Final Proposal

Giovanni Lunetta

3/20/23

1. Research question:

Are there specific types of 311 complaints that are more common in certain neighborhoods of NYC, and is there a relationship between neighborhood demographics and the types of complaints received?

2. Methods:

a. Data collection: I will use the publicly available NYC 311 request dataset, which contains information on all 311 service requests received by the city. I will limit our analysis to all of 2022.

b. Data cleaning and preparation: I will clean and preprocess the data to ensure it is ready for analysis. This will include handling missing data, normalizing data types, and extracting relevant features such as neighborhood, complaint type, and demographic data.

c. Descriptive analysis: I will start by visualizing the overall distribution of 311 complaints by type and neighborhood. I will also examine trends in complaint volume over time.

d. Inferential analysis: To test our hypothesis that certain complaints are more common in certain neighborhoods, I will use statistical tests such as chi-squared tests and logistic regression to analyze the relationship between neighborhood demographics and the types of complaints received.

e. Visualization: I will use maps and other visualizations to help communicate our findings and identify spatial patterns in 311 complaints.

f. Expected outcomes: I expect to identify patterns and trends in the types of 311 complaints received in different neighborhoods of NYC. I also expect to find evidence of a relationship between neighborhood demographics and the types of complaints received. These findings could help city agencies better target resources and improve service delivery to residents.

3. Model:

a. Clustering: I could use clustering algorithms such as k-means to group neighborhoods together based on the types of complaints received. This could help identify patterns and similarities across different areas of the city.

b. Classification: I could build a classification model to predict the type of complaint based on the neighborhood and other relevant variables, such as time of day or demographics. This could help understand which types of complaints are most likely to occur in which neighborhoods.

4. Conclusion:

Analyzing NYC 311 data can provide valuable insights into neighborhood-level trends in quality of life concerns and service delivery. By exploring patterns and relationships between demographic and complaint data, we can help identify areas in need of attention and provide evidence-based recommendations for improving service delivery in NYC.