of more indicators and a training over a longer period of time.

The results of this study are relevant to urban planning and allocation of ambulance resources, specifically in the city of The Hague. Further development of these types of models might enable policy makers to redistribute ambulance resources in an efficient way and to identify socio-economic indicators that should be altered through targeted policy to relieve ambulance demand.

In conclusion, the research question was be answered by stating that socio-economic indicators could be used as predictors for ambulance calls in the city of The Hague. However, the data and models shown in this report are not sufficiently accurate to use for the allocation of ambulance resources.

2 RELATED WORK

There is a need to identify potential socio-economic factors that influence ambulance demand. In this section, the state of the art literature on these potential predictors is discussed. General information on the Dutch emergency response and ambulance call system is also provided. Understanding the distribution of ambulance demand and identifying predictors for a better distribution enables the accurate prediction of ambulance calls.

2.1 AMBULANCE CALLS IN THE HAGUE

Ambulance calls through the emergency response number 112 are divided into A1, life-threatening and urgent, A2, non-life-threatening but urgent and B, non-life-threatening and non-urgent situations [20]. The ambulance response time is dependent on this characterisation as well as the region's availability of ambulances.

The Dutch RIVM (National Institute for Health and Environment) is tasked with the distribution of ambulances and ambulance postings. The RIVM distributes ambulances and postings yearly based on a frame of reference. The distribution is established based on mathematical models and equations, using data from previous years. A recent development in this frame of reference suggested adding new variables to the equation, with special attention to how many citizens can be reached within 12 minutes [16]. This model, however, does not include socio-economic factors.

The region 'Haaglanden', of which The Hague is a (major) part, has been facing slight increases in demand for ambulance services from 2018 to 2019, but also faces an increase in ride length for all types of situations [20]. Though the trend is small it can be noted over several years: the increase in ride length is apparent in 2019 and 2018 while the increase in demand is apparent over the period 2012-2019 [20, 19, 18, 17].

This increase in demand and ride length poses a challenge for Haaglanden, and in particular its major city The Hague. The Hague should attempt to distribute the ambulances and postings that it has in an efficient and fair way, providing their citizens with a safe and responsive healthcare system.

2.2 SOCIO-ECONOMIC INDICATORS

Predicting ambulance calls or demand is not a novel concept, one early work by [2] already included socio-economic factors of residents in a prediction for ambulance demand. This work concluded that low-income and migrant groups tended to use the ambulance service more often. The analysis also found that areas with large proportions of elderly or children were more likely to use the ambulance service [2]. In more recent work, Noulas predicted ambulance calls across North West England using social indicators. Higher regional levels of deprivation based on a UK-specific score were shown to imply higher volumes of ambulance calls. This deprivation score was established on the basis of income, crime levels, accessibility to education, health deprivation, disability, barriers to housing and quality of the living environment. The work also concluded that daytime populations are a better predictor of ambulance calls than residential population [14]. Lastly, a Belgian study by Philips investigated patient contacts and found that a few subgroups are more likely to seek emergency support: men, people who do not speak the native language of the country, persons with an African nationality and persons who do not have medical insurance [15].

Another indicator that can be used to predict demand for ambulances, is the overall health of the residents or visitors of a neighborhood. Research by Eurostat emphasizes the relation between income and health [12]. One of the reasons for this relation is the higher rate of smokers and lower rate of eating (fresh) fruits and vegetables in low-income groups. This research also concluded that people with a higher level of education are more prone to a healthy life(style) including exercising more regularly, which is a major contributor to good health [12]. Acciai also emphasises the relation between health and level of income. Specifically, this research linked type and physicality of work with level of income and health [1]. Goldman stresses the worrisome relation between level of education and health [13]. Booth adds environmental inequalities and lack of green space to the risk factors for health and shows that large numbers of inhabitants with a migration background, lack of social cohesion and appearance of violence in neighbourhoods are stresses for physical and psychological health [3].

2.3 PROBLEM STATEMENT

As described above, there is a significant amount of literature, explaining the connection between different socio-economic factors and health. This literature implies that these factors could correlate with more need to call an ambulance. What we do not know as of yet, is if we could use this knowledge about the socio-economic factors to predict where ambulance calls will be more frequent in a city like The Hague. Knowing the predictive value of ambulance calls could help the emergency services in the city to prepare for expected stress on their services in different parts of the city. A list of the socio-economic as they have been introduced in the previous section can be found in Appendix A. Here, we limited the list to those factors that can be directly measured.

Research question: In this report, we investigate the predictive value of socio-economic indicators for ambulance calls in the neighbourhoods of The Hague.

Examining the predictive value of these indicators could enable policy makers in The Hague to distribute their ambulances and postings in an effective and fair way, to counter the increase in demand and ride time.

3 EXPLORATORY DATA ANALYSIS

In this chapter, the process of retrieving, scraping and cleaning the data will be discussed. Subsequently, the data will be inspected through exploratory data analysis, which will result in strong suggestions for modelling.

The initial list of socio-economic indicators is based on on the literature review in 2, this data was retrieved from open data sources and than cleaned to identify potential relations with number of emergency calls through the use of scatter plots. Each of these scatter plots gave an idea of the correlation between the variables and the emergency calls. The results of the analysis of the initial variables can be found in Appendix A. Based on the results of the scatter plot analysis, a new subset of indicators was created. This subset contained only the variables that showed a form of correlation with emergency calls. One outlier was removed for this analysis: the neighbourhood Kortenbos, which had an usually high number of emergency calls.

After the scatter plot analysis, a geospatial analysis was performed to analyse the spatial distribution of the independent and dependent variables. The Geo-spatial analysis was performed by plotting the distribution of the variables using choropleths based on shape files of the Hague. Coloured spatial density plots were also created for these variables in order to show the relation between the variable and density of emergency calls in one plot. Lastly, point data analysis on the emergency call data was performed to create a kernel density plot of emergency calls in The Hague. Multiple KDE plots were created based on emergency calls made during different time intervals. This made it possible to identify areas in which the amount of emergency calls are influenced by the time of day.

The chapter is finalised with a discussion on the outliers and the limitations of the analysis.

3.1 DATA RETRIEVAL

Four sources of data for 2017 and 2018 were used to gather all information:

- CBS data: most indicators were retrieved from the public data of CBS. This data provided a reliable wide range of data reported by the Dutch government and most indicators for this analysis were found in this dataset. The data was filtered to only include that of The Hague. The indicators were selected based on the findings in chapter 2. The precise selection of variables can be found in A, accompanied by basic statistical information and sources of these data [7, 8].
- Den Haag in Cijfers data: some indicators analysed in the initial EDA were available through Den Haag Cijfers. This included the number of citizens in each

neighbourhood of the Hague working in the professions considered physically most stressing and the location of religious facilities in the city [11].

- Emergency calls data: the emergency call data used describes the origin and time of each call to emergency services between 2017 and September 2020, of which a subset concerning only 2017 and 2018 was taken. The location of the call was specified through longitude and latitude. We gained access to this data through our course professor: Trivik Verma.
- Shape files of The Hague: to create visual, spatial representations of the data shape files were used. These are open source data, which we received from our course professor as well.

3.2 DATA SCRAPING

The following alterations to the existing dataset were necessary for further analysis:

- Some variables were not directly available, if this was the case they were generated through basic calculations on existing data. This was the case for the percentage of citizens under 14 and the percentage over 65 and is highlighted in A.
- For the year 2018, data on mean income per neighbourhood was missing for 29 neighbourhoods. 29 missing values on a total of 114 neighbourhoods is significant. As the data on this variable was available for the years 2011-2017, the missing data was generated by means of linear regression. A linear regression model was trained on the mean income per neighbourhood from 2011 to 2017, the resulting linear regression line was then extrapolated to 2018 to generate data for the missing neighbourhoods in that year. An example of the model for the neighbourhood Bosweide in The Hague can be found in figure 3.1.
  The resulting value for Bosweide was €49.600 per year.

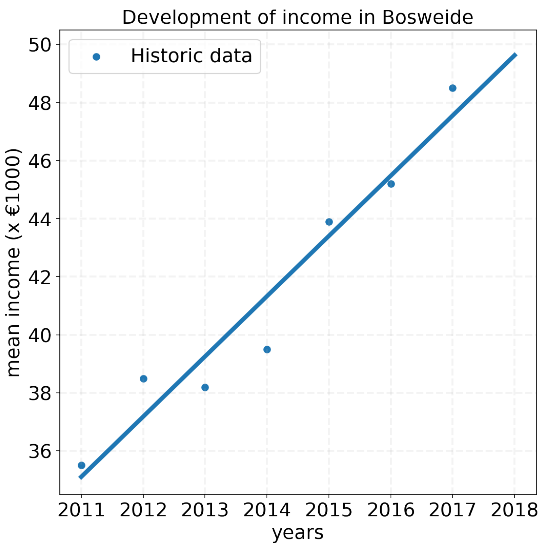


Figure 3.1: Regression model of the mean income in the neighbourhood Bosweide

The complete results of the linear regression analysis can be found in Appendix B

3.3 DATA CLEANING

Data cleaning was carried out using the pandas library in Python. This resulted in four different dataframes. The first two dataframes respectively concerned 2017 and 2018 with the neighbourhoods in The Hague as rows and the socio-economic indicators, number of ambulance calls and geographic information as columns. All numerical data was saved as a float, all text data was saved as a string.

The latter two dataframes contained the information on each ambulance call in respectively the year 2017 and 2018. Each row represented one call to the emergency service, the columns show the exact time, date and position of the call.

The dataset contained four neighbourhoods where nobody lives, these are Oostduinen, Vliegeniersbuurt, De Reef and Tedingerbuurt. As these neighbourhoods do not have any data with regards to demographics, they are unworkable for the analysis so they were excluded from the models. The rows for these neighbourhoods were removed from the dataframe.

3.4 INVESTIGATING RELATIONS

After the dataset was as complete and possible and clean, exploratory data analysis could be performed on the initial list of indicator variables. For each indicator variable, the correlation between the indicator variable and the number of emergency calls was analysed through the use of scatter plots. The scatter plots and explanations can be found in appendix A.

The scatter plot analysis resulted in the identification of nine potential variables useful for the prediction of emergency calls. They can be categorised as followed:

- **Demographic variables:** Number of inhabitants, % of inhabitants under 14, % of inhabitants over 65 and % of non-western citizens
- **Economic variables:** Mean income per inhabitant, % of household in the lowest 40% national income, % of households in the highest 20% national income
- **Other variables:** Number of crimes, Urban density level

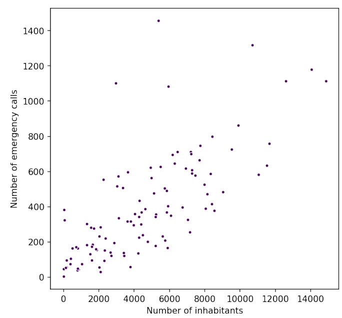
These nine variables all showed a clear form of correlation with the emergency call variable except for the % of inhabitants under 14 years old. This variable does not show any clear sign of linear correlation. It does however show a form of poly-nominal correlation. Because of this and the relevance of this variable based on the literature the variable was included in this selection. The remaining indicators that were analysed did not show any clear form of correlation with number of emergency calls. Further analysis will therefore only look at these nine variables. The scatter plots of the selected predictor variables can be seen in fig 3.2.

Although all nine variables in fig 3.2 show correlation with emergency calls their degree of correlation differs strongly. The number of inhabitants, number of crimes and degree of urbanity show some of the clearest forms of correlation while the mean income per inhabitants, the percentage of households in the lowest 40% national income, the percentage of households in the highest 20% national income and the percentage of non-western citizens are already less correlated with emergency calls. The percentage of inhabitants under 14 and the percentage of inhabitants over 65 show the least correlation. This is especially interesting for the percentage of people over 65 since it was expected that more old inhabitants, with vulnerable health, would result in a higher the number of emergency calls.

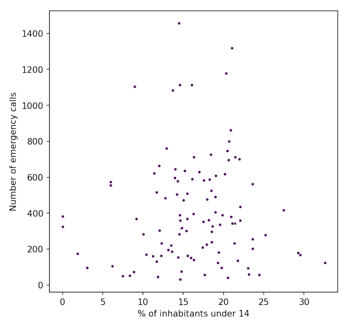
Given the scatter plots the variables can be divided into positively correlating variables and negatively correlating variables. Positively correlating meaning that if the independent variable increases the dependent variable seems to increase as well and negatively correlating meaning that if the independent variable increases the dependent variable decreases.

The positively correlating group consists of the number of inhabitants, number of crimes, the percentage of households in the lowest 40% national income, the percentage of non-western citizens and the degree of urbanisation. The degree of urbanisation is seen as positively correlating because 1 is the highest degree of urbanisation and 5 the lowest. So the higher the degree of urbanisation the higher the amount of emergency calls.

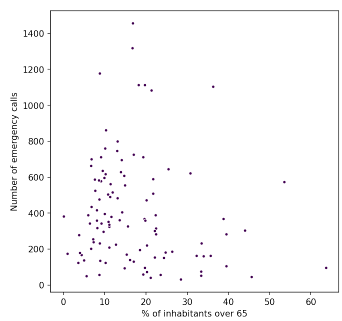
The negatively correlating group consists of the mean income per inhabitant, the percentage of inhabitants over 65 and the percentage of households in the highest 20% national income.



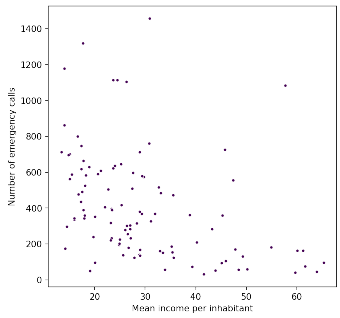
(a) Number of inhabitants



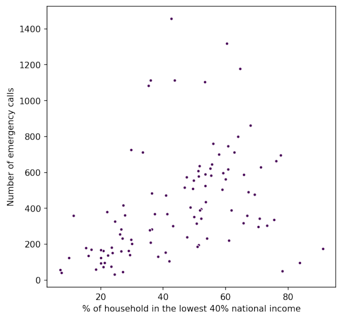
(b) Percentage of inhabitants under 14



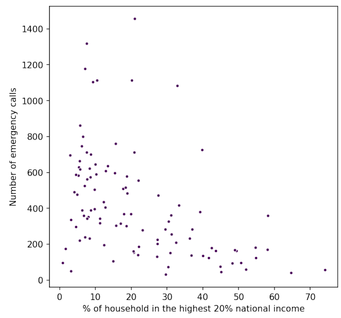
(c) Percentage of inhabitants over 65



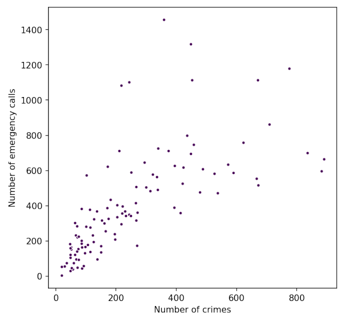
(d) Mean income per inhabitant



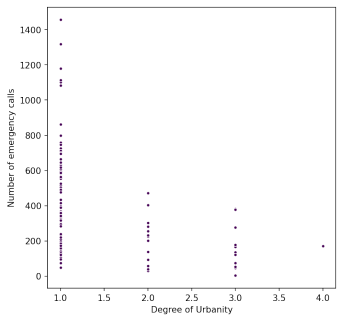
(e) Percentage of households in the lowest 40% of national income



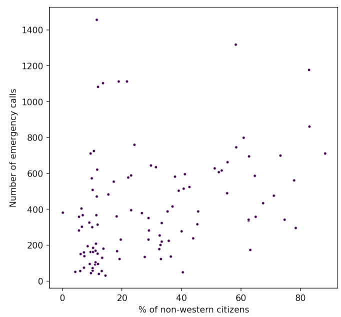
(f) Percentage of households in the highest 20% of national income



(g) Number of crimes



(h) Degree of urbanity



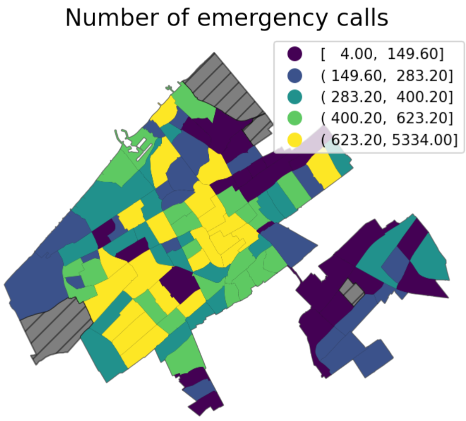
(i) Percentage of non-western citizens

Figure 3.2: Correlation between the different predictors and number of ambulance calls

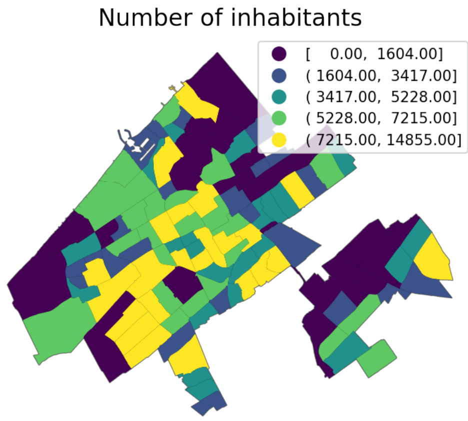
3.5 SPATIAL ANALYSIS

In this section the Geo-spatial distribution of the independent variables and the dependent variable will be analysed. This has been done for each variable by creating a choropleth of The Hague using shapefiles and plotting the value of the given variable on top. This resulted in 10 different choropleths of The Hague in which each neighbourhood has a specific colour that represents the value of the plotted variable. The choropleths can be seen in figure 3.3. The Viridis colour scheme was used to make the plots colour-blind friendly. In this colour scheme yellow represents the highest values and dark purple represents the lowest. neighbourhoods that had missing values for the depicted indicator have been made dark-grey with stripes.

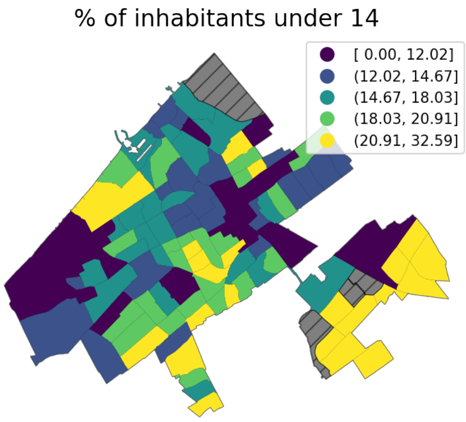
Notice that the colour bar in sub-figure 'j' is continuous. This is to show how uniform the degree of urbanity is in the city of The Hague where almost every neighbourhood has a degree of urbanity one: representing the highest level of urbanisation.



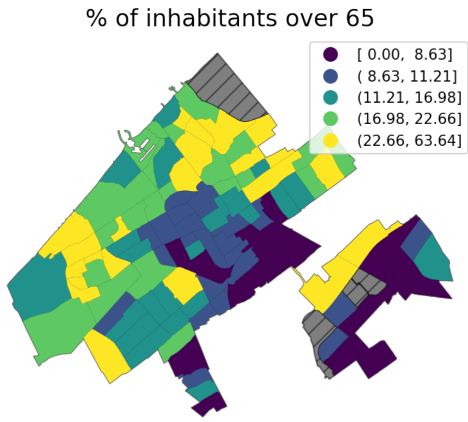
(a) Number of emergency calls



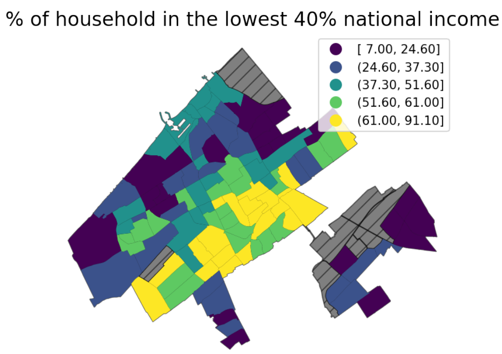
(b) Number of inhabitants



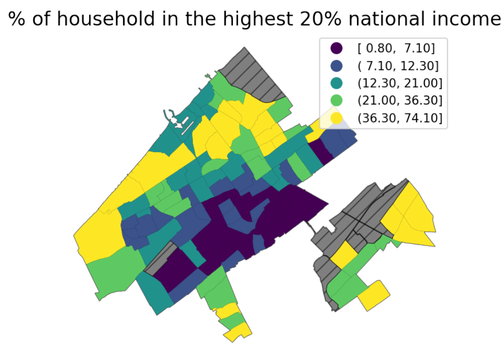
(c) Percentage of inhabitants under 14



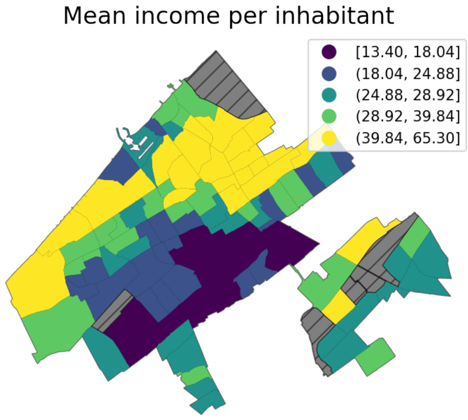
(d) Percentage of inhabitants over 65



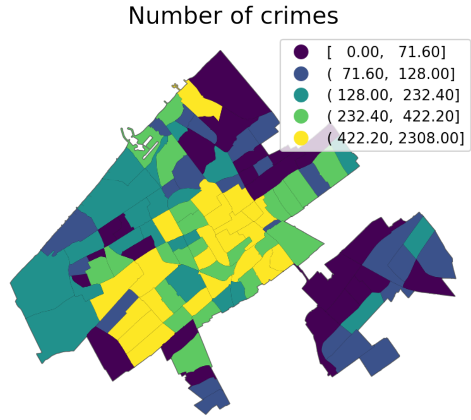
(e) Percentage of households in the lowest 40% of national income



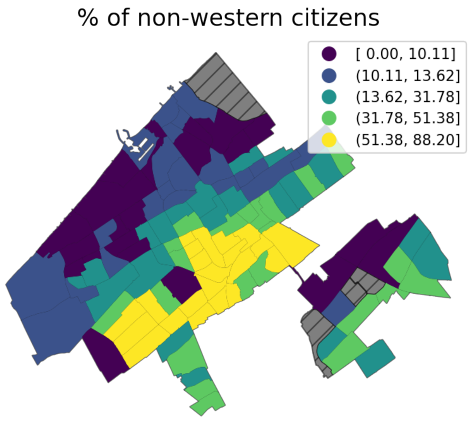
(f) Percentage of households in the highest 20% of national income



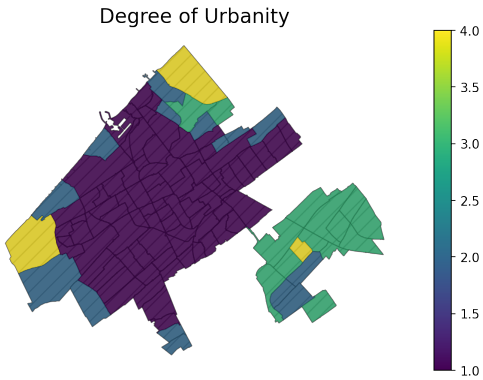
(g) Mean income per inhabitant



(h) Number of crimes



(i) Percentage of non-western citizens



(j) Degree of urbanity

Figure 3.3: spatial distribution of variables to the density of ambulance calls

The choropleth for the emergency calls shows a high concentration of calls around the city centre, the area south west of the city centre and parts in the north west towards the sea. Almost the same areas can be identified in the choropleth for the number of inhabitants and the choropleth for the number of crimes which was to be expected given the results from the scatter plot analysis. The area at sea is a popular beach area for people from The Hague and tourists. This might offer a possible explanation for the high amount of emergency calls in that area even though not many people live there and crime number average.

When observing the spatial distribution of mean income per inhabitant, the percentage of households in the highest 20% national income and the percentage of households in the lowest 40% national income a clear division between high income neighbourhoods and low income neighbourhoods can be identified. While high income neighbourhoods are located in the north/north west areas of The Hague low income neighbourhoods are clearly located in the south/south west of The Hague. When comparing the amount of emergency calls in high income neighbourhoods to those in low income neighbourhoods a small difference can be seem. It seems that low income neighbourhoods have a bit more emergency calls than high income neighbourhoods given the fact that low income neighbourhoods contain more light green areas than high income neighbourhoods. The same conclusions can be drawn for the percentage of non-western citizens which shows the same clear division between north and south as mean income.

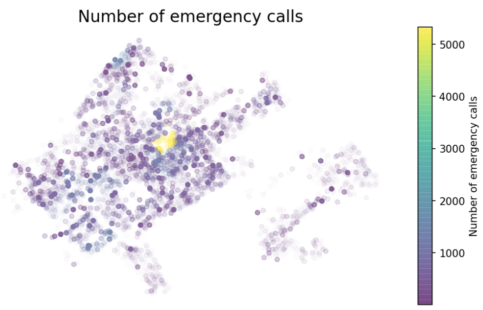
The choropleth for the percentage of inhabitants under 14 doesn't seem to show any strong trends except for the neighbourhoods in the south-east part of The Hague which are almost all yellow. However when the distribution of inhabitants under 14 per neighbourhood is compared to the amount of emergency calls per neighbourhood there does not seem to be a clear relation.

The choropleth of the percentage of inhabitants over 65 shows shows that they are mostly located in the north/north-west of The Hague which is comparable to the spatial distribution of mean income. The big difference however is that the amount of emergency calls seems to be quite low when only looking at the northern most neighbourhoods (excluding the beach area) even though those areas have a high percentage of inhabitants over 65. The choropleths and scatter plots seem to suggest that older people have need of emergency calls less even though the initial literature suggested otherwise.

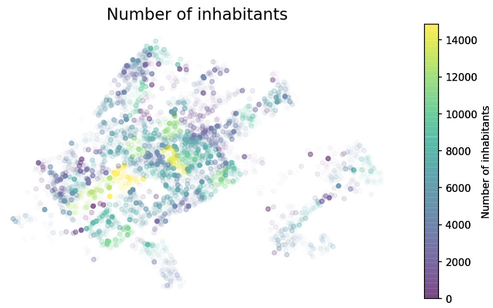
Based on the choropleth analysis the following conclusion could be made, namely that the amount of emergency calls seems to be extremely dependent on the amount of people living in a neighbourhood. Although crime also seems to be a good predictor this conclusion was not made because it is highly plausible that the high crime numbers are themselves also dependent on the amount of people in a neighbourhood. In less densely populated areas the effect of the other socio-economic predictors seems to shine through slightly, especially for the income related variables which might also be related to the percentage of non-western citizens in those neighbourhoods. The age of inhabitants seems to be somewhat useful as a predictor when looking at the percentage of inhabitants over 65 and quite useless when looking at the percentage of inhabitants under 14.

Each plot in figure 3.4 represents the spatial distribution of one of the nine variables in relation to the number of emergency calls. The colour of the dots determines the value of the variable in that area which is based on the same value distribution as the plots in figure 3.3. This means that dark purple represent the lowest values while bright yellow represents the highest values. The amount of emergency calls in each dot is represented by its transparency. The more transparent the dot is the lower the amount of emergency calls there are in that area of The Hague. The less transparent the dot the higher the number of emergency calls. These plots show the comparison made between the different choropleths in figure 3.3 except that each figure immediately shows the relation with the amount of emergency calls per area.

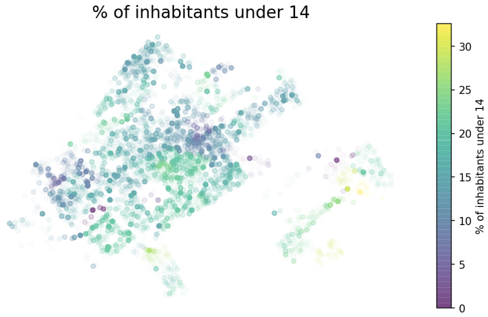
Both the choropleth and spatial distribution plot for the degree of urbanisation showed that almost 80% of The Hague is in the highest level of urbanisation. Since this variable contained almost no variation in it's values it was not very interesting to research. This variable has therefore been dropped from the predictor list.



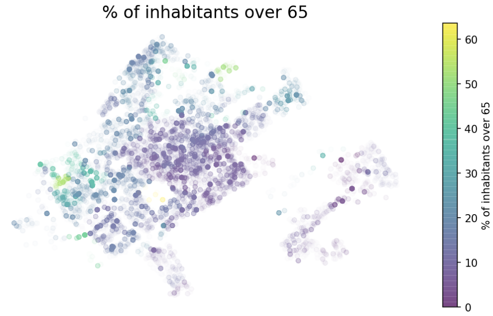
(a) Number of emergency calls



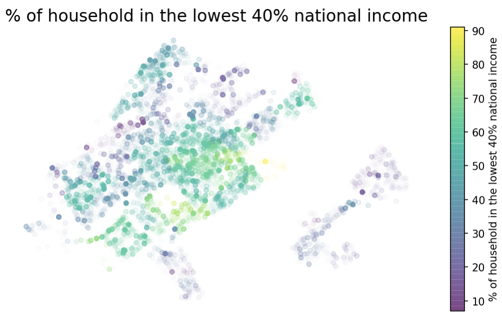
(b) Number of inhabitants



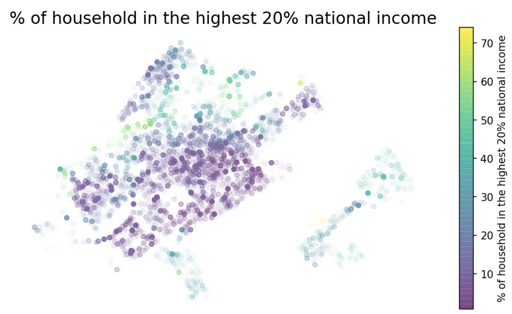
(c) Percentage of inhabitants under 14



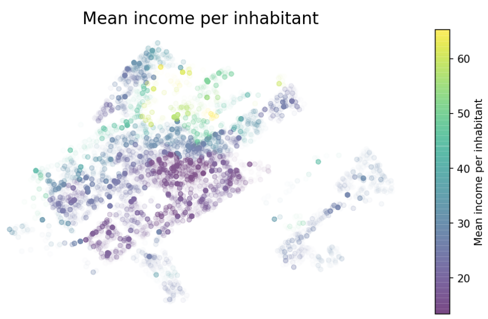
(d) Percentage of inhabitants over 65



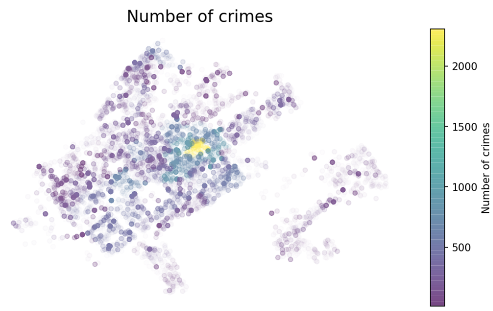
(e) Percentage of households in the lowest 40% of national income



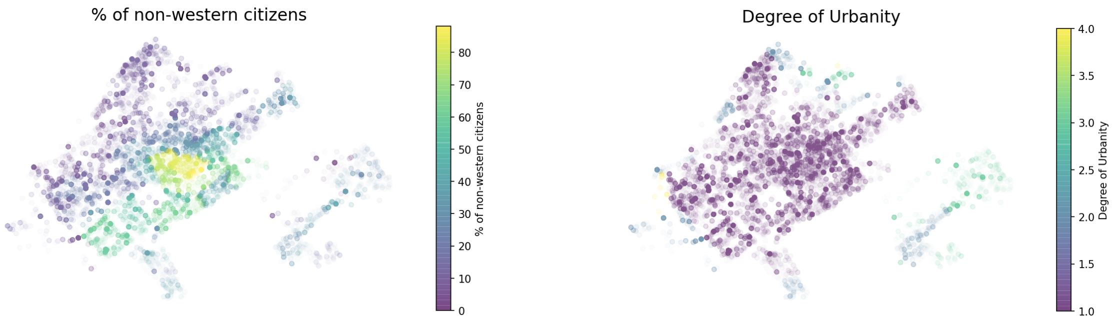
(f) Percentage of households in the highest 20% of national income



(g) Mean income per inhabitant



(h) Number of crimes



(i) Percentage of non-western citizens **(j)** Degree of urbanity

Figure 3.4: sSpatial distribution of variables to the density of ambulance calls

Due to the big dependence on number of inhabitants in a neighbourhood, a choropleth in which the amount of calls is standardised to the amount of emergency calls per 1000 citizens was also made. This choropleth can be seen in figure 3.5. In this choropleth there is a lot less yellow in the city centre. There is however more yellow in the areas north of the city centre which are neighbourhoods with a relatively high percentage of people over 65, the neighbourhood at the beach is still yellow and new hotspots have appeared in the eastern and western neighbourhoods of The Hague. Be mindful however that the neighbourhoods that are now yellow are mostly neighbourhoods that contain low amount of inhabitants. All this choropleth shows is that these neighbourhoods contain relatively high number of emergency calls given their number of inhabitants.

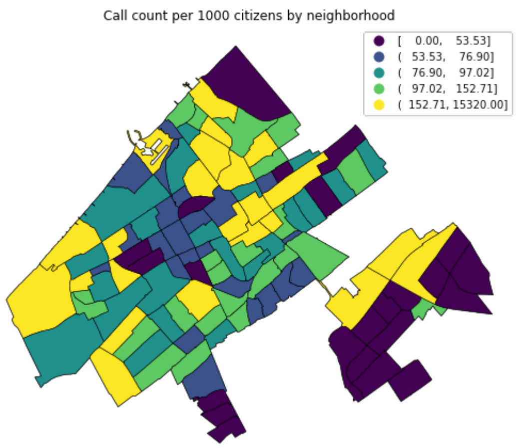


Figure 3.5: Standardised choropleth of the emergency call distribution per 1000 inhabitants

3.6 KDE ANALYSIS

The emergency call data set also provided the time of each emergency call which allowed for the opportunity to analyse time dependent emergency calls and their location.

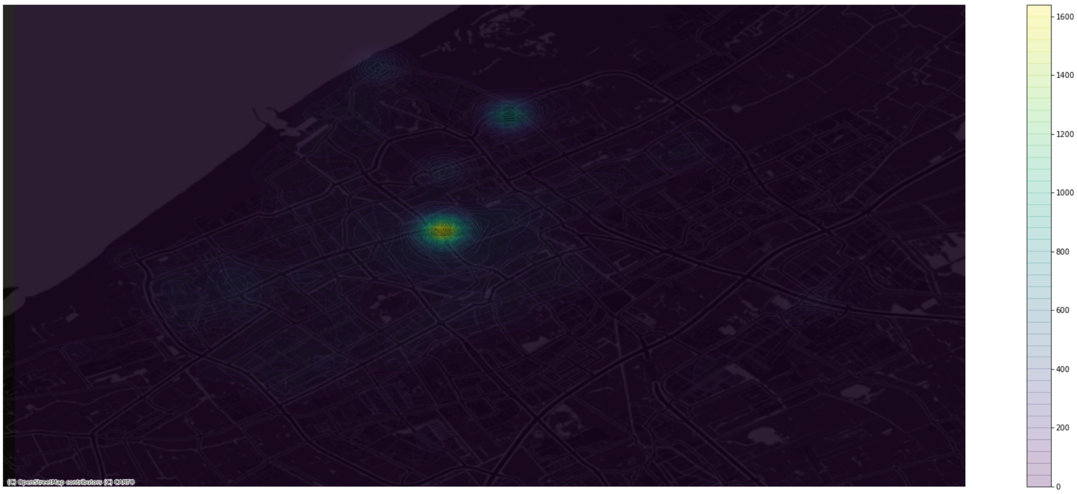
In order to analyse this a kernel density plot was made containing all the emergency calls that were made during a specific time interval. Since the difference between night and day was most interesting, an interval containing the calls between 22:00 pm and 06:00 am and an interval containing the calls between 6:00 am and 22:00 pm was made representing night and day respectively.

The KDE plots can be seen in figure 3.6. The KDE plot of calls during the day shows a small but large concentration of calls made in the city centre and some smaller concentrations of calls in the areas north of the city centre. The rest of the calls seem to be quite evenly spread out over the rest of The Hague.

The KDE plot of calls during the night shows the same four concentrations of calls as during the day although more widely spread out. The most interesting part in this plot however, is that the amount of calls in the rest of the city seem to be higher. This might be because people work in the city centre during the day and go their for shopping etc while at night most people are at home and thus back in their own neighbourhood.

Although it might seem that there are more calls made during the night because the distribution seems to more evenly spread out, this is not the case. When looking at the absolute values of emergency calls, about 80% of the calls is made during the day and only 20% of the calls happen during the night. This difference can also be seen in the corresponding colour bar next to the KDE plots which shows the value per colour.

For further analysis KDE plots for six hour time intervals from midnight to midnight have also been made. These can be found in appendix C. The KDE plot from 18:00 to 24:00 shows a lot of resemblance with the KDE plot from 22:00 to 06:00. This strengthens the idea that the high concentration of calls in the city centre during the day is due to work while after work the calls are more spread out since most workdays are over at 6 pm.



(a) KDE plot of emergency calls between 06:00 and 22:00

(b) KDE plot of emergency calls between 22:00 and 06:00

Figure 3.6: KDE plots of emergency calls during the day and during the night

3.7 OUTLIERS

Two outlier neighbourhoods were present in the data. The first one was Kortenbos, which had 5334 ambulance calls in 2017, while the neighbourhood with the second highest number of ambulance calls (Scheveningen Badplaats) had 1456. Since this outlier skewed the data for all indicators, it was not taken into account for further analysis. Conclusions made at the end will therefore not be applicable to this neighbourhood.

The second neighbourhood that contained an outlier was Zuidwal, where the number of crimes committed was significantly higher than in any other neighbourhood. With a number of crimes of over 2308, it stood out from the rest of the neighbourhoods, where the maximum number of crimes was commit ed in 895 crimes per year. Due to the significance of this outlier, it was not taken into account for the remainder of the analysis based on crimes. Conclusions made at the end will therefore also not be applicable to this neighbourhood and further research should be done in order to identify the reason for this extreme outlier. The second outlier was only excluded in the study of Crimes as a predictor, but otherwise included. Further research should be done in order to identify the reason for these extreme outliers.This will be further discussed in section 5.2.

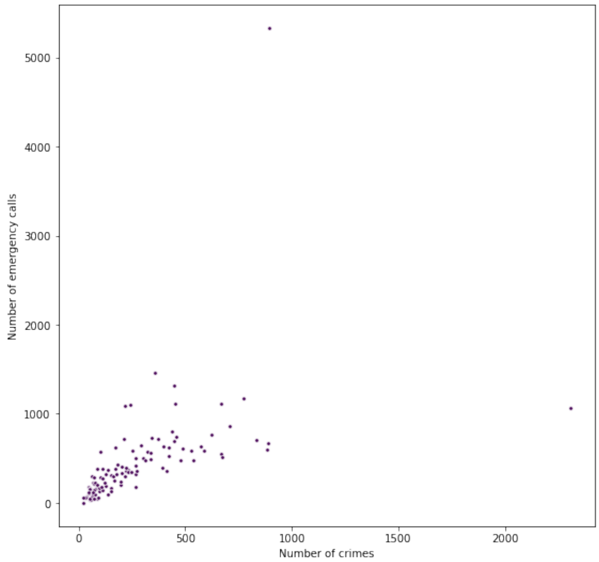


Figure 3.7: Outliers in the dataset

3.8 LIMITATIONS

The identified missing values were dealt with in two different ways while performing the exploratory data analysis which has implications for the study:

- Income data: In 2018 there were 29 neighbourhoods that had no mean income data. This missing data was extrapolated using a regression model that was trained on data from 2012 to 2017. This estimate will deviate a bit from the actual values for 2018, therefore all conclusions our study makes that are based on the mean income are to be taken with a grain of salt.
- Empty neighbourhoods: Some neighbourhoods have no inhabitants. Given the fact that there are ambulance calls coming from these areas, it is already clear that the socio-economic indicators of a neighbourhood can never fully predict the amount of ambulance calls.

Next to missing data, our study is influenced by a number of methodological choices:

- Socio-economic factors of inhabitants were used, but ambulance calls can be made by visitors of the neighbourhood. In 2017 and 2018, people did not stay home all day. We are thus ignoring part of the population that request ambulances.
- Predicting the number of calls was done without taking into account the shifting pattern in calls between day and night. This relates to the previous point in that we do not consider daytime population but only residents.

- Only linear, polynomial and multi regressions were used. This severely limits the scope of our analysis, as other models may be more appropriate to predict ambulance calls.

4 ANALYSIS

In this chapter, the predictors selected after the exploratory data analysis were studied further and the most promising ones were used to develop a final model to predict ambulance demand in the neighbourhoods of the Hague. The results of the model were then analysed.

4.1 EXPERIMENTAL DESIGN

The eight predictors selected after the exploratory data analysis were used individually to train two supervised machine learning algorithms. These algorithms were trained on data from 2017 and tested on data from 2018. Given the evident outlier in emergency calls (Kortenbos) and in crimes (Zuidwal) that greatly affect the quality of the analysis as mentioned in the previous chapter, all models described in this chapter were trained and tested separately with and without those outliers. One model used a linear regression, the other a second-degree polynomial regression. Given how the percentage of population under 14 seemed to follow a third-degree polynomial, an algorithm was developed specifically for that case as well. The models were then compared with each other using their Root Mean Squared Error (RMSE) and their coefficient of determination (R^2). The former is fully explained by its name, the latter isn't. The coefficient of determination uses the division of the Mean Square Error (MSE) of the prediction by the variance σ^2 of the data. This division compares the mean square error to the variance of the data. After all, if a dataset has a high variance, it is to be expected that the model will have a higher MSE. The coefficient of determination is then $R^2 = 1 - \frac{MSE}{\sigma^2}$, so that a higher error results in a lower score. The best R^2-score is as close to 1 as possible.
Two predictors reached higher scores than the others on both datasets (with and without outliers) and were thus selected for the final model: they are the number of crimes and the population of each neighbourhood. On the data without outliers, the number of crimes delivered a score of $R^2_{Crimes} = 0.55$ and $RMSE_{Crimes} = 156$, while the population size reached $R^2_{Pop} = 0.47$ and $RMSE_{Pop} = 227$. The RMSEs tell us that the prediction based on crimes is off by a bit less than one call every second day, while the prediction based on population size is off by a bit more than one call every second day. The third best predictor on that dataset, the percentage of households in the lowest 40% of national income, only reaches a score of $R^2 = 0.24$ ($RMSE = 276$), hence why it was not considered for the final model. The individual scores of the two selected predictors dropped to $R^2_{Crimes} = 0.36$, $RMSE_{Crimes} = 424$ and $R^2_{Pop} = 0.21$, $RMSE_{Pop} = 470$ when the models were trained with the outliers, but they remained the ones with the highest scores. The final model consists of a multiple regression using these two predictors and

was trained and tested with the same data as the models above. It used the following equation 4.1:

$$y = \beta_0 + \beta_{1,1} \cdot x_1 + \beta_{1,2} \cdot x_1^2 + \beta_{2,1} \cdot x_2 + \beta_{2,2} \cdot x_2^2 \tag{4.1}$$

With x_1 the number of crimes and x_2 the population size. All weights and constants β_i were determined by the model. The final, multi regression model had a score $R^2_{Final} = 0.61$ and $RMSE_{Final} = 192$ without outliers. For reference, that score dropped to $R^2_{Final} = 0.39$ and $RMSE_{Final} = 405$ when the two outliers are included. This meant that, on average, the final model was off by about one call every second day. The map of recorded ambulance calls in 2018 is shown with the map of predicted calls by the model without outliers for 2018 in Figure 4.1.

4.2 IN-DEPTH ANALYSIS OF THE FINAL MODEL

The final model based on multi regression lead us to believe that socio-economic indicators play a role in ambulance demand and should thus be used in future predictive models, although more research is needed. This argument is based on the nature of the variables used for the model and on its relative accuracy.

4.2.1 Nature of the selected variables

Using population size to predict ambulance calls did not offer insight into individual socio-economic aspects of neighbourhoods. However, population size was only the second best predictor that was found, and merely improved the accuracy of the model. The number of crimes, though correlated with population size, was our best predictor for ambulance demand and the backbone of our model. One may also think that the model's high R^2-score was caused by the correlation between the chosen predictors and population density. In order to test that theory, the population density of each neighbourhood in 2017 and 2018 was analysed for prediction power just like the others predictors at the beginning of this chapter, with the same machine learning algorithms. The highest-ranking model reached its highest score on the datasets without outliers following a polynomial regression at $R^2_{PopDen} = 0.27$ and $RMSE_{PopDen} = 261$. The former is largely inferior than that of the predictors taken for the final model, the latter only slightly superior. This showed that the strong correlation of the predictors with population density is not the reason for the results achieved with the multi regression model.

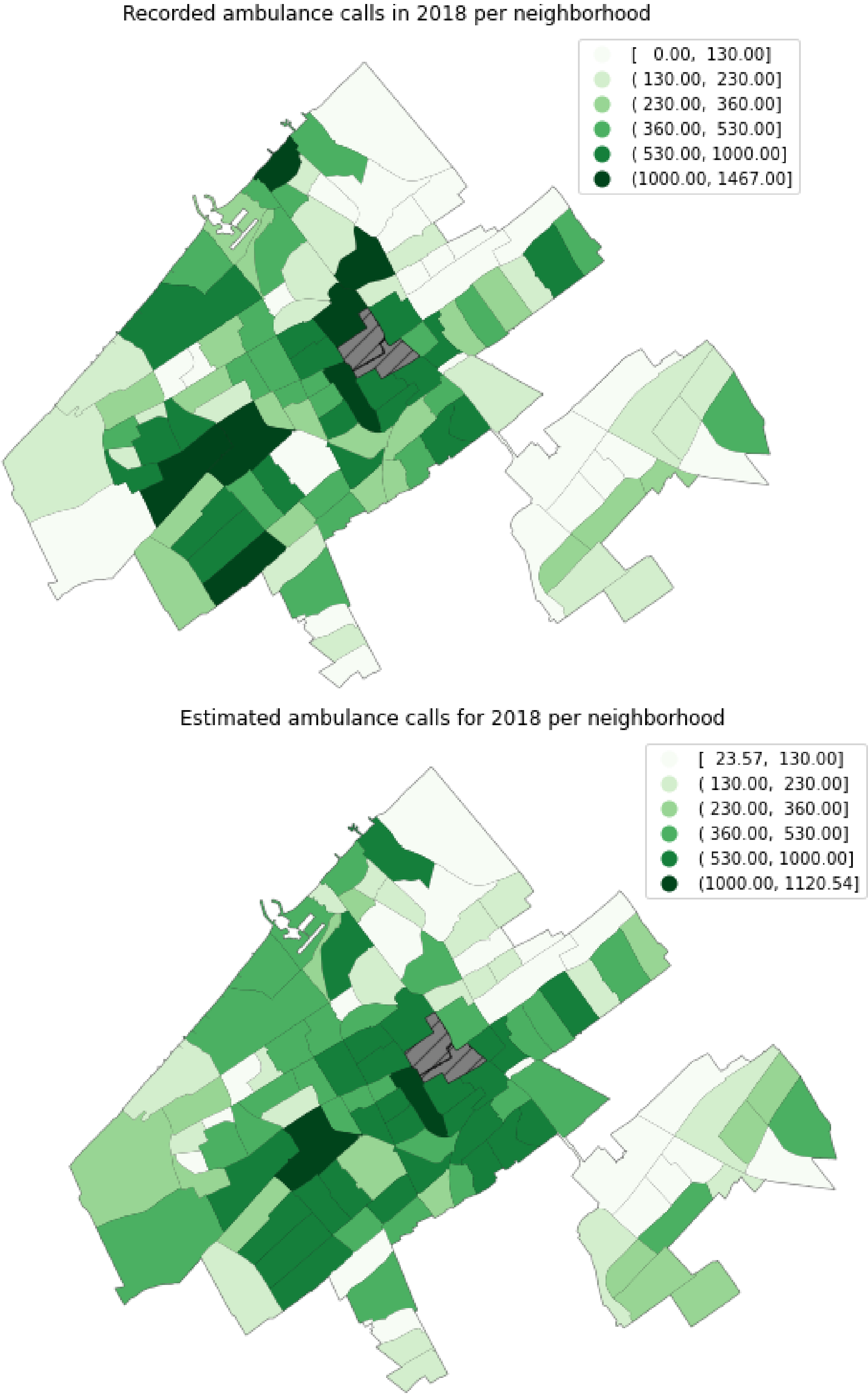


Figure 4.1: Map of ambulance calls in the Hague in 2018 (upper) and predicted ambulance calls for the Hague in 2018 with the multi-linear model based on crimes and population size (lower). The hashed, grey neighbourhoods are Kortenbos and Zuidwal, which were excluded of the analysis due to their unusually high number of ambulance calls and crimes respectively.

4.2.2 Accuracy of the model

For a model trained on only two predictors, the R^2 and $RMSE$ of the final model was sufficient to provide some veracity to our argument. To better analyse the accuracy of the model, the difference between the recorded calls and the predicted calls for each neighbourhood in 2018 is shown in Figure 4.2. This was the subtraction of the recorded calls from the prediction. The differences were divided into four categories: differences of less than one call per month (white), differences between one call per month and call per week (light colours), differences between one call per week and one call per day (darker colours) and differences of more than one call a day (darkest colours). Positive differences (purple) imply that the number of calls was overestimated, while negative differences (orange/brown) imply that the number of calls was underestimated.
The western side of city seems to be more prone to underestimation while the eastern side of the city seems more prone to overestimation. There is no clear pattern in the distribution of absolute error. Quite a few predictions were off by less than one call per week, which is reassuring. Furthermore, only five predictions were off by more than one call per day, although it should be noted that the largest underestimate of calls shown in brown is of more than nine hundred calls over the entire year, which is the equivalent of 2.6 calls per day. Most of the predictions seemed to be off by more than one call per week but less than one call per day (light colours excluding white). As mentioned before, the final model has an $RMSE$ of 192, signifying an estimate error of about one call every second day.

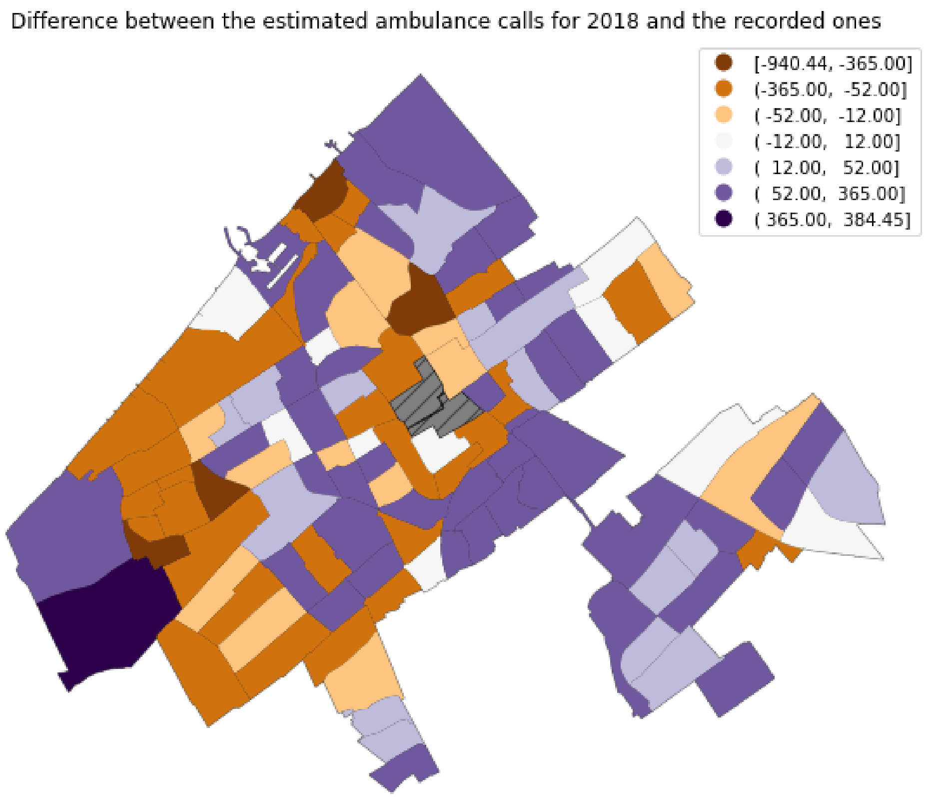


Figure 4.2: Difference between the predicted ambulance calls for 2018 and the recorded ones. Lighter colours show smaller differences, darker colours show larger ones. Dark grey, hashed neighbourhoods are the outliers Kortenbos and Zuidwal which were not included in training and testing.

5 CONCLUSION & DISCUSSION

An efficient use of emergency resources is crucial to the proper functioning of a city. Ambulances, police officers and firefighters must be able to reach any citizen in need quickly. Resources for these services in the Netherlands are distributed on the base of prediction models that do not employ socio-economic predictors [16]. In this study, socio-economic factors were analysed to investigate their predictive potential for ambulance calls in neighbourhoods of the Hague.

After an initial literature-based selection of twenty socio-economic factors (see: Appendix A for full list) and exploratory data analysis and modelling, two suitable predictors stood out: the number of crimes committed in the neighbourhood and the population size. The multi-linear model trained on data from 2017 and tested on data from 2018 reached a score of $R^2_{Final} = 0.61$ and an $RMSE = 192$ on the dataset without 2 outliers. These promising results reflected the literature as discussed in Chapter 2, where crimes were also mentioned as a good predictor for ambulance calls [14]. However, numerous predictors such as mean income and migration background of inhabitants were discarded, even though they were mentioned in previous work [2, 14, 15] as suitable predictors.

Some flaws of the model can be explained through reflection on the literature discussed in Chapter 2. The most prominent explanation that the literature can offer is that factors concerning daytime population of a neighbourhood are a better predictor than factors of residential population [14]. Using the daytime population was outside of the scope and available data for this report.

5.1 LIMITATIONS

There are a few limitations of the work and model provided here, enumerated below:

- Two outlier neighbourhoods were excluded from the training and testing of the model. These were Kortenbos (number of emergency calls in 2017: 5334, while the second highest neighbourhood (Scheveningen Badplaats) had 1456) and Zuidwal (number of crimes in 2017: 2308, while the second highest neighbourhood (Kortenbos) had 895). Due to their extreme values these variables disrupted the model, resulting in an $R^2_{Final} = 0.39$ when they were included. These neighbourhoods should be carefully considered by policy makers when redistributing ambulance resources because they clearly have worrisome characteristics.

- Income data was mentioned in previous literature as a good predictor for number of ambulance calls [2, 14]. In the employed dataset, data on mean income was partly missing. This issue was resolved by filling the gaps using a linear regression model (as explained more extensively in Chapter 3). Mean income was not included in the final model because it did not show sufficient predictive power. If this data would have been complete these results might have been different.

- Three types of models were tested: linear regression, polynomial regression and multiple regression. Employing more models for testing might have resulted in finding a better-performing model.

5.2 FUTURE RESEARCH

A few suggestions for future research can be made based on the results of this analysis.

- As suggested previously, observing daytime population instead of residential population would be a strong addition to the model. This idea is also supported by the analysis of the KDE-plots for daytime and nighttime calls seen in Figure 3.6 and work by [14].

- Training a similar model on more data might improve reliability of the predictions. More data could either be added by including more years or by including more cities. This second option could not only improve the results of the model but also strongly increase the chances that these results would actually be employed for ambulance resource distribution. Ambulance resources are distributed on the national level [16], therefore research aiming to reorganise this distribution should preferably also be carried out on this scale.

- Simultaneously, a higher quality of data might result in the selection of more variables with more predictive power. A simple way to include higher quality data would be to redo the analysis, including income data, once this data for 2018 has become publicly available.

- During the analysis, it was often noted that many of the variables were dependent on population size and density. In future research, it might be interesting to include variables that are less dependent on these factors, like the number of businesses, amount of green space or number of other facilities. Environmental factors such as green space were also mentioned by [3] as a factor of impact for health.

- Testing the data on more types of models might result in a different model selection, with potentially better results. Also, in this research the only type of relations that were suggested between number of emergency calls and the indicators were linear and polynomial relations. Exploring more types of relations could result on new insights on predictive power of indicators.

- The dataset on emergency calls provided a categorisation of calls based on the severeness of the case. This classification was not included in this analysis but could add to an efficient and fair distribution of ambulance resources in which severe cases are treated faster.

In conclusion, the findings show that some socio-economic factors such as the number of crimes in a neighbourhood could be used to improve the prediction of ambulance demand in the Hague. However, many other indicators, even those widely mentioned in the literature, seemed to have little predictive power in this case. More research with a wider scope is required to establish better predictions, that could one day be employed to distribute ambulance resources in an efficient way.

BIBLIOGRAPHY

[1] F Acciai. "The age pattern of social inequalities in health at older ages: are common measures of socio-economic status interchangeable?" In: *Public health* 157 (2018), pp. 135–141.

[2] Carole A Aldrich, John C Hisserich, and Lester B Lave. "An analysis of the demand for emergency ambulance service in an urban area." In: *American Journal of Public Health* 61.6 (1971), pp. 1156–1169.

[3] Jaime M Booth et al. "Barrios, ghettos, and residential racial composition: Examining the racial makeup of neighborhood profiles and their relationship to self-rated health". In: *Social Science Research* 69 (2018), pp. 19–33.

[4] CBS. *Fysieke arbeidsbelasting werknemers; bedrijfstak*. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83159NED/table?dl=1D4AA. Online; retrieved at 01-11-2020. 2020.

[5] CBS. *Geregistreerde misdrijven; wijken en buurten 2017*. https://data.overheid.nl/dataset/1004-geregistreerde-misdrijven--wijken-en-buurten-2017. Online; retrieved at 01-11-2020. 2020.

[6] CBS. *Geregistreerde misdrijven; wijken en buurten 2018*. https://data.overheid.nl/dataset/1154-geregistreerde-misdrijven--wijken-en-buurten-2018. Online; retrieved at 01-11-2020. 2020.

[7] CBS. *Kerncijfers wijken en buurten 2017*. https://www.cbs.nl/nl-nl/maatwerk/2017/31/kerncijfers-wijken-en-buurten-2017. Online; retrieved at 01-11-2020. 2017.

[8] CBS. *Kerncijfers wijken en buurten 2018*. https://www.cbs.nl/nl-nl/maatwerk/2018/30/kerncijfers-wijken-en-buurten-2018. Online; retrieved at 01-11-2020. 2018.

[9] CBS. *Opleidingsniveau naar gemeenten, wijken en buurten*. https://www.cbs.nl/nl-nl/maatwerk/2020/17/opleidingsniveau-naar-gemeenten-wijken-en-buurten. Online; retrieved at 01-11-2020. 2020.

[10] CBS. *Opleidingsniveau naar wijk en buurt, 1 oktober 2017*. https://www.cbs.nl/nl-nl/maatwerk/2019/30/opleidingsniveau-naar-wijk-en-buurt-1-oktober-2017. Online; retrieved at 01-11-2020. 2019.

[11] Den Haag in cijfers. *Thema's - Buurten*. https://denhaag.incijfers.nl/jive. Online; retrieved at 01-11-2020. 2020.

[12] Eurostat. *Quality of life indicators - health - Statistics Explained. Eurostat statistics explained.* https://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_health#Gender.2C_age_and_income-related_differences. Retrieved at 22-11-2020. 2020.

[13] Dana Goldman and James P Smith. "The increasing value of education to health". In: *Social science & medicine* 72.10 (2011), pp. 1728–1737.

[14] Anastasios Noulas et al. "Foursquare to the rescue: Predicting ambulance calls across geographies". In: *Proceedings of the 2018 International Conference on Digital Health*. 2018, pp. 100–109.

[15] Hilde Philips et al. "Out of hours care: a profile analysis of patients attending the emergency department and the general practitioner on call". In: *BMC family practice* 11.1 (2010), p. 88.

[16] RIVM. *Doorontwikkeling referentiekader ambulancezorg 2020 (2020-0009)*. https://www.rivm.nl/bibliotheek/rapporten/2020-0009.pdf. Retrieved at 22-11-2020. 2020.

[17] RIVM. *Referentiekader spreiding en beschikbaarheid ambulancezorg 2016 (2016-0093)*. https://www.rivm.nl/bibliotheek/rapporten/2016-0093.pdf. Retrieved at 22-11-2020. 2016.

[18] RIVM. *Referentiekader spreiding en beschikbaarheid ambulancezorg 2017 (2017-0109)*. https://www.rivm.nl/bibliotheek/rapporten/2017-0109.pdf. Retrieved at 22-11-2020. 2017.

[19] RIVM. *Referentiekader spreiding en beschikbaarheid ambulancezorg 2018 (2018-0128)*. https://www.rivm.nl/bibliotheek/rapporten/2018-0128.pdf. Retrieved at 22-11-2020. 2018.

[20] RIVM. *Referentiekader spreiding en beschikbaarheid ambulancezorg 2019 (2019-0157)*. https://www.rivm.nl/bibliotheek/rapporten/2019-0157.pdf. Retrieved at 22-11-2020. 2019.

A LIST OF VARIABLES

In this Appendix, we listed all variables defined in chapter 2.3 that were directly measurable. Table A.1 contains the description of each variable. Notice here that each variable is measured for 2017 and for 2018, so two datasets were created with these exact variables. Table A.2 shows where we retrieved the data from, how the data was cleaned, and their most important statistics in 2017. We only show the statistics for 2017 here because this is the year we train our model on, 2018 data is only used to test our model.

Variable	Description
1. Number of Emergency calls	The amount of ambulance calls in each neighbourhood over the course of 1 year.
2. Number of inhabitants	The number of inhabitants living in each neighbourhood.
3. % of inhabitants under 14	Percentage of inhabitants in each neighbourhood under the age of 14.
4. % of inhabitants over 65	Percentage of inhabitants in each neighbourhood over the age of 65.
5. Mean income per inhabitant	Average yearly income per neighbourhood time €1000.
6. % households in the lowest 40% national income	Percentage of households in each neighbourhood that belongs to the lowest 40% of the entire Netherlands.
7. % households in the highest 20% national income	Percentage of households in the neighbourhood that belong to the group of households with the 20% highest income in the Netherlands.
8. Number of crimes	Number of committed crimes in each neighbourhood. Crimes that were taken into account for this variable are all crimes that are expected to potentially lead to the calling of an ambulance: property crime, destruction of property, and violence.
9. % of non-western citizens	Percentage of inhabitants in each neighbourhood with a non-western background. Non-western backgrounds include African, Latin-American, Asian (with a exception for Indonesia and Japan), and Turkish.
10. Degree of Urbanity	The degree of urbanity is determined by the density of addresses in each neighbourhood. The higher the urbanity, the lower the number, with 1 being the highest urbanity with over 2500 addresses per square km. A degree of 2 means 1500-2500 addresses per square km, 3 1000-1500 addresses, 4 500-1000 addresses and 5 less than 500 addresses per square km.

11. Level of education	The percentage of citizens that finished respectively lower education, medium education or higher education. The distinction between lower, medium and higher education can be found in the datafile from cbs [10].
12. % of men	The percentage of citizens per neighbourhood that are male.
13. Number of religious facilities	The number of religious facilities in the neighbourhood, this includes churches, mosques, and other prayer facilities.
14. Number of citizens working physically intensive jobs	The number of citizens per neighbourhood that work in the top three most physically intensive jobs, as described by CBS [4]. These include jobs in Transport & Storage, Trade, and Healthcare.

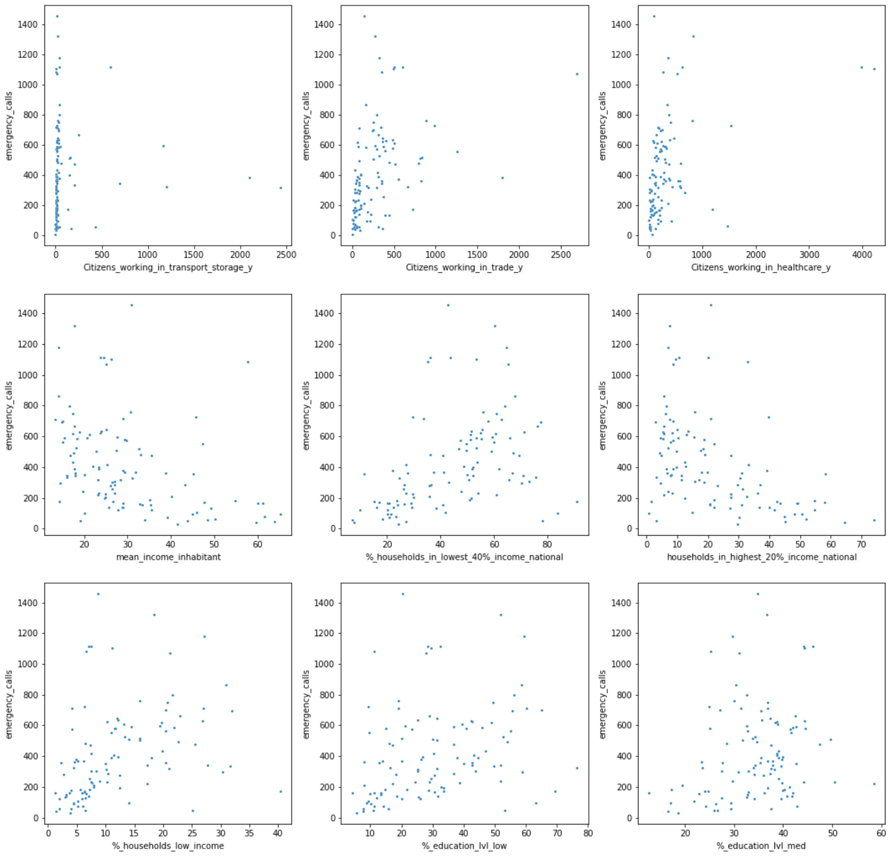
Table A.1: Variables and their descriptions

Variable	**Data source**	**Data Cleaning**	**Statistics for 2017**
1. Number of Emergency calls	Course material	Only selected calls from 2017-2018.	Number of calls: 52003
2. Number of inhabitants	CBS [7] [8]	-	min: 0, max: 14855, mean: 4966
3. % of inhabitants under 14	CBS [7] [8]	Constructed from the number of inhabitants and number of inhabitants under 14.	min: 1.9, max: 32.6, mean: 16.7
4. % of inhabitants over 65	CBS [7][8]	Constructed from the number of inhabitants and number of inhabitants over 65.	min: 0.9, max: 63.6, mean: 16.9
5. Mean income per inhabitant	CBS [7] [8]	Missing values in 2018 were constructed using linear regression as explained in chapter 3.2 and appendix B.	min: 13.4, max: 65.3, mean: 29.4
6. % households in the lowest 40% national income	CBS [7] [8]	-	min: 7, max: 91.1, mean: 44.8
7. % households in the highest 20% national income	CBS [7] [8]	-	min: 0.8, max: 74.1, mean: 21.5
8. Number of crimes	CBS [5] [6]	The total crimes were constructed as a sum of property crime, destruction of property, and violence.	min: 29, max: 2308, mean: 284.9
9. % of non-western citizens	CBS [7] [8]	-	min: 5.4, max: 88.2, mean: 30.4

10. Degree of Urbanity	CBS [7] [8]	-	min: 1, max: 4, mean: 1.3, roughly 80% of all neighbourhoods are in category 1.
11. Level of education	CBS [10][9]	Combination of the percentage of inhabitants that finished respectively low, medium and higher education.	-
12. % of men -	CBS	CBS [7] [8]	-
13. Number of religious facilities	Den Haag in Cijfers [11]	-	-
14. Number of citizens working physically intensive jobs	Den Haag in Cijfers [11]	Combination of citizens working in Transport & Storage, Trade, and Healthcare.	-

Table A.2: Variables and their properties

Figure A.1 shows the correlation between variables 2-14 we defined in the tables above and the number of ambulance calls. Notice that we plotted the three physical jobs separately, as well as the percentage of education. This has been done to investigate if either of these correlate well with ambulance calls.
As can be seen in the plots, there are a number of variables that do not correlate in a distinct manner with ambulance calls. These variables were not taken into consideration for our prediction model, leaving variables 2-10 as potential predictors for ambulance calls.


emergency_calls
Citizens_working_in_transport_storage_y
Citizens_working_in_trade_y
Citizens_working_in_healthcare_y
mean_income_inhabitant
%_households_in_lowest_40%_income_national
households_in_highest_20%_income_national
%_households_low_income
%_education_lvl_low
%_education_lvl_med

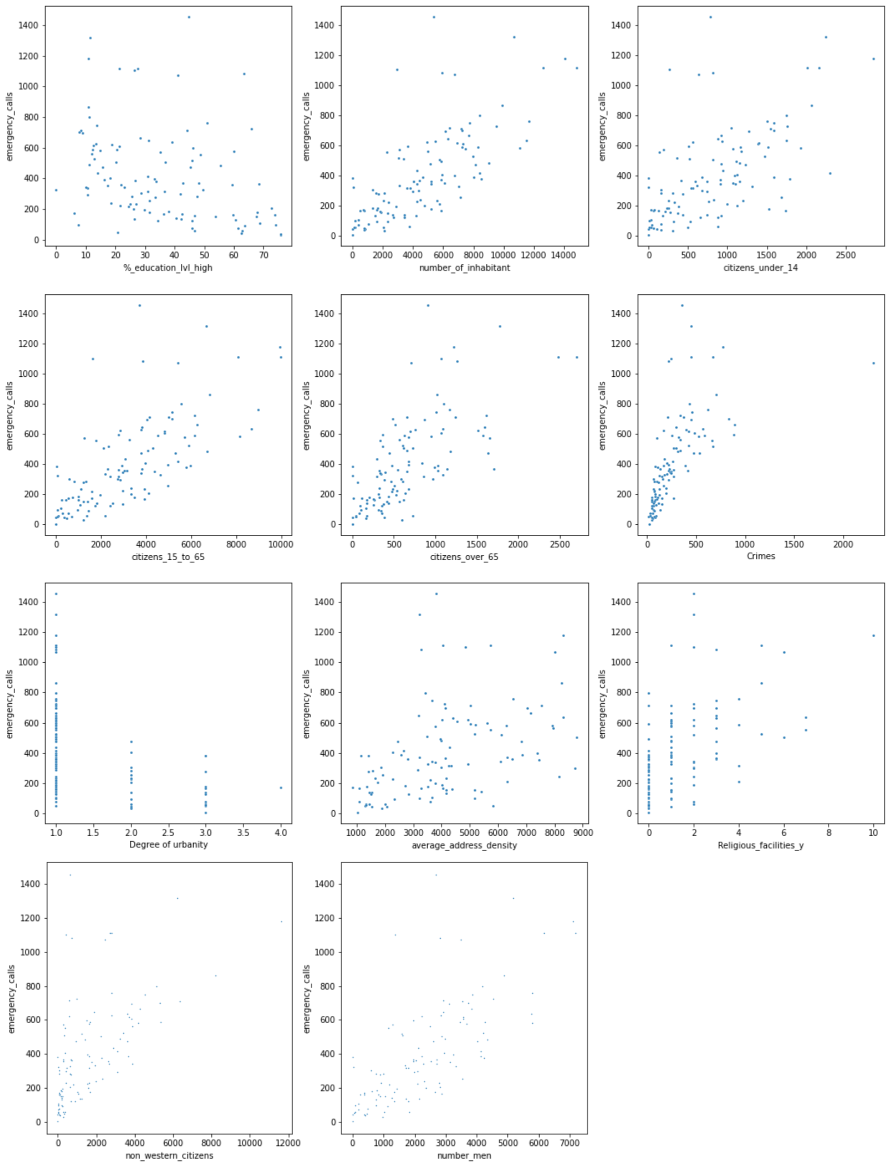
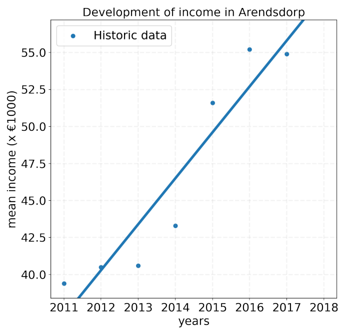


Figure A.1: Scatterplots showing the correlations between our defined indicators and number of ambulance calls

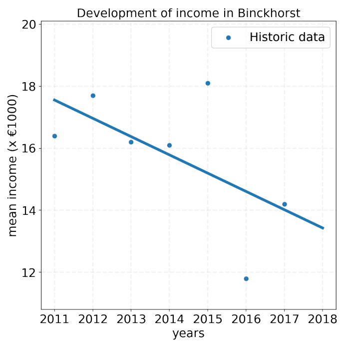
B LINEAR REGRESSION MODEL FOR MISSING INCOME DATA

We discuss the outcome of the linear regression model that was created to fill in the missing values for mean income in 29 neighborhoods in The Hague. Figure B.1 shows the model for these neighborhoods. What can be seen is that the development of the income differs per neighborhood, some have an upward trend, indicating that the neighborhood has become richer over the years, others have declined. In general, a strong trend can be found in the development of the income, but exceptions are present. For instance in the neighborhood Zorgvliet in figure B.1ab, the income fluctuated quite a bit over the years, resulting in a relatively flat linear regression line.

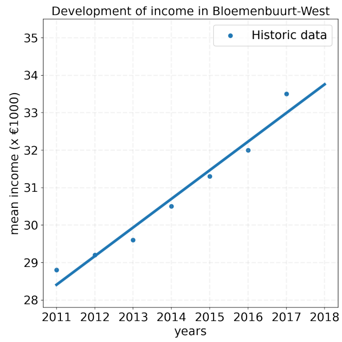
Nonetheless, overall we can say that the extrapolated data we use in the model for 2018 is representative of the neighborhoods.



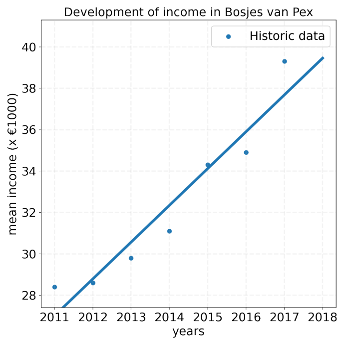
(a) Linear regression Arendsdorp



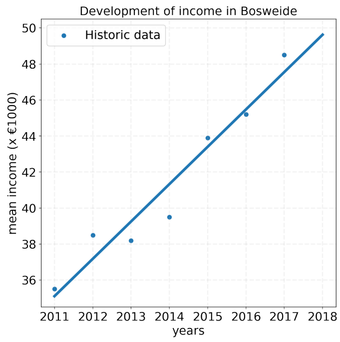
(b) Linear regression Binckhorst



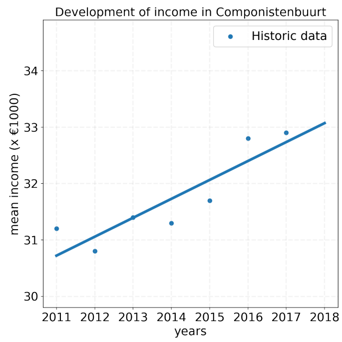
(c) Linear regression Bloemenbuurt-West



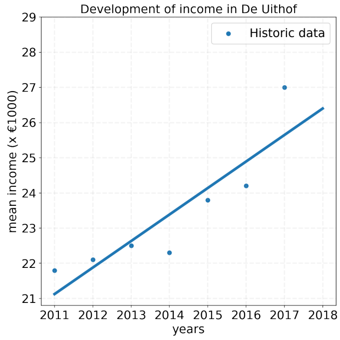
(d) Linear regression Bosjes van Pex



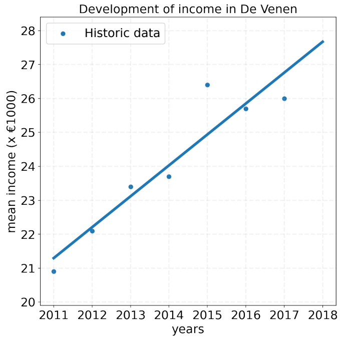
(e) Linear regression Bosweide



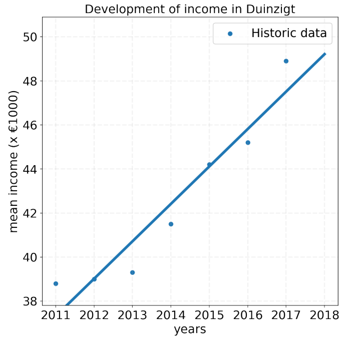
(f) Linear regression Componistenbuurt



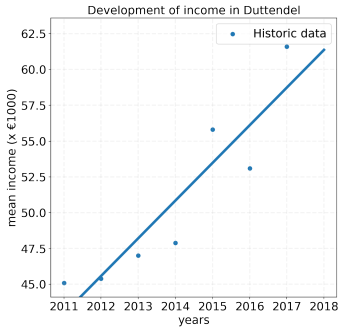
(g) Linear regression De Uithof



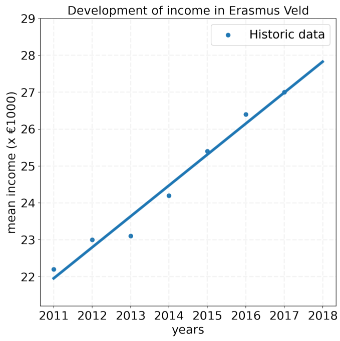
(h) Linear regression De Venen



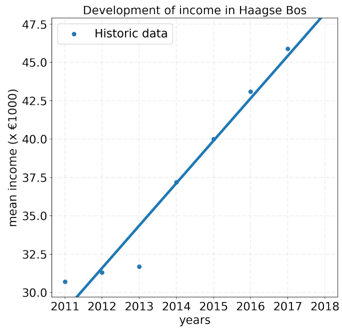
(i) Linear regression Duinzigt



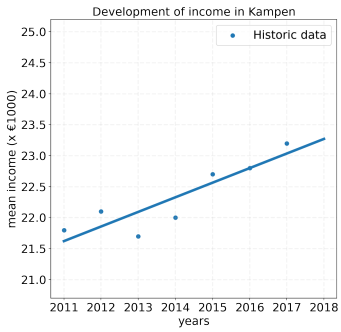
(j) Linear regression Duttendel



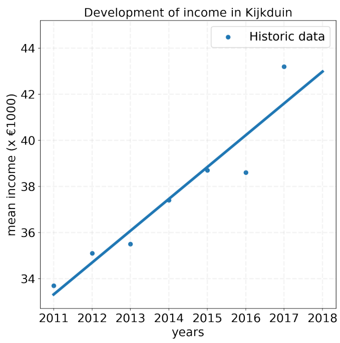
(k) Linear regression Erasmus Veld



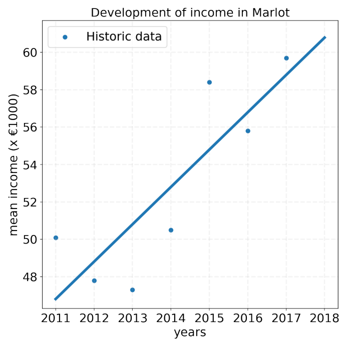
(l) Linear regression Haagse Bos



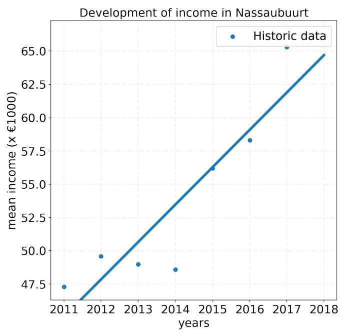
(m) Linear regression Kampen



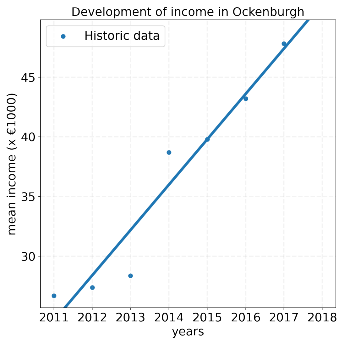
(n) Linear regression Kijkduin



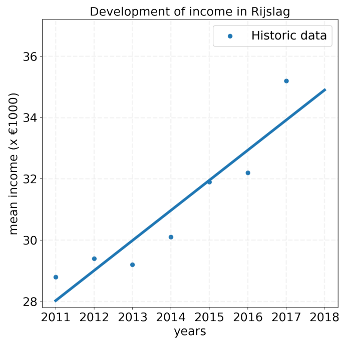
(o) Linear regression Marlot



(p) Linear regression Nassaubuurt

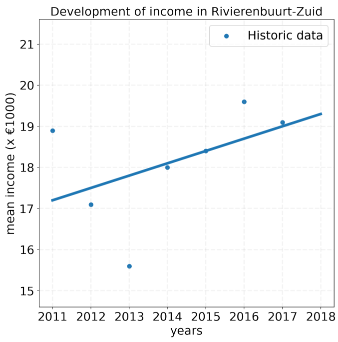


(q) Linear regression Ockenburgh

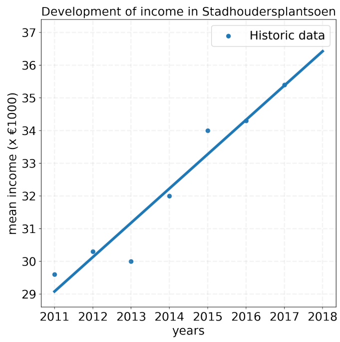


(r) Linear regression Rijslag

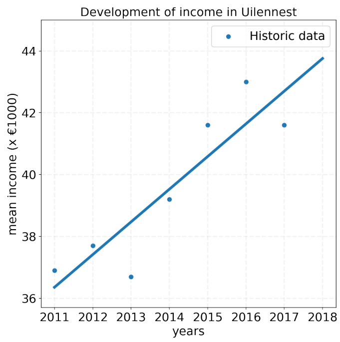
Figure B.1: Linear regression for missing data for income



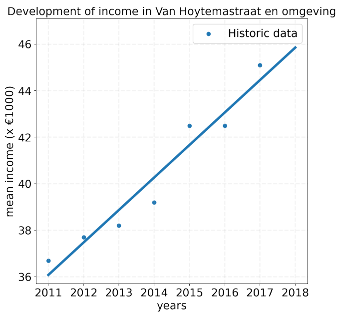
(s) Linear regression Rivierenbuurt-Zuid



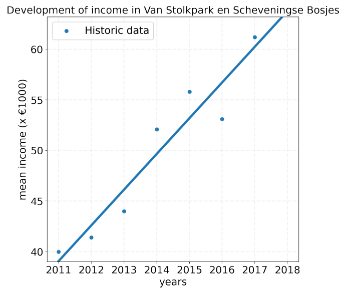
(t) Linear regression Stadhoudersplantsoen



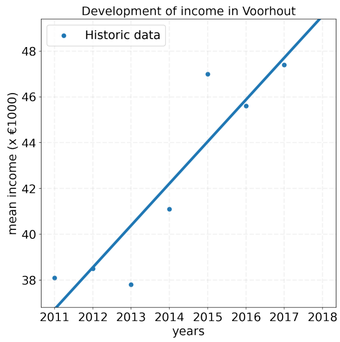
(u) Linear regression Uilennest



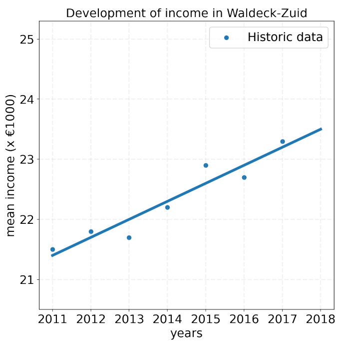
(v) Linear regression Van Hoytemastraat en omgeving



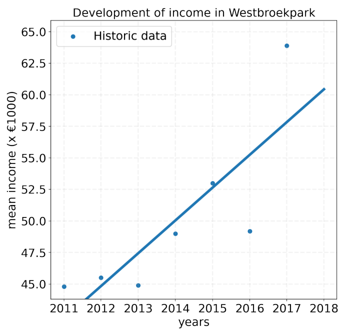
(w) Linear regression Van Stolkpark en Scheveningse Bosjes



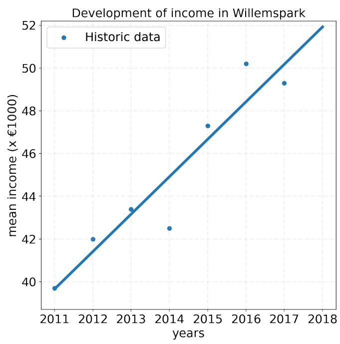
(x) Linear regression Voorhout



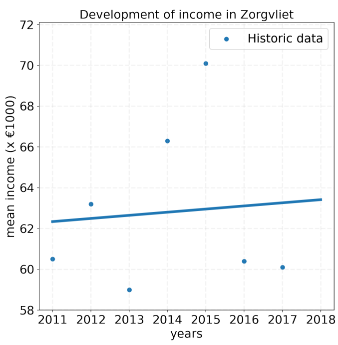
(y) Linear regression Waldeck-Zuid



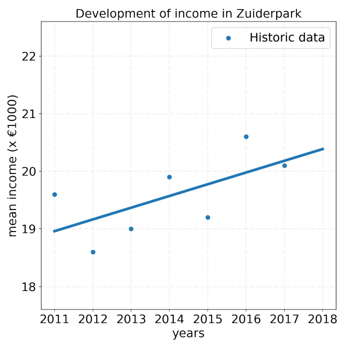
(z) Linear regression Westbroekpark



(aa) Linear regression Willemspark



(ab) Linear regression Zorgvliet



(ac) Linear regression Zuiderpark

Figure B.1: Linear regression for missing data for income

C KDE PLOTS

This appendix will discuss the KDE plots that were made for each six hour time interval. The time intervals ranged from 00:00 - 06:00, 06:00 - 12:00, 12:00-18:00 and 18:00-00:00. The respective KDE plots can be seen in figures C.1, C.2, C.3 and C.4.
Figure C.1 shows a clear concentration of emergency calls in the city centre while the rest of the calls seem to be quite evenly spread out throughout the rest of The Hague.
Figure C.2 shows the same concentration of emergency calls in city centre, however a small but clear concentration of emergency calls is also visible north east of the city centre. Two smaller but still visible concentrations of calls can be identified south-west of the city centre and north-west of the city centre towards the beach. Next to that the overall distribution of calls throughout the city seems to be lower.
Progressing through the day, fig C.3 shows that the distribution of calls get even more concentrated in the city centre and the areas north of the city centre. There are still calls in the rest of the city but their amounts are very small compared to the amount of calls in the identified concentrated clusters.
Fig C.4 shows a much more evenly spread out number of calls throughout the city again. However, the concentrated clusters of emergency calls in the city centre and the areas north of the city centre are still visible.

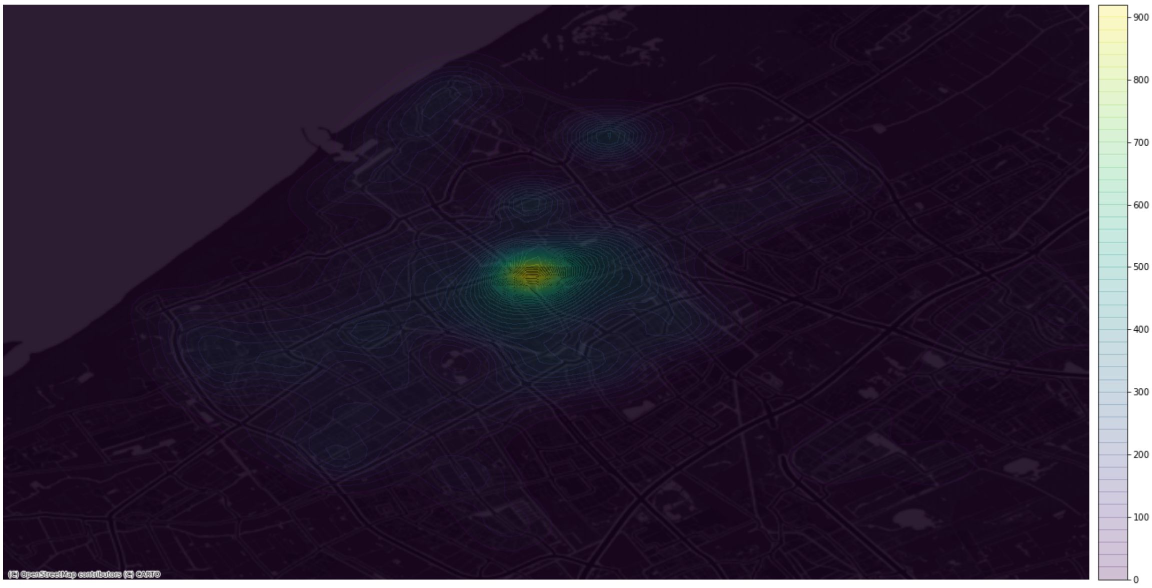


Figure C.1: KDE plot of emergency calls made between 00:00 and 06:00

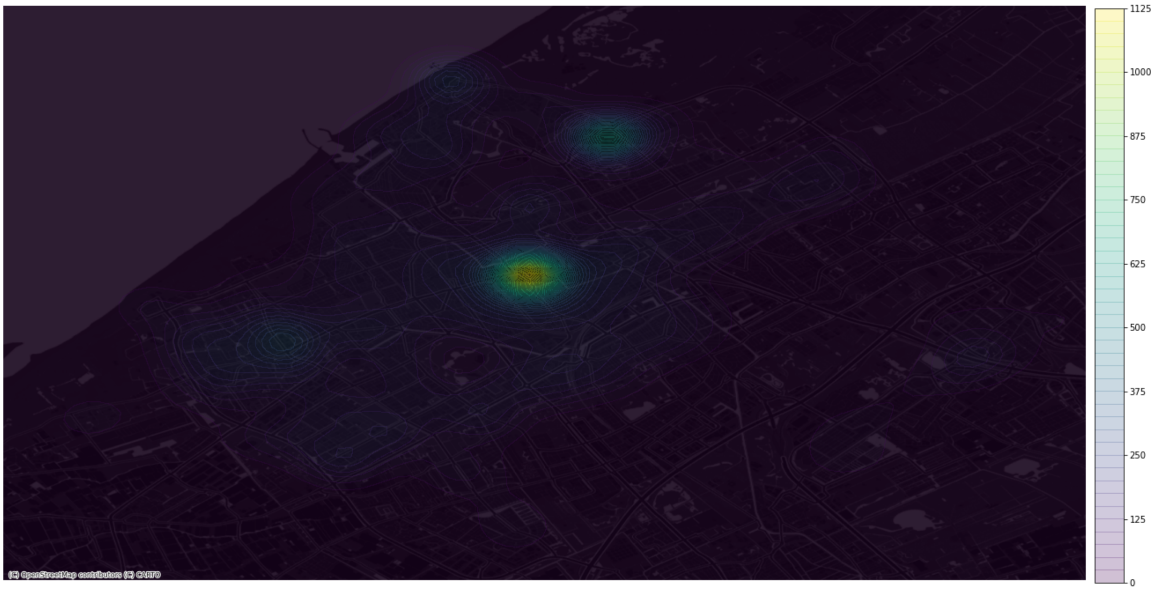


Figure C.2: KDE plot of emergency calls made between 06:00 and 12:00

Figure C.3: KDE plot of emergency calls made between 12:00 and 18:00

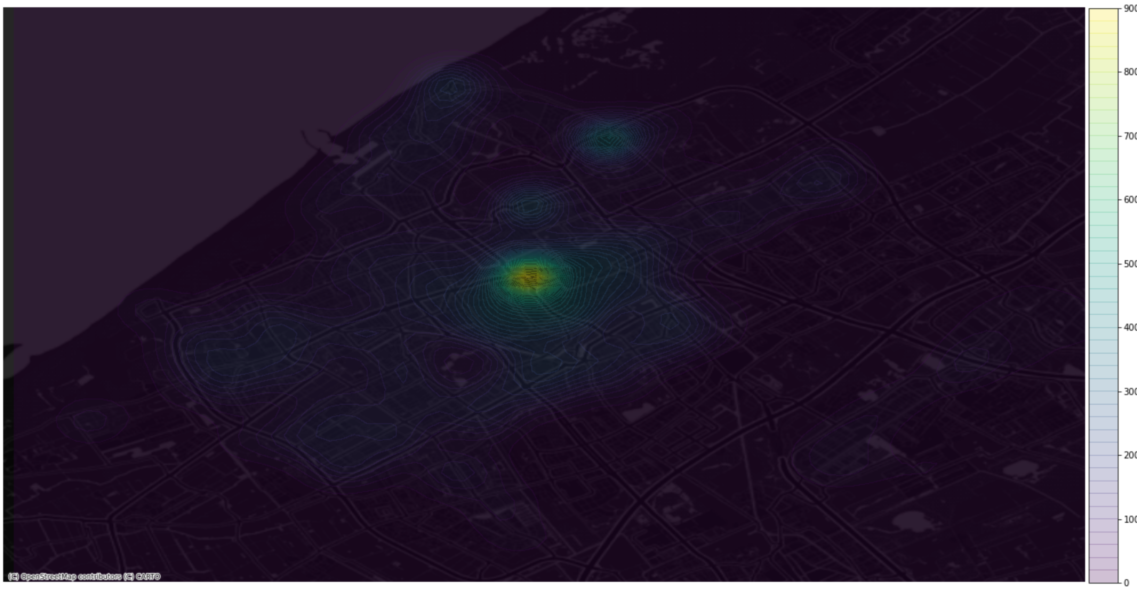


Figure C.4: KDE plot of emergency calls made between 18:00 and 00:00

Based on the KDE plots it can be concluded that a significant amount of emergency calls are made in the city centre no matter the time of day. However the calls seem to be more evenly spread out during night time (18:00 to 06:00).This might be explainable due to the fact that less people work, shop or hang out in the city centre during those times because they are probably at home instead of in the city centre.