APANPS5205 - Project Report
Social and Economic Factors Influencing New Yorkers despair

Group: FP-3
Names: Robin Klaver, Jose Medina, Xinyu Zhang, Tirth R Gandhi

Introduction

In March 2024[1], the Citizens Budget Commission released a new survey revealing that only 30% of the 6,600 households surveyed rated conditions in the city as "excellent or good." This marks a significant decline from the previous survey conducted in 2017, where 50% of the respondents expressed satisfaction with the city's quality of life.

The survey results showed varying levels of satisfaction across boroughs, with Manhattan leading at 40% of residents satisfied with the quality of life. Brooklyn, Queens, and Staten Island followed, while the Bronx had the lowest satisfaction rate, with only 21% of residents rating the quality of life as "good or excellent."

Safety concerns have increased, worrying many New Yorkers about the future. Only 37% of respondents believe public safety in their neighborhood is "excellent or good," a drop from 50% in 2017. Similarly, just half of New Yorkers feel safe riding the subway during the day, compared to over 80% in 2017. The survey also indicates increased flight risk; only half of the respondents plan to stay in the city over the next five years, down from nearly 60% in 2017.

Research Problem

How do social and economic factors influence 311 service requests in New York City.

Research Questions:

- Are residents from neighborhoods with higher income or education levels more likely to file complaints due to better knowledge of government channels? (Spatial Analysis, Cluster)
- Do complaints get resolved faster in higher-income, higher-education neighborhoods? (Regression)
- What does text mining reveal about the resolution time and complaint type? (Sentiment analysis)

Overview of the used data

Our largest dataset contains information about issues and resolutions on a local government level, specifically the city of New York. While the total dataset contains 36.8M rows and 41

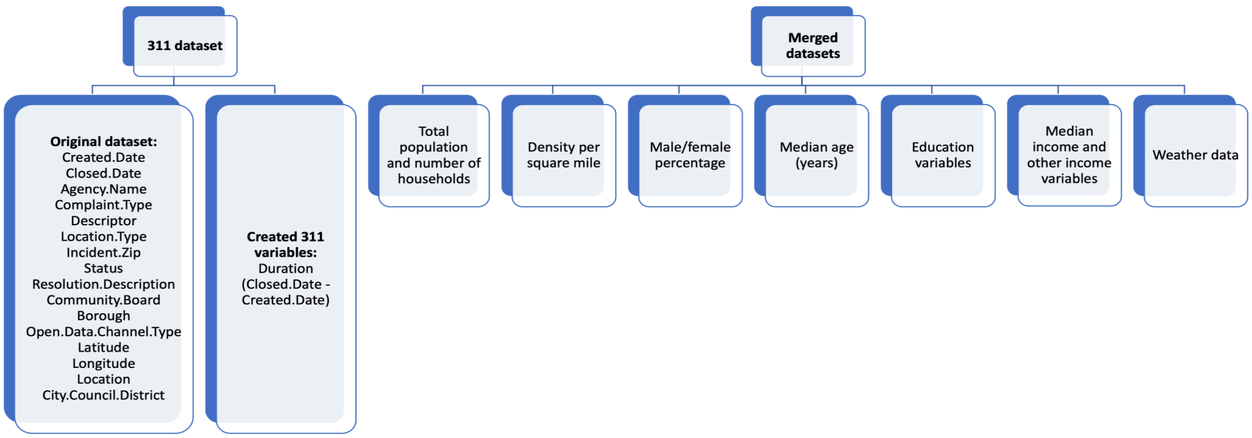
[1] https://www.nytimes.com/2024/03/19/nyregion/new-yorkers-poll-survey.html

columns, for this research only the full year of 2023 was used. In addition, several external datasets were integrated to complement the data:

- “Census_total_population_2020.csv” is a Demographic and Housing Characteristics dataset from the U.S. Census Bureau containing the total population per zip code in 2020. The dataset consists of 1827 rows with zip codes and the associated total population. This for the total population, to get an estimate of complaints per resident.
- “Density_opendatasoft.csv" is a geographical repository maintained by Opendatasoft. This source used information from the U.S. Postal Service, U.S. Census Bureau, National Weather Service, American Community Survey, and the IRS to construct datasets. We used this dataset as recent density data was not available from Census. However, multiple zip-codes were randomly cross-checked with other sources and the density data seems to accurately represent the density per square mile. The dataset contains 1812 rows with zip-codes. This data will be used as an input for multiple analysis as the number of certain types of complaints, like residential noise complaints, could be impacted by the density.
- “Census_2022_gender_age.csv” is an American Community Survey 5-Year Estimates Data Profiles dataset from the U.S. Census Bureau containing information about gender and median age per zip code. Although the survey runs till 2022, the data is from 2020. The data is based on a survey sample, and the listed values are in between the 90 percent margin of error lower and upper confidence bounds. The dataset consists of 1827 rows and will be used in analysis to determine whether these characteristics are associated with the number of complaints.
- “CensusReporterEducation.csv” is an educational dataset from the U.S. Census Bureau and provides information on the education level of residents aged 25 years and older. The dataset categorizes education level from no schooling completed to advanced education in all levels. The dataset contains 250 rows and 52 columns and will be used to determine whether these characteristics are associated with the number of complaints and resolution time.
- “acs2022_5yr_B19001_86000US11414.csv” is also sourced from the U.S. Census Bureau and provides income data from 2018-2022. It contains the number of households in 16 income levels, starting at 10k until 100k in 5k increments, followed by 100k to 200k in 25k increments, and 200 and above. The dataset has 250 rows and 36 columns and will be used to determine the effect of income on the number of complaints and the resolution time.
- “Median_household_income_2022.csv” is an American Community Survey 5-Year Estimates Subject dataset from the U.S. Census Bureau containing information about the median income in the last 12 months 2022. Just like the dataset with information about gender and age, the data is from 2020 despite the survey running till 2022. This data is also based on a survey sample, and the listed values are in between the 90 percent margin of error lower and upper confidence bounds.
- “weather.csv” is sourced from Visual Crossing Weather, a leading provider of weather data and enterprise analysis tools. The dataset contains New York Cities' precipitation in inches and average temperature in Fahrenheit for every day of 2023. The dataset will be

used to determine whether certain types of complaint can be predicted using weather data.

An overview of the 311-data and integrated datasets can be found in the following illustration. An overview of all used files is enclosed in appendix II. The next chapter will contain more information about the efforts to prepare the data and newly created variables.



Efforts to prepare the data

After importing the “311_Service_Request all 2023.csv” containing a full year of 311-data, multiple variables were checked for data consistency. For example, "Police.Precident" and "Police.Precincts" both contained consistently incorrect data. Moreover, multiple variables like Community.Districts and Borough.Boundaries were dropped because they were not described. Also, the Zip.Codes variable does not reflect the actual relevant zip code and was dropped. Lastly, the variables "Facility.Type", "Due.Date", "BBL", "Vehicle.Type", Taxi.Company.Borough", "Taxi.Pick.Up.Location", "Bridge.Highway.Name", "Bridge.Highway.Direction", "Road.Ramp", "Bridge.Highway.Segment" are irrelevant for our current research and are dropped.

To identify missing values, blank cells were replaced with NA. As all analysis is based on Incident.Zip as the key identifier, 38,605 rows with missing zip-codes were deleted. As the dataset has 3,225,374 rows in total this was not deemed problematic for further analysis. Next, the Duration variable was created by subtracting the Closed.Date from the Created.Date and by defining it in minutes as the unit of measurement.

The csv-files containing population, density, and gender/age were joined first. Not all zip-codes could be joined with population and density data, resulting in dropping 15,465 rows with NA-values in both variables. Next, “CensusReporterEducation.csv” was imported and the variables (B15003001-B15003025) were renamed to meaningful variable names. As this data is highly dimensional, multiple summarizing variables were created: ratio of no completed education, ratio of lower school completed, ratio of high school or alternative credential completed, and ratio of residents finishing at least a Bachelor degree. As this data is also used in several analyses, 2,188 rows with missing values were omitted. Similarly, income data was imported and renamed. As this dataset is also highly dimensional, several new new columns were added: "ratio_less_than_50K", "ratio_50K_to_99_999", "ratio_100K_to_149_999", "ratio_200K_or_more". As this structure, using multiple variables for income, might be suboptimal for certain analyses, another dataset was merged containing the median income in the last 12 months. In total, 3,131 rows were omitted because no income data was available for the Incident.Zip code. Lastly, the weather data was merged by extracting the data from the variable containing a data and a time.

Lastly, after assessing remaining data from the multiple datasets, 73 variables were dropped because they were deemed irrelevant for further analysis. This includes 15 variables from the 311-dataset, containing mainly information about the location. However, Incident.Zip is mainly used for our analysis. Also, 58 variables in the education and income data were dropped because the previously calculated summarizing variables will be used for further analysis.

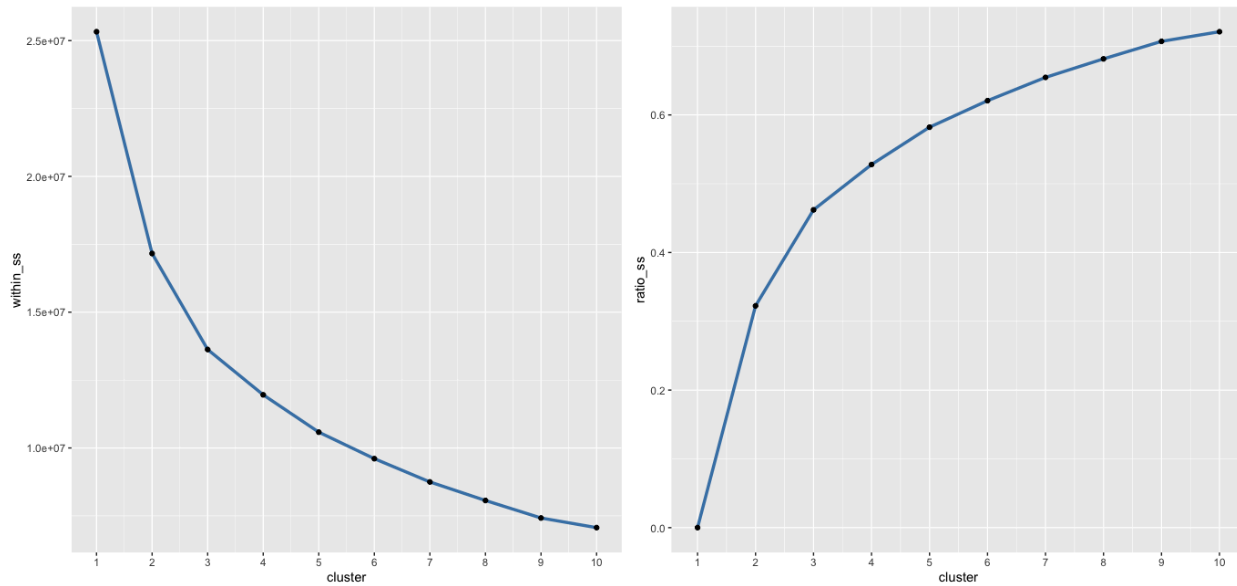
Data analysis methods

Clustering residents

To identify reported complaints with similar demographic backgrounds, a cluster analysis was performed. The needs based variables used for clustering are density per square mile, as density might influence the type and number of complaints. Residential noise complaints might for example be higher for residents in apartment buildings. Also, the percentage of male residents, median age, education data, and median income were added. Please note that the percentage of female residents was not included, as it represents the remaining percentage after accounting for males and could cause this dimension to be overweight.

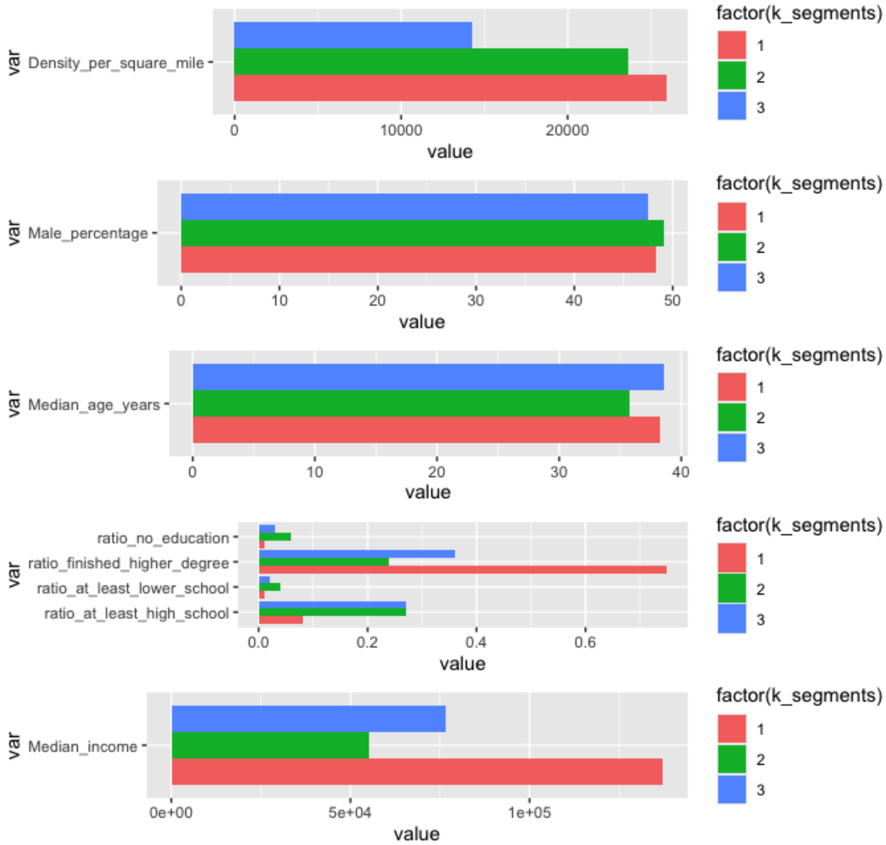
Lastly, the clustering was performed both with and without the variables temperature and precipitation. One might argue that temperature and precipitation influence certain resident needs and should therefore be included. On the other hand, this only applies to a subset of complaint types and is from a different category than variables describing demographic backgrounds. For the sake of simplicity, this report will describe the clustering without the weather variables. The weather variables came out exactly the same across clusters, and it seemed the sum of squares plots were also affected by adding those variables. The sum of squared plot and clusters for the variant with weather data are enclosed in Appendix I.

To perform the clustering, the data was first checked for missing values, of which there were none. Subsequently, the data was standardized to make sure certain variables are not overweighted. A data-driven approach, using the total within cluster sum of squares and a ratio plot, was taken to determine the number of clusters. The number of clusters is inferred from a change in the graph, known as the elbow. Alternatively, the ratio of between cluster sum of squares and total sum of squares is computed and the elbow is assessed. Both plots support both a 2 and 3 cluster solution. The 2-cluster solution is slightly clearer in the graph, however, the 3-cluster solution captures more nuanced distinctions. Therefore, we proceeded with a 3-cluster solution.



The 3-cluster solution results in the following cluster sizes and characteristics, as inferred from the cluster graph below. The number of complaints per resident is calculated by first grouping the data on zip code, then the number of complaints per resident is calculated by dividing the total number of complaints per zip code by the zip-code population. The average number of complaints per resident is then calculated by grouping on clusters.

Cluster	Total number of complaints	% of total complaints	Average number of complaints per resident	Characteristics
1	614,170	19%	0.312	Highest density, high median age, highly educated, high median income.
2	1,064,138	34%	0.382	High density, lowest median age, lowest education, lowest median income
3	1,487,677	47%	0.316	Lowest density, higher median age, higher educated, higher median income



In summary, we can classify the clusters as follows:

- **Cluster 1:** “Manhattan professionals” who are highly educated, live in high-density neighborhoods, and file the lowest number of complaints per resident.
- **Cluster 2:** “Urban Dwellers” in high-density neighborhoods (for example apartment buildings), who are on average lower educated, have a lower median income, and file the most complaints per resident.
- **Cluster 3:** ““Wealthy Suburban Residents” Represents complaints from residents in lower-density neighborhoods with a high income, education and median age.

Clusters as an input for regression analysis and decision trees

Subsequently, the clusters were added to the dataset for regression analysis and decision tree modeling to evaluate their impact on resolution duration and number of complaints per resident, assessing if the clusters significantly influence the models. In order to achieve this, a new subset of the data was created. The data is grouped by Incident.Zip with the median duration and number of complaints per resident as dependent variables. The number of complaints per resident variable was calculated by dividing the number of complaints per zip code by its population. The independent variables in the following analyses are the same variables as used for clustering.

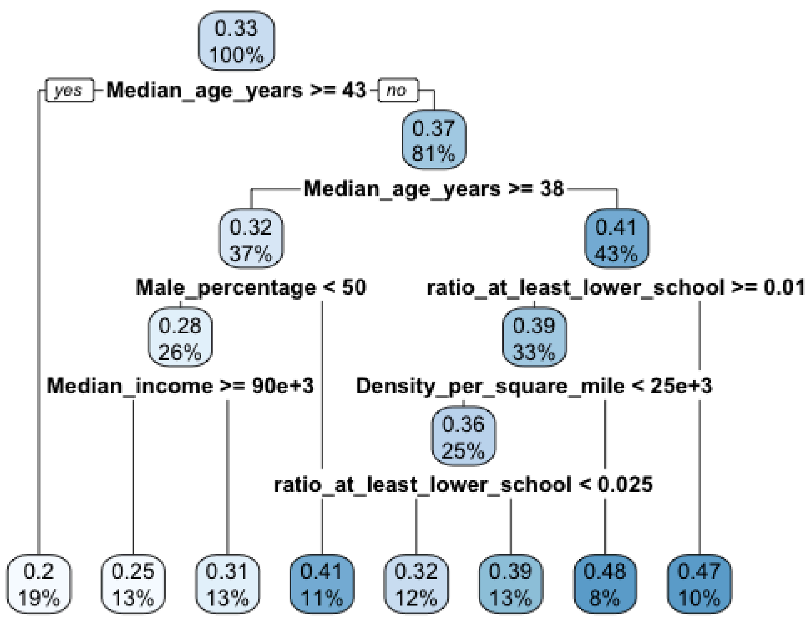
Duration of complaint resolution

The model doesn't seem to adequately capture the duration of complaint resolution. The adjusted R-squared value is negative and none of the variables is significant. The decision tree also has an RMSE of 30 hours. Therefore, it seems that these variables don't adequately explain a significant difference in complaint resolution time. This data might be too aggregated, as resolution times are heavily influenced by the complaint type.

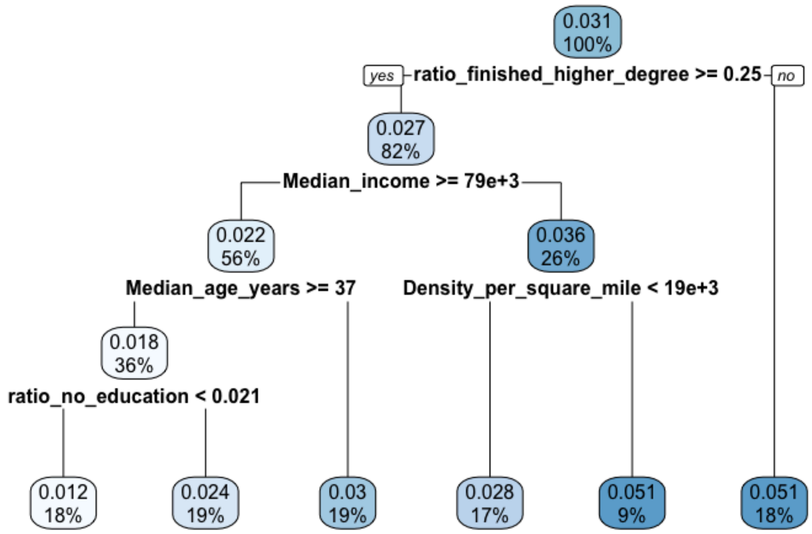
Number of complaints

The regression model for the number of complaints per resident has an adjusted R-squared of 19%, with a RMSE of 0.18. The male percentage was found to be significantly associated with increased complaints per resident, while median income and median age showed a significant association with fewer complaints per resident. Median age was found to be the most significant variable. Cluster was not found to be significantly impacting the number of complaints.

A decision tree had a lower RMSE of 0.16. The most important variable in this tree was a median age above 43 years, which was associated with a lower number of complaints. Other factors involved the male percentage, education levels and density as shown in the decision tree below.



By creating a subset of the data focusing solely on residential noise complaints, the RMSE significantly decreased to 0.014. For this complaint category, education level was the most important factor, with more noise complaints observed in less educated areas. The cluster does again not appear to be one of the most important factors in the decision tree model.



Regression analysis on resolution duration for different complaint types

We used a linear regression model to analyze the relationship between the characteristics of different postal code areas and the complaint resolution time (Duration). First, we extracted the specific hour of complaint creation (Created.Hour) from the Created.Date variable. Based on this time, we categorized the complaint times into five time periods: Early Morning (0:00-8:00), Morning (8:00-12:00), Afternoon (12:00-17:00), Evening (17:00-22:00), and Night (22:00-24:00). Each complaint was assigned to its respective time period category (Time.Period), and corresponding dummy variables (e.g. Time.PeriodEarlyMorning) were created. Similarly, based on the Open.Data.Channel.Type variable, we categorized the complaint submission channels into ONLINE, UNKNOWN, PHONE, MOBILE, and OTHER, and created corresponding dummy variables (e.g. Open.Data.Channel.Type_ONLINE). By aggregating the total number of complaints within each postal code area (Incident.Zip), we calculated the number of complaints per resident (Complaints_per_Resident).

Since the processing times for different types of complaints vary, we decided to analyze the resolution times separately for each type of complaint to eliminate the impact of this variation. Based on the most common complaint types in the dataset, we chose to conduct in-depth analyses for Illegal Parking, Noise - Residential, and Blocked Driveway complaints.

Regression model variables

In our regression model, we included input variables such as population density, male percentage, median age, education levels (no education, at least lower school, at least high school, higher degree), median income, number of households, complaints per resident, five time periods of complaint creation and five complaint submission channels and output variable median duration. This model aims to analyze the relationship between these factors and the complaint resolution time across different postal code areas. It is worth noting that to avoid multicollinearity issues, the model automatically selects a reference category and excludes it. The coefficients of the reference category serve as the comparison point for all other categories in the model. For our analysis, Time.PeriodNight and Open.Data.Channel.Type_OTHER are the baseline categories for Time.Period and Open.Data.Channel.Type, respectively. Therefore, in the regression results, their coefficients are displayed as NA.

Regression result - Illegal Parking

In the analysis of the duration of illegal parking complaints, we identified several significant influencing factors. An increase in the proportion of residents with at least a lower school education, at least a high school education, and a higher degree significantly reduces the complaint duration. These effects are significant at the 5%, 1%, and 0.1% significance levels, respectively. Regarding income, median income is significant at the 5% level. For every one dollar increase, the complaint duration increases by approximately 0.0003903 minutes. For the time periods, complaints submitted in the morning, afternoon, and evening, compared to those submitted at night, have significantly shorter complaint durations. These reductions are significant at the 1%, 1%, and 5% significance levels, respectively. The R-squared value of 0.421 means that these predictors can explain 42.1% of the variation in the duration of illegal parking complaints. This indicates that the model has some ability to understand and predict the duration of illegal parking complaints.

Regression result - Noise – Residential

Similarly, for Noise – Residential complaints, an increase in the proportion of residents with at least a high school education and those with a higher degree significantly shortens the complaint duration. These effects are statistically significant at the 1% and 0.1% levels, respectively. For every additional dollar in median income, the duration increases by approximately 0.0002202 minutes, significant at the 5% level. Afternoon complaints are resolved faster than night-time complaints. This result is significant at the 5% level. The model explains approximately 30% of the variability in complaint duration (R-squared = 0.2998). The model has a moderate explanatory power. The significant F-statistic and its p-value confirm that the model is statistically significant overall.

Regression result - Blocked Driveway

In Blocked Driveway complaints, higher population density is associated with a decrease in complaint resolution time. Speciy significant at the 5% level. An increase in the proportion of residents with at least a lower school education, at least a high school education, and a higher degree significantly reduces the complaint duration. These effects are significant at the 5%, 5%, and 0.1% significance levels, respectively. For every additional dollar in median income, the duration increases by approximately 0.0004383 minutes. This finding is statistically significant at the 5% level. The R-squared value of 0.4083 indicates that approximately 40.83% of the variability in the duration of blocked driveway complaints is explained by the model.

Sentiment analysis

Complaint Resolution Analysis

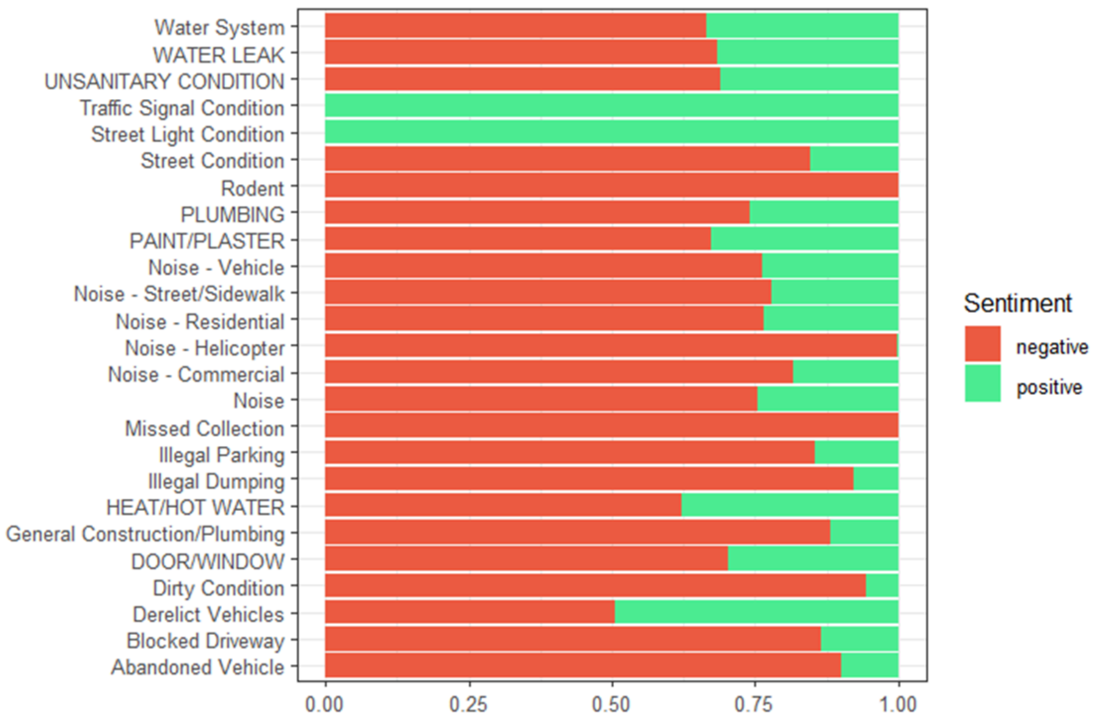
Our project aims to uncover insights about resolution times and complaint types through text mining of NYC311 service requests. NYC311 handles thousands of inquiries, comments, and service requests daily. The dataset we analyzed includes information such as complaint type, responding agency, and geographic location. By focusing on the Resolution Description, which captures the last action taken by the responding agency, we can glean valuable information about the nature and efficiency of complaint resolutions.

The Resolution Description, which provides a narrative of the final actions taken on service requests, typically contains around 20 words and 135 characters. For example, a typical resolution description might be: "The Department of Housing Preservation and Development responded to a complaint of no heat or hot water and was advised by a tenant in the building that heat and hot water had been restored. If the condition still exists, please file a new complaint."

Complaint Type serves as the primary categorization of incidents or conditions reported by residents. Defined by the responding agencies, these categories can expand to meet evolving customer demands. As of the data download date, there are 208 distinct complaint types. This detailed categorization allows us to analyze the data comprehensively, identifying trends and areas for improvement.

Binary Sentiment with Top 25 Complaint Types

For our initial analysis, we adopted a straightforward approach, using the Bing Binary Sentiment Lexicon to categorize words based on their positive or negative valence. We analyzed the proportion of positive and negative words across the top 25 complaint types.

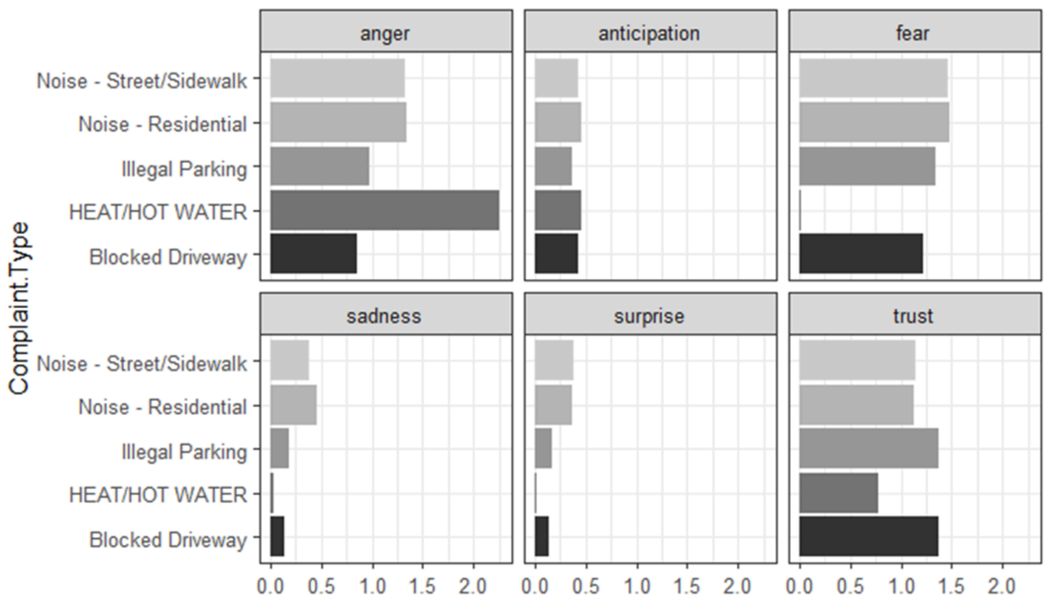


Unsurprisingly, complaint types like Rodent, Missed Collection, and Noise were overwhelmingly negative. The sentiment distribution across various complaint types reveals key areas of resident dissatisfaction. Complaints about Water System, Water Leak, and unsanitary conditions show a high proportion of negative sentiments, indicating substantial resident dissatisfaction with water-related issues and sanitation. These categories represent critical areas for improvement to address resident concerns effectively. On the other hand, complaints about Traffic Signal Condition and Street Light Condition display a positive sentiment distribution. However, this is primarily due to the fact that the Resolution Description tends to be a polite instruction on how to submit a complaint to the appropriate agency, rather than a depiction of a positive experience.

Noise-related complaints, whether from vehicles, streets, residential areas, helicopters, or commercial areas, predominantly carry negative sentiments. This persistent frustration highlights the need for better noise management across various environments. Additionally, high negative sentiment proportions in HEAT/HOT WATER, Illegal Parking, and Blocked Driveway complaints underscore ongoing issues that significantly impact residents' daily lives.

Emotion Lexicon with Top 5 Complaint Types

In the second part of our analysis, we utilized the NRC Emotion Lexicon from the TidyText library to categorize words based on the emotions they convey. This lexicon classifies words into eight primary emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, along with positive and negative sentiments. This approach allowed us to explore the emotional layers present in the complaint descriptions. As in the previous section, we analyzed the relationship between emotions and Complaint Type, providing a more detailed understanding of how specific issues resonate emotionally with residents.

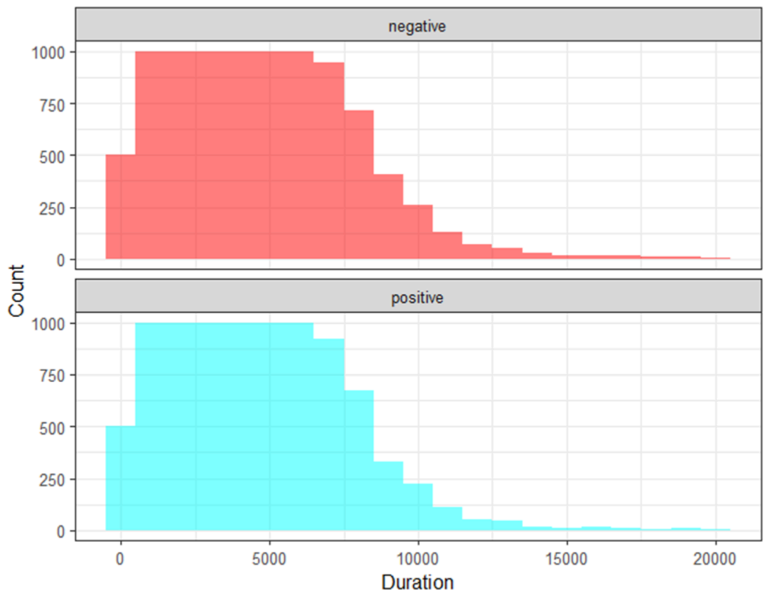


Examining complaint types across different emotion categories reveals distinct patterns of resident dissatisfaction and concern. In the "Anger" category, the high frequency of HEAT/HOT WATER complaints indicates significant frustration among residents about heating and hot water issues. In the "Anticipation" category, although less frequent, complaints about HEAT/HOT WATER and Blocked Driveway suggest these issues are linked to residents' expectations and hopes for timely resolutions. The "Fear" category emphasizes the impact of Noise complaints, suggesting these issues might compromise residents' sense of safety and security.

The “Sadness” and "Surprise" category have fewer complaints overall, but Noise complaints still stand out, indicating unexpected noise disturbances are significant. Finally, the "Trust" category is dominated by HEAT/HOT WATER and Blocked Driveway complaints, underscoring the importance of reliable heating and water services in maintaining residents' trust. Overall, the analysis suggests that improving services related to heating, hot water, parking, and noise management could significantly enhance resident satisfaction and trust.

Binary Sentiment and Resolution Time

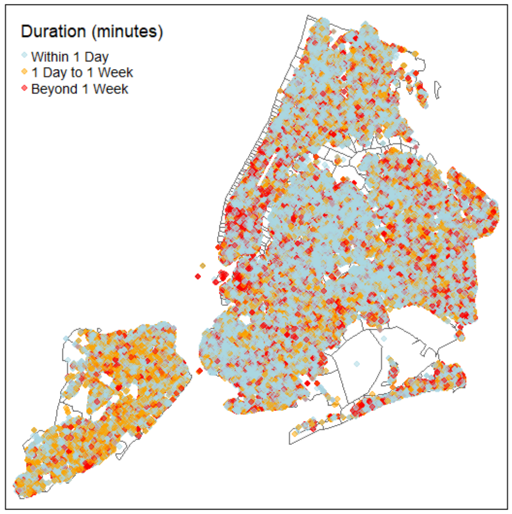
The median resolution time for a complaint in our dataset is 7hr (411 minutes), while the average time is 11 days (15,934 minutes). Our final analysis aimed to identify a correlation between the proportion of positive sentiment and resolution time. We utilized the Binary Sentiment Lexicon to simplify the analysis and plotted the results in a histogram, comparing negative and positive sentiments.



The histogram displays a correlation between resolution times and resident sentiments, highlighting both positive and negative feedback. Most complaints are resolved relatively quickly, with the highest frequency of resolution times between 0 and 3 days (5000 minutes). In this range, the proportion of positive sentiments is slightly higher, indicating resident satisfaction with prompt complaint resolution. Although negative sentiments are present, they are in smaller proportions compared to positive sentiments, suggesting that faster resolutions generally yield more favorable responses. As resolution times extend beyond 3 days (5000 minutes), negative sentiments become more prominent, especially noticeable in the 7 days (10,000) to 10 days (15,000+ minutes) range.

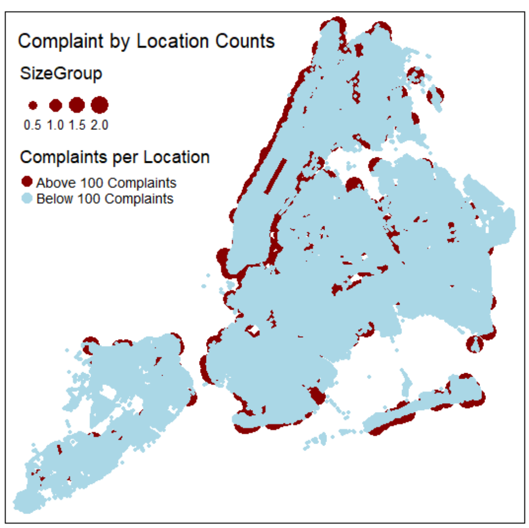
Spatial analysis

Our spatial analysis has revealed that complaints are filed throughout NYC at least once. The duration for complaint length can be grouped into two or three different categories depending on how highly Staten Island is weighted in future analysis. Wealthier neighborhoods in highly dense neighborhoods tend to have lower duration times, however repair work can skew the results.



Staten Island makes up the bulk of complaints that last beyond a week while the other boroughs have substantially fewer of those complaints. If Staten Island is a focus of less concern then two duration groups would suffice which is consistent with our cluster analysis. Complaints taking longer than one week to resolve revolve around repair work like elevator, floors, doors and windows, water leaks, or other hazardous conditions along with vehicle complaints like helicopter noise and for hire vehicles. Between a day and week involve hot water, garbage collection and disposal, noise, vehicles, and other hazardous conditions. When we shift the analysis to complaints per location, there is no correlation between how long complaints last and how often a complaint is filed.

When we examine the amount of complaints by location, we find that two thresholds of more than 100 complaints and less than 100 complaints would suffice.



The most problematic areas are near bodies of water where more people are gathered which is consistent with our prior findings and near public spaces like parks, even the few darkred spots in-land are near public spaces or business districts when you cross-verify with Google Maps. When we examined the aggregate characteristics of female and male percentages, ratio of people who earn over 200k by location, and ratio of higher education completion, complaints from more problematic locations tended to be slightly worse off than non-problematic locations. The regression analysis supports my findings.

Conclusions and recommendations

Complaints are more frequently filed in areas with lower socioeconomic status, lower education levels, and lower median income. High-income and well-educated areas, such as "Manhattan professionals," file the fewest complaints per resident, followed by "Wealthy Suburban Residents," with the most complaints coming from "Urban Dwellers" in high-density neighborhoods. This is also confirmed by spatial analysis. Geographically, Staten Island accounts for the majority of time-consuming maintenance complaints, with most complaints concentrated around bodies of water like lakes and the Hudson River, as well as in densely populated centers such as commercial areas and public spaces. Despite the initial assumption that higher income and education levels would lead to more complaints due to better knowledge of government channels, the data does not support this.

For complaint resolution times, higher education levels consistently lead to faster resolutions across all complaint types. This suggests that educated residents are better at communication and problem-solving, facilitating more efficient complaint handling. Conversely, higher median income is associated with longer resolution times, possibly due to the complexity or higher expectations of complaints from wealthier areas. Complaints submitted at night take longer to resolve, likely due to reduced staff availability. Higher population density correlates with quicker complaint resolutions, probably because of more efficient resource allocation and streamlined processes.

Text mining and sentiment analysis revealed that complaints about water systems, waste management, and unsanitary conditions are predominantly negative, indicating significant resident dissatisfaction. Heat/Hot Water complaints are often associated with anger and fear, highlighting concerns about safety. Quicker resolutions generally lead to more positive sentiments, showing the importance of timely responses.

Overall, the data indicates that lower socioeconomic factors are more directly associated with higher complaint frequencies, while higher education levels improve complaint resolution efficiency. Wealthier areas may require more targeted resources to handle complex complaints effectively.

Practical implications for decision makers

Focus on Timely Resolutions: The correlation between quick resolution times and positive resident sentiment suggests that efforts should be concentrated on reducing the time it takes to address complaints. Implementing more efficient processes and increasing resource allocation for high-frequency complaint types can help achieve this goal.

Targeted Interventions: The high levels of dissatisfaction with water-related issues and noise disturbances indicate areas that require targeted interventions. Investments in infrastructure improvements, stricter enforcement of noise regulations, and better maintenance of water systems can address these persistent issues.

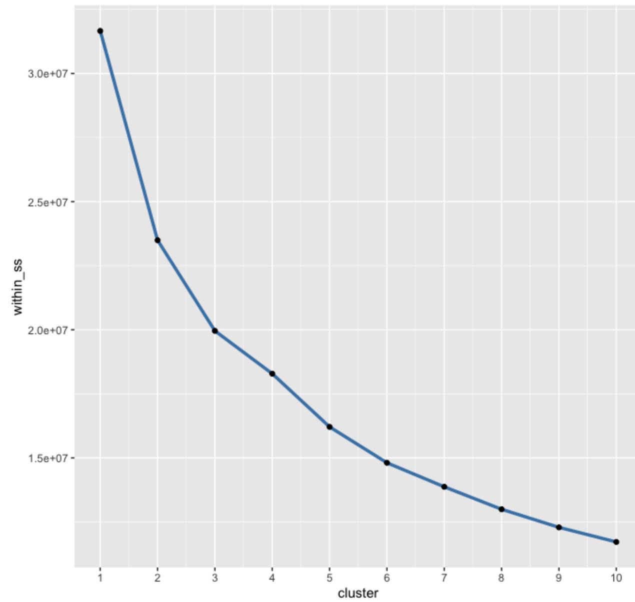
Leverage Education to Improve Communication: Since higher education levels are linked to faster complaint resolutions, initiatives to educate residents on how to effectively communicate their issues can be beneficial. Providing resources on the 311 website on how to file complaints and interact with city services can empower residents, leading to quicker and more efficient resolutions. For clusters that file the most complaints, the city could conduct community workshops to educate residents about how to effectively use 311 services, understand their rights, and engage with local governments.

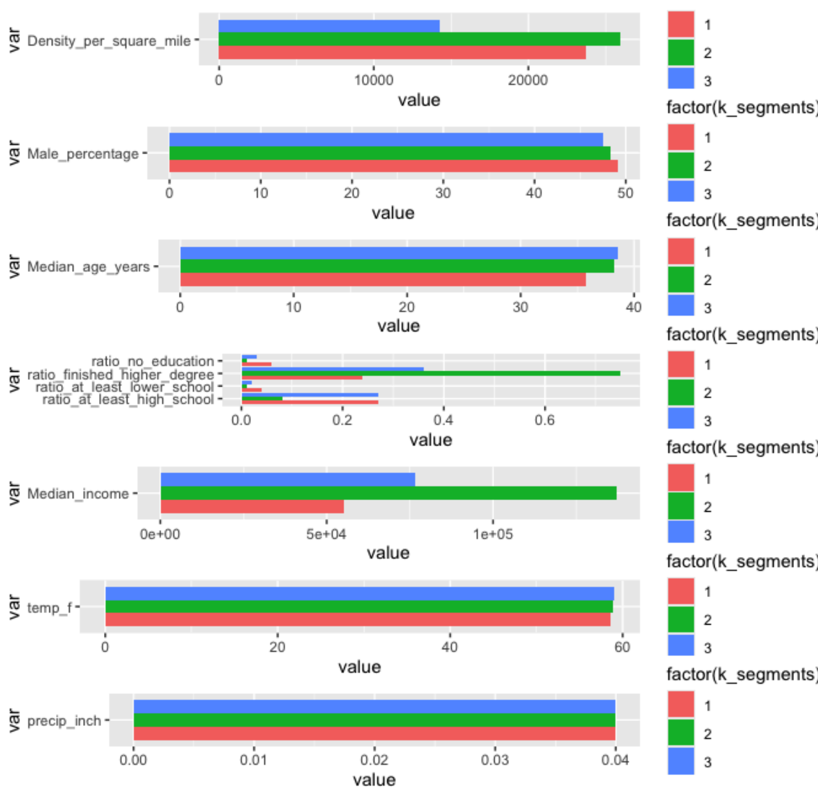
Optimize Resource Allocation Based on Time of Day: The data shows that complaints filed at night take longer to resolve. To address this, the city can optimize resource allocation by ensuring adequate staffing levels during night hours. This approach will help reduce the backlog of complaints and improve response times for issues reported overnight.

References

Paolicelli, A. (2024, March 20). *Survey finds New York City residents unhappy with quality of life*. Survey: New York City residents unhappy with quality of life. https://ny1.com/nyc/all-boroughs/news/2024/03/20/survey-finds-new-york-city-residents-unhappy-with-quality-of-life

Appendix I: Sum of squared plot and clusters including weather data





Appendix II: Enclosed files

R Scripts

1. 311_analysis_data: cleaning and merging of the dataset used for analysis
2. Sentiment Analysis
3. SpatialAnalysis
4. Cluster analysis
5. Regression_model

CSV Files

1. 311_analysis_data: dataset used for analysis
2. Census_total_population_2020.csv
3. Density_opendatasoft.csv“
4. Census_2022_gender_age.csv
5. CensusReporterEducation.csv
6. acs2022_5yr_B19001_86000US11414.csv (household income)
7. Median_household_income_2022.csv
8. weather.csv

Reports

1. Final presentation Social and Economic Factors Influencing New Yorkers despair.pdf
2. APANPS5205 - FP 3 - Project Proposal.pdf

Recording

1. Presentation recording.mp4