FINAL PROJECT REPORT

BAD LANDLORDS AND PROPERTY VIOLATIONS

(Deliverable 5)

Our project concentrates on property violations and bad landlords in the city of Boston. 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 contains all our initial project findings as well as 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.

Base Project

- Defined and addressed the bad landlord and property violations issue
- Analyzed datasets provided
- Filtered data to create visualizations
- Searched for other available datasets
- Came up with metrics to point out bad landlords
- Answered a few base questions:
    - What are the types of violations? How many violations for each type?
    - Where are the violations in Boston? Distribution of % of violations across Boston
    - How old are the buildings? How is the renovation status? How many property violations per year?

Extension Project

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

To answer the following questions:
Do affordable houses have fewer/more violations?
Why are violations concentrated in certain areas?

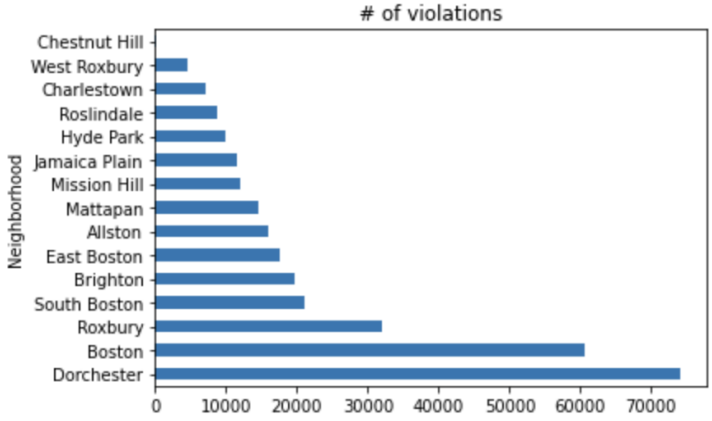
In the following report, we have tried to gather as much information as we could, for the interactive map. In the interactive map, we have displayed:

- 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
- Top 3 worst landlords for each neighborhood
- Number of residents per neighborhood

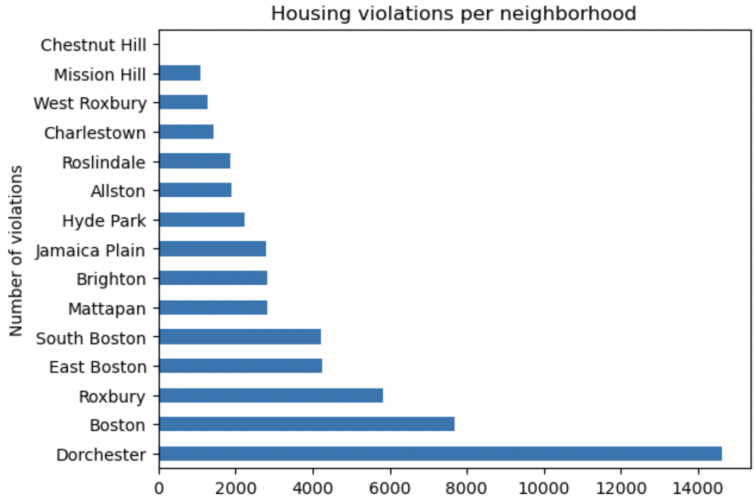
We have also worked towards answering the questions, which are built on the questions we answered for deliverable 2 (“Where are the violations in Boston? Distribution of % of violations across Boston”, “What are the types of violations? How many violations for each type?”).

Violations, their types, and their distribution

We plotted the number of violations (all types) as found in the RentSmart dataset. The graph below shows Dorchester to have the highest number of violations with the total number of violations in Boston being over 220,000.

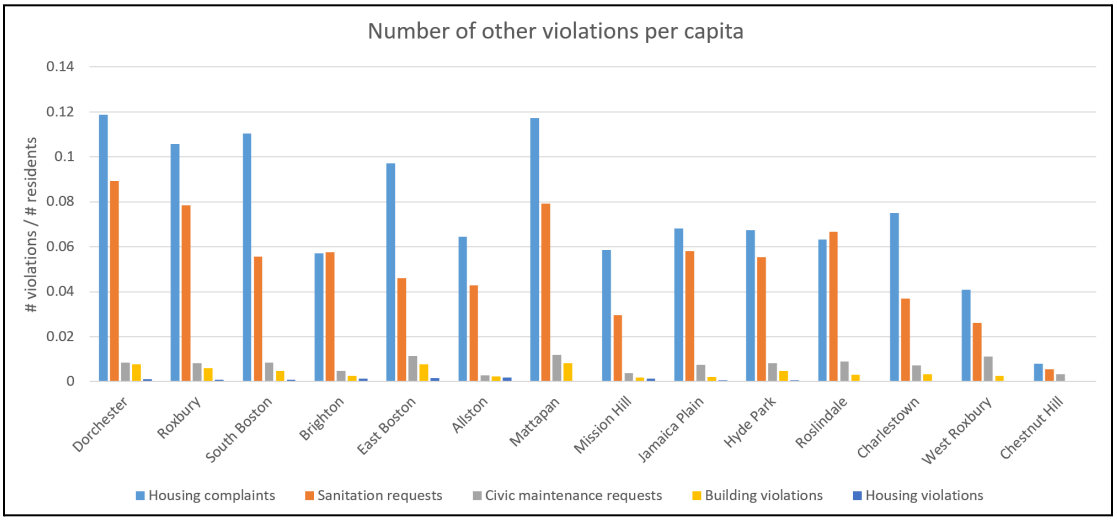


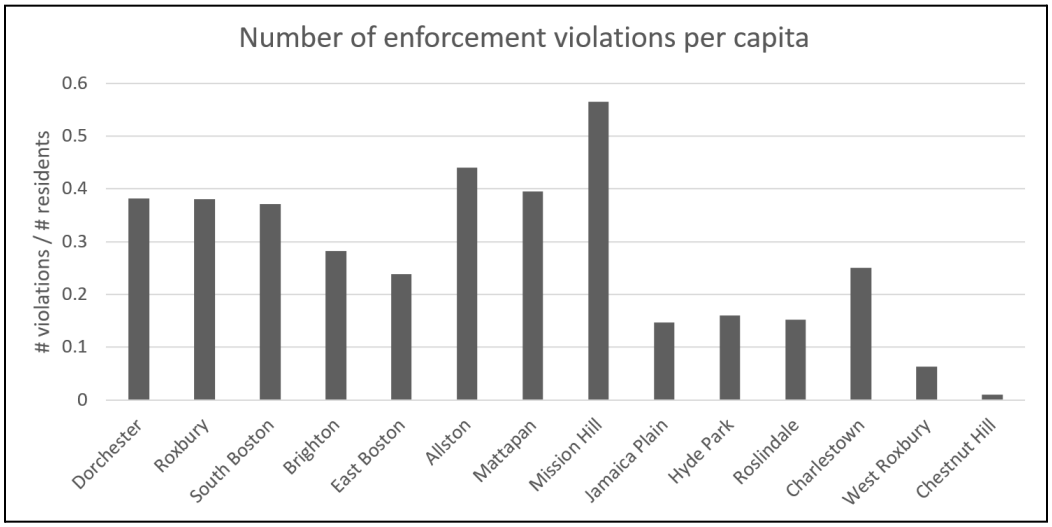
We then plotted the number of housing violations in Boston which also showed Dorchester to have the highest number. The total number of housing violations in Boston is a little over 54,800.



Types of violations

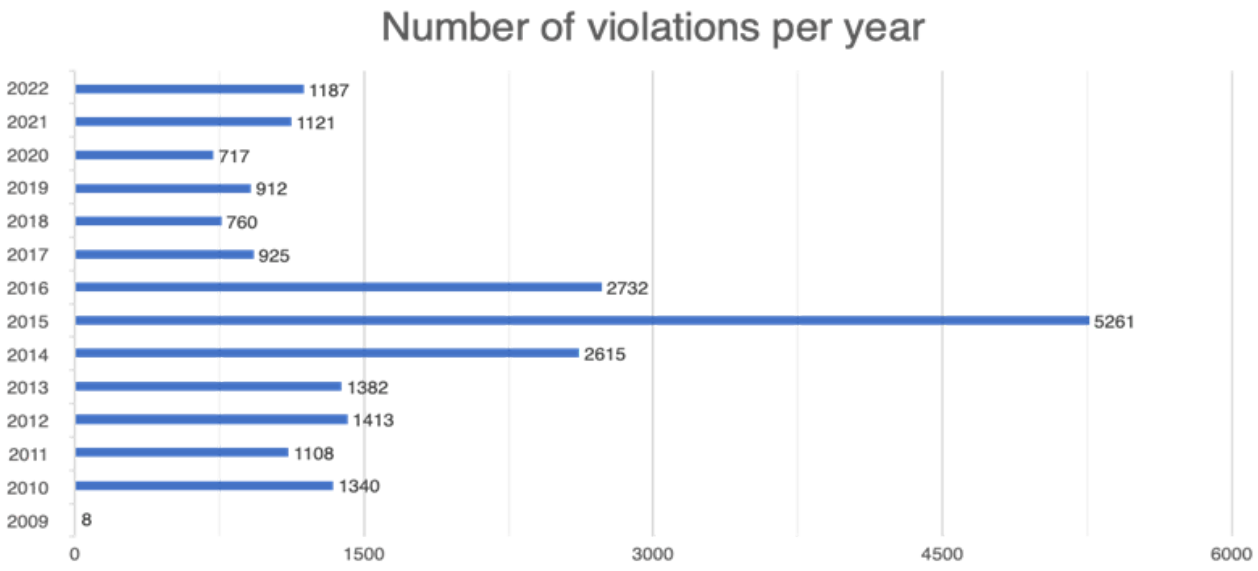
This analysis gives us important insights as to which neighborhoods have the highest number of violations (and what kinds of violations) adjusted for the population in that neighborhood. For example, Mission Hill had the highest number of enforcement violations.





Numbers of violations per year

We can observe that the number of violations per year reached the peak in 2015, which is above 5000. From 2013 to 2017, the number of violations experienced a rapid increase and decrease. In other times of this period, the number of violations was stable at around 1000.

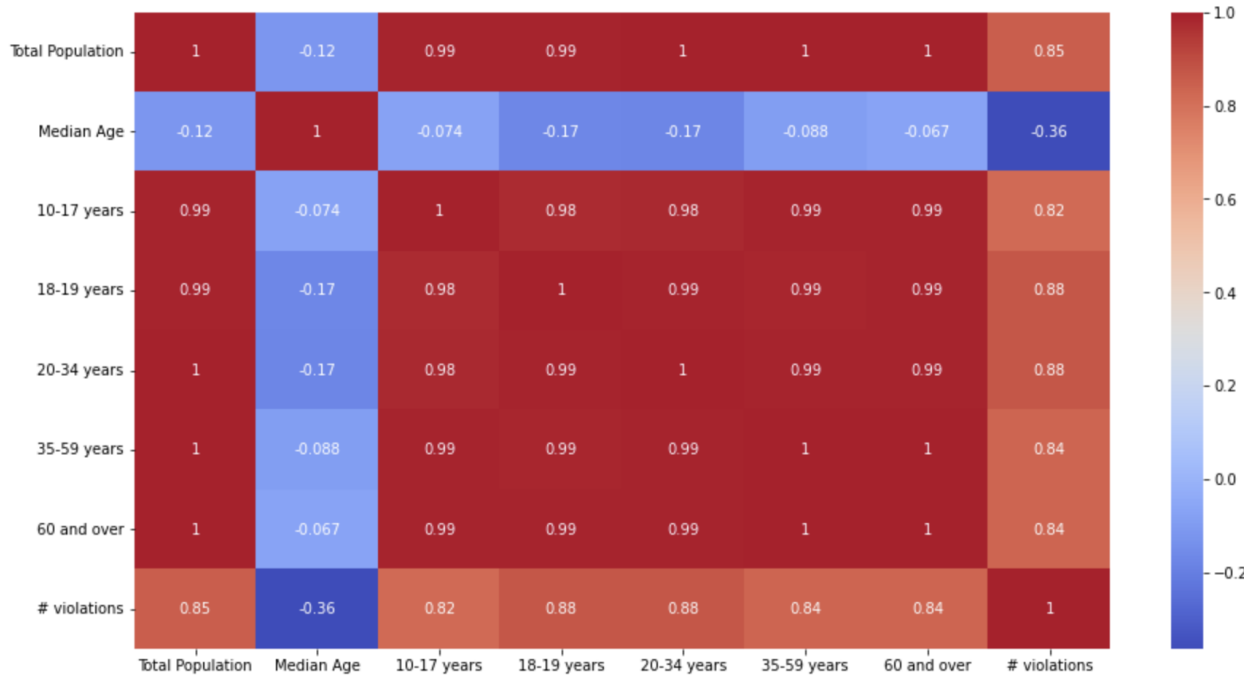


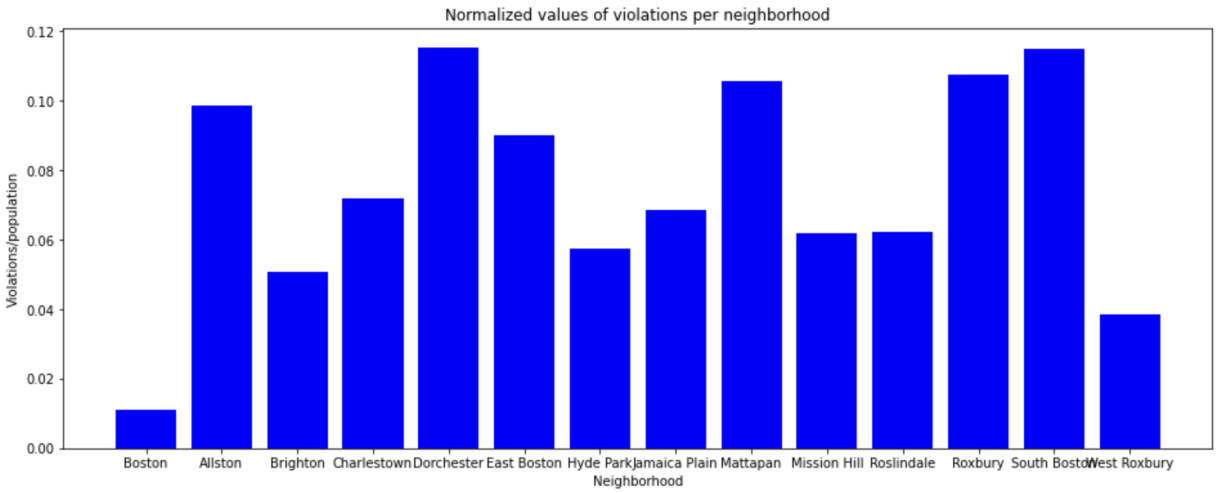
Effects of factors on number of housing violations - Violations after normalization

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. We have only taken into consideration housing violations and housing complaints.

Population

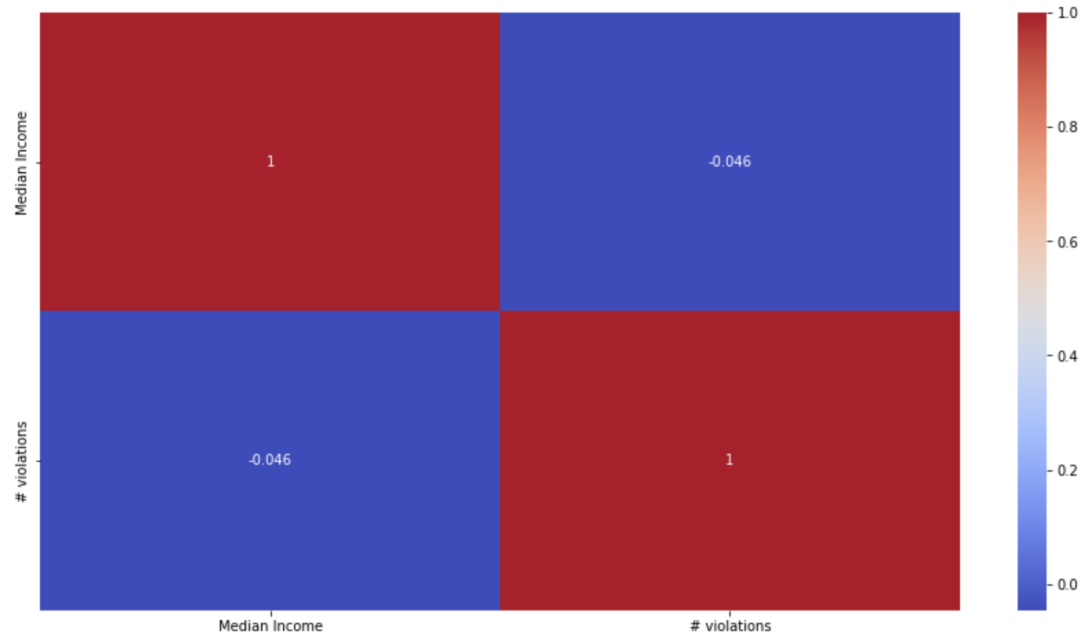
We realized that a larger number of people in one neighborhood and a relatively smaller number of people in another neighborhood could be a factor. Therefore, using the population dataset for Boston, we normalized the number of violations by dividing it by the population in that neighborhood and plotted 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.





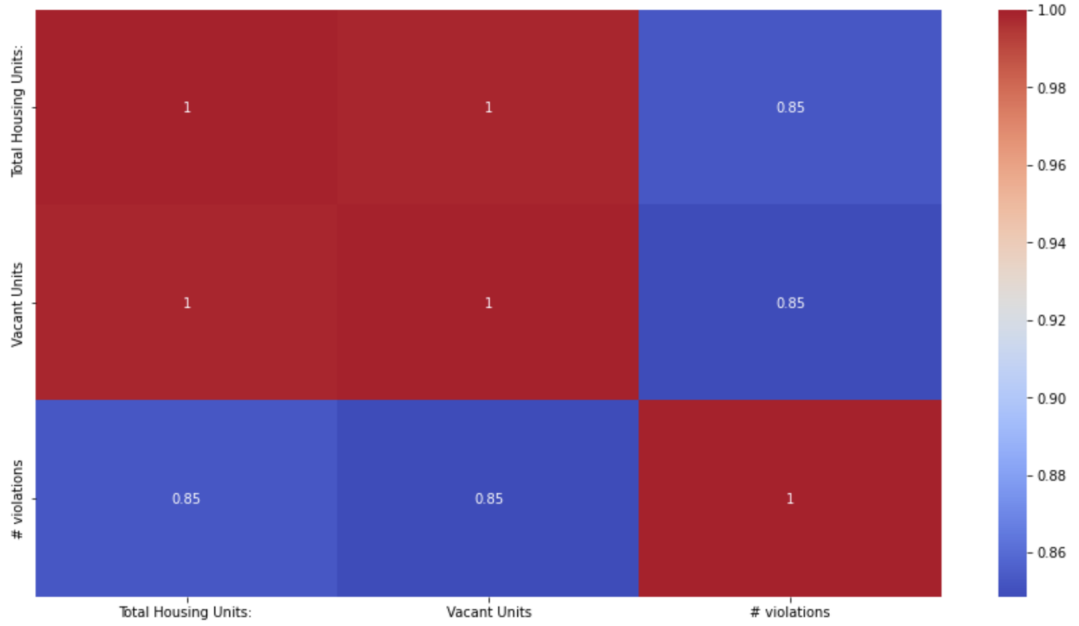
Household Income

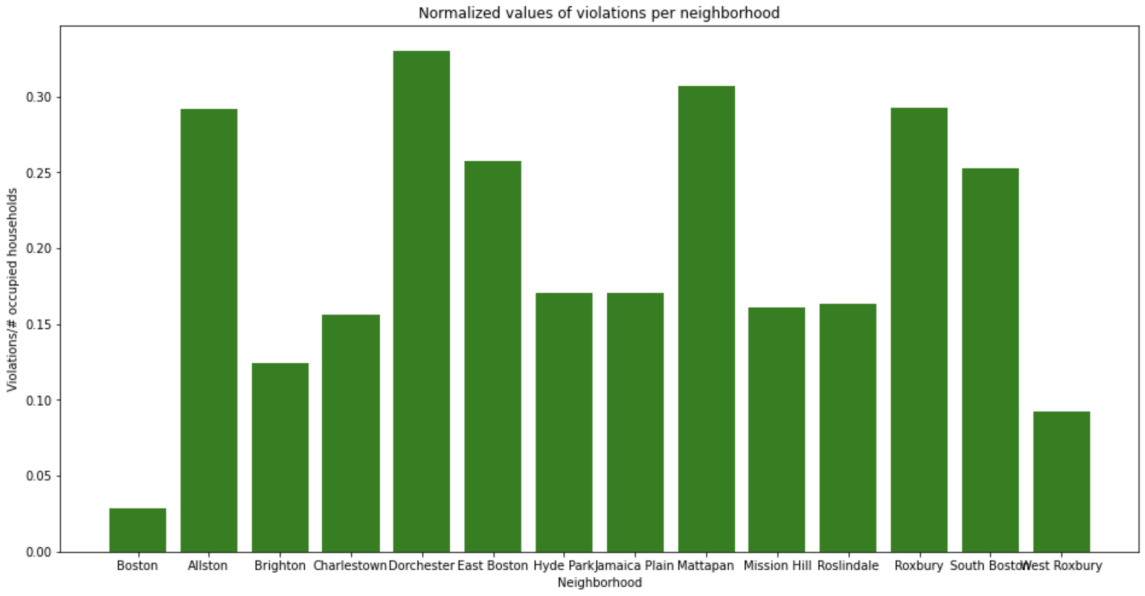
We wanted to see if household income affected the number of violations in separate neighborhoods. However, since the correlation factor was negative, we did not normalize our results based on this.



Number of occupied households

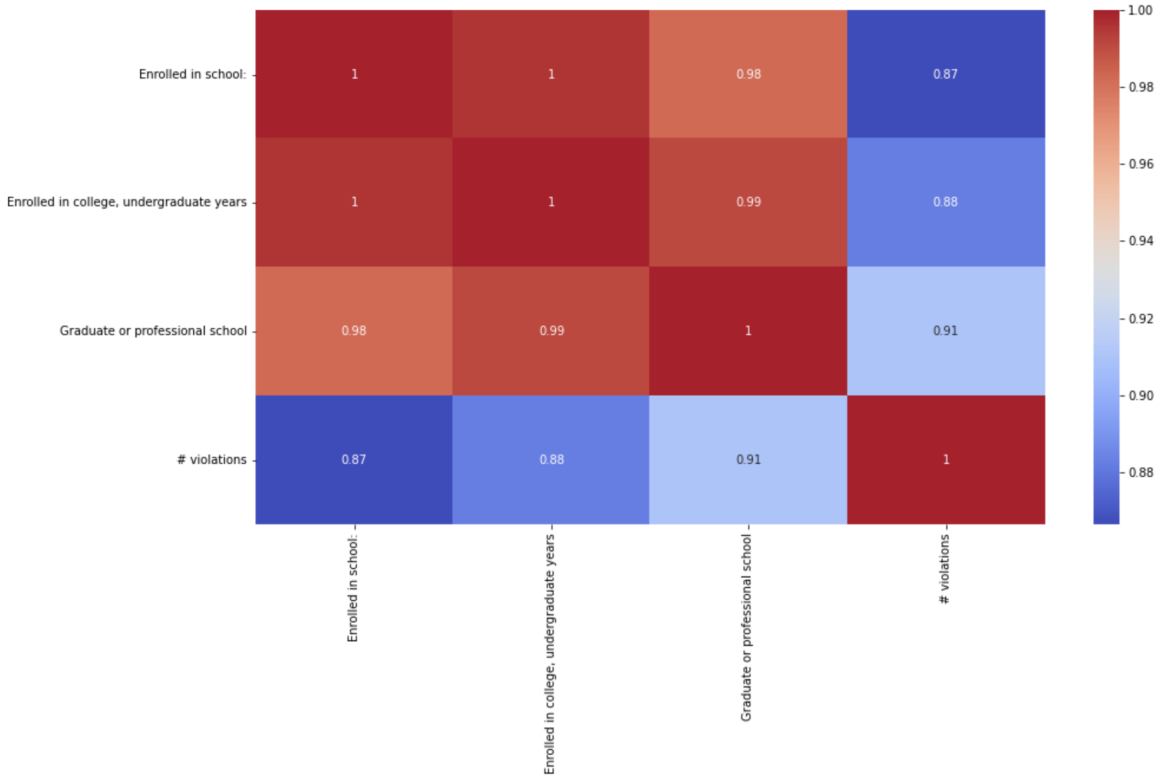
Another interesting aspect to look at, apart from population, is the number of households. Population definitely gives us a sense of the number of violations in terms of the number of people, however, a better insight would be given by the number of occupied households. The following graph shows the 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.

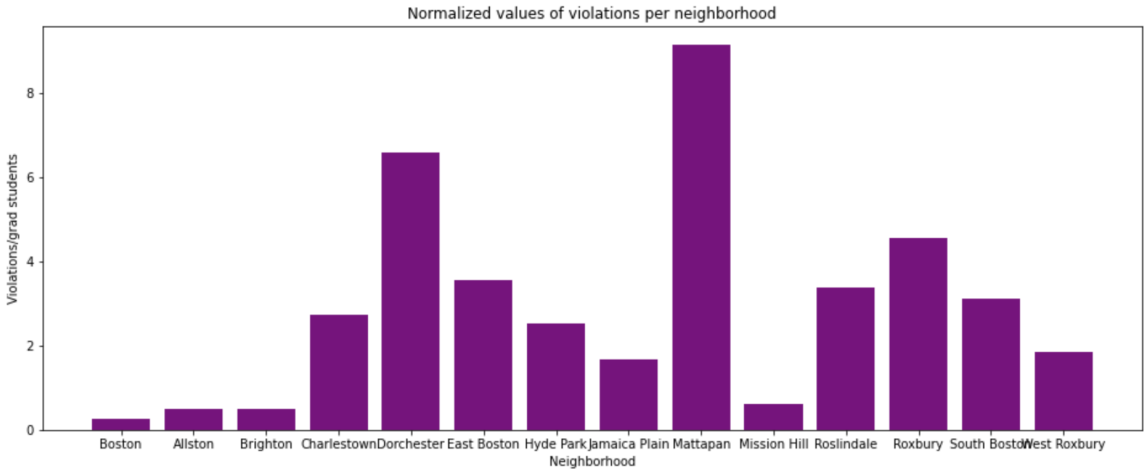




Number of students enrolled in school

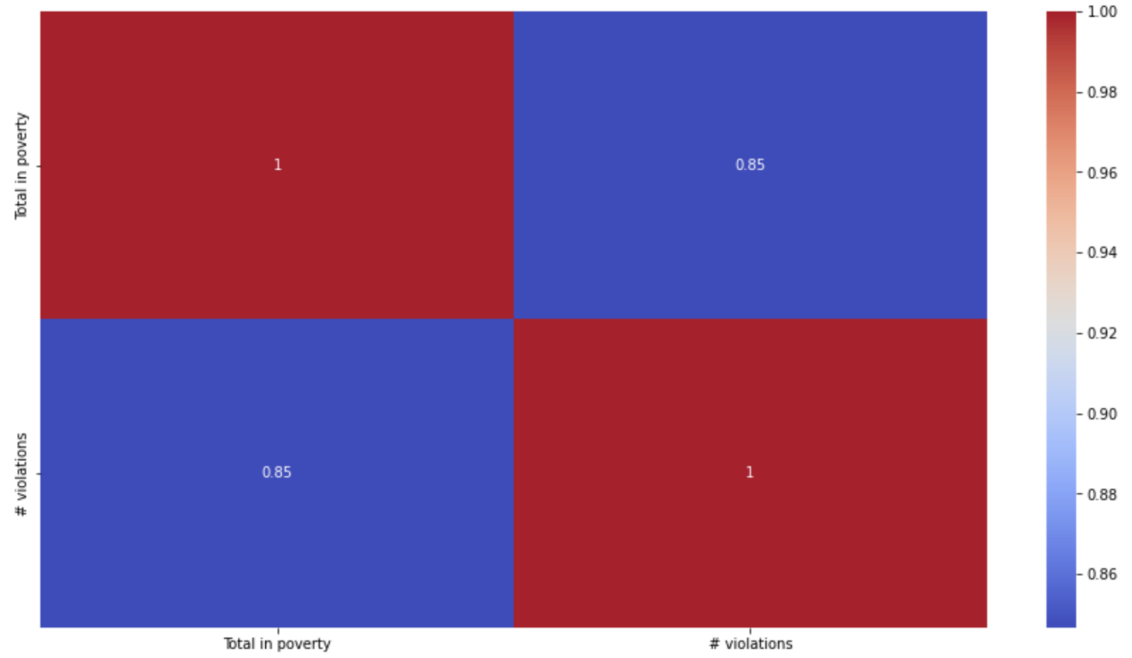
We wanted to see if the number of students in a neighborhood affected the number of violations. Therefore, we generated a correlation matrix for students which gave the highest correlation between graduate/professional students and number of violations. Therefore, we normalized the number of violations 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.

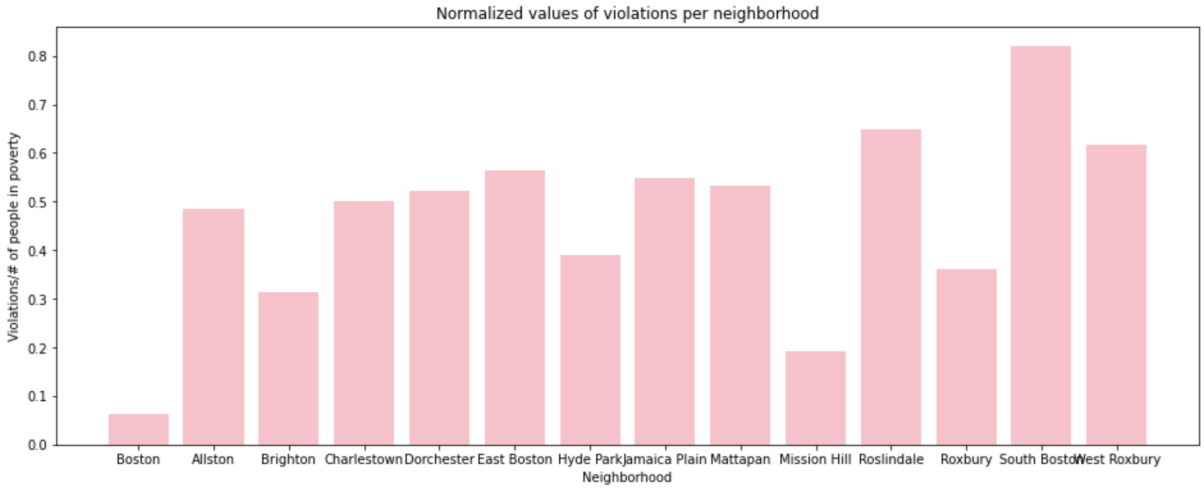




Number of people in poverty

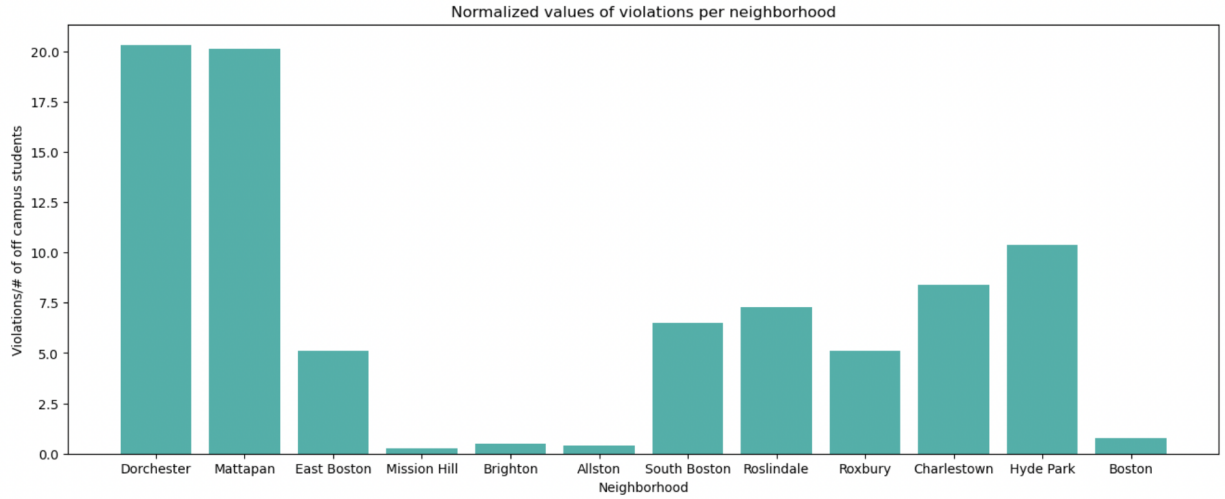
An interesting factor we took into account was the number of people in poverty. Since the household income did not give favorable results, we did not expect much from this either. To our surprise, after plotting the correlation matrix and getting a correlation coefficient of 0.85, we decided to normalize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

We used an online geo-decoding service with Geopy which took zipcodes of addresses of off campus students in the provided datasets (since no neighborhood info was provided) and returned the neighborhood to which they belonged. The following graph is a representation of the normalized violations when divided by the number of off campus students living per neighborhood. The highest seems to be in Dorchester and Mattapan (indicating that for a higher number of violations, there are fewer students living) while the lowest is in Allston, Brighton, Mission Hill, and Boston indicating a large population of students living in these neighborhoods.

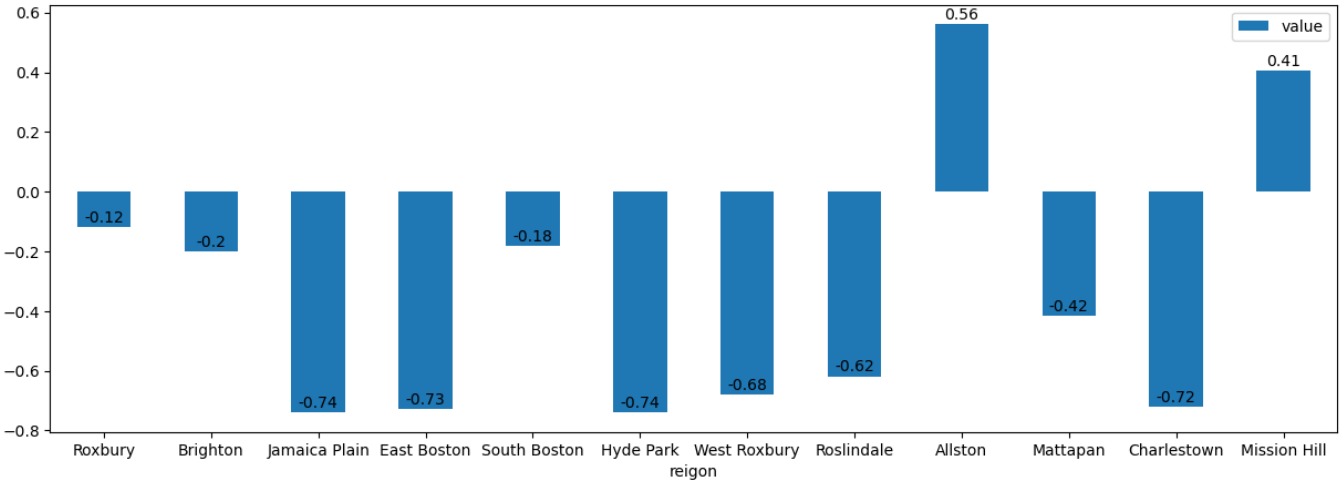


Results

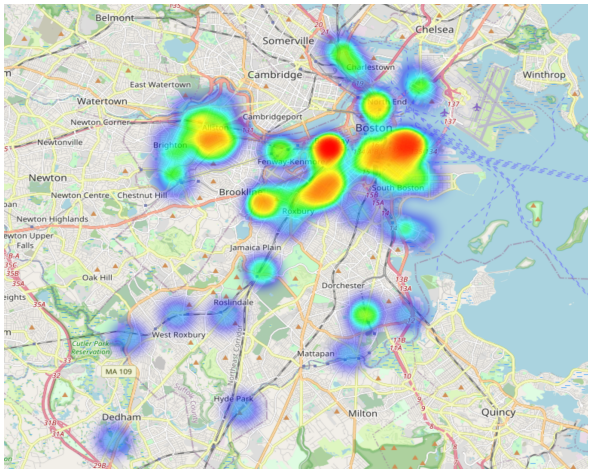
Dorchester shows a consistently high number of violations after normalization by multiple factors. Thus, we now know that the neighborhood of Dorchester has the most number of housing violations.

The correlation between violation and the house price

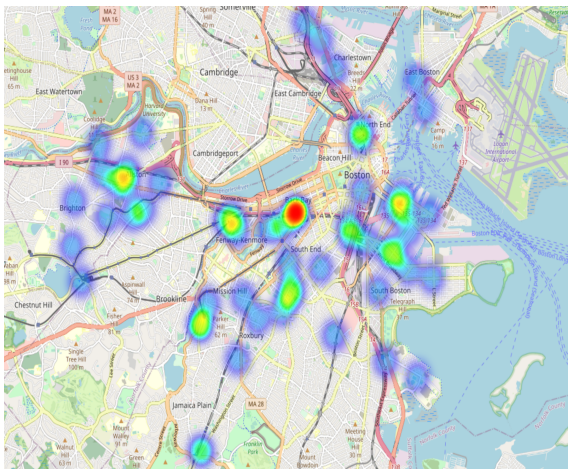
We analyzed the data provided by **Zillow** to get the mean price of houses each year from 2018 to 2022 for neighborhoods in Boston. From the RentSmart dataset, we could get the same data about the number of violations. The bar chart shows the correlation between violations and the house price. Anything 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, Allston and Mission Hill's violation number is positively related to house price while most other areas are negatively related to the house price.



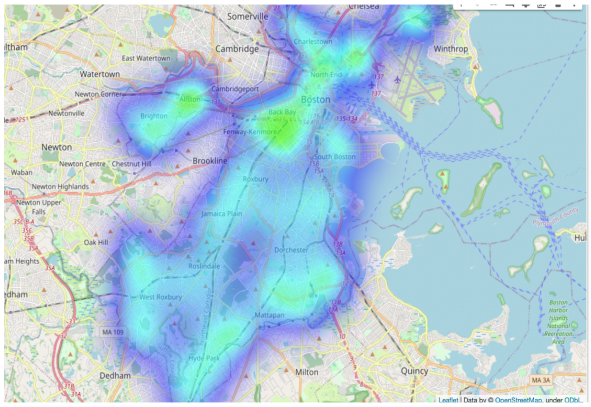
Comparison of distribution of AMI and violation



AMI 0.5-0.8



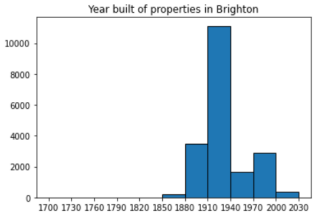
AMI >0.8

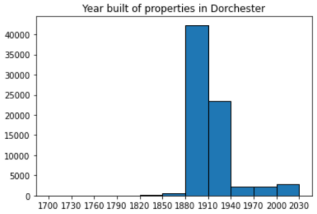


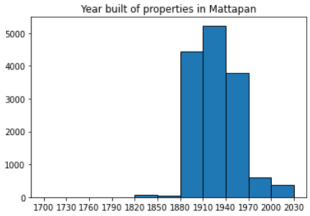
Above maps show the distribution of AMI and the violation heatmap. The AMI map depends on the point level which is transformed from street address using google geoapi. The observation is for AMI category 0.5-0.8 the heatmap is about consistent with the violation heatmap. It is easy to find that Allston 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.

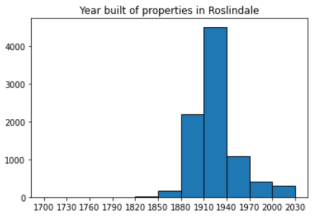
Age of properties across Boston grouped by Neighborhood

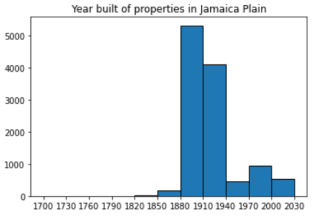
We answered the key questions - “What years were the units built in, in the violations report?” and “Status of renovation for such units”. Since deliverable 3, we rearranged all analysis so that they are grouped by neighborhood. Here’s the updated analysis on age & renovation status of the properties.

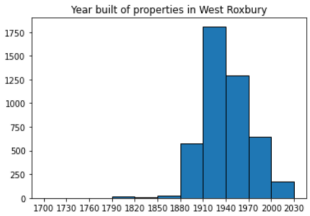


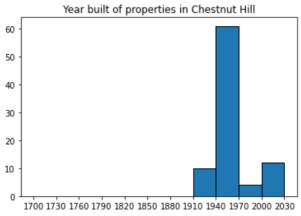


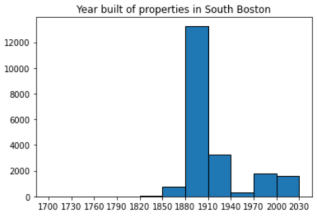


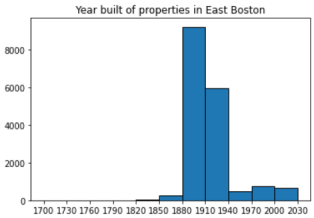


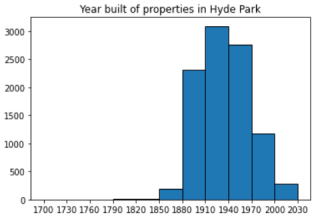


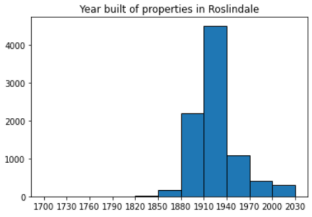


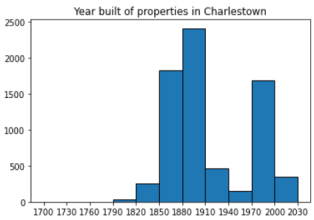


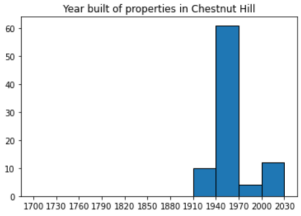






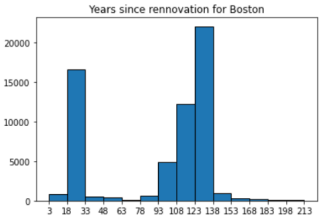


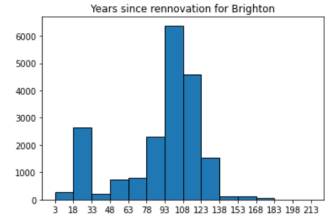


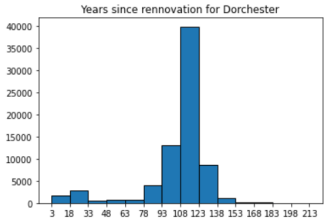


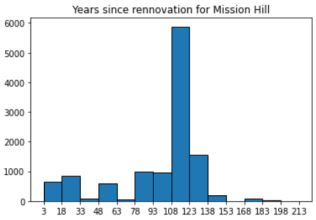
Renovation status of properties across Boston grouped by Neighborhood

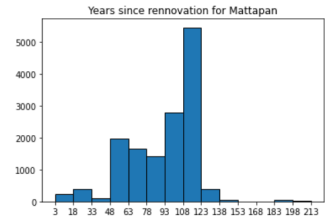
(Demonstrated by number of years since last renovation/remodel)

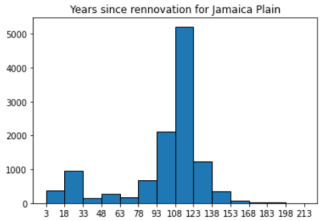


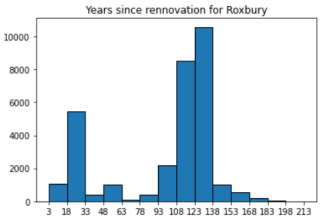


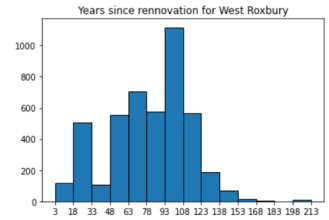


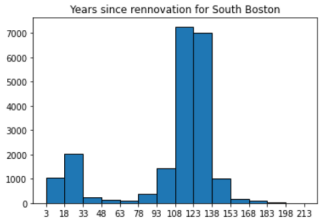


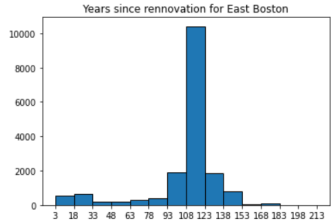


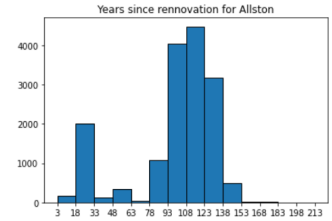


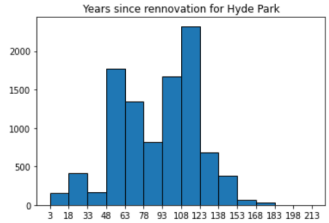


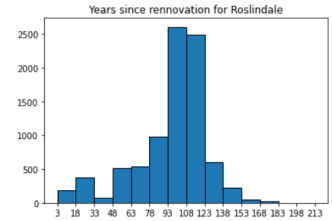


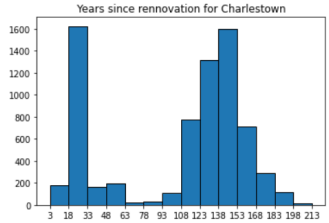


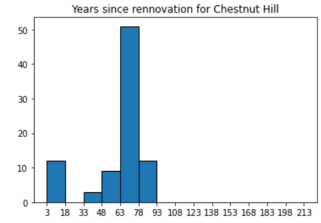












It is worth noting that the largest peak for each neighborhood, wherever they appear on the histogram, usually indicates units that did not receive any renovation since construction, as that applies to the majority of the properties. Smaller peaks, if exist, that appear before the

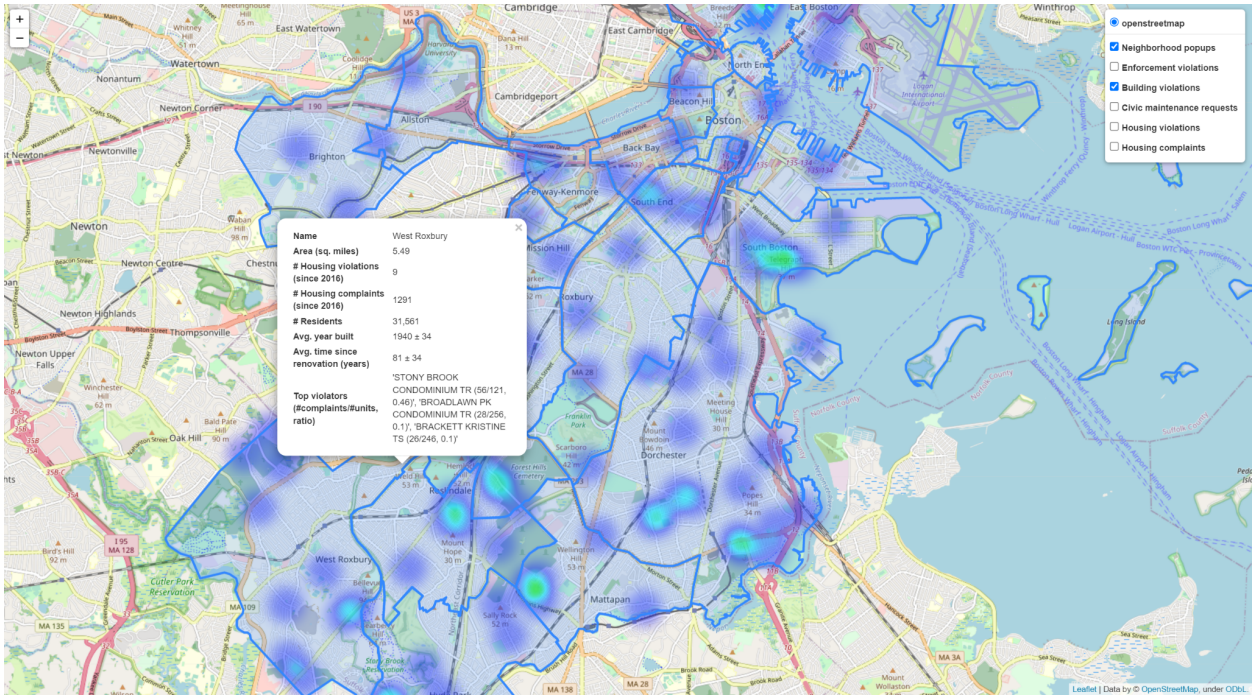
major peak are those that did receive renovation/ remodeling after construction. For example, units in Charlestown had a large-scale renovation around 20 years ago, as indicated by the corresponding histogram.

Interactive map

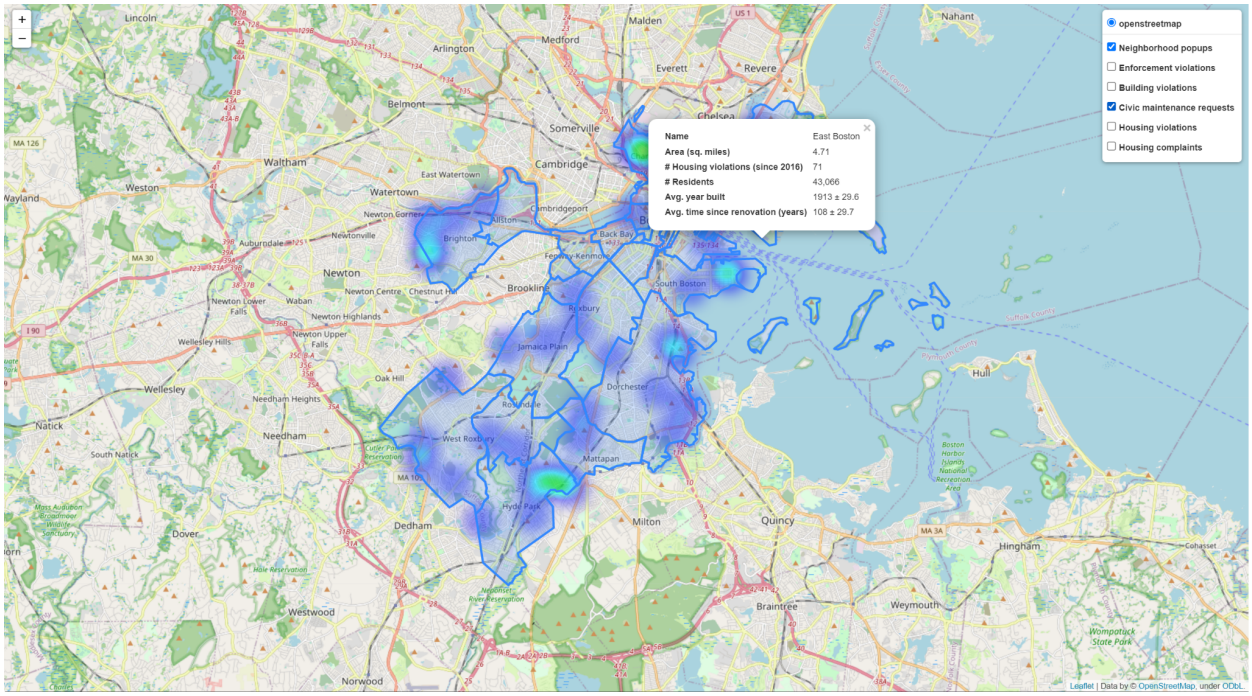
The interactive map can be accessed here. **Instructions on how to view it:**

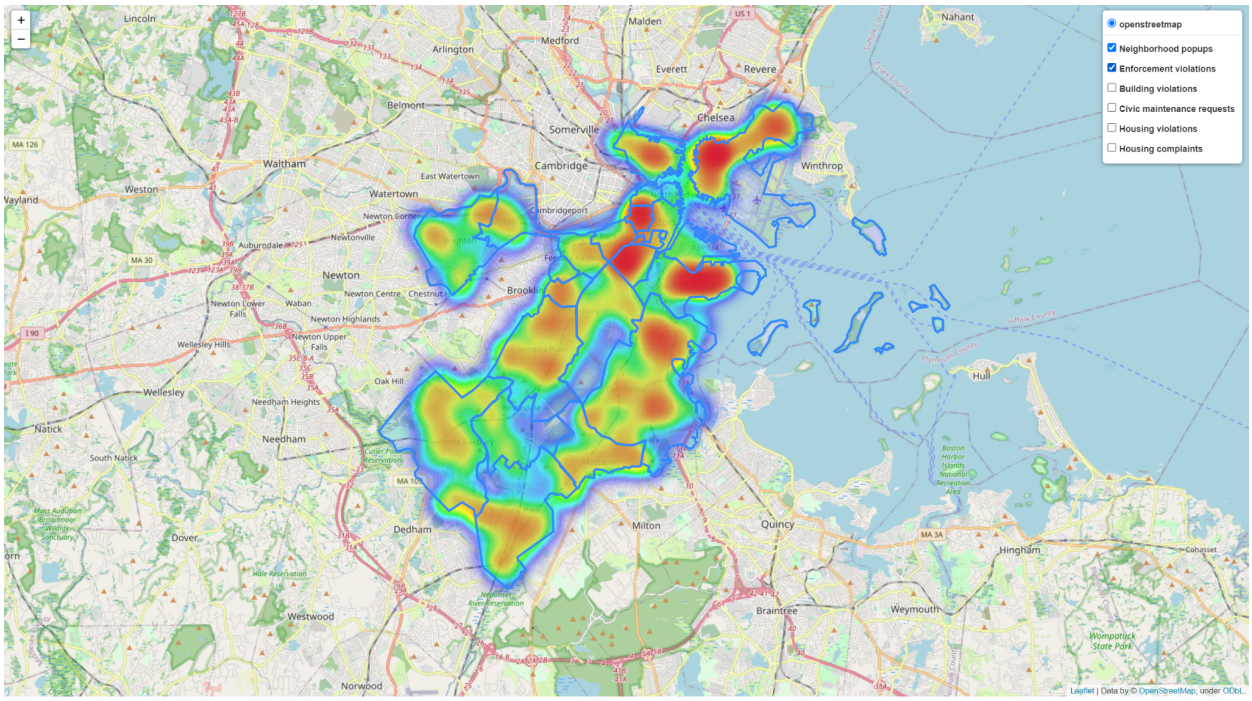
- Download the .html file on your computer
- Open the .html file with your browser

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. The most interesting part about the map is that for each neighborhood, it shows the top 3 worst landlords which we calculated by dividing the number of violations for each owner by the number of properties they owned. 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.


openstreetmap
Neighborhood popups
Enforcement violations
Building violations
Civic maintenance requests
Housing violations
Housing complaints
Name East Boston
Area (sq. miles) 4.71
Housing violations (since 2016) 71
Residents 43,066
Avg. year built 1913 ± 29.6
Avg. time since renovation (years) 108 ± 29.7
Leaflet | Data by © OpenStreetMap, under ODbL.


openstreetmap
Neighborhood popups
Enforcement violations
Building violations
Civic maintenance requests
Housing violations
Housing complaints
Leaflet | Data by © OpenStreetMap, under ODbL.

Conclusion

- From what we observed in the above slides, we can clearly see that **Dorchester** has the highest number of violations. The number of off campus students heavily influences the number of violations in each neighborhood with the ratio of violations to students being the lowest in **Allston, Brighton,** and **Mission Hill** indicating a higher number of students.
- Allston and Mission Hill's violation number is positively related to house price while most other areas are negatively related to the house price.
- From the construction time for properties in different neighborhoods we could not observe an obvious correlation between time of construction and number of violations. The neighborhood with the most violations, Dorchester, didn't have a different distribution on construction/renovation time.

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.
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.

Challenges encountered

Trends were harder to observe when multiple results were displayed on the same graph
For the trend in availability of affordable housing analysis, simply grouping by zip code resulted in 20+ lines and the trend could not be easily observed. We fully grouped the zip codes with respect to the number of units available.

Most datasets contain NaNs across the columns but not all of them can be handled with a simple df.dropna().
With knowledge learned in class, we separated the NaNs into a few types: For NaNs in must-have cells such as geographical locations, the datapoint was dropped. For NaNs that would not influence processing of data, they are replaced with the mean value of the column. For empty cells in categorical data, they were simply ignored.

Assumptions

- While trying to understand affordability inventory, we did not have a "rent" or "value" column, hence we assumed "declared value" of construction works to be correlated with value of property, i.e. construction cost implies value of property.

- We assume very less bias in the datasets: Even if data is skewed over neighborhoods, we assume the data to give a general, approximate global picture of each neighborhood.
- Zillow provided the latest data on housing prices.
- The numbers for the number of properties owned by landlords is up to date.

Sources

- https://data.boston.gov/dataset/building-and-property-violations1/resource/800a2663-1d6a-46e7-9356-bedb70f5332c - Building and property violations: Contains different types of violations and where they occurred.
- https://data.boston.gov/dataset/311-service-requests/resource/81a7b022-f8fc-4da5-80e4-b160058ca207 - 311 Service Requests: Contains all types of service requests, the place at which the violation occurred, the department it is concerned with, and whether it is overdue or not.
- https://data.boston.gov/dataset/rentsmart - Rentsmart: Gives property violations information for addresses along with the owner information.
- https://data.boston.gov/dataset/boston-buildings-inventory/resource/391a32e6-d4bb-48d3-a990-cb35a5768a40 - Boston Building Inventory: Information about buildings, their type, number of units, owners, etc.
- https://drive.google.com/drive/folders/1FaPi_Ml4YU_p_i9p_vRDAFctuXejgiic - Student data for all colleges in Boston (from which we retrieved the number of off campus students).
- https://data.boston.gov/dataset/neighborhood-demographics/resource/d8c23c6a-b868-4ba4-8a3b-b9615a21be07 - Neighborhood Demographics: To get number of populations/number of people per neighborhood.
- https://data.boston.gov/dataset/property-assessment/resource/4b99718b-d064-471b-9b24-517ae5effecc - Property Assessment Dataset: To get the number of properties owned by a landlord.
- https://www.zillow.com/research/data/ - Zillow housing dataset: To get the prices of households in different neighborhoods.