

## **PROJECT REPORT**

(Deliverable 4)

Our project concentrates on property violations and bad landlords. We have tried to focus on the distribution of such violations across Boston. We have also tried to find out bad landlords through the number of violations. This report refines our observations from deliverable 2 and 3 (along with suggestions from the client), expands the scope of our project and adds additional insights and improvements.

### **Extension Project**

Interactive map with the kind of violations occurring in the residence.

Come up with more metrics to determine a bad landlord.

To answer the following questions:

Do affordable houses have fewer/more violations?

Why are violations concentrated in certain areas?

In the following report, we have tried to gather as much information as we could, for the interactive map. In the interactive map, we have displayed:

- Number of 4 types of violations per capita
- Average age of buildings per neighborhood for different time periods
- Average time since last renovation for different time periods

We are yet to add:

- Top 10 violators (according to number of violations) - normalized based on number of properties owned
- Number of properties with violations per neighborhood

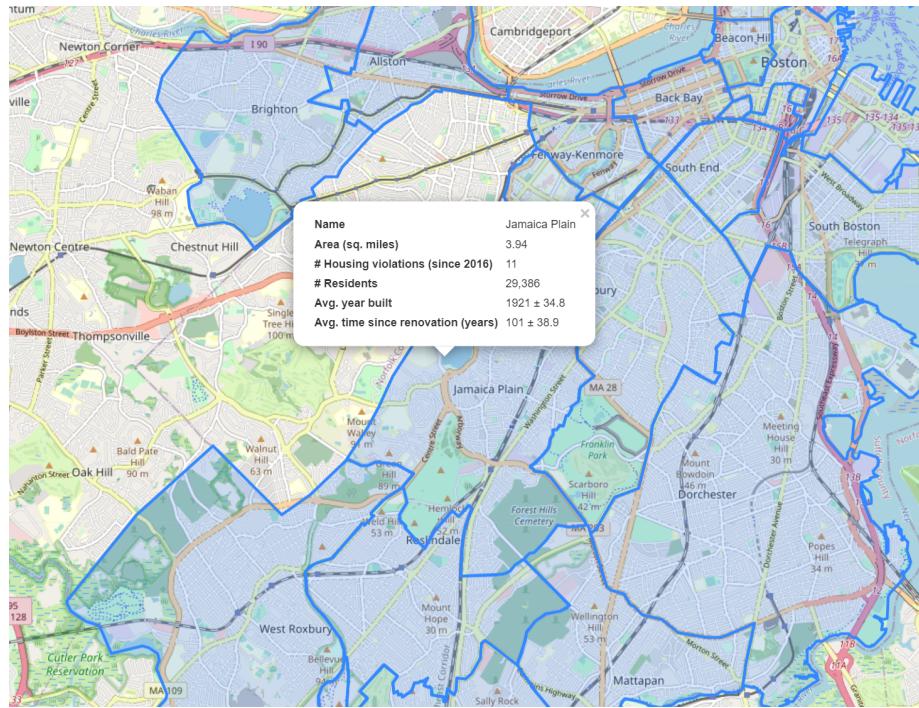
We have also worked towards answering the questions, which are built on the questions we answered for the deliverable 2 ("Where are the violations in Boston? Distribution of % of violations across Boston", "What are the types of violations? How many violations for each type?").

### **Interactive map**

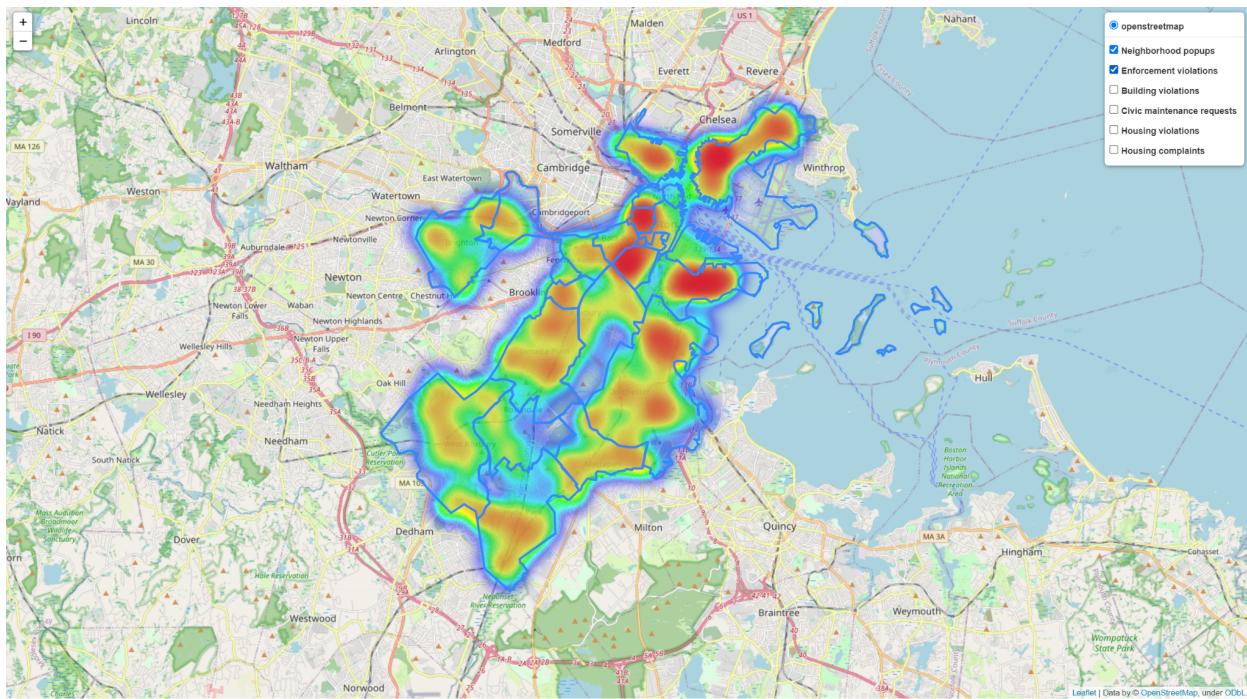
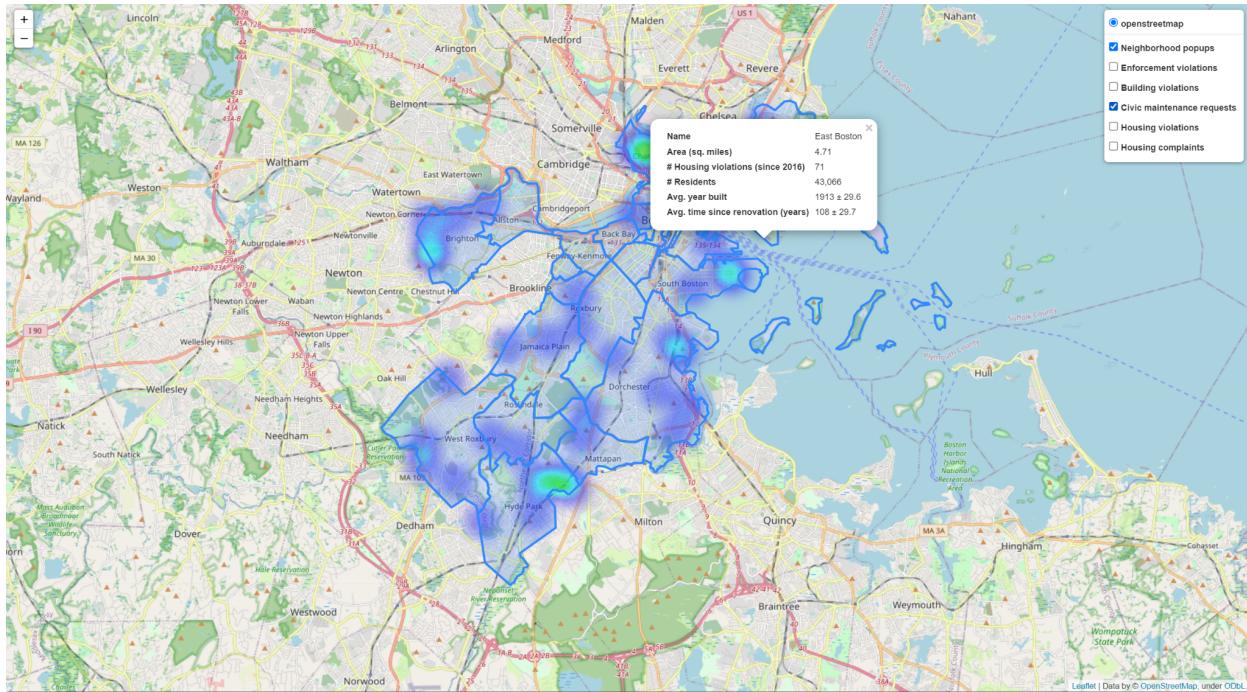
The interactive map can be accessed [here](#). **Instructions on how to view it:**

- Download the .html file on your computer
- Open the .html file with your browser

With the interactive map, you can view, for each neighborhood, the number of: housing violations, housing complaints, population, and also the: area (sq. feet), avg. year the buildings were built, and avg. time since the renovation of the buildings in the neighborhood. This is implemented with location pop-ups and can be seen in the figure below.



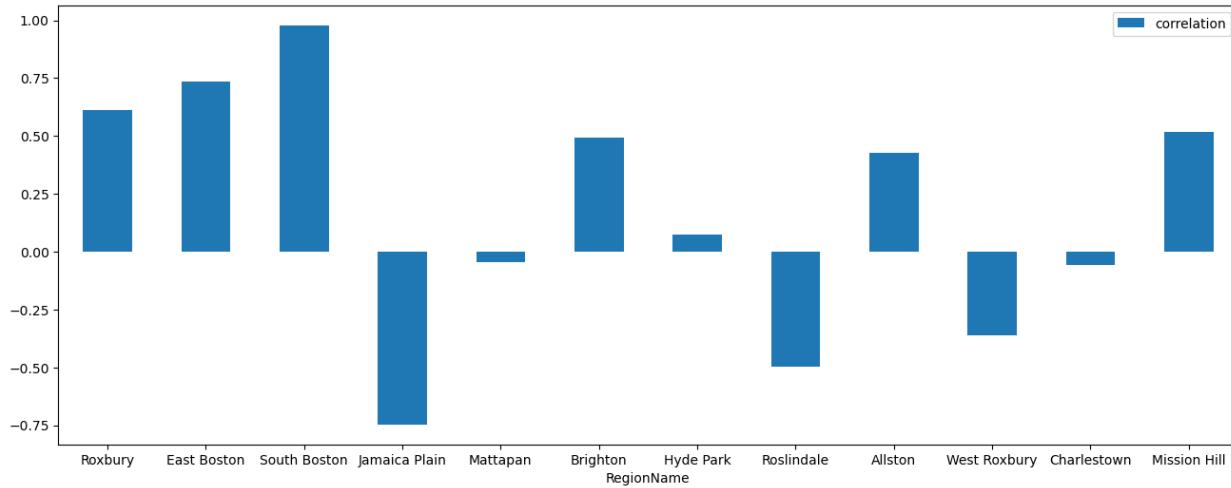
Moreover, with layer-control on the top-right corner, one can select the heatmap to show for the type of violation made, and a demonstration of this is shown below.



### The correlation between violation and the house price

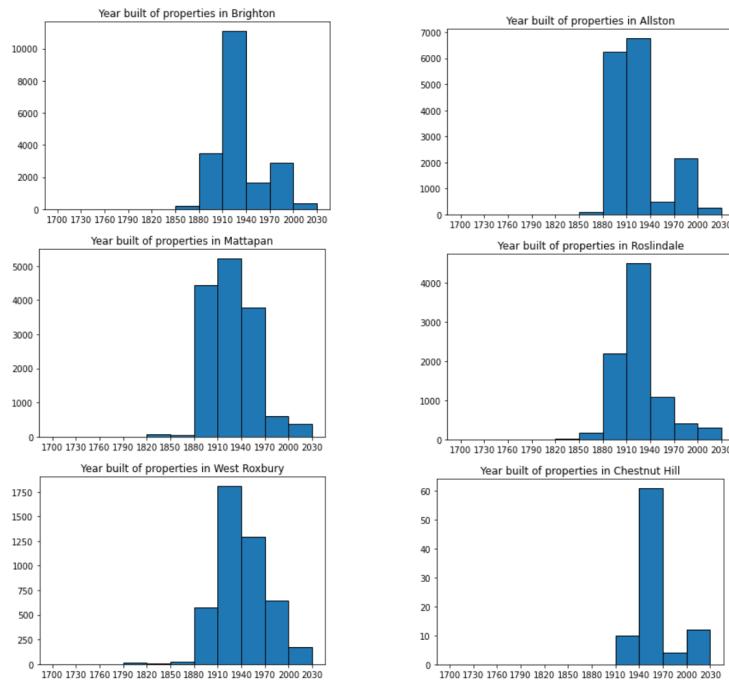
We analyzed the data provided by Zillow to get the mean price of houses each year from 2018 to 2022 for neighborhoods in Boston. According to the data in Rentsmart, we could get the

same data about the number of violations. The bar chart shows the correlation between violations and the house price. Anything over 0 indicates they are positively related and anything below 0 means they are negatively related (there is no association of house prices with number of violations). From the bar chart, we can observe the situation varies in different neighborhoods. Typically, South Boston's violation number is highly related to house price while the number of violations in Jamaica Plain is negatively related with home prices.



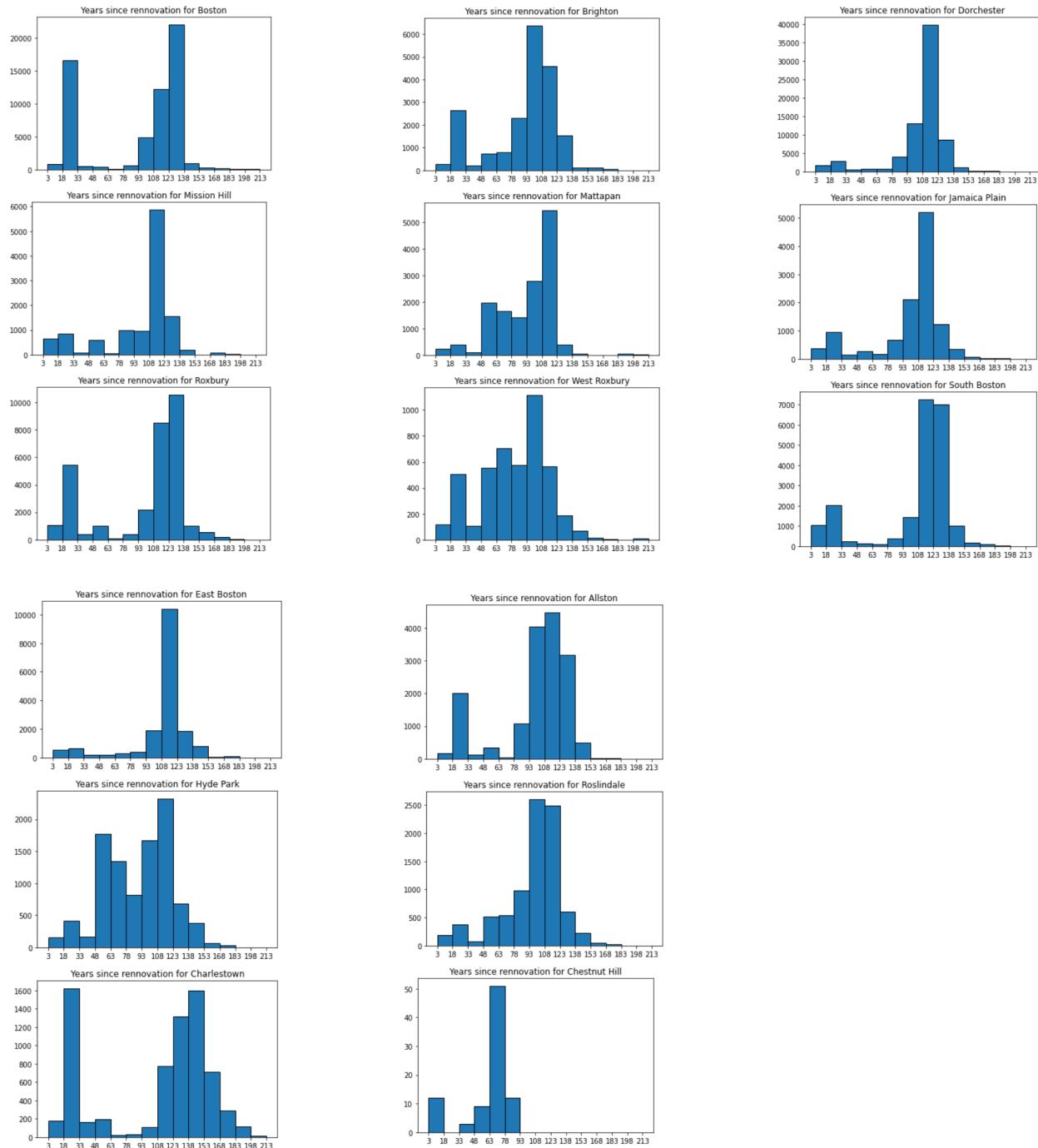
### Age of properties across Boston grouped by Neighborhood

We answered the key questions - “What years were the units built in, in the violations report?” and “Status of renovation for such units”. Since deliverable 3, we rearranged all analysis so that they are grouped by neighborhood. Here’s the updated analysis on age & renovation status of the properties.



## **Renovation status of properties across Boston grouped by Neighborhood**

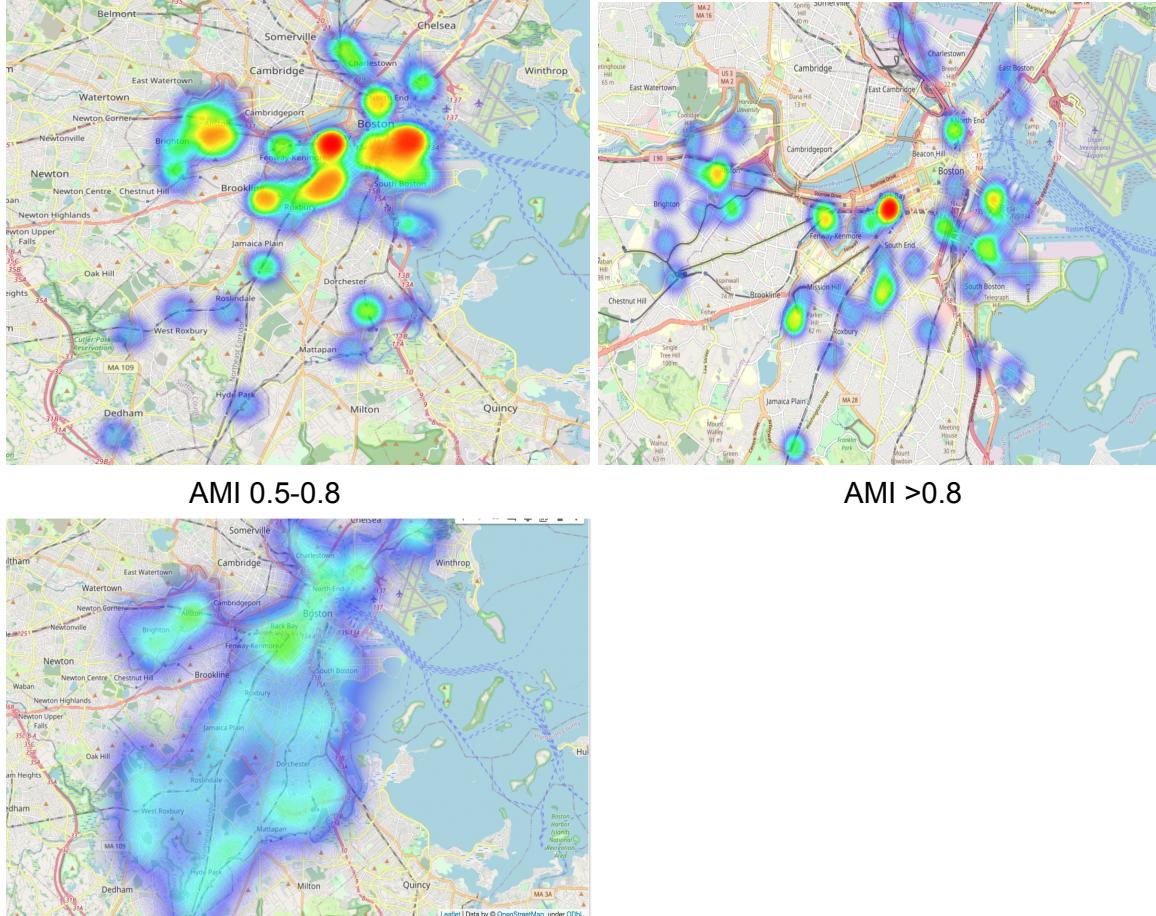
(Demonstrated by number of years since last renovation/remodel)



It is worth noting that the largest peak for each neighborhood, wherever they appear on the histogram, usually indicates units that did not receive any renovation since construction, as that applies to the majority of the properties. Smaller peaks, if exist, that appear before the major

peak are those that did receive renovation/ remodeling after construction. For example, units in Charlestown had a large-scale renovation around 20 years ago, as indicated by the corresponding histogram.

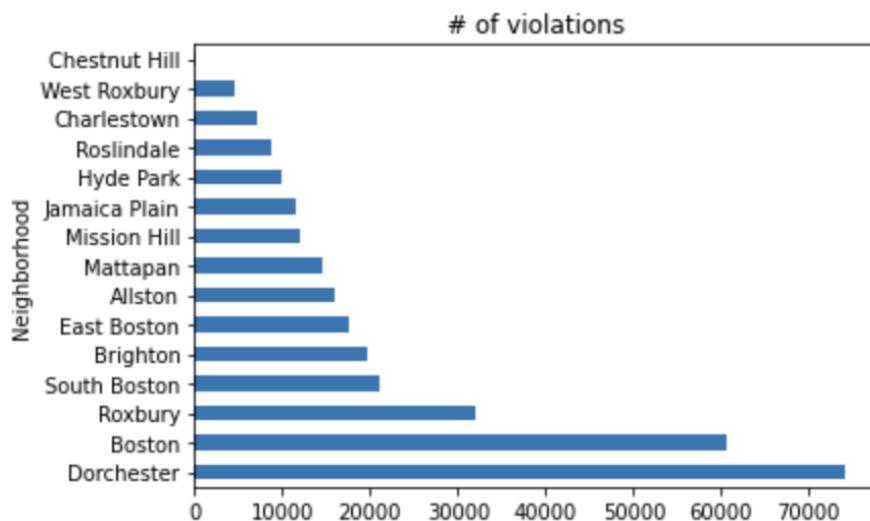
### **Comparison of distribution of AMI and violation**



### **Distribution of violations/complaints/requests**

Above maps show the distribution of AMI and the violation heatmap. The AMI map depends on the point level which is transformed from street address using google geoapi. The observation is for AMI category 0.5-0.8 the heatmap is about consistent with the violation heatmap. It is easy to find that Allston and Backbay have a higher intensity of distribution, and in north end and south boston they have a relatively consistent distribution. For category >0.8, the data points are less and therefore, a little less relevant to the violation map.

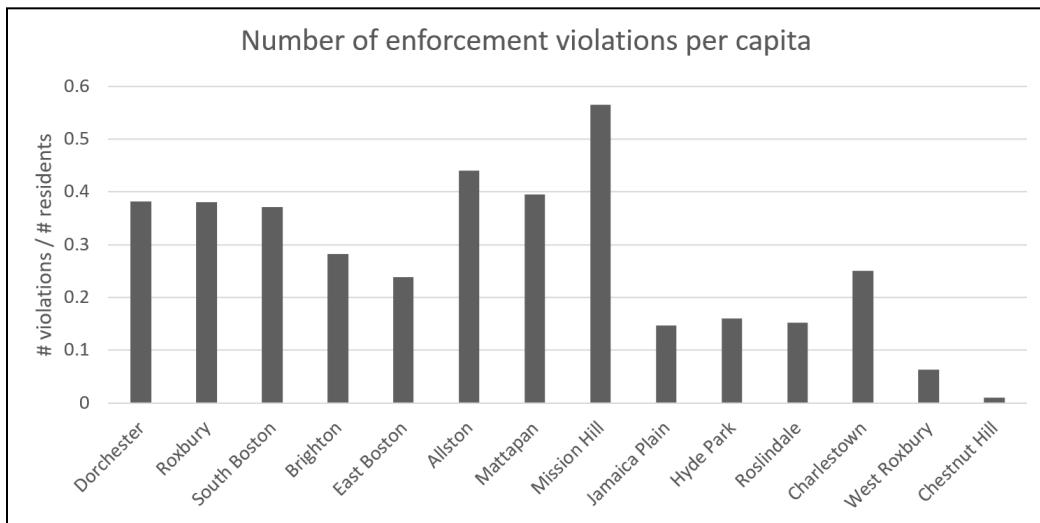
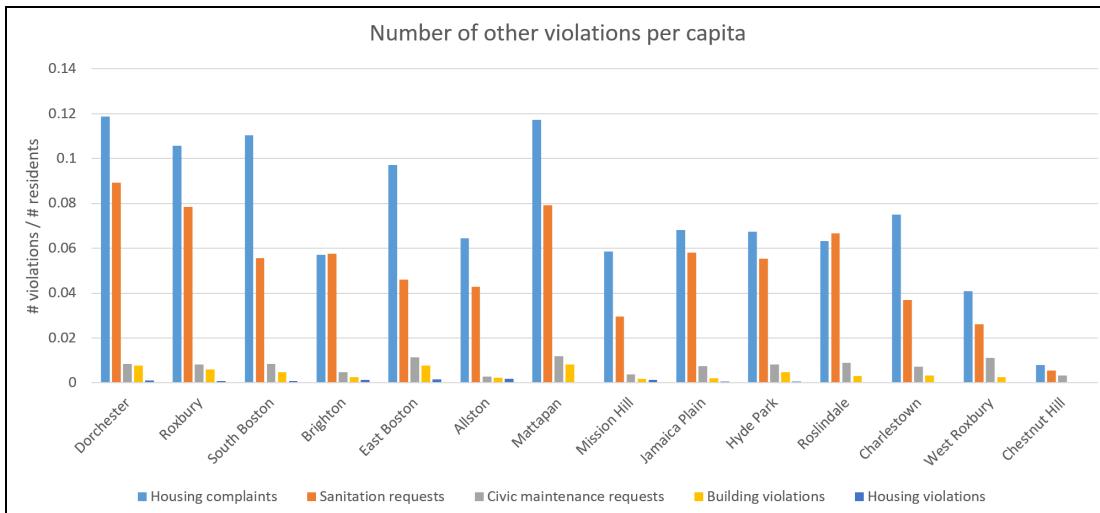
## Types of violations and their distribution



Dorchester has the highest number of violations. The overall total number of violations in Boston is over 220,000.

## Types of violations (divided by population per neighborhood)

We further analyzed the violations into the types of violations in each neighborhood using the census information of the population in different neighborhoods and the RentSmart dataset. This analysis gives us important insights as to which neighborhoods have the highest number of violations (and what kinds of violations) adjusted for the population in that neighborhood. For example, Mission Hill had the highest number of enforcement violations.



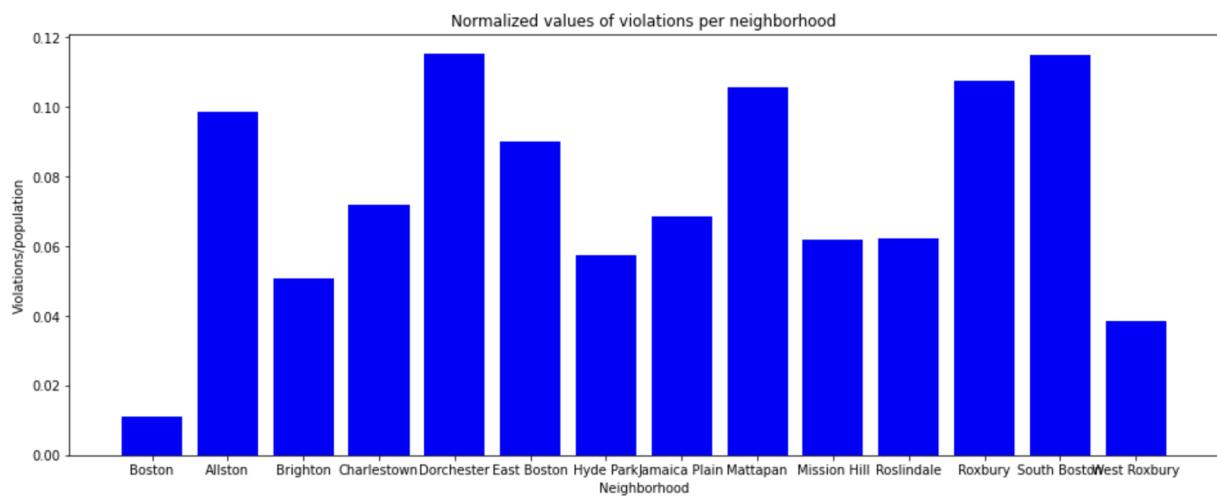
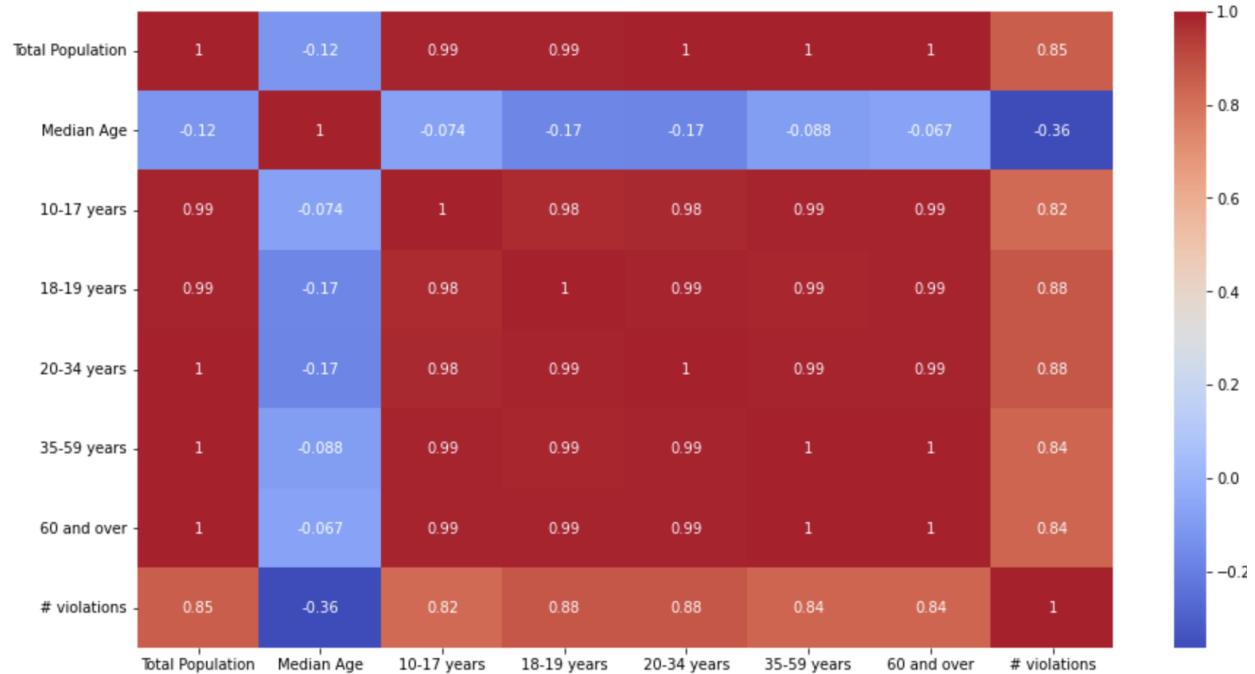
### Effects of factors on number of violations - Violations after normalization

We found out that Dorchester had the maximum number of violations according to the RentSmart database. Now, we needed to find out why these violations were concentrated in certain neighborhoods like Dorchester, South Boston, Beacon Hill, Back Bay, etc. To do so, we took various factors into consideration and plotted the correlation matrix for these multiple factors and normalized the number of violations for them.

#### Population

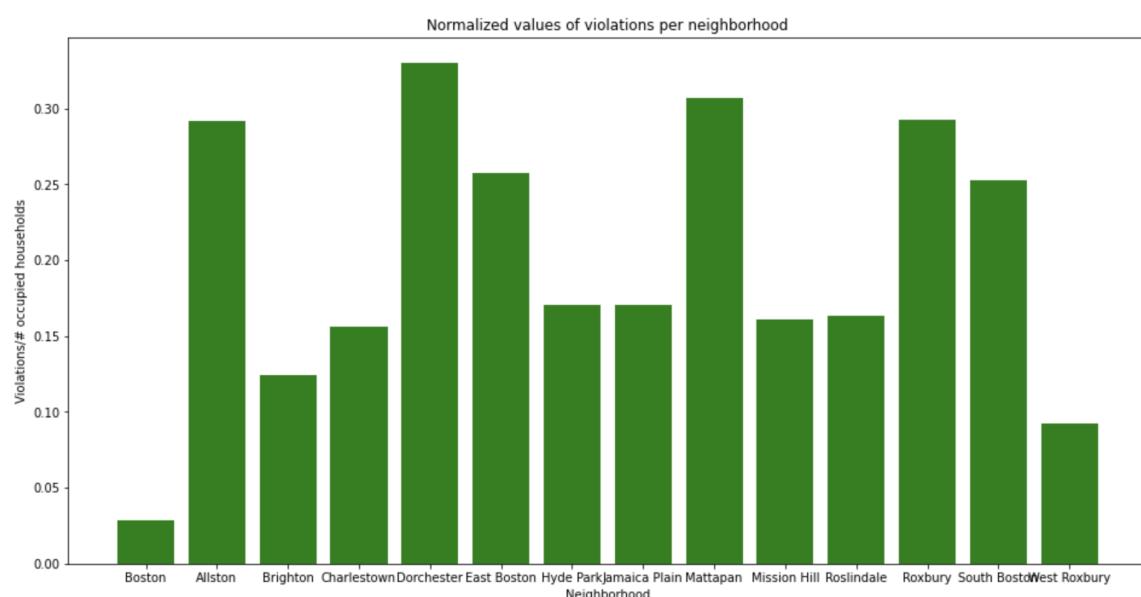
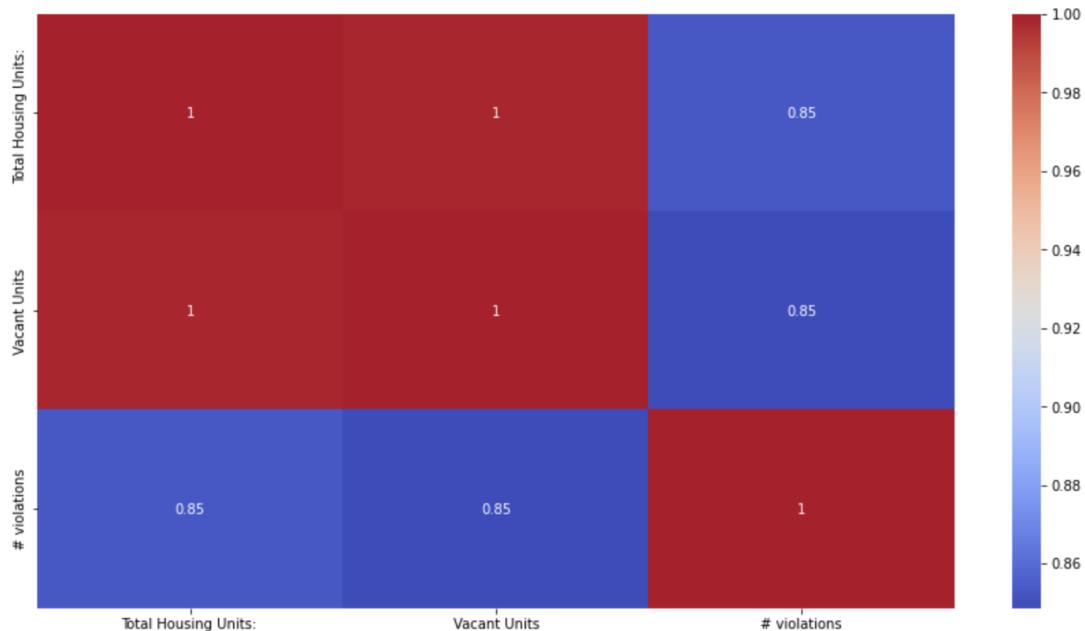
We realized that a larger number of people in one neighborhood and a relatively smaller number of people in another neighborhood could be a factor. Therefore, using the

population dataset for Boston, we normalized the number of violations by dividing it by the population in that neighborhood and plotted the following graph. The graph shows that the ratio of violations to the population is highest at Mission Hill, followed by Mattapan and Dorchester, therefore giving us more insights about the role of population in the distribution of violations. Lesser population and greater violations give high numbers.



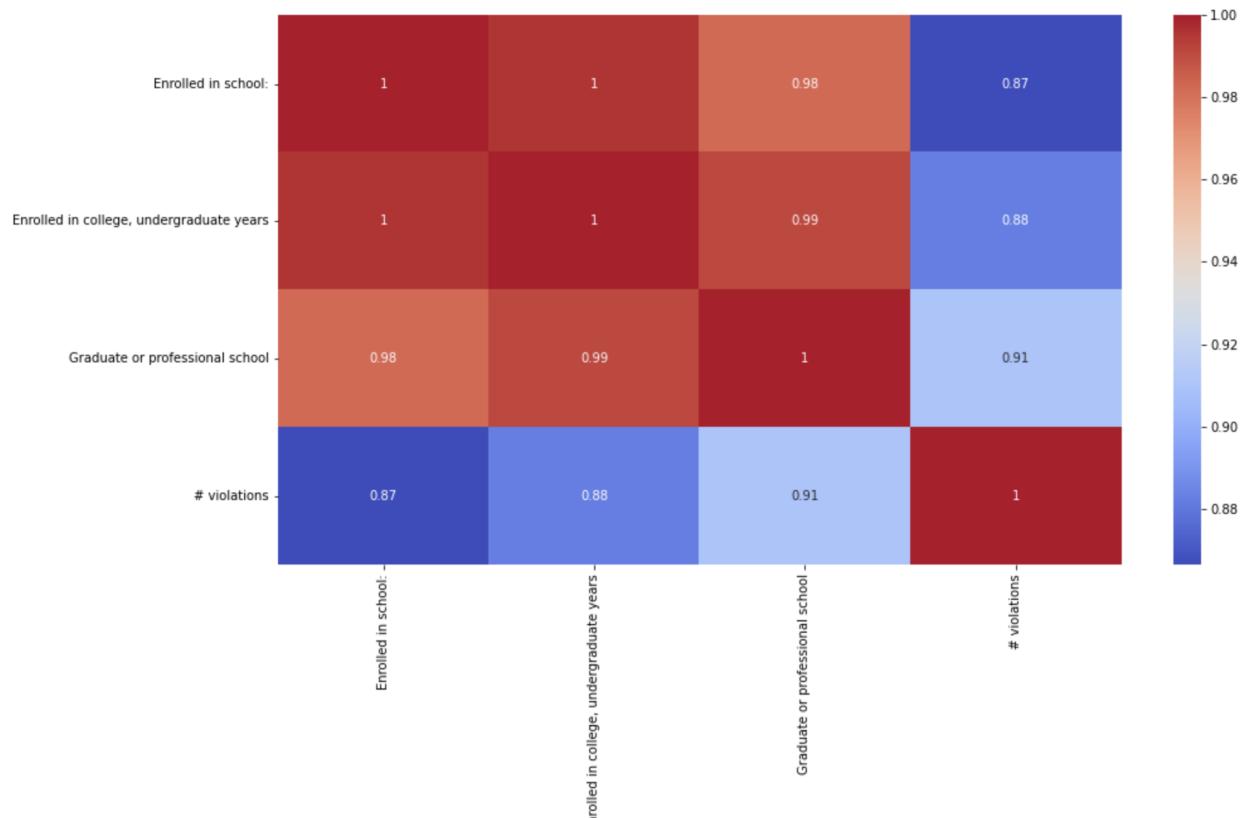
## Number of occupied households

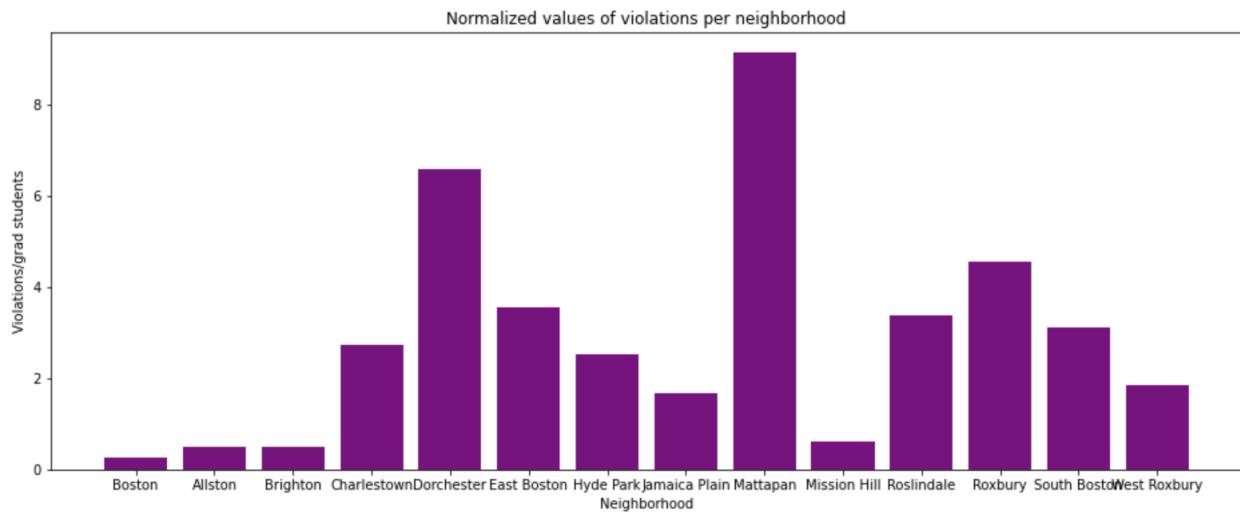
Another interesting aspect to look at, apart from population, is the number of households. Population definitely gives us a sense of the number of violations in terms of the number of people, however, a better insight would be given by the number of occupied households. The following graph shows the correlation matrix between number of violations and number of vacant and total households. Given this relationship, we normalized the number of violations by the number of occupied households (total households - vacant units). According to that, Dorchester, Mattapan, Allston and Roxbury have the highest ratio of violations to occupied households.



## Number of students enrolled in school

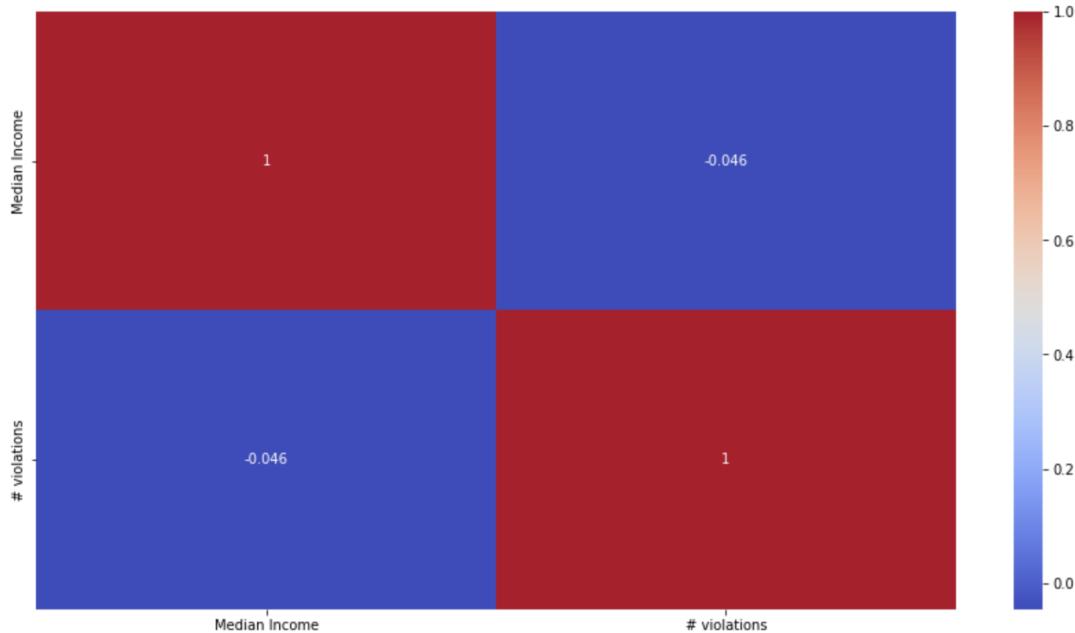
We wanted to see if the number of students in a neighborhood affected the number of violations. Therefore, we generated a correlation matrix for students which gave the highest correlation between graduate/professional students and number of violations. Therefore, we normalized the number of violations by dividing by the number of graduate students. According to the graph, Mattapan has the highest ratio i.e. smaller number of students but more violations.





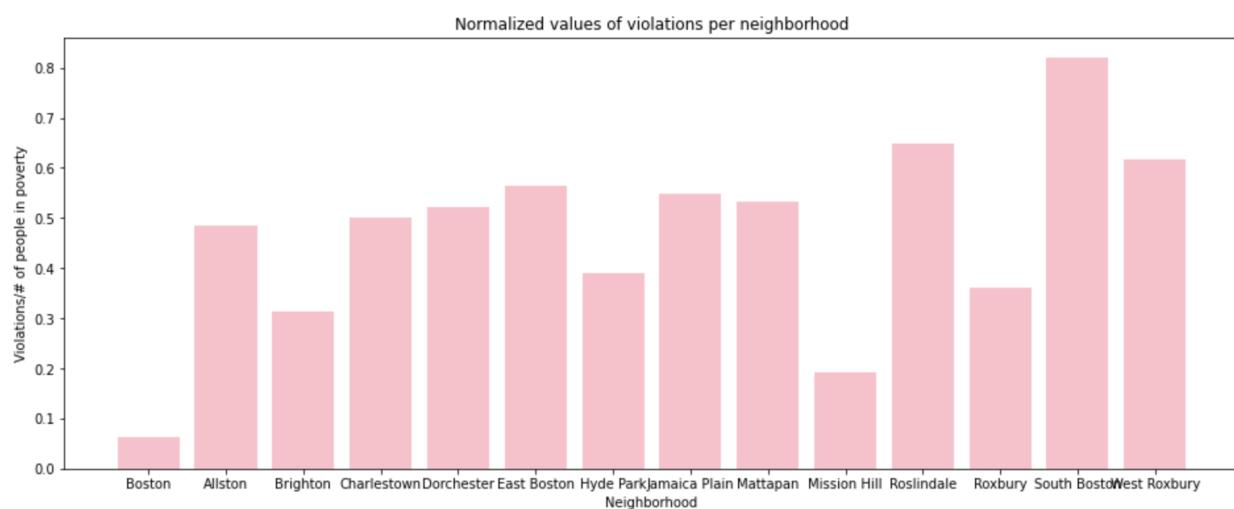
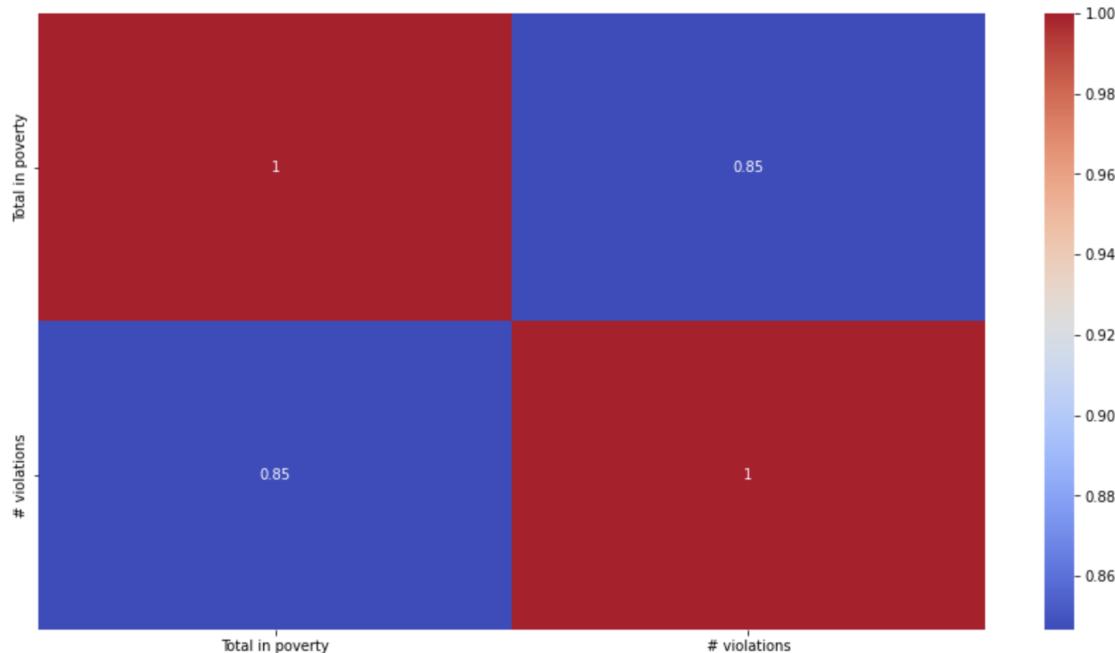
## Household Income

We wanted to see if household income affected the number of violations in separate neighborhoods. However, since the correlation factor was negative, we did not normalize our results based on this.



## Number of people in poverty

An interesting factor we took into account was the number of people in poverty. Since the household income did not give favorable results, we did not expect much from this either. To our surprise, after plotting the correlation matrix and getting a correlation coefficient of 0.85, we decided to normalize the results. According to this, South Boston had the highest ratio (lesser number of people in poverty, more the number of violations).



### Number of students living off campus

Currently under progress. We plan to utilize an online geo-decoding service with Geopy to decode street addresses of students in the provided datasets. (since no neighborhood info was provided) We will have the information by the final presentation and incorporate analysis on correlation between off-campus students and housing violations/complaints.

### Conclusion

The interactive map gives a lot of insight into the housing violations for each neighborhood. It also gives the number of residents per neighborhood so it is easier for the user to compare the number of violations with the census. Our final report will also contain the top 10 worst landlords for each neighborhood.

I believe this report justifies how to factor in a multitude of metrics to get the number of violations as well as bad landlords. It also appropriately gives information about the distributions of the same through a variety of heatmaps and visuals.