Bad Landlords II: Councilor Breadon

Team 1

Risheet Nair(‘24) - risheetn@bu.edu
Chen Gong(‘23) - gc0216@bu.edu
Chengyu Zhang(‘23) - chengyuz@bu.edu
Rohan Sawant(‘24) - rohan16@bu.edu

1 Introduction:

The purpose of this project is to increase the accessibility and transparency of the Boston planning, zoning, and development process so her constituents and residents across Boston can better understand the evolving landscape of the city. The end goal is to build a trackable system for Boston property violations to then determine which landlords are bad landlords.

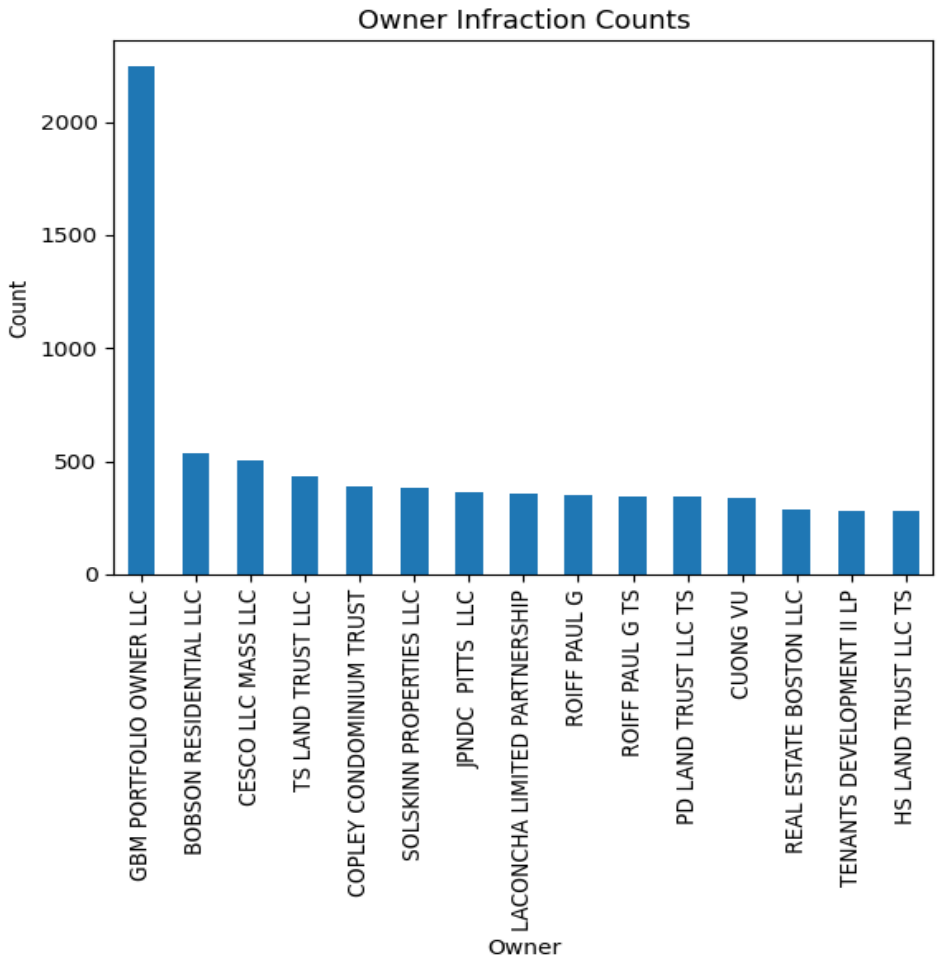
The impact that we hope to have by the end of the semester is to provide usable data visualizations that make it possible to create a threshold for what constitutes a bad landlord. It is currently tough to classify which violations are considered extremely severe and enough to mark a landlord as bad. With the analysis we've done and have left to do, we would like to make the line a little bit clearer so as to figure out what ordinance is needed to define the threshold for the bad landlord list.

In addition to this goal, we chose to pursue an extension project in which we analyzed how different social factors influence the number and types of violations committed by landlords.

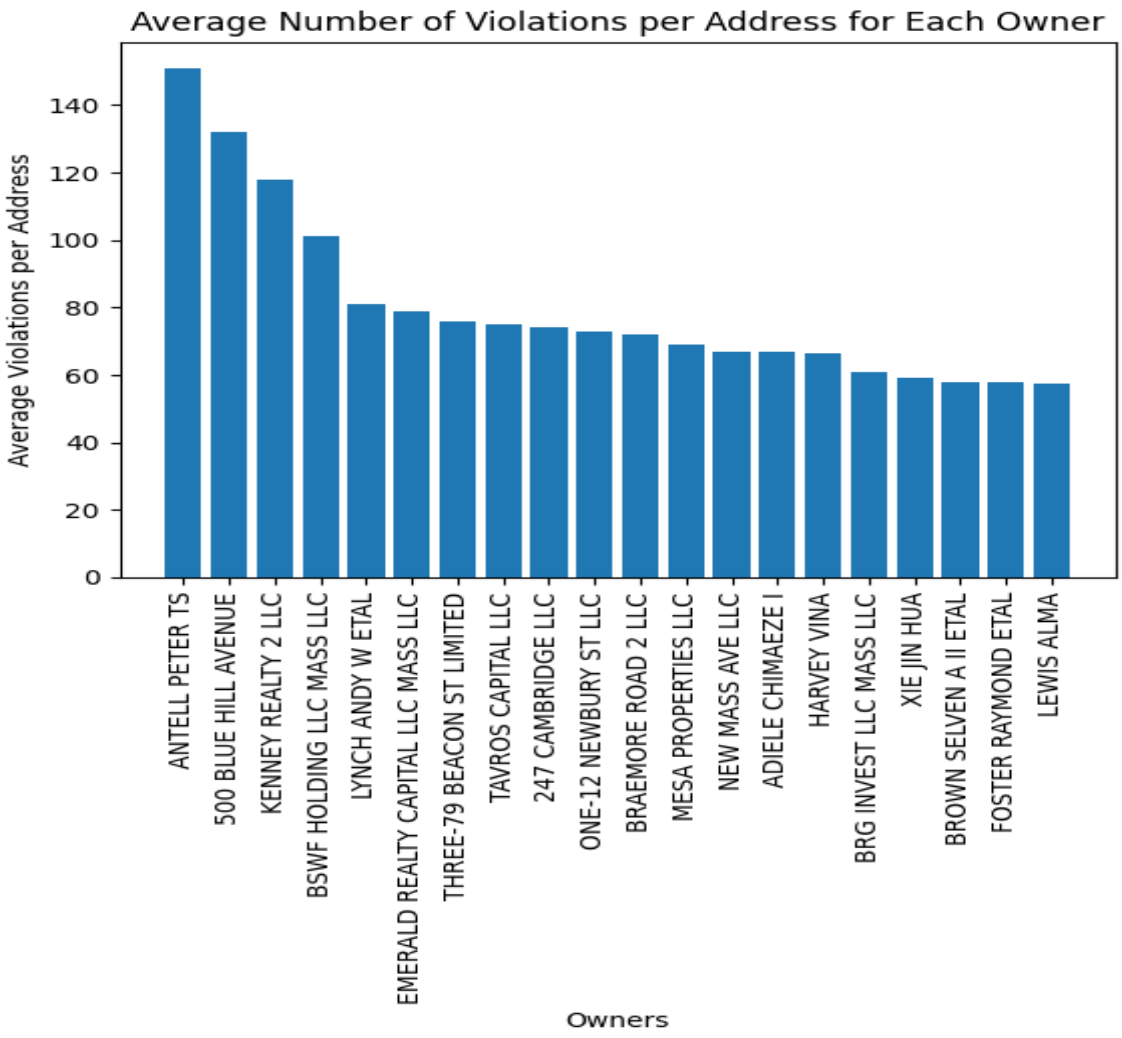
The primary datasets we used to do our base analysis and extension project were: rentsmart, buildings and property violations, affordable housing stock data, income restricted housing inventory, and the climate ready boston social vulnerability dataset.

2 Base Analysis:

First is the owner infraction count. We calculated the violation numbers of each owner. And the violation number is as follows. The x axis represents the owner name(the name of the landlord). And the y axis represents the violation number.

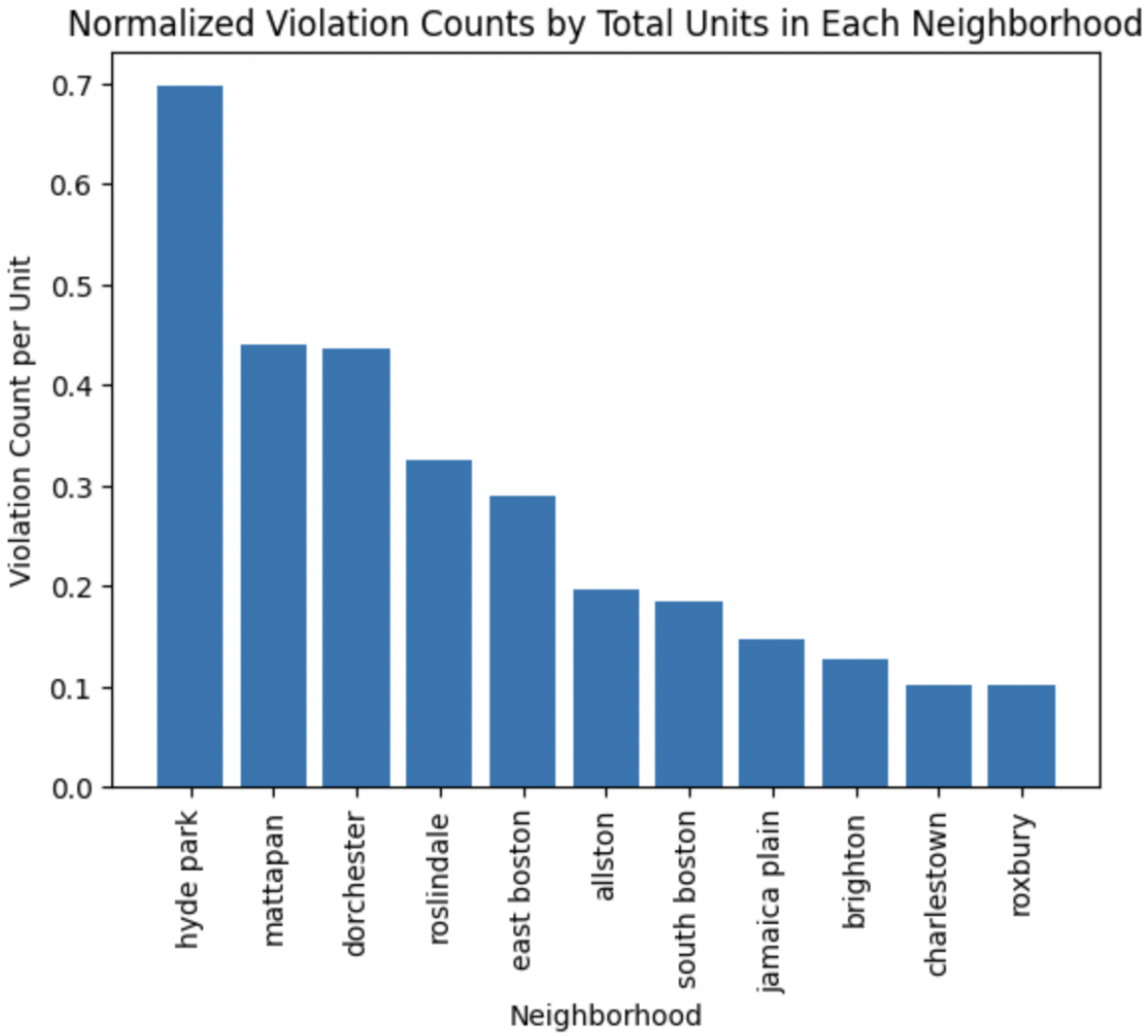


Then we normalize the data by each owners' address count, which means how many violations per house/apartment. We can notice that the ranking changes a lot. The x axis represents the Owner name(the name of landlord). And the y axis represents the violation number per address.

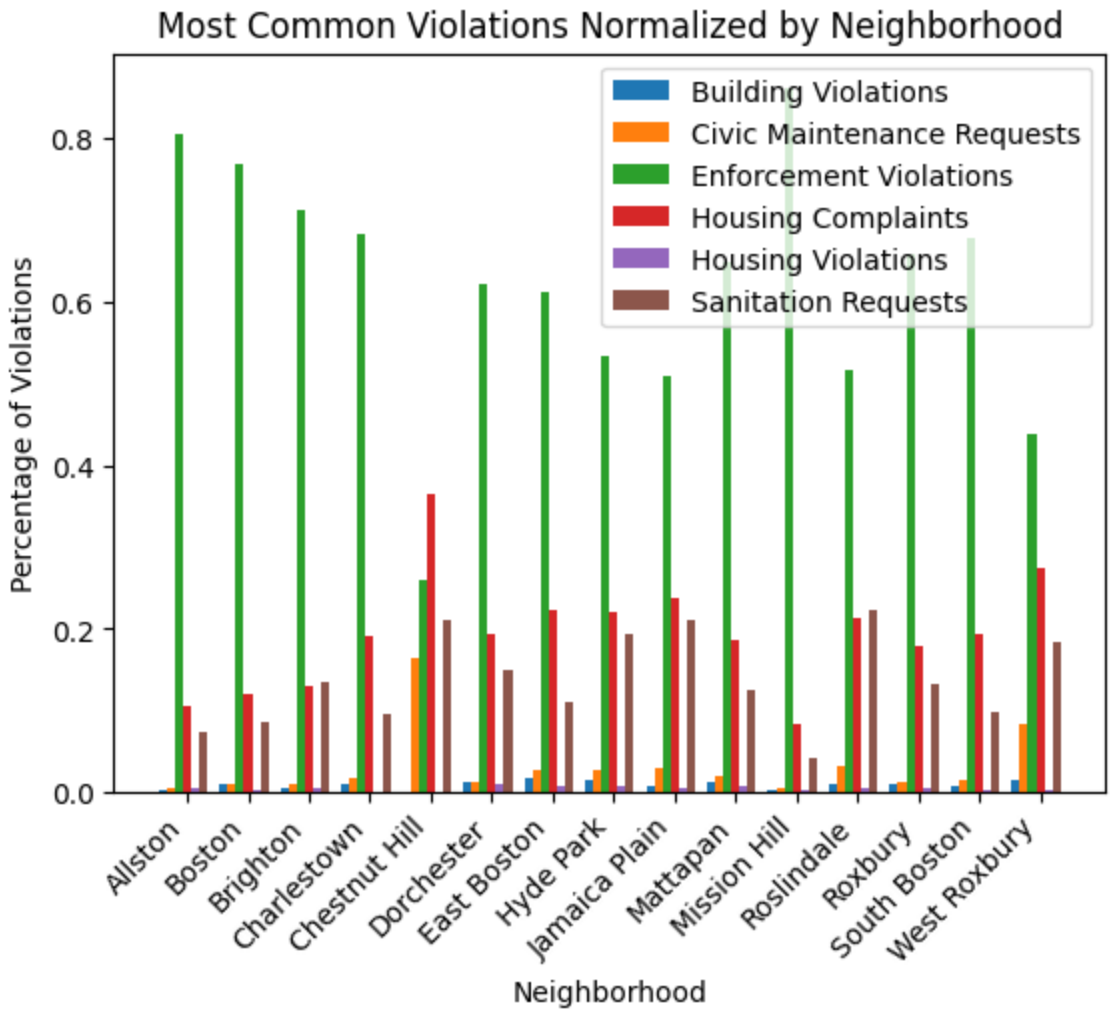


From the graph above we can conclude that Antell Peter TS is the property manager with the greatest number of violations per unit owned.

We now display the normalized violation counts by neighborhood as follows:



The above graph was normalized by how many total units were in each neighborhood. So we can now see that Hyde Park and Mattapan have the highest number of violations based on total units in each respective neighborhood.



The above graph shows the most common violations normalized by Neighborhood. We can see each neighborhoods' different violation types in percentage. The x value represents the Neighborhood, and violation types are represented by different colors. The y value of this plot is the percentage of each violation type.

Key Questions:

1: What is the spectrum of violations and severity in regards to worst landlords classifications?: Based on the above graph, enforcement violations are by far the most common types of violations that occur in almost every neighborhood. These violations are closely followed by housing complaints. However, it was tough to mark a threshold for severity as the data we had access to was limited.

2: What landlords are non-compliant in terms of overall volume?: As we saw from the normalized violation count bar graph, the two major landlords with the greatest number of violations are Antell Peter TS and Kenney Realty 2 LLC. This data was normalized by the average number of units owned per address for each landlord. The neighborhoods in which the largest number of violations occur by number of total units are Hyde Park and Mattapan. Particularly, in Hyde Park, there is a large percentage of properties that have been reported for a violation. We found it wasn't possible to merge the datasets that indicated which landlords committed particular violations in which neighborhoods, especially in terms of severity so that may be something to look into for next semester.

3: Which landlords are impacting the vulnerable communities the most?: This was a part of our extension project so this issue is addressed in the next section.

3 Extension Analysis:

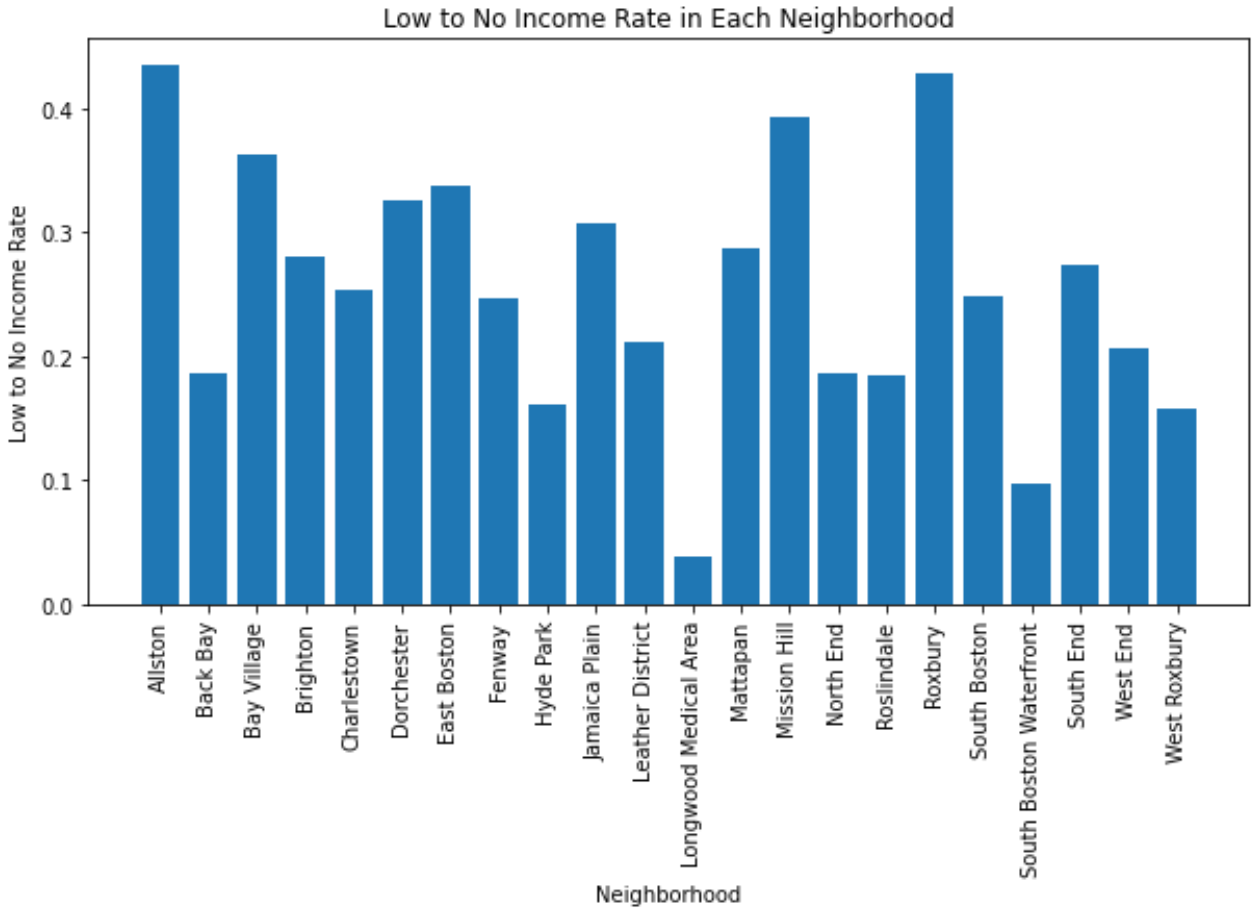
Our Extension proposal is determining the influence of various social factors on the number/types of violations that landlords commit. Some factors that we might consider are: the number of children/elderly people, the number of individuals who don't speak English as their first language, the number of minorities who live in a particular neighborhood, etc.. Since housing is a part of society, it is possible that landlords' violations might be influenced by the social environment of the neighborhood. This is what we hope to analyze.

Some questions for analysis might include: What social factors influence the number violations most and what are the social features of different neighborhoods?

By analyzing the Climate_Ready_Boston_Social_Vulnerability dataset, we find the social environment of different neighborhoods.

We began by analyzing how many low to no income people reside in each neighborhood. A lack of financial resources impacts a household's ability to prepare for a disaster event and to support friends and neighborhoods. For

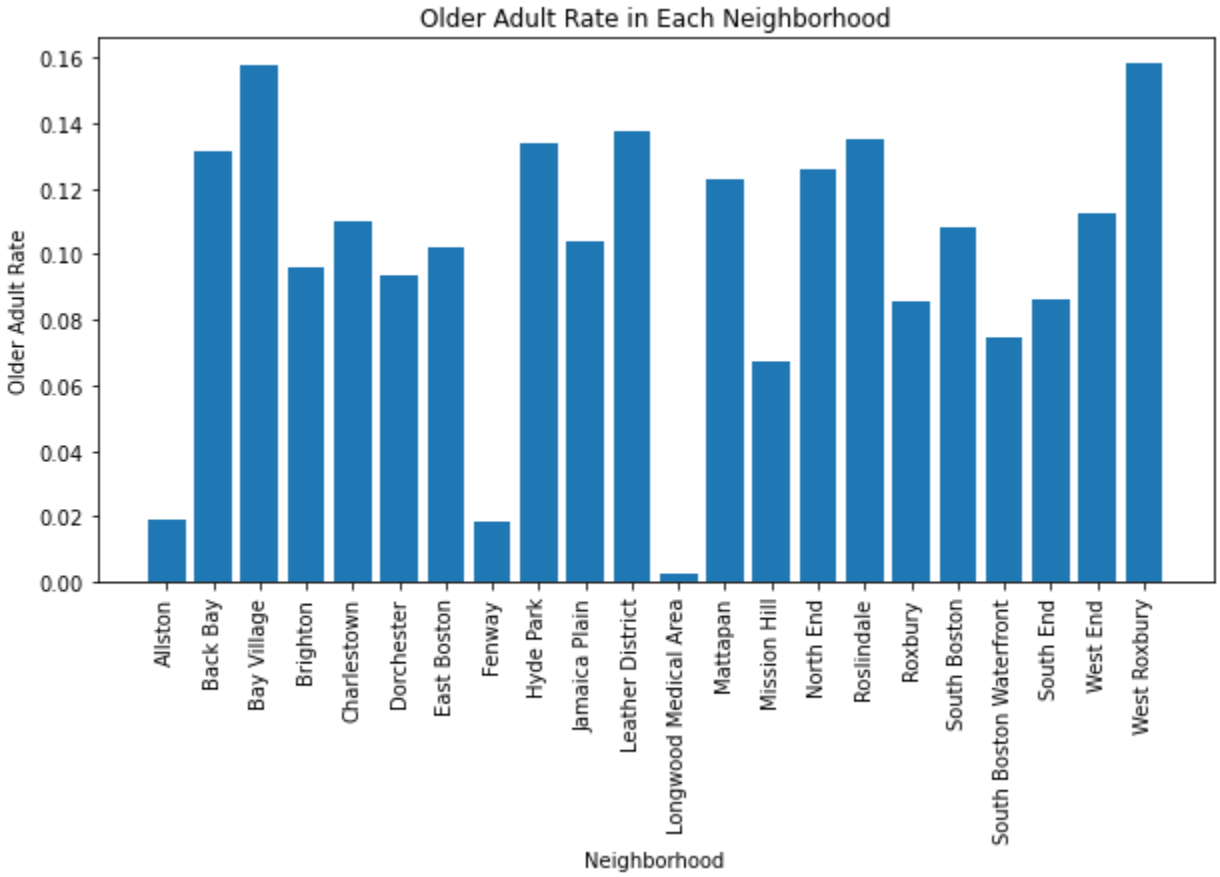
example, residents without televisions, computers, or data-driven mobile phones may face challenges getting news about hazards or recovery resources. Renters may have trouble finding and paying deposits for replacement housing if their residence is impacted by flooding. Homeowners may be less able to afford insurance that will cover flood damage. Having low or no income can create difficulty evacuating in a disaster event because of a higher reliance on public transportation.



Based on the graph above, we can see that Allston, Roxbury, and Mission Hill have the highest density of low to no income people.

We then investigated the distribution of older adult rates across various neighborhoods in Boston using the Climate Ready Boston dataset. By calculating the ratio of older adults to the total population in each neighborhood, we generated a bar plot that displays the older adult rate for each area.

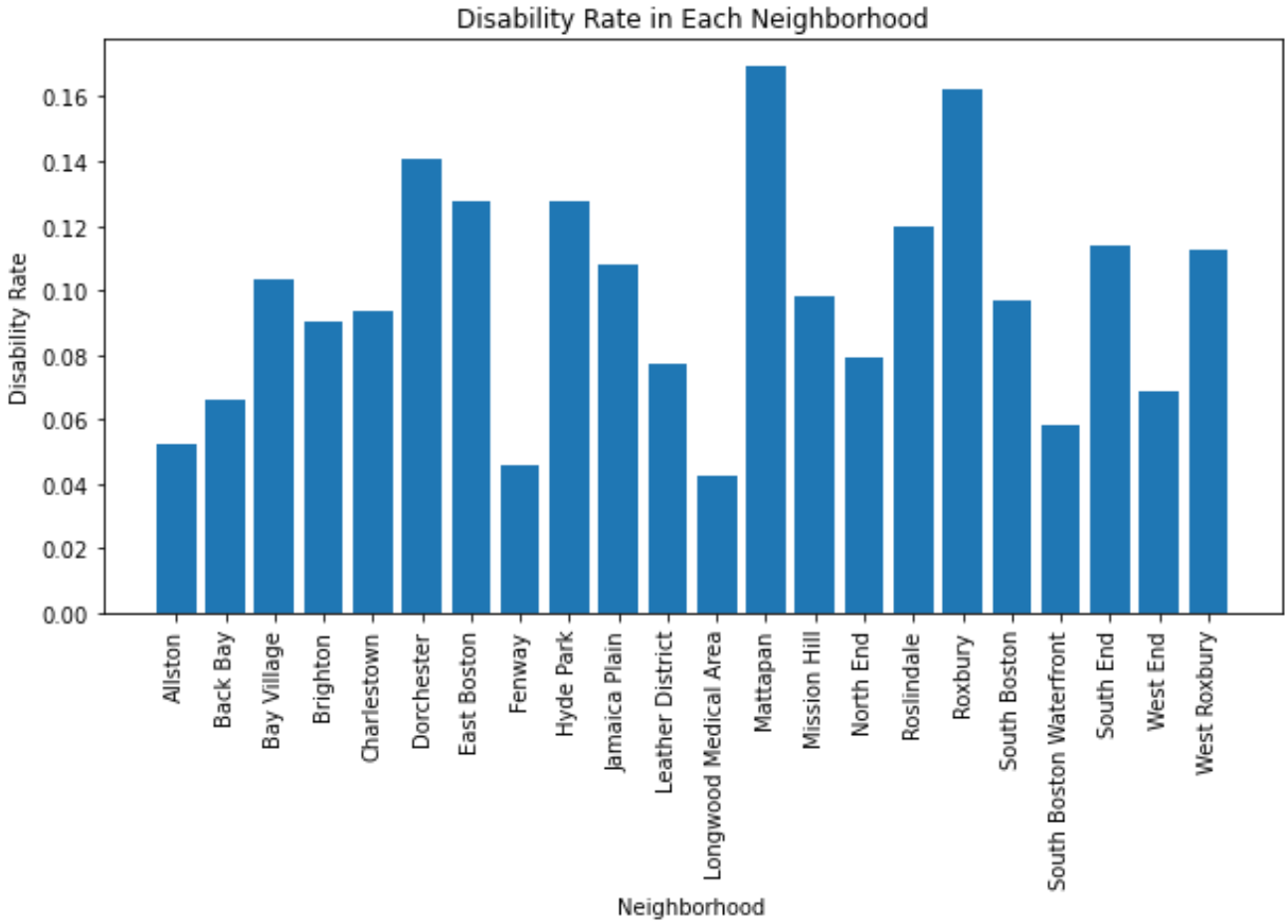
The older adult rate plot highlights the disparities in the proportion of senior residents among different neighborhoods. This information is valuable for policymakers, local organizations, and healthcare providers, as it emphasizes the need to allocate resources and develop programs tailored to the unique needs of elderly populations in each community. Understanding the distribution of older adult rates can help stakeholders create targeted initiatives such as accessible healthcare services, age-friendly infrastructure, and social support programs to enhance the quality of life for older residents and address their specific needs.



From the above graph, we conclude that West Roxbury, Bay Village, and Leather District have the highest proportions of older residents.

Additionally, we have examined the distribution of disability rates across Boston neighborhoods using the Climate Ready Boston dataset. By normalizing the number of disabled people based on the population of each neighborhood, we created a bar plot that illustrates the disability rate in each area.

The disability rate plot unveils the varying concentration of disabled residents among Boston neighborhoods. This crucial insight helps policymakers, local organizations, and support services identify areas that may require more focused efforts and resources to address the unique challenges faced by disabled residents. By understanding the distribution of disability rates, stakeholders can devise targeted strategies such as accessible infrastructure, adaptive programs, and specialized services that cater to the specific needs of disabled individuals, promoting social inclusion and reducing vulnerability within these communities.

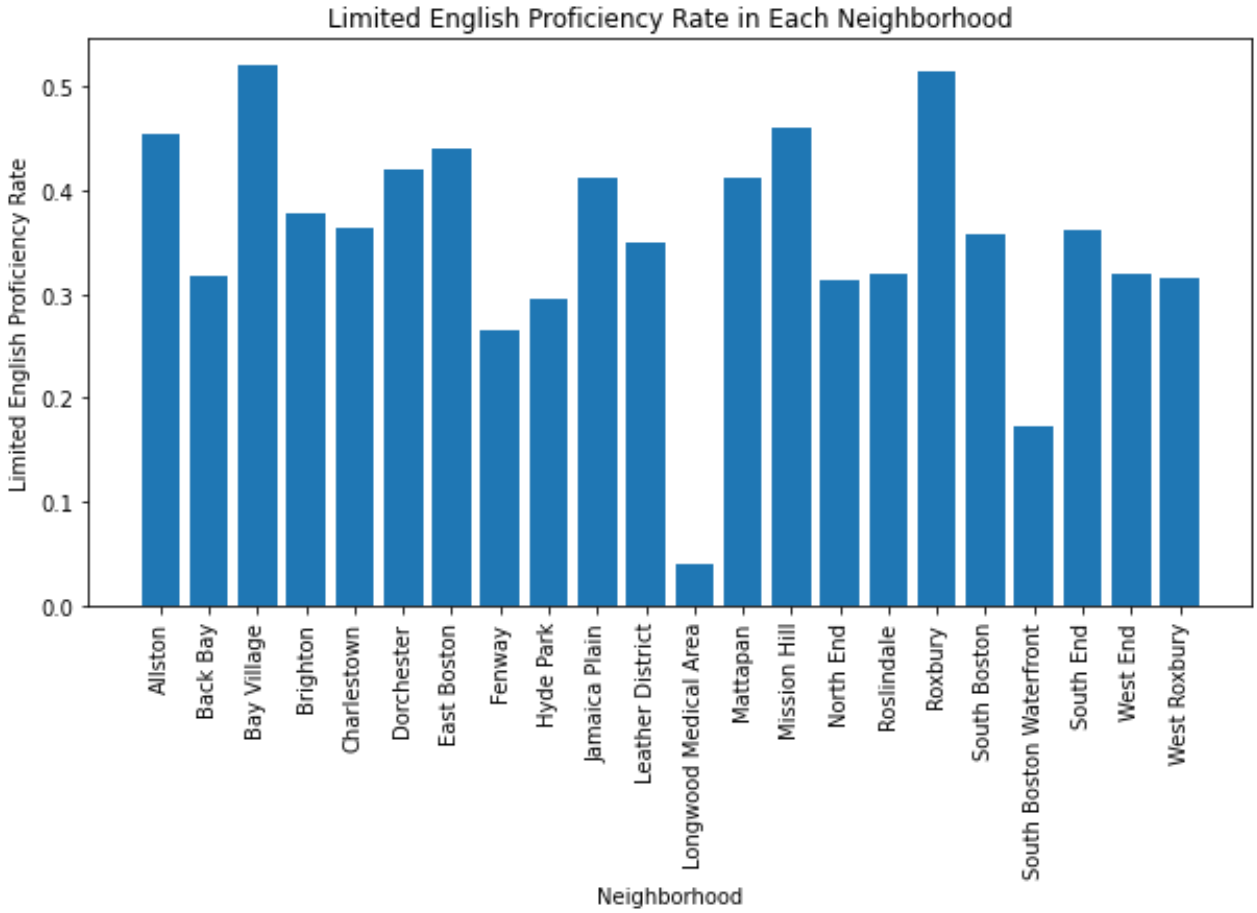


The top 3 neighborhoods that have a high disability rate are Dorchester, Mattapan, and Roxbury based on the visualization above.

We have analyzed the Climate Ready Boston dataset to investigate the distribution of limited English proficiency (LEP) rates across various neighborhoods in the city.

By normalizing the LEP values based on the population of each neighborhood, we created a bar plot that displays the LEP rate in each area.

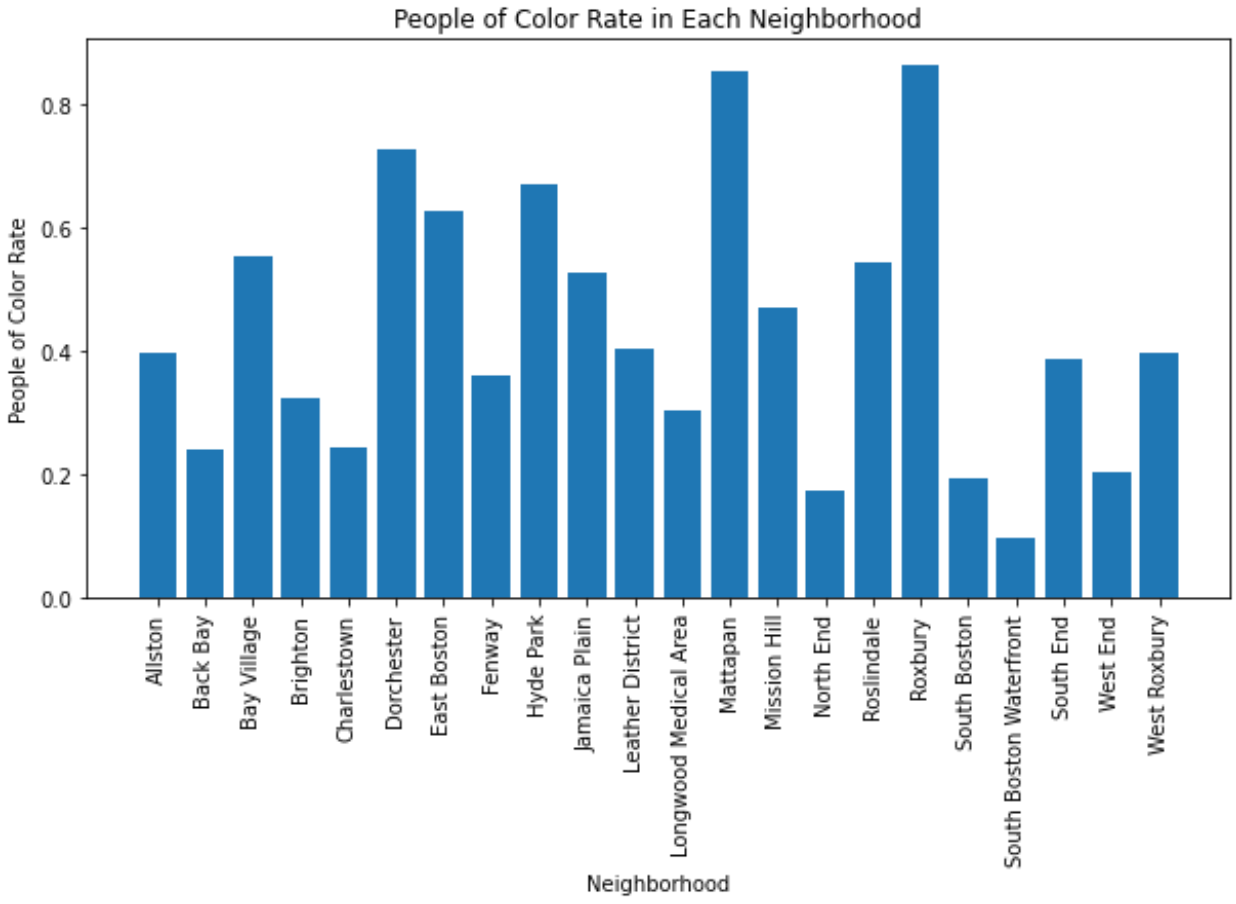
The resulting LEP plot revealed significant variations in the rates of limited English proficiency among Boston neighborhoods. Some neighborhoods exhibited a higher concentration of residents with LEP, whereas others showed lower rates. This information is crucial for policymakers and local organizations, as it highlights the need to allocate resources and support tailored to the linguistic needs of each community. By understanding the distribution of LEP rates, stakeholders can better target their efforts in providing language services, educational programs, and other initiatives to promote social inclusion and mitigate vulnerability for residents with limited English proficiency.



Based on the visualization above, Mission Hill, Bay Village, and Roxbury have the highest rate of residents with limited English proficiency.

We have also analyzed the dataset to study the distribution of people of color rates across various neighborhoods in the city. By normalizing the number of people of color based on the population of each neighborhood, we created a bar plot that presents the people of color rate in each area.

The people of color rate plot uncovers significant variations in the proportions of racially diverse populations among Boston neighborhoods. This information is essential for policymakers, local organizations, and community leaders, as it underscores the need to recognize and address the specific challenges faced by these communities. By understanding the distribution of people of color rates, stakeholders can develop targeted initiatives that promote racial equity, social inclusion, and access to resources and opportunities. These efforts may include cultural competence training, educational support programs, and policies that foster diversity and inclusivity within the city.



Based on the visualization above, Roxbury, Mattapan, and Dorchester have the highest population of POC.

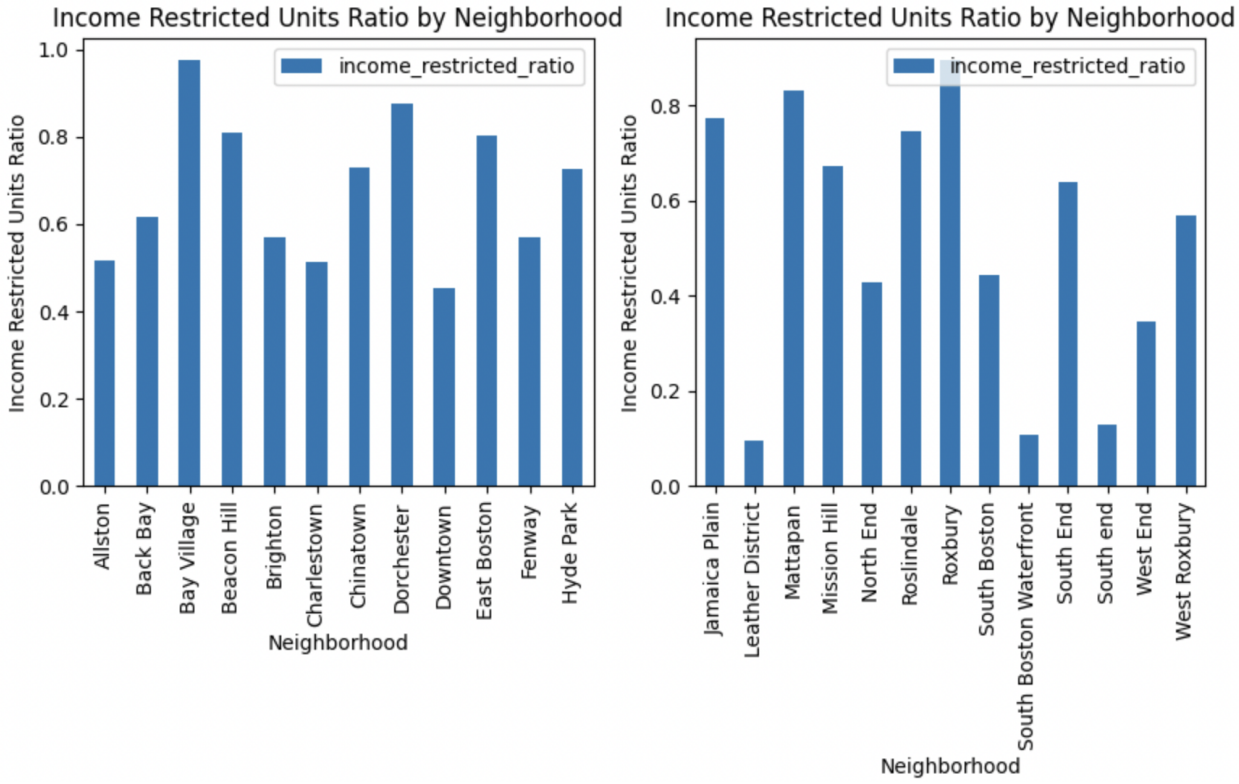
Conclusions:

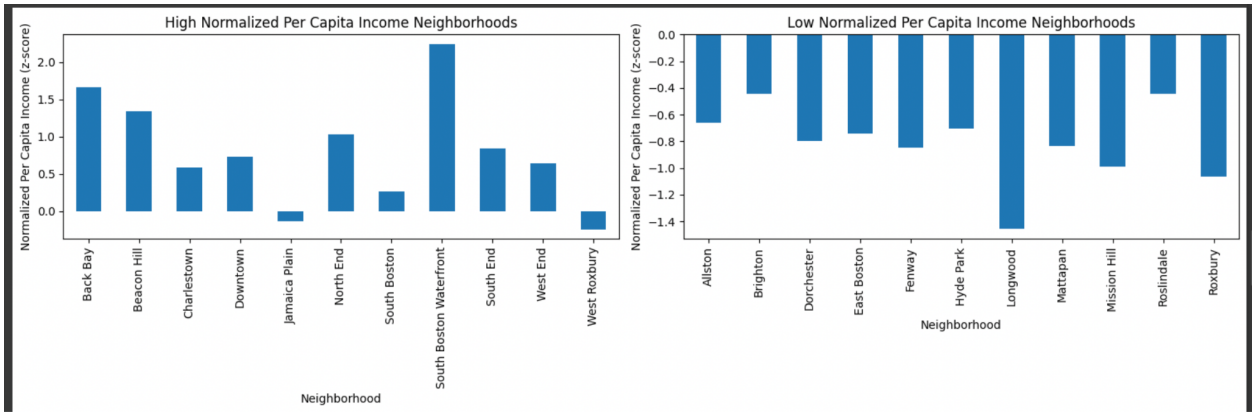
Also we can see from the above visualizations, the social features that impacted discrimination the most were:
The number of low-to-no income people, the number of older adults and disabilities, Limited English Proficiency by Neighborhood and The POC population by neighborhood.

The main conclusion that we can draw from the above visualizations is that Roxbury is a neighborhood with an extremely large rate of low to no income residents, residents with disabilities, residents with limited English proficiency, and POC population. This is a significant observation to make because as we saw from the base analysis, Roxbury is one of the top neighborhoods with a large rate of violations per unit. We guess it's because people with less income are less likely to pay for their rent and are more likely to live in housing that is poorly maintained, overcrowded, so are the POC population which we can also see in the example of Mattapan. Mattapan has a large population of disabled residents and a large POC population. As before, we see that Mattapan has a large rate of violations per unit. And the example of Bay Village and Roxbury, where the violation of housing complaints are the most common compared with other regions, tells us that the number of older adults and disabilities are also important factors to consider. Those people may be more strict about the environment and conditions which will cause violations like housing complaints and sanitation requests.

4 Overarching Questions Answered

One of the tasks we wanted to achieve was to look at the number of evictions and housing court cases in Boston. For this we concluded that the data was not sufficient enough to perform any meaningful analysis. This was because a lot of the data was either pre-Covid or during Covid which does not indicate the changing demographics and times. Therefore, we shifted our focus to other questions such as is there a correlation between violation counts and income restricted housing in certain neighborhoods and which neighborhoods have the highest number of violations based on number of units. Additionally, we wanted to see whether landlords were unfairly targeting marginalized communities in terms of number of violations.





The first graph above shows the normalized number of income restricted units per neighborhood. The second graph shows a z-score normalized income per capita by neighborhood.

Once again, we can see that Roxbury and Mattapan are problematic neighborhoods. While they both are extremely income restricted, the people who live there have a very low per capita income. This may indicate a trend of landlords attempting to push out residents who don't clear a certain threshold of income especially with the rising rent prices.

As we saw in the extension analysis, there are instances of landlords taking advantage of marginalized communities specifically in the neighborhoods of Mattapan and Roxbury(see previous section for more).

5 Work Contributed

Risheet: Normalized violation counts by neighborhood, income restricted housing analysis and normalization, rentsmart dataset analysis and visualization creation, landlord violation analysis and graphs, created the README on GitHub
Chen: extension project analysis and graph creation, affordable rental units analysis, AMI analysis, eviction analysis.
Rohan: extension project analysis and map creation, neighborhood violation analysis and graph creation
Chengyu: preliminary eviction analysis, AMI analysis, service request analysis, violation type analysis and heatmap creation

6 Codebase Navigation

The GitHub repository for our team can be navigated as follows:

- analysis: This directory has all the Jupyter notebooks with the data analysis that we did. Each notebook is appropriately named to indicate what type of analysis was done in each file.
- data: This directory has all the csv and data files that were either found or given to us at the beginning of the project. Again, the files are appropriately named to indicate what data is stored in each respective csv.
- deliverables: This directory holds all of our checkpoints, reports, presentations, and deliverables as part of the project.

Running the code simply requires running each cell in each Jupyter notebook. The notebooks have all the necessary libraries being imported before they are used. One thing to keep in mind is that different notebooks use different csv files for the data but the file names indicate which dataset is being used.