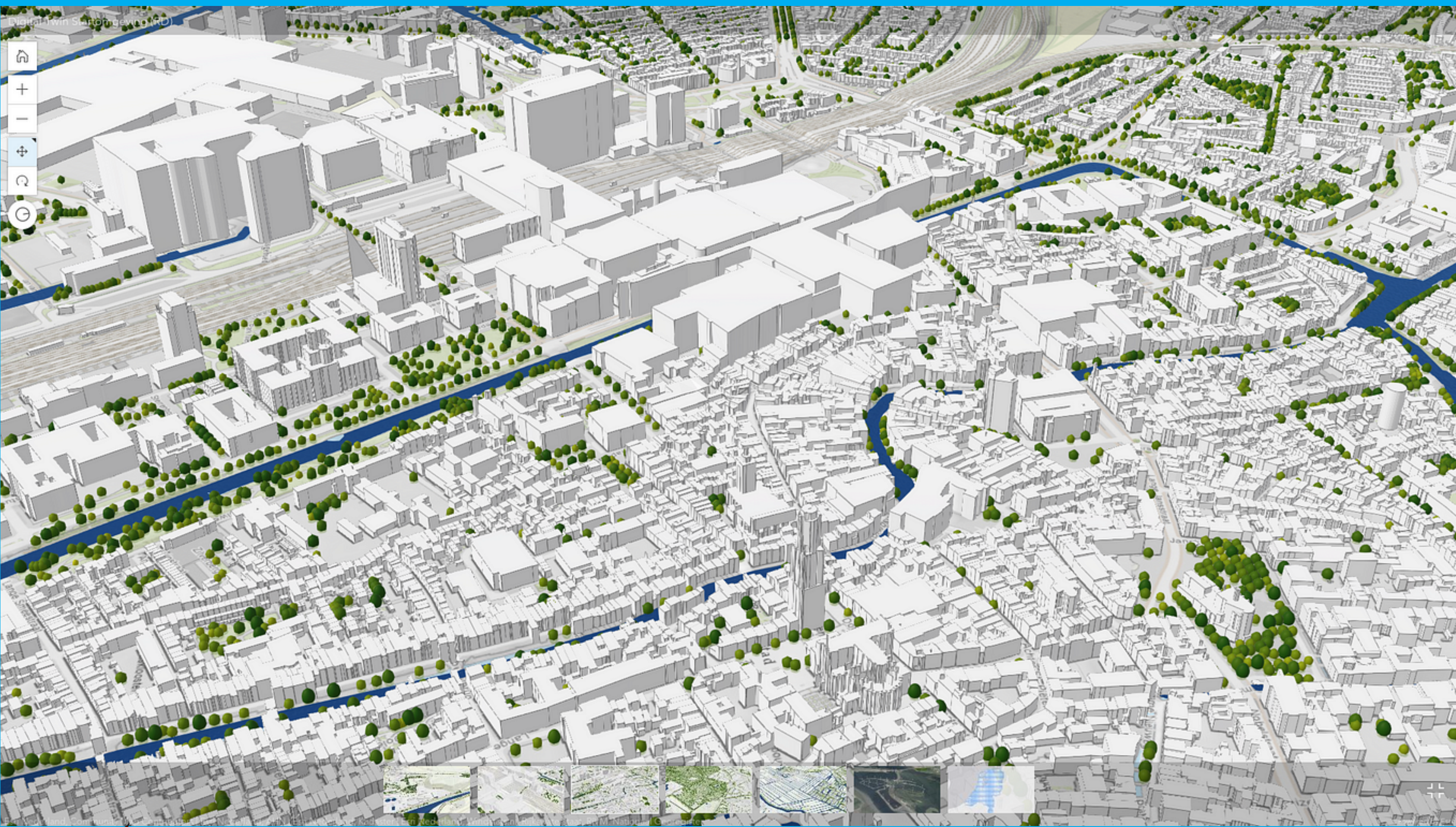
Identifying the Health Vulnerability in a City

EPA-122A Group 8

February 2, 2024

Lieke Bikker (4720679)
Jasha Kuster (5661757)
Pepijn de Haan (4956451)
Lohit Gandham (5733766)



TUDelft

ABSTRACT

Existing studies found geographical correlation between socio-economic status and healthcare. Asaria et al.'s research [2], which was conducted in the United Kingdom, identified a social gradient in hospital admissions. On a larger scale, Marmot [8] concluded that the availability of emergency health services showed variation among individuals with different socio-economic statuses when comparing Western- and Eastern-Europe on a national level. This was also found on a smaller scale, as Shahidah et al. [1] concluded that there was a spatial gradient under ambulance calls in Singapore, Asia. Focusing on the Netherlands, Storm et al.'s article [11] delved into another aspect of consideration, as they outlined various factors serving as risk factors for health inequality in the Netherlands. Notably, certain institutional aspects, such as the absence of collaborative efforts between the public health sector and social (youth affairs, education) or physical (housing, spatial planning) policy sectors, were highlighted.

However, the social gradient of medical services within Dutch cities remains unknown at this point. As such, the diverse city of Rotterdam was chosen to investigate the potential relation between socio-economic factors and healthcare. This was done using K-Means clustering, an unsupervised machine learning algorithm. The following research question was answered in this report:

"What is the relationship between socio-economic status of neighbourhoods and the number of ambulance calls in Rotterdam, Netherlands?"

To address our research question, we utilized a 112 Call Dataset obtained from 112-Nederland [10], containing information on emergency calls in the Netherlands from 2017 to 2020. Additionally, we used the Central Bureau of Statistics (CBS) neighborhood data [4], using key socio-economic indicators such as income, population density, migration background, and hospital distance for our analysis.

The analysis found a complex relationship between socio-economic factors and ambulance service use, challenging assumptions as high socio-economic status and proximity to healthcare facilities did not consistently correlate with lower ambulance call frequencies; instead, healthcare accessibility and socio-economic diversity emerged as critical factors influencing emergency service patterns.

The insights of this study hold significant relevance for policymakers and healthcare planners aiming to establish emergency healthcare systems that are both equitable and efficient, particularly in cities characterized by diverse socio-economic landscapes, as they emphasize the significance of comprehensive healthcare accessibility and underscore the necessity for urban health policies to account for the diverse nature of healthcare demand.

Contents

Abstract 1

1 INTRODUCTION 3

2 RELATED WORK 5

3 EXPLORATORY DATA ANALYSIS (EDA) 7
- 3.1 Overview of the study . . . 7
- 3.2 Description of the data . . . 7
- 3.3 Nature of Selected Variables. . . . 8
- 3.4 Description of research site and timeframe . . . 9
- 3.5 Exploratory Data Analysis . . . 10
- 3.6 Limitations on data and study . . . 15
- 3.7 Errors and missing values in data . . . 15

4 ANALYSIS 17
- 4.1 Experimental Design: . . . 17
- 4.2 In-Depth Analysis of Clustering Process. . . . 17
  - 4.2.1 Specifying the number of clusters used . . . 17
  - 4.2.2 Analysis per cluster . . . 18
  - 4.2.3 General findings from cluster analysis . . . 19
- 4.3 Patterns and Supporting Points . . . 20
  - 4.3.1 Socio-Economic Diversity and Ambulance Service Utilization . . . 20
  - 4.3.2 Population Density, Households under the social minimum and Non-western inhabitants . . . 21
  - 4.3.3 Healthcare Accessibility and Service Utilization . . . 21
  - 4.3.4 General findings from patterns and supporting points . . . 21
- 4.4 Limitations of the Analysis . . . 21
  - 4.4.1 Clustering Limitations . . . 21
  - 4.4.2 Methodological Constraints . . . 22
  - 4.4.3 Generalizability of Findings . . . 22
  - 4.4.4 Implications of Limitations . . . 22

5 DISCUSSIONS / CONCLUSIONS 23

Bibliography 24

1 INTRODUCTION

Among others, a countries' emergency services can be argued as one of the vital services a country possesses. It provides a sense of security for its citizens, aids in maintaining peace and prosperity and provides quite a substantial measurement for socio-economic welfare. While each of the three main-emergency services (police, ambulance and fire department) are all argued to be of similar importance within their respective fields of expertise, this analysis will lay its focus on the medical services and the influence of socio-economic status. This is done because it's argued that levels of health in societies follow a social gradient; the lower the social-economic position, the lower the health [8].

In literature it is argued that a social gradient is seen between Western- and Eastern-Europe [8]. What is not yet discussed for Europe however is if these kind of socio-economic inequalities could also present themselves on a far smaller scale. Since there is a general consensus that neighborhoods within cities can have apparent unequal distributions from a socio-economic perspective [6], this analysis will investigate if this can also be seen through medical services. Our research will focus on Europe, in particular Rotterdam, Netherlands, to investigate whether a spatial correlation can be found between ambulance calls and socio-economic status; to the best of our knowledge no study has been focused on this particular topic in Rotterdam.

To conduct this analysis a dataset with emergency calls in the Netherlands over the 2017 - 2020 period was provided. From there the city of Rotterdam was chosen as it was expected to provide a well-suited case for several reasons: Rotterdam is the second-largest city in the Netherlands, it is highly diverse counting over 170 nationalities and the larger part of Rotterdam households are lower-income [5]. All in all, if the social-gradient as stated above is indeed visible on the city-scale, it would likely be visible there. To guide the analysis the following research question is used:

"What is the relationship between socio-economic status of neighborhoods and the number of ambulance calls in Rotterdam, Netherlands?"

We hypothesise that neighborhoods with a variable values that correlate with a low socio-economic status are more likely to have a higher number of ambulance calls. This hypothesis is based on the literature discussed in chapter 2.

The data that was gathered to answer this research question originated from an 112 Call Dataset, provided by the teaching team of Spatial Data Science, obtained from 112-Nederland [10]. This dataset gave us information about the emergency calls in The Netherlands in the timeframe 2017-2020. The other dataset that was used for this research is the Central Bureau of Statistics (CBS) neighborhood data [4]. This dataset provided us with key characteristics of each neighborhood in The Netherlands, of which we used a few to represent socio-economic status - relating to income, population density, migration background and hospital distance. More elaboration on the actual used variables in our analysis can be found in Chapter 4.

Reflecting on what we found when answering the research question we gained some first insights through the Exploratory Data Analysis (EDA). Here we see the neighborhoods with the most ambulance calls, and gain insights in the other variables' distributions. In Chapter 4, the cluster analysis showed several key findings: 1) the socio-economic variables we studied likely do not influence ambulance call rates, 2) population density and low socio-economic status are likely correlated and 3) healthcare demands are likely influenced by a lot of other variables not considered in this study.

All in all, answering the research question as follows: the relationship between socio-economic status of neighborhoods and ambulance calls likely does not follow a social gradient scale for the city of Rotterdam. Seeing people from neighborhoods with a lower economic status might not experience a lesser availability of emergency healthcare services than people in other neighborhoods. However, it is likely that healthcare demand is influenced by other factors as well, and ambulance call rates do not reflect total healthcare demand.

These findings are argued to be of significant importance due to the way in which they reflect what drives ambulance demand. Further research on this could help the municipality streamline their emergency services and optimise capacity utilisation.

Lastly, we reflect on how the answer to our research question might broaden, clarify or challenge existing theories. In the beginning of this introduction and chapter 3 several authors' viewpoints were introduced. Showing the cost of healthcare as a result from inequalities among neighborhoods in the United Kingdom [11], similar examples on a country scale in Europe[5] and a similar example of ambulance calls in Singapore [8]. The answer to research question in this study, while leaving some room for interpretation and further study, challenges these findings. We would advise further research to solidify the patterns (or lack thereof) that were found.

2
RELATED WORK

A study by Asaria et al [2], conducted in the United Kingdom found, that a so called social gradient could be observed in hospital admissions. Meaning socio-economic inequalities result in higher levels of morbidity and lower levels of life expectancy. Their study concludes on the notion that health systems are likely to save costs if they were to improve situations in more deprived neighborhoods. While this approach has a financial focus, our study will use these findings as a basis for further investigating the social aspects. Another study by Marmot concluded that availability of emergency health services varied across people from different socio-economic statuses on the country level when comparing Western- and Eastern-Europe [8].

Moreover, this is an interesting research avenue as [1] found a spatial gradient under ambulance calls in Singapore, Asia. Their conclusion was that there was geographical correlation in the ambulance calls, and that the risk of making an ambulance call decreased for neighborhoods with an increased socio-economic status. Their study lays a foundation for further research into geographical correlation regarding other public health systems, and help policies target areas and people at a higher risk and / or a higher ambulance deployment effectiveness. Based on the literature introduced,

Building on these articles, our study dives into the question whether a social-gradient, in terms of wealth and medical services, can be witnessed at the city-level (specifically Rotterdam) in the Netherlands; providing a topic not yet answered in scientific literature.

The importance of this topic is argued to result from its sociological foundation as social health is argued to be a constituent of social capital [9]. Meaning if we perceive all people in a society as equal then medical services should (at least strive to) be equal among this society as well. From this sociological point of view, the idea that a person who is of higher socio-economic status, has better access to medical services than people of lower socio-economic status would indicate a fundamental inequality within society.

Diving a little deeper into the context of the issue at hand, we will first look at what is known in relation to the social-gradient of socio-economic statuses and access to healthcare services in the city of Rotterdam. Reasoning from an overarching point of view, it is generally known that there are significant inequalities between neighborhoods in certain cities in the Netherlands. Mainly in the larger cities such as Amsterdam and Rotterdam [5]. While quite general, this is a first step in building the analysis; since inequality provides a case on which to test if this affects the access to medical services.

Moving on, another point of reasoning is discussed in an article by Storm et al. [11]. They list several situational aspects as risk factors for health inequality in the Netherlands. Some of these, mostly institutional, aspects are a lack of intersectoral collaboration between the public health sector and social (youth affairs, education) or physical (housing, spatial planning) policy sectors. The latter will be taken under consideration in this study, identifying possible relations between spatial planning and access to medical emergency services.

Besides the aspects described there is also enough we do not know yet in regard to the social-gradient of medical services within cities in the Netherlands. Being so because of little existing research on this specific topic. The questions that need to be answered mainly relate to the extend in which we would see characteristics of inequalities in healthcare services within the city of Rotterdam. Do relatively poorer neighborhoods call the ambulance more frequently than affluent neighborhoods? Is there a relationship between migration background and how often the ambulance services are called? In this article, we attempt to provide an answer to the research question relating to these questions:

"What is the relationship between socio-economic status of neighborhoods and the number of ambulance calls in Rotterdam, Netherlands?"

The research question was setup as a guidance for this analysis. We intend to answer it by making use of spatial data science. An approach in which the data is first cleaned and analysed to look for first hints of causalities between variables. After which a cluster analysis will be conducted to possibly gain further understanding through unsupervised statistical cluster allocation. All in all, answering this question hopefully allows us to draw a conclusion, with a certain amount of uncertainty, on the extend to which the so-called social-gradient can be witnessed among neighborhoods in the city of Rotterdam and if there is a relation between socio-economic status and the availability of health services.

In the end, if outcomes from this analysis would indicate such an equality indeed exists this could provide incentives for policy aimed at resolving the issue. Also, the outcome of this study could give rise to other research into spatial correlation between public health systems and socio-economic status.

3
EXPLORATORY DATA ANALYSIS (EDA)

3.1. Overview of the study

In this section of the article, an Exploratory Data Analysis (EDA) is conducted to gain some first insights in the structure of the data in regard to certain variables of interest. Additionally, this could provide some first insights in the way these variables behave in relation to each other within the city of Rotterdam.

The first step in this process was to plot some statistical depictions of the data in graphs and pairplots; with the intention of gaining understanding in how the data, as a whole, is structured. After this the next step was seeing how the data relates to different neighborhoods in Rotterdam; introducing the spatial component. This is done through plotting the variables of interest on the geometry maps of Rotterdam in general and in choropleths. All these steps will be explained thouroughly throughout the chapter, the next chapter on the analysis will build further upon these results.

3.2. Description of the data

For this study we were requested to provide insights in the ambulance response time, in a specified area within the Netherlands, and its possible relations to socio-economic status. The dataset provided was the 112_calls_dataset. This dataset includes the 112 emergency calls from the whole Netherlands over the 2017 - 2020 period, among other information it also included long- and latitude of the call and priority levels. Since the focus of this study is on the city of Rotterdam and the relation between neighborhoods, these were filtered from the data. Additionally, used the following variables from this dataset:

1. **pmeId**: Unique identifier for each call.
2. **pmeTimeStamp**: Time when the call was registered..
3. **pmeStrippedMessage**: Cleaned version of the original message.
4. **pmePrio** & **pmePrioLevel**: Priority of the message and its normalized level.
5. **pmeDienst**: Service type (A=Ambulance, B=Brandweer/Firefighter, P=Police).
6. **pmeLifeLiner**: Indicates if extraction by air was involved (True/False).
7. **gemName**: Name of the municipality (Gemeente).
8. **pmeLatitude** & **pmeLongitude**: Geographical coordinates of call (point).

This data provided a substantial amount of information to start the analysis. However, two things were missing: the geometry of Rotterdam's neighborhoods and some socio-economic data. For the geometry the 'Wijk- en Buurtkaart' was used by the Central Bureau of Statistics (CBS)[4], this data was openly available on the internet. For the socio-economic variables the same source was used to gain insights in distribution of incomes per inhabitant and population density. This dataset contains many 'key' variables about every neighborhood in The Netherlands; the dataset contained over 40 variables ranging from age distribution to income to surface area.

We decided to use the following variables (renamed) from the CBS dataset mentioned above to give an indication of the socio-economic status of a neighborhood:

1. **INK_INW2 –> Income_per_inhabitant**: The mean personal income per person based on the total population in private households.

2. **BEV_DICHTH –> Population_density**: The (unrounded) population density per square kilometer of land is determined by dividing the (unrounded) population on January 1st by the (unrounded) land area.

3. **P_N_W_AL –> percentage_of_non_western_inhabitants**: A person with a migration background originating from one of the countries in the continents of Africa, Latin America, and Asia (excluding Indonesia and Japan) or Turkey.

4. **P_SOCMINH –> percentage_households_living_under_the_social_minimum**: Households at or under the social minimum.

5. **AF_ZIEK_E –> Distance_to_hospital**: The average distance to a hospital in kilometeres (km).

6. **BU_NAAM**: Neighborhood name.

7. **geometry**: Geographical coordinates of neighborhoods (polygon).

3.3. Nature of Selected Variables

In the exploration of the relationship between the socio-economic status of neighborhoods in Rotterdam and the patterns in ambulance calls and response times, a diverse array of variables was meticulously selected to construct a comprehensive and insightful analysis, shortly introduced above. These variables were not chosen arbitrarily; each serves a specific purpose in painting a detailed picture of the socio-economic landscape and its interaction with emergency healthcare services. The variables which will used in our analysis will be elaborated on below, and the corresponding EDA will follow in next the section.

1. Call Count (Ambulance Calls)

The 'Call Count' variable represents the total number of ambulance calls in each neighborhood, computed using a spatial join. It is a critical indicator of the healthcare demand in an area. High call counts may signify higher healthcare needs, potentially driven by various factors including population density, prevalence of chronic diseases, or even socio-economic stressors that exacerbate health issues. Conversely, lower call counts could indicate either a healthier population or a lack of access to or utilization of emergency services, possibly due to socio-economic barriers or cultural differences in seeking healthcare. This variable is pivotal in understanding how different areas in Rotterdam interact with emergency services.

2. Income per Inhabitant

Income per inhabitant is a direct measure of the economic status of a neighborhood. It provides insight into the purchasing power of residents. Higher income levels might correlate with better health outcomes due to factors such as access to quality food, healthcare services, and healthier living conditions. In contrast, lower-income areas might depend more on public healthcare services, including ambulance services, and could face challenges such as overcrowding or under-resourcing.

3. Population Density

Population density impacts the demand for and provision of emergency services. Densely populated areas are likely to have a higher frequency of ambulance calls due to the sheer number of people present. Moreover, high population density can exacerbate traffic congestion, potentially impacting response times. In socio-economic terms, densely populated areas might also represent diverse community compositions, including vulnerable populations who could have different healthcare needs.

4. Percentage of Non-Western Inhabitants

This variable reflects the cultural and ethnic composition of a neighborhood, providing insights into the diversity of the population. The percentage of non-western inhabitants can influence healthcare patterns due to varying health behaviors, cultural attitudes towards seeking medical help, and potential language barriers. Understanding the cultural makeup of a neighborhood is crucial in tailoring emergency services to be culturally sensitive and effective.

5. Percentage of Households Under the Social Minimum
This variable indicates the percentage of households living under the social minimum, shedding light on the extent of economic hardship in a neighborhood. Households under the social minimum are likely to experience various stressors that can affect health, such as poor nutrition, living conditions, and chronic stress. These factors can lead to increased reliance on ambulance services. Additionally, this variable helps in understanding the social determinants of health and their impact on emergency healthcare demands.

6. Distance to Hospital
The proximity of a neighborhood to the nearest hospital is crucial in emergency situations. Shorter distances are expected to lead to quicker response times, which can be critical in life-threatening conditions.

Each variable in this analysis was chosen to provide a multifaceted understanding of the interplay between socio-economic factors and ambulance service patterns in Rotterdam. The 'Call Count' variable offers direct insights into the demand for ambulance services. In contrast, variables like income per inhabitant, population density, cultural composition, economic hardship, and hospital proximity provide a deeper socio-economic context. Together, these variables paint a comprehensive picture of how socio-economic conditions shape the health needs and emergency service utilization of different neighborhoods. This nuanced understanding is crucial for policymakers and healthcare providers to ensure equitable and effective distribution of healthcare resources, particularly in emergency medical services.

3.4. Description of research site and timeframe

As was briefly introduced in the introduction and beginning of this chapter the focus of this study will be on the city of Rotterdam, the second largest city of the Netherlands with citizens from over 170 different nationalities [5]. Within this city the main object of study will be the relation among and between different neighborhoods. While it is always possible to look at a smaller level, meaning even at differences within neighborhoods, this study argues that neighborhoods posses a collective level of socio-economic status that allows the measurement of possible differences in a city with the size of Rotterdam.

We decided to choose 2018 as the year of study for this analysis. The reasoning behind this is explained in the limitations section of this chapter.

3.5. Exploratory Data Analysis

To begin the EDA, the first step was to visualize the distribution of the ambulance calls among the neighborhoods as shown in **Figure 4.1**.

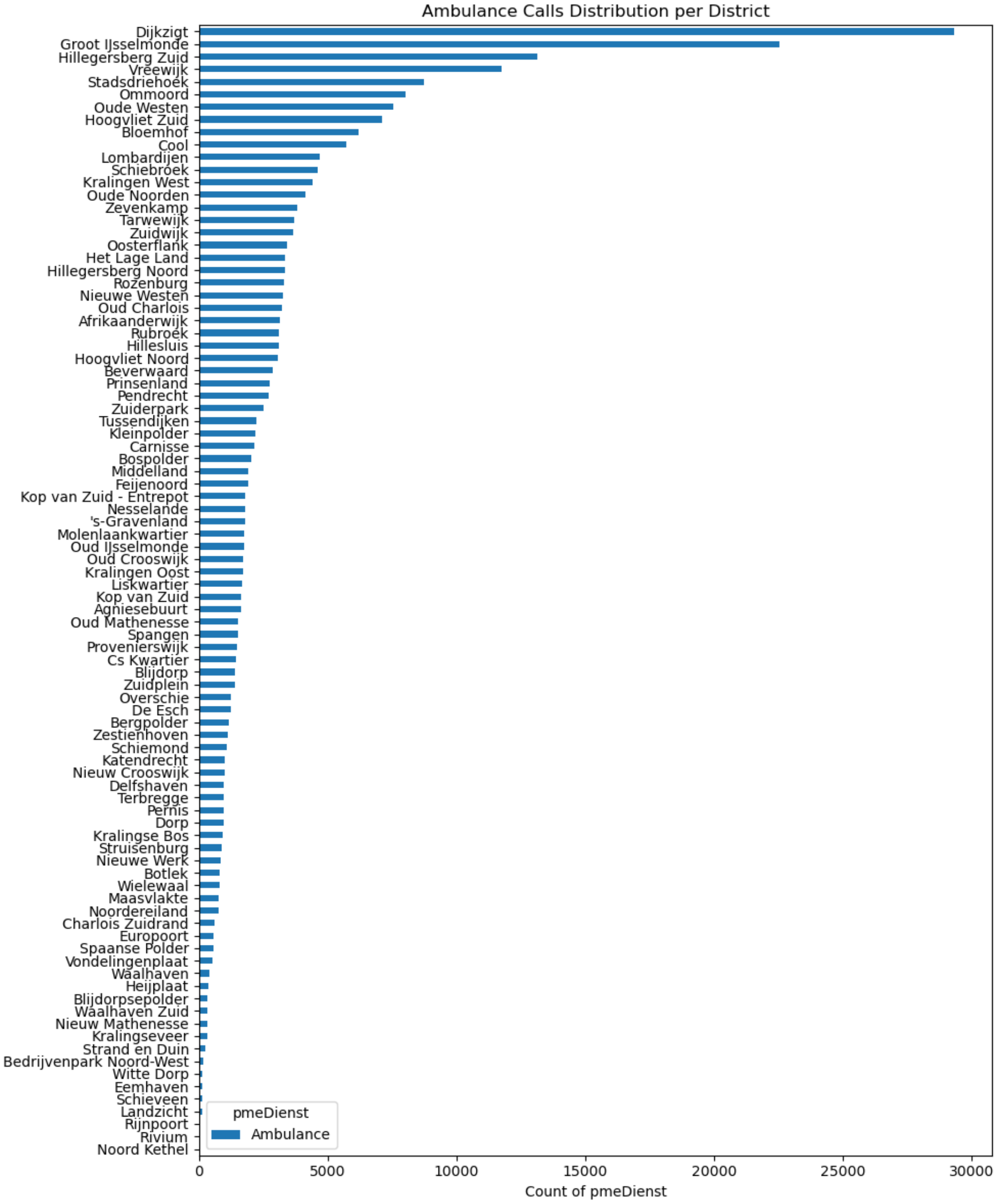


Figure 3.1: Barplot on ambulance calls distribution across neighborhoods in Rotterdam

Figure 3.1 provides some first insights as to which neighborhoods see what amount of ambulance calls. For most of the neighborhoods the distribution does not provide any eye-catching results, that is except for the four neighborhoods at the top: 1) Dijkzicht, 2) Groot IJsselmonde, 3) Hillegersberg Zuid and 4) Vreewijk. Some further investigation on Google-maps on these areas showed that all four have a relatively large hospital situated within these neighborhoods. With the Erasmus hospital, being the largest of Rotterdam, in Dijkzicht; the location with the most ambulance calls by far. This provided a first insight into were some of the higher values for neighborhoods might originate from. Perhaps this indicates a correlation between distance to hospital and ambulance call frequency.

The second step was to gain a better understanding in the socio-economic distributions among the city of

Rotterdam. Therefore a series of graphs were plotted: 1) the income per inhabitant, 2) percentage of non-western inhabitants, 3) percentage of households under social minimum, 4) population density, and 5) distance to hospital. The first three and last two variables were merged in the graph to provide a more information dense graph. This works because of the scaling on the X-axis.

1. **Income per inhabitant**.

Figure 3.2 shows the income level per inhabitant in blue. Most calls were made to neighborhoods where the income per inhabitant was between 15.000 and 30.000 euros. After a peak of around 28.000 the average income of the neighborhoods ambulance calls were made to starts to decline quite sharply.

2. **Percentage of non-western inhabitants**.

Figure 3.2 shows the percentage of non-western inhabitants in the neighborhoods where calls were dispatched in orange. It seems to be quite uniform, with the exception of neighborhoods where 30-40 percent of the inhabitants are from non-western backgrounds.

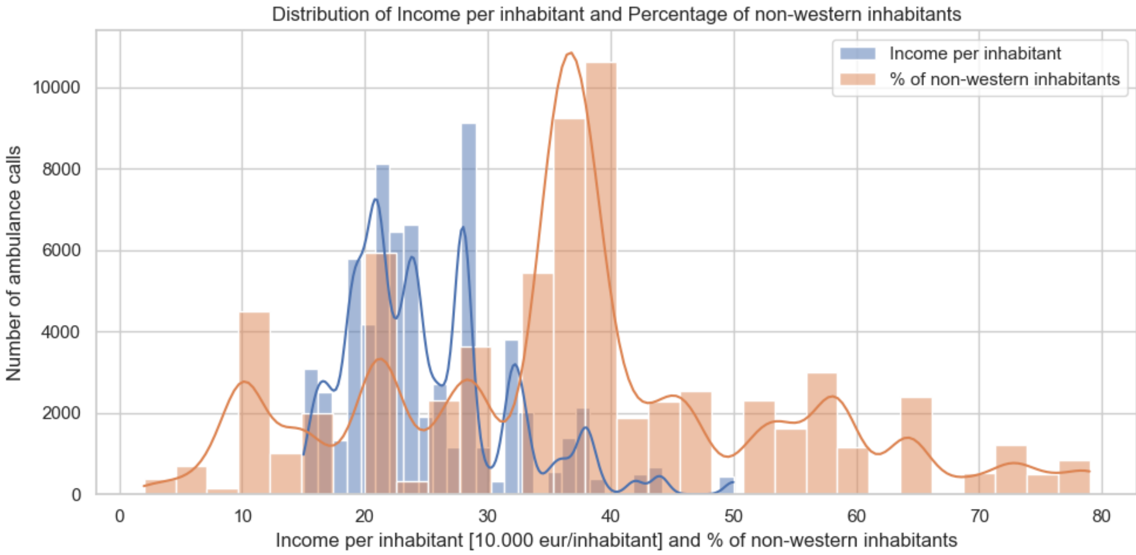


Figure 3.2: Distribution of income

3. **Percentage of households under social minimum**.

Figure 3.3 shows the percentage of households that live below the social minimum of the neighborhoods to which ambulances were dispatched in blue. This distribution seems to be quite random; a lot of calls were placed to neighborhoods where a low percentage of households live under the social norm, but there are also peaks for neighborhoods where a higher percentage live under the social norm.

4. **Population density**.

Figure 3.3 shows the population density of the areas where calls were made to in orange. Most calls were made to neighborhoods with a low population density, with the number of calls being moderate from a population density of around 7.500/km2 and onwards.

5. **Distance to hospital**.

Finally, **figure 3.3** (in green) clearly shows that most ambulance calls were made to neighborhoods that were close to a hospital, often within 2 km.

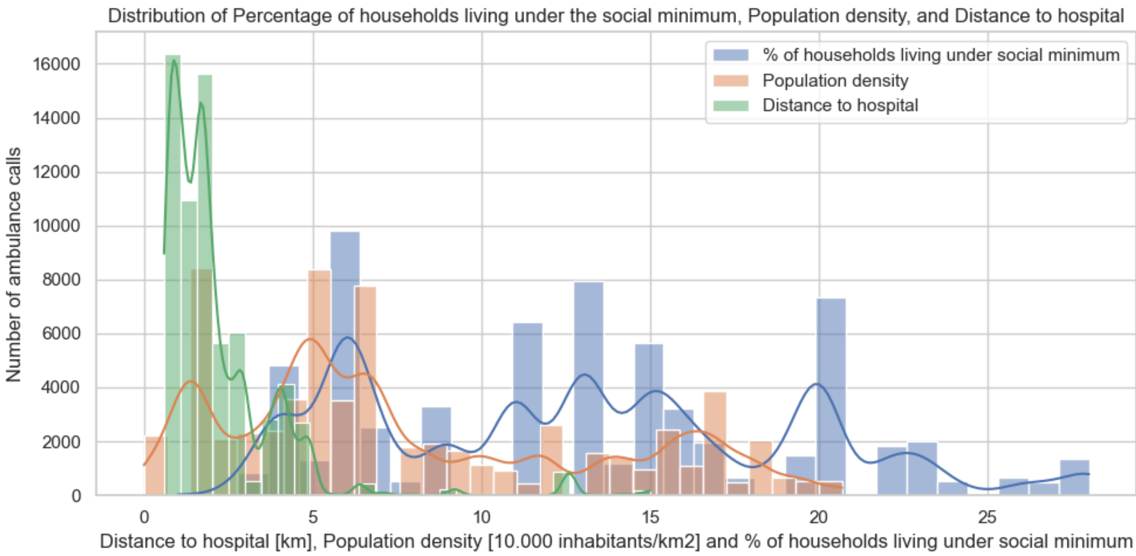


Figure 3.3: Distribution of households under social minimum

Before taking a look at the spatial distribution of these variables, it is important to look at the different correlations of the variables between themselves. We do this using a pairplot. The pairplot, as the name suggest, 'pairs' every variable with every other variable. This allows for a quick visualisation of potential correlations.

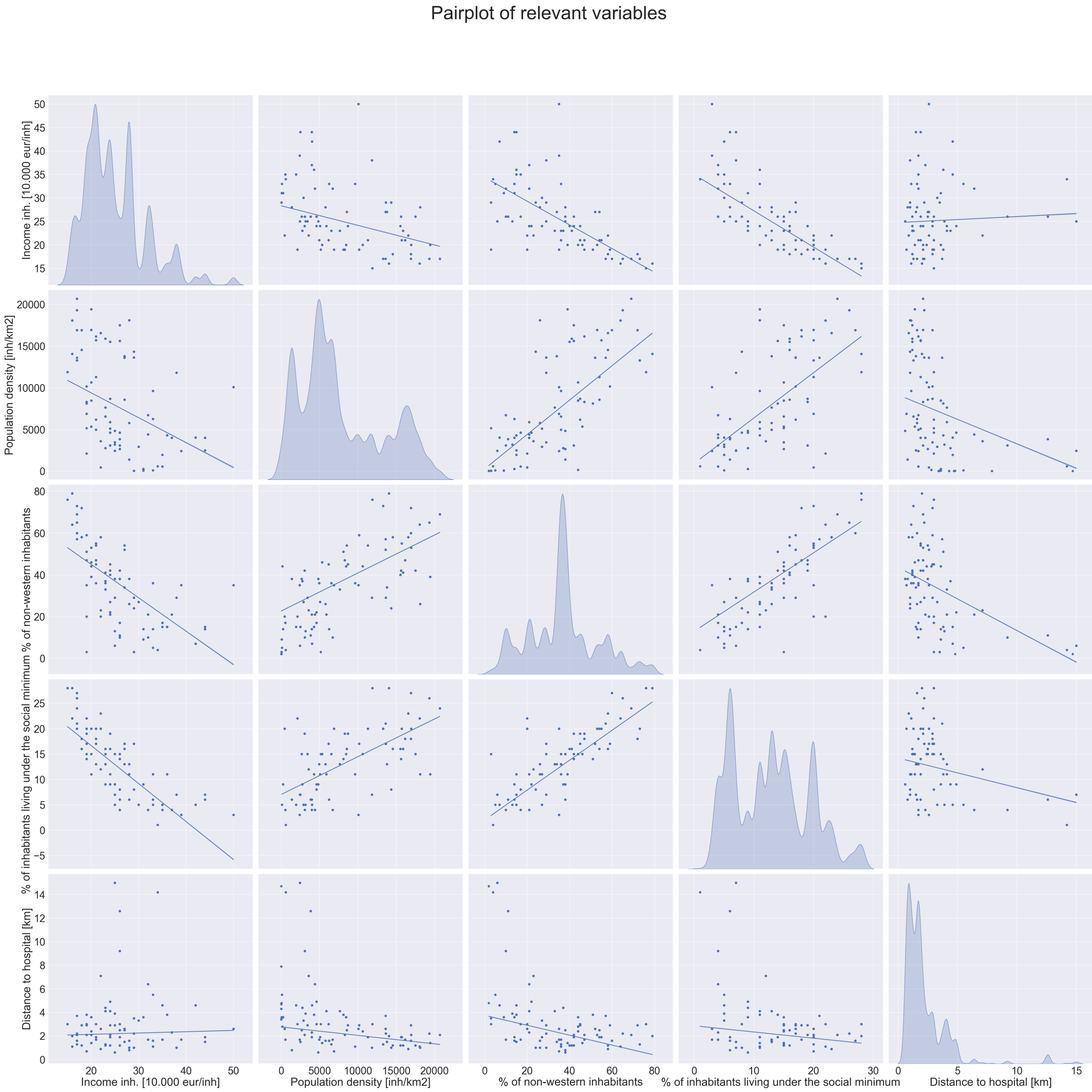


Figure 3.4: Pairplot

In **figure 3.4**, every dot represents a neighborhood with a value for variables x and y. This allows us to conclude that there is a relationship between income, percentage of people living under the social minimum and people from a non-western background. Neighborhoods with a higher percentage of non-western inhabitants often have a higher percentage of households that live under the social minimum and by extension have a lower average income per inhabitant. Interestingly, these neighborhoods often also have a higher population density. There seems to be no significant relation to the distance to a hospital in those neighborhoods. In fact, there seems to be no relation between the distance to a hospital and the other variables, with the exceptions of some outliers that are very far removed from the hospital.

Figure 3.5 shows different choropleths of the variables that we are analysing. The figure shows that it seems there is no relation between the number of ambulance calls in a neighborhood and the income per inhabitant, as there are neighborhoods with both high and low income that have a high number of calls. The same could be said for the spatial distribution of the population density compared to the ambulance calls, although lower density areas seem to get a little bit more calls. Distance to the hospital does seem to be correlated to the number of calls, as neighborhoods closer to the hospital tend to get more ambulance calls. This is all in line with the findings that were discussed earlier. The bottom two choropleths in **figure 3.5** show the areas that are socio-economically underdeveloped and compares it with the number of ambulance calls. Here a relatively weak correlation seems to arise for the south/southeastern part of Rotterdam and a neighborhood to the north. However, this correlation seems to be quite weak.

The ambulance calls seem to group in four areas; the north, the center, the south/southeast and one neighborhood in the south (away from the main city). This is interesting as the groupings of the other variables seem to differ more. There is no one variable that seems to directly correlate with the number of ambulance calls in Rotterdam.

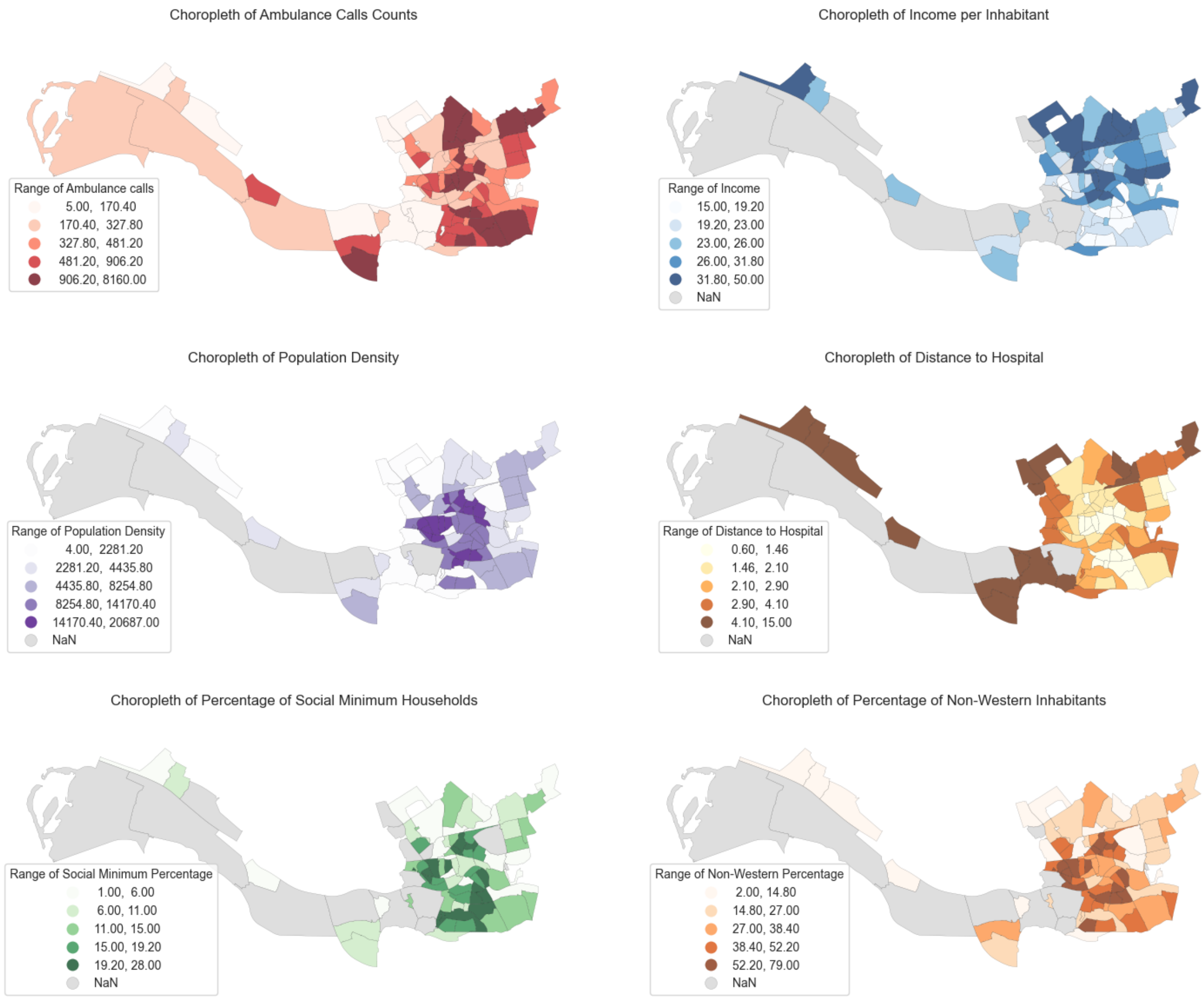


Figure 3.5: Choropleth of relevant variables

One last observation that stood out in all of the choropleths above is the amount of NaN-observations in the strip reaching to the West from the city center. To gain a better understanding into why this could be so, through spatial analysis, the population-density was plotted onto a basemap of Rotterdam in **figure 3.6**. This helped to visualize the actual geographical context of the area. Showing that this area consists primarily out of the industrial, harbour area known as the Port of Rotterdam. Since these areas are mostly not residential and also difficult to encompass within certain socio-economic characteristics it is argued the NaN-values in these neighborhoods will likely not influence the study's results significantly.

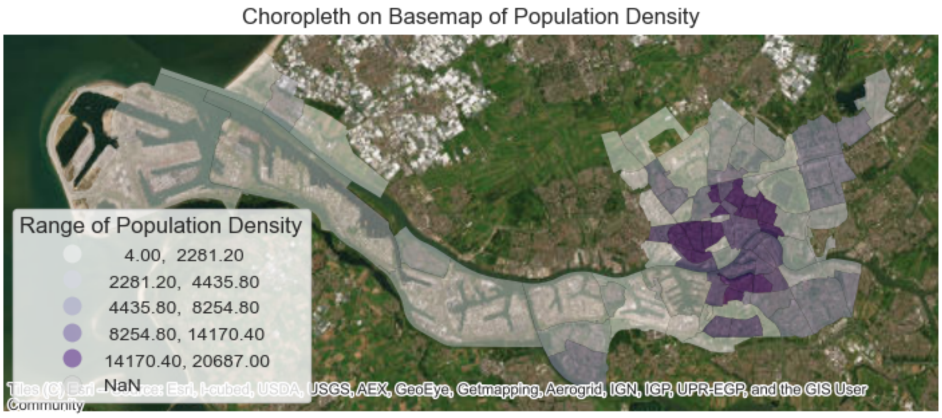


Figure 3.6: Choropleth on geographical context

3.6. Limitations on data and study

Now the EDA, through descriptive statistics and the choropleths, has been presented it is also important to define what limitations have been identified at this stage.

The first, and most prominent, limitation identified is the NaN-values which appear for all of the variables (except the ambulance calls) in quite some neighborhoods outside the center of Rotterdam. How we dealt with this is explained further in **section 3.7**.

A second limitation was the year for which to apply the socio-economic data. Since the ambulance calls data was for the 2017 - 2020 period, a fitting timeframe for the socio-economic data had to be applied. One option would have been to take yearly averages of the same period, however we chose to use a year in the middle; this resulted in 2018 being the year of measurement. The reasoning behind this decision is that socio-economic changes don't generally occur at a pace wherein this would have significantly changed within one- or two years, therefore making a year in the middle suitable. Also, after a quick inspection this year had a reasonable number of ambulance calls and is before the Covid-19 outbreak, which is known to have a great influence on the healthcare and way of living. The year 2018 therefore represents a 'usual' and representative year in our opinion.

Another limitation of the study is the interplay of other variables possibly affecting the ambulance calls besides socio-economic status. Certain factors such as neighborhoods with a significantly older population or were chemically hazardous fumes are present from nearby industries would not be seen as primarily socio-economic in the direct sense; meaning we don't take this from the variables. Indirectly they can be seen as socio-economic, reasoning that neighborhoods nearby industries are often lower-income for example. For this study we would have liked to include such factors as control variables. Sadly no suitable data could be found for this application.

Finally, it is important to note that we often use healthcare access and demand and ambulance call frequency interchangeably. The hypothesis that one would expect neighborhoods with a lower socio-economic status would have a higher call frequency is based on the assumption that healthcare demand and ambulance call frequency are related, which wasn't investigated.

3.7. Errors and missing values in data

One important factor, skewing the data, was found in the neighborhoods with the highest rate of ambulance calls. This could result in an omitted variable bias, wherein this skew is reasoned to be induced by socio-economic variables included in the study [7]. Therefore we looked into other possibilities. Finding that in Dijkzicht, Groot IJsselmonde, Hillegersberg Zuid and Vreewijk the high number of ambulance calls likely results from the hospitals situated in those neighborhoods. How this can be so is not quite clear however, since the data states from were an ambulance call was placed. While there is the easy conclusion that a hospital likely has high amounts of medical emerengies in and around its vicinity, it is not entirely logical that these would all consists of ambulance calls. This is one aspect to take into account throughout the further analysis but which is also open to further investigation.

Another limitation in the data, which mainly becomes visible from the choropleths, is the amount of NaN values in certain neighborhoods and for certain variables. Mainly some of the larger neighborhoods, positioned in the strip of Rotterdam extending itself from from the center in a Westward direction, misses data for all variables but the ambulance calls. Therefore drawing conclusions on the relation between ambulance calls and some socio-economic characteristics of these neighborhoods will not be possible in this study. An important side note here is that these areas consist mainly out of areas with low population density. Being situated near the Maas delta, were ships enter the industrial area of the Port of Rotterdam. This was also visualized in **figure 3.6**.

4
ANALYSIS

4.1. Experimental Design:

As seen in the previous section we conducted exploratory data analysis to find any hints of causation or correlation among our selected variables visually. Now we will perform an analysis that investigates the relationship between selected socio-economic factors and ambulance call count in Rotterdam neighborhoods, hypothesizing that socioeconomic status significantly influences emergency service patterns.

The primary argument of this analysis is to explore the relationship between the socio-economic status of neighborhoods in Rotterdam and the patterns in ambulance call counts. The experimental design involved clustering neighborhoods using K-Means based on socio-economic indicators and ambulance service data. This unsupervised learning approach was chosen to identify inherent groupings in the data without presupposing any specific relationship. As this is argued to provide added value on the EDA findings; seeing the clusters that were manually set up, based on the variable data, could be limiting our vision. This unsupervised method assigns the cluster based on a statistical algorithm and thus might provide a new perspective on what we already have.

4.2. In-Depth Analysis of Clustering Process

In the scientific investigation exploring the relationship between the socio-economic status of neighborhoods in Rotterdam and ambulance service patterns, a detailed clustering process was employed using the K-Means algorithm. This process began with the selection of key variables: Call Count (total ambulance calls per neighborhood), Income per Inhabitant, Population Density, Percentage of Non-Western Inhabitants, Percentage of Households Under the Social Minimum, and Distance to Hospital. These variables were chosen for their potential to illuminate the socio-economic landscape's impact on healthcare demands[3].

4.2.1. Specifying the number of clusters used

The optimal number of clusters was determined using the elbow method as shown in **figure 4.1** based on which the decision for the optimal number of clusters often involves a bit of subjectivity [12]. The key for us was to find a balance where increasing the number of clusters doesn't significantly decrease the within-cluster sum of squares (WCSS). We choose 4 clusters as a reasonable choice as the plot shows a noticeable bend or "elbow" around that number. This would suggest that adding more clusters beyond 4 doesn't provide as much benefit in terms of decreasing the WCSS. However, if the plot doesn't show a clear elbow or if the decrease in WCSS continues significantly beyond 4 clusters, it might be worth considering a different number which is not the case.

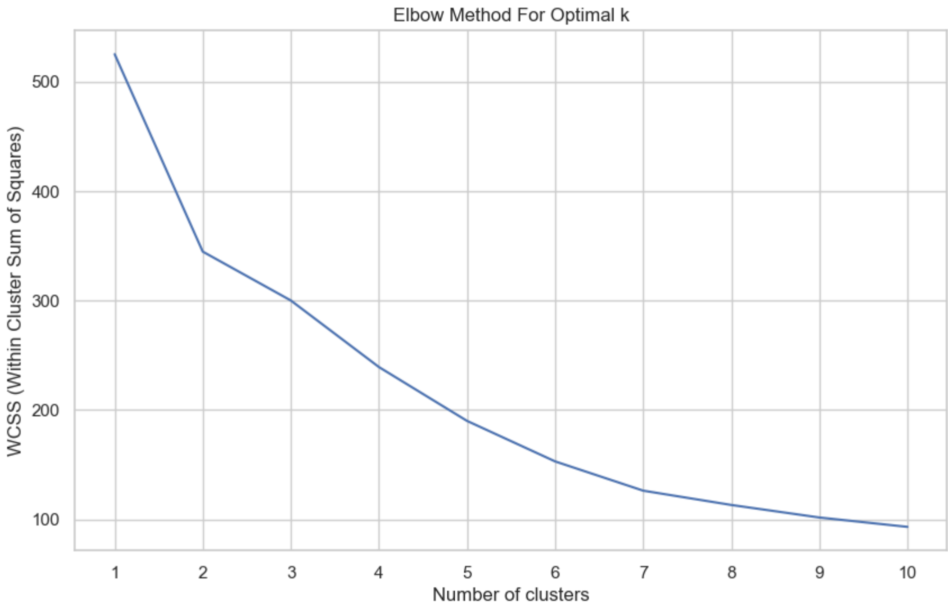


Figure 4.1: Elbow plot to determine the number of clusters to be initialized for clustering

4.2.2. Analysis per cluster

Upon applying K-Means clustering, four distinct clusters emerged, each reflecting unique characteristics of Rotterdam's neighborhoods as shown in the bar plot in **figure 4.2**. A detailed analysis of these clusters is mentioned below:

Cluster 0

This cluster might represent neighborhoods with high socio-economic status, much like cluster 2. The average income per inhabitant is highest compared to other clusters, and there is a lower percentage of households living under the social minimum. The population density in these areas is also quite low, as is the percentage of non-western inhabitants. This could indicate that this cluster represents the more well-off suburban neighborhoods. It has an average distance to the hospital. However, the gap between households under the social minimum and the high income might indicate that this cluster is more segregated, with a higher wealth gap. The number of ambulance calls is on the lower side, which could be attributed to its higher socio-economic status.

Hypothesis Correlation: The lower ambulance call rate in these neighborhoods could indicate a correlation with the higher socio-economic status.

Cluster 1

This cluster clearly denotes neighborhoods with low socio-economic status. It has a high population density, lowest income of all clusters, highest percentage of households living under the social minimum and highest percentage of non-western inhabitants. However, it also has a low average number of calls. This is interesting as it does not support our hypothesis. It also has an average to low distance to the hospital.

Hypothesis Correlation: The low socio-economic status of this cluster and also low number of calls do not align with our hypothesis.

Cluster 2

This cluster might also represent neighborhoods with high socio-economic status. The average income per inhabitant is relatively high compared to other clusters, and it has the lowest percentage of households living under the social minimum. The population density in these areas is also lowest of all clusters, as is the percentage of non-western inhabitants. This could indicate that this cluster represents the more well-off suburban neighborhoods. This is supported by the distance to the hospital, which is high, as hospitals are often in the middle of the city. The number of ambulance calls is lowest of all clusters, which could be attributed to its higher socio-economic status.

Hypothesis Correlation: The lower ambulance call rate in these neighborhoods could indicate a correlation with the higher socio-economic status.

Cluster 3
This cluster includes the neighborhoods with most ambulance calls. The neighborhoods have a mixed socio-economic status with an average median income, lower population density, average percentage of households living under the social medium and a medium-high percentage of non-western inhabitants. The neighborhoods in this cluster are closest to the hospital.

Hypothesis Correlation: This cluster provides some evidence that supports our hypothesis, but given the large number of ambulance calls one would expect the socio-economic status to be a lot worse. Hence, it contradicts the hypothesis.

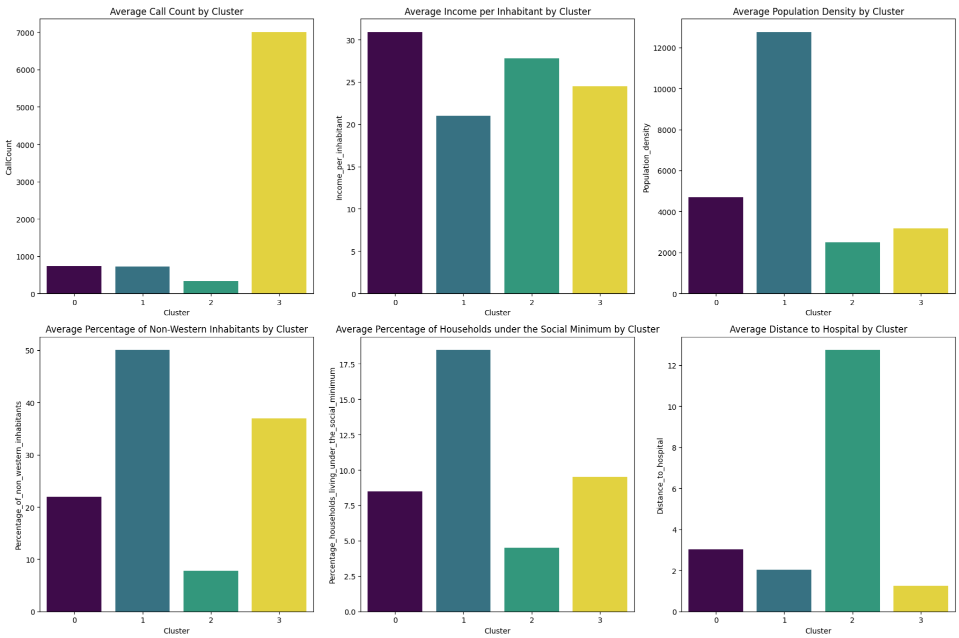


Figure 4.2: The Cluster Analysis

4.2.3. General findings from cluster analysis

The cluster analysis does not suggest a relationship between the socio-economic status of neighborhoods and the amount of ambulance calls. Generally, clusters 0-2 all have a low number of ambulance calls while differing a lot in socio-economic status. The one cluster with a lot of calls has mixed socio-economic status.

However, it's important to consider other factors such as healthcare accessibility, population age distribution, and specific health challenges in each area, which might also influence these findings. This analysis underscores the multifaceted nature of urban healthcare dynamics and the importance of considering socio-economic factors in emergency healthcare planning and resource allocation.

Moreover, the cluster analysis that we did can be better visualized using the choropleth graph in **figure 4.3**. Each color represents a distinct cluster, corresponding to the groupings based on socio-economic factors and ambulance call counts.

In Cluster 0 (i.e., Dark Purple) the spatial distribution in these areas might be more evenly distributed or located in specific parts of the city or outskirts of the city. If these neighborhoods align with higher socio-economic status, the fewer ambulance calls would corroborate the hypothesis. It could suggest better overall health, lower population density, or better access to healthcare facilities.

In Cluster 1 (i.e., Dark Blue) the spatial distribution for these neighborhoods might be primarily located in densely populated areas like the city center. These areas coincide with neighborhoods known for lower socio-economic status. However, the number of ambulance calls in these areas is low, contradicting the hypothesis.

In Cluster 2 (i.e., Green), the spatial distribution of these neighborhoods is on the far outskirts, with low population densities. This cluster may represent areas with high socio-economic characteristics. The relationship between ambulance calls and socio-economic status in these areas seems to support the hypothesis.

Cluster 3 (i.e., Yellow) only has two neighborhoods, but has the higher number of calls by far. The socio-economic factors are mixed, so it might suggest that these two neighborhoods are hotspots for ambulance calls.

The map in **figure 4.3** reveals how socio-economic factors and ambulance call frequencies are geographically distributed, highlighting areas of potential concern or interest. As of now, there seems to be a weak relationship between socio-economic factors and ambulance call frequencies.

One more interesting observation are the NaN-observations in the western strip extending from Rotterdam's city center (this was also seen in **figure 3.6**), identified in our choropleth map in **figure 4.3**, is attributed to the area's industrial nature, primarily the Port of Rotterdam. This region, being largely non-residential and industrial, does not fit the socio-economic and residential criteria of our analysis. Consequently, the lack of data in these zones is understood to have a minimal impact on the study's outcomes, highlighting the necessity of contextual geographical understanding in urban data analyses.

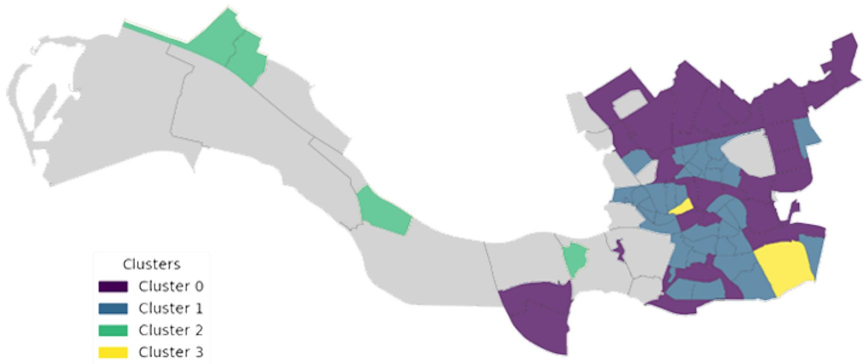


Figure 4.3: Choropleth showing the clusters in Rotterdam

4.3. Patterns and Supporting Points

This section starts by providing supporting points for the argument as stated in **section 4.1**, reflecting on how the data and output do- or do not support this argument. Secondly, patterns in the data, providing evidence for each supporting point will be provided.

4.3.1. Socio-Economic Diversity and Ambulance Service Utilization

- *Pattern:* The relationship between socio-economic factors and ambulance service utilization is complex and does not follow a straightforward pattern. While Cluster 2, with favorable socio-economic conditions, exhibits the lowest number of ambulance calls, the data from Clusters 0 and 1, which vary significantly in socio-economic statuses, show an identical number of ambulance calls. This suggests that other factors, possibly including healthcare accessibility, population density, or community health initiatives, might also play crucial roles.

- *Example:* Cluster 0, which has a lower average income per inhabitant at 30.18 and a high ambulance call count of around 900, contrasts with Cluster 2, which despite its lower income, shows fewer ambulance calls. This discrepancy suggests that income alone may not be a direct predictor of ambulance service utilization.

- *Outlier:* Cluster 2 stands out as an anomaly, having low income yet the lowest ambulance call count, which could be partly explained by its lower population density, indicating that demographic factors may significantly influence ambulance service demand.

- *Representation and Evidence:* The observed data underscore the nuanced impact of socio-economic diversity on ambulance service utilization. Contrary to a simple direct correlation, the findings highlight the multifaceted influence of socio-economic status, where variables such as income and population density interplay to shape healthcare service patterns. This complexity suggests that socio-economic status alone cannot predict ambulance service usage accurately without considering the broader socio-demographic context.

4.3.2. Population Density, Households under the social minimum and Non-western inhabitants

- *Pattern:* Higher population density correlates with a higher percentage of non-western inhabitants percentage and the percentage of households under the social minimum.
- *Example:* Cluster 1, with the highest population density (12,165), also has a high percentage of non-western inhabitants and percentage of households under the social minimum.
- *Representation and Evidence:* This pattern suggests that areas with more people are mostly non-western inhabitants and often have more households living under the social minimum, illustrating a relationship between population density and the number of non-western inhabitants and percentage of households under the social minimum.

4.3.3. Healthcare Accessibility and Service Utilization

- *Pattern:* The analysis revealed that healthcare accessibility, beyond mere physical distance to healthcare facilities, plays a crucial role in ambulance service utilization. Neighborhoods with varied socio-economic statuses showed differences in ambulance call frequencies that could not be solely attributed to income levels, population density, or proximity to hospitals.
- *Example:* In comparing Clusters 1 and 2, both of which are relatively economically divergent, the similarity in ambulance call frequencies suggests that factors such as the availability of primary healthcare services, public health initiatives, and community health literacy may significantly influence emergency healthcare demands.
- *Representation and Evidence:* This observation emphasizes the importance of comprehensive healthcare accessibility—including the availability of preventive care and health education—in shaping emergency service patterns. It implies that effective management of healthcare demands requires a broad approach that integrates socio-economic support systems, enhances healthcare literacy, and ensures equitable access to a range of healthcare services.

4.3.4. General findings from patterns and supporting points

The findings from the analysis challenge simplistic assumptions about the determinants of ambulance service utilization, highlighting the nuanced roles of healthcare accessibility and socio-economic factors. It becomes clear that neither socio-economic status, population density, nor physical proximity to hospitals can independently predict ambulance call patterns. Instead, a complex interplay of healthcare accessibility, community health literacy, and preventive care availability emerges as critical in understanding emergency healthcare demands. This complexity calls for multifaceted public health strategies that address the underlying socio-economic disparities, improve healthcare literacy, and ensure equitable access to comprehensive healthcare services.

4.4. Limitations of the Analysis

This study aims to provide a comprehensive understanding of the relationship between socio-economic factors and ambulance service patterns in Rotterdam. While the use of K-Means clustering and data visualization techniques has offered significant insights, there are inherent limitations to this analysis that must be acknowledged.

4.4.1. Clustering Limitations

The choice of K-Means clustering, while useful for identifying patterns within the data, has its limitations. K-Means assumes spherical clusters and may not capture the true complexity of socio-economic patterns.

Additionally, the method is sensitive to outliers, which can skew the results and affect the interpretability of the clusters.

4.4.2. Methodological Constraints

The analysis was also limited by the methodological choices made. The elbow method used to determine the number of clusters involves subjectivity, as the "elbow" is not always clear-cut, which may lead to different numbers of clusters if interpreted by another researcher. Furthermore, the lack of a temporal dimension means that changes in socio-economic status and ambulance service patterns over time were not considered, potentially missing out on trends and shifts in the data. Finally, the choice not to omit or handle the neighborhoods that could be considered outliers will have significantly influenced the analysis.

4.4.3. Generalizability of Findings

Another limitation is the generalizability of the findings. The study focuses on Rotterdam and may not be applicable to other cities or regions with different socio-economic landscapes or healthcare systems. The relationship observed in Rotterdam may differ significantly from those in other urban or rural areas.

4.4.4. Implications of Limitations

These limitations have implications for the study's findings. While the analysis suggests that there is not a strong relationship between socio-economic status and ambulance call counts, this conclusion is drawn within the context of the data and methods used. It is important for future research to address these limitations, perhaps by incorporating additional socio-economic variables, exploring alternative clustering algorithms, and including a longitudinal analysis to provide a more complete picture of the dynamics at play.

Recognizing these limitations is essential for interpreting the results accurately and for informing the direction of future research. The findings should be viewed as a contribution to a larger conversation about the intersection of socio-economic factors and healthcare services, rather than definitive conclusions.

5 DISCUSSIONS / CONCLUSIONS

Prior research, such as that by [8], has underscored a geographical correlation between healthcare and socio-economic status, suggesting that areas with lower socio-economic standings often face higher health risks and service demands. However, a gap exists in understanding how these dynamics play out in an urban context like Rotterdam, where socio-economic diversity is marked, and healthcare infrastructure is relatively advanced.

The core question guiding this investigation was: *"What is the relationship between the socio-economic status of neighborhoods and the number of ambulance calls in Rotterdam, Netherlands?"*. This question seeks to bridge the literature gap by applying data-driven analysis to the urban context, providing insights into the socioeconomic determinants of emergency healthcare service patterns.
The analysis revealed that the relationship between socio-economic factors and ambulance service utilization is complex and not straightforward. High socio-economic status and proximity to healthcare facilities did not consistently correlate with lower ambulance call frequencies, challenging conventional assumptions. Instead, healthcare accessibility and socio-economic diversity emerged as critical factors influencing emergency service patterns.

These findings suggest that emergency healthcare service utilization cannot be predicted solely based on socio-economic status or geographic proximity to healthcare facilities. Instead, they highlight the importance of comprehensive healthcare accessibility and the need for urban health policies that consider the multifaceted nature of healthcare demand.

By demonstrating the nuanced relationship between socio-economic factors and ambulance call patterns, this study fills a critical gap in the literature. It expands existing knowledge by illustrating that in an urban setting like Rotterdam, socio-economic diversity and healthcare accessibility play more significant roles in emergency healthcare utilization than previously understood.

The most plausible explanation for these findings is the comprehensive nature of Rotterdam's healthcare infrastructure, which may mitigate the impact of socio-economic disparities on emergency service utilization. Additionally, public health initiatives and community healthcare programs could play a role in evening out the demand for emergency services across socio-economic strata.

While the study's methodology was robust, alternative explanations such as data limitations or unmeasured variables (e.g., individual health behaviors, the severity of emergencies) could also influence the findings. Moreover, the assumption that ambulance call rate and healthcare demand are directly linked might be flawed.

This research addresses an important gap in the literature by providing a nuanced understanding of the socio-economic determinants of ambulance service utilization in an urban context. The insights gained are crucial for policymakers and healthcare planners striving to create equitable and efficient emergency healthcare systems in cities with diverse socio-economic landscapes.

The study's findings challenge simplistic assumptions about healthcare demand and socio-economic status, revealing the importance of accessibility and the complex interplay of factors influencing emergency service utilization. This research contributes to a more comprehensive understanding of urban healthcare dynamics, offering valuable insights for the development of targeted, effective public health strategies and policies.

Bibliography

[1] Nur Shahidah Marcus Eng Hock Ong Arul Earnest, Say Beng Tan. Geographical variation in ambulance calls is associated with socioeconomic status. *Academic Emergency Medicine*, 19:180–188, 2012. doi: doi:10.1111/j.1553-2712.2011.01280.x.

[2] M. Asaria, T. Doran, and R. Cookson. The costs of inequality: whole-population modelling study of lifetime inpatient hospital costs in the english national health service by level of neighbourhood deprivation. *Journal of Epidemiology and Community Health*, 2016.

[3] Paula Braveman and Laura Gottlieb. The social determinants of health: it's time to consider the causes of the causes. *Public Health Reports*, 129(Suppl 2):19–31, 2014. doi: 10.1177/00333549141291S206.

[4] Centraal bureau voor de statistiek. Wijk- en buurtkaart 2018, 2024. URL `https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2018`. Accessed: January 10, 2024.

[5] Marloes M. Hoogerbrugge Martijn J. Burger. Neighborhood-based social capital and life satisfaction: the case of rotterdam, the netherlands. *Urban Geography*, 39:1484–1509, 05 2018.

[6] Mehdipanah R. Gullón P. Triguero-Mas M. Cole, H. V. Breaking down and building up: Gentrification, its drivers, and urban health inequality. *current environmental health reports*, 8:157–166, 2021.

[7] P. A. Jargowsky. Omitted variable bias. *Encyclopedia of social measurement*, 2:919–924, 2005.

[8] M. Marmot. Epidemiology of socioeconomic status and health: are determinants within countries the same as between countries? *Annals of the New York Academy of Sciences*, 896:16–29, 1999.

[9] C. Muntaner, J. Lynch, and G. D. Smith. Social capital, disorganized communities, and the third way: understanding the retreat from structural inequalities in epidemiology and public health. *Political And Economic Determinants of Population Health and Well-Being*, pages 427–450, 2020.

[10] PDOK. 112 Calls Dataset, 2022. URL `https://112-nederland.nl/`. Accessed: January 10, 2024.

[11] den Hertog F. Van Oers H. Schuit A. J. Storm, I. How to improve collaboration between the public health sector and other policy sectors to reduce health inequalities?–a study in sixteen municipalities in the netherlands. *International journal for equity in health*, 15:1–14, 2016.

[12] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering*, volume 336, page 012017. IOP Publishing, April 2018.