

Deliverable 4 - Team 3

Project Description and Goal:

This project, assigned by the Office of City Councilor Breadon of District 9, is centered around identifying risky landlords along with correlating factors for the different types of housing violations that these landlords commit. The interest from the City Council's side is to increase transparency for potential property buyers/renters when making the ever important decision in where they will live. By constructing a way in order to distinguish landlords by how bad/good they are, not only will prospective tenants benefit, but it will also allow the city council to identify which landlords need the strongest punishments for the violations they commit.

Primary Questions to Answer: These are 2 of the questions that Councilor Breadon and co. would like us to answer.

1. What are the types of violations? How many violations for each type?
2. Whether affordability affects the amount of property violations?

Extension Project: Does the number of bad landlords or types of violations impact how people move into certain areas/neighborhoods of Boston?

1. Questions to Answer:

- a. Big Picture: Do a higher number of bad landlords, or do a specific type of violation have impacts on people moving into certain neighborhoods of Boston?
- b. Smaller Scope: Do individual landlords see a decrease in revenue/number of properties they are able to rent out based on how bad of a landlord they are/number of violations they commit?
- c. Looking at whether based on the answers above, the number of violations/bad landlords in vulnerable groups/areas is getting better/worse over years compared to general Boston?

2. Motivations Behind Extension Questions:

The main motivation behind trying to look at the questions above is to look at whether consumers pay attention to a landlord's past track record, and how it affects the consumer's (buyer's) decisions as well as the ability of landlords to rent out their properties. In other words, does a prospective buyer care about their prospective landlord's past track record? Do bad landlords see long term effects of committing a high amount of significant violations rather than just paying fines? We plan to look at this within a large and small scope by looking at this question on an area by area basis as well as a landlord by landlord basis respectively. However, in order to do this, there are many factors that we will have to normalize in order to make this analysis fair.

3. Plan to answer these questions while making sure the analysis is fair:

1. Defining what a bad landlord is:

- a. Although one landlord may have committed more violations than another landlord, we must take into account the severity of a violation. If landlord A has committed 2 minor violations whereas landlord B committed 1 major violation, it might be reasonable to conclude that landlord B is the worse landlord among the two. As a result, we must quantify the severity of each type of violation and then calculate a score to quantify how good/bad a landlord is.
- 2. Removing affordability/price as a factor:
 - a. Oftentimes people will move into the cheapest area that they can, or rather, the cheapest property that satisfies their needs. Because of this, we do not want our analysis to be skewed towards cheaper areas of Boston since they might have a higher amount of people moving into those areas compared to more expensive options. As a result, when looking at the larger scope question, we plan to take this into account by bracketing areas by AMI and comparing all the neighborhoods within each AMI bracket to each other.
- 3. Using percent of possible income as landlord rent metric:
 - a. Since landlords own a different amount of properties, it would not be fair to compare them based on the total amount of income that they have. As a result, to compare the amount of income that they have, it would be most fair to look at landlords with at least 5 properties, and measure the amount of rent they are able to obtain based on the percent of money they obtain divided by the total possible rent income.
- 4. Using Zillow Data and matching addresses to landlords:
 - a. Based on the feedback on the presentation, we plan to use Zillow housing data as well as look at other sources in order to look at property sales and through this, calculate the rates over time that people continue to rent/buy properties in each area of Boston.

Data collection:

- 1. Primary Questions:
 - a. Rentsmart
Dataset(<https://data.boston.gov/dataset/rentsma...>)
 - b. Building and Property Violations
Datasets(<https://data.boston.gov/dataset/building-and-property-violations1>)
 - c. Public Works Violations
Dataset(<https://data.boston.gov/dataset/public-works-violations>)
 - d. 2022 Property Assessment
Documentation(<https://dataverse.harvard.edu/file.xhtml?fileId=6416049&version=1.0>)
 - e. 2010-2021 Code, Building, and Property Violations
Dataset(<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TD9YOY>)

2. Extension Project:
 - a. Zillow Housing Data(<https://www.zillow.com/research/data/>)
 - b. Rentsmart
Dataset(<https://data.boston.gov/dataset/rentsmart/resource/dc615ff7-2ff3-416a-922b-f0f334f085d0>)
 - c. 2022 Property Assessment
Documentation(<https://dataverse.harvard.edu/file.xhtml?fileId=6416049&version=1.0>)
 - d. 2010-2021 Code, Building, and Property Violations
Dataset(<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TD9YOY>)
 - e. Boston Buildings
Inventory(<https://data.boston.gov/dataset/boston-buildings-inventory/resource/391a32e6-d4bb-48d3-a990-cb35a5768a40>)

Results Obtained/Questions answered:

Primary Questions:

1. What are the types of violations? How many violations for each type?
 - As we see below, using the Rentsmart Data in conjunction with the Property Violations Dataset provided from Harvard, we were able to determine that these were the frequencies of each type of violation over an 11 year period. Enforcement violations clearly dominated, as it alone is more than all the others combined. However, with an emphasis on housing, we can also see that housing complaints are quite prevalent compared to many of the other violations excluding enforcement violations.

With Enforcement Violations

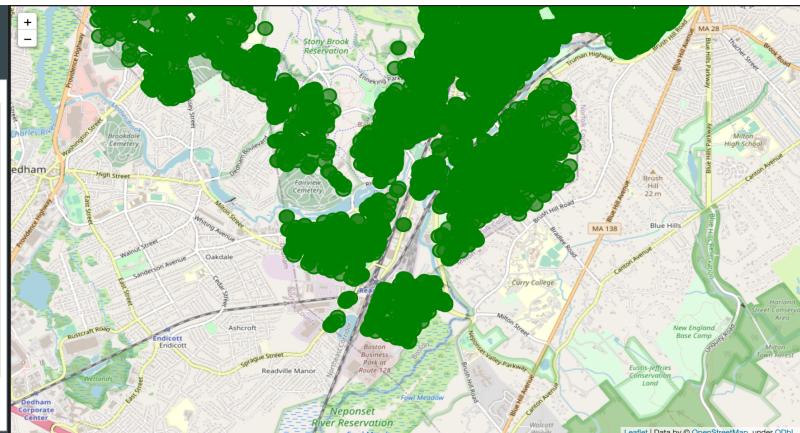
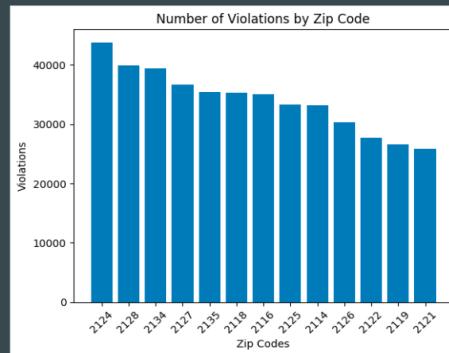
- With the data used to create the graph above, it allowed us to focus on more specific metrics, such as analyzing violations per zip code. Although not explicitly stated in the question, we felt this was important to include as it highlights the areas in Boston suffering most from property violations, and that there should be a heavy emphasis by the City of Boston to help correct the behavior of bad landlords in these areas. Along with this, we can visualize this with a cluster map showing the center of these violation-stricken areas.

Without Enforcement Violations

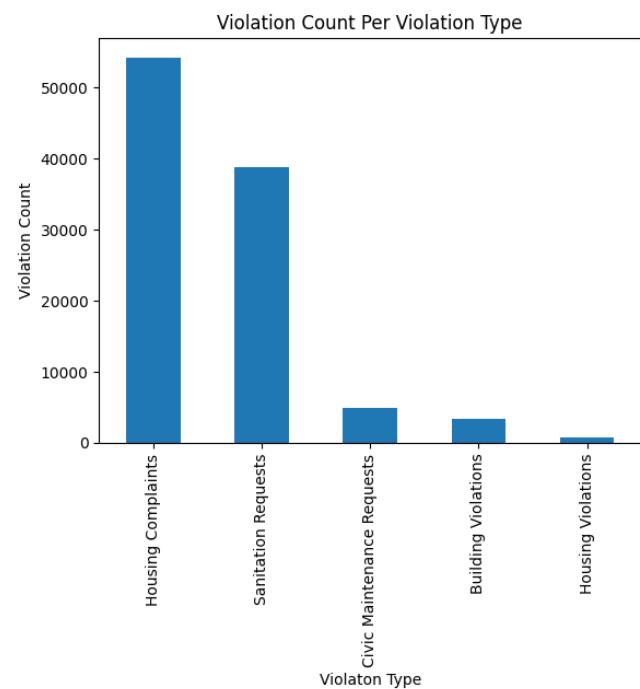
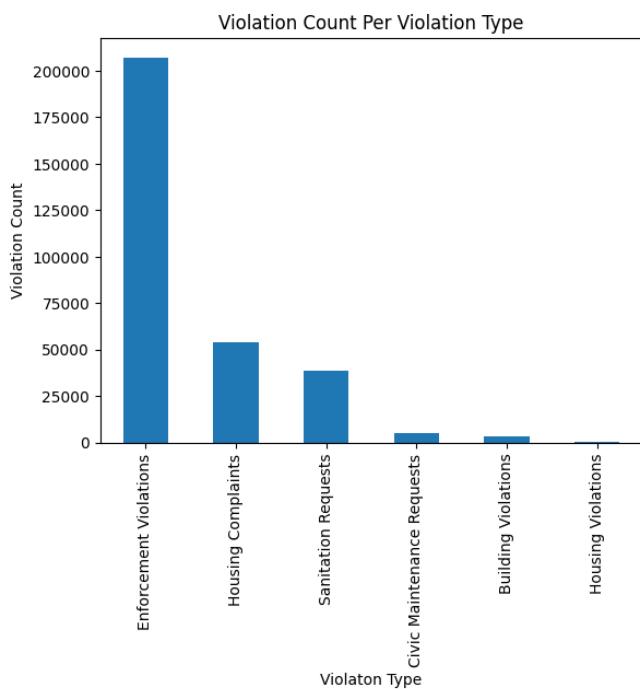
Violations by Zip Code

13 zip codes hold 76% of the violations

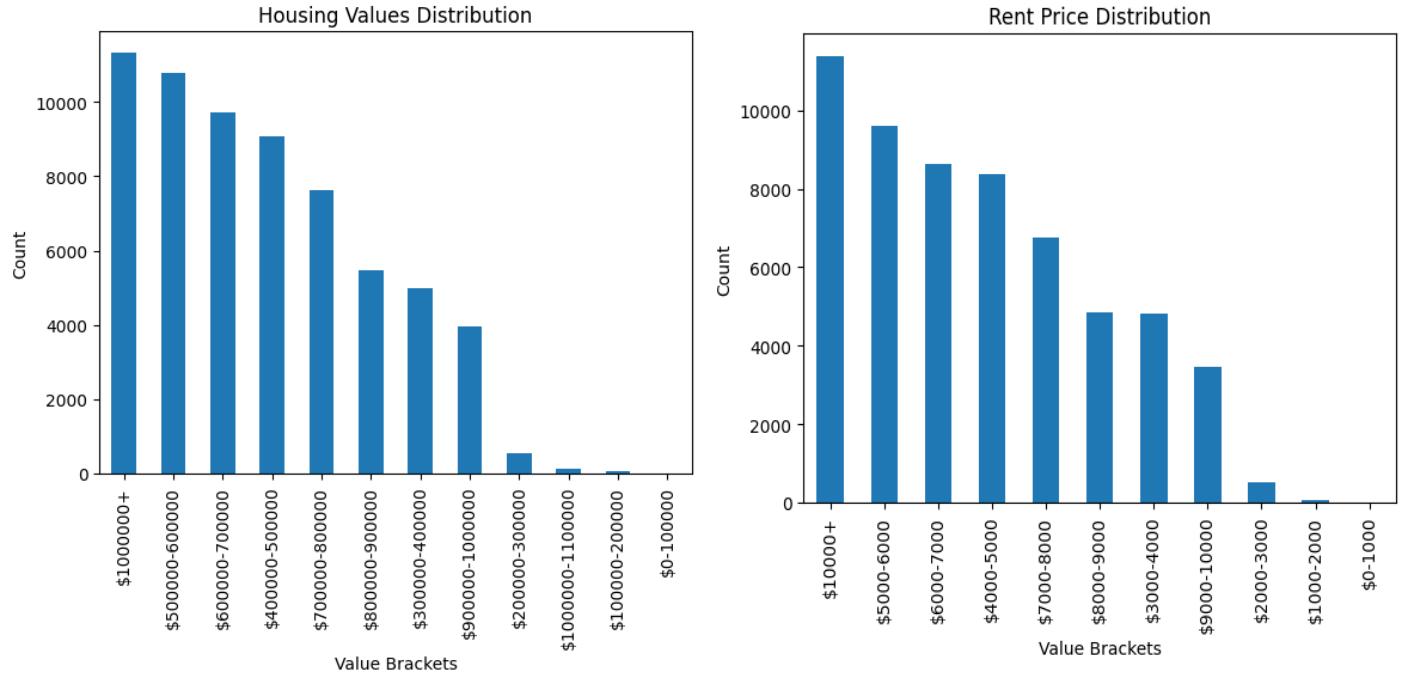
- 2124 (Dorchester)
 - Total: 43,777, 7.22%
- 2128 (East Boston)
 - Total: 39,860, 6.58%
- 2134 (Allston)
 - Total: 39,462, 6.52%
- 2127 (South Boston)
 - Total: 36,653, 6.05%
- 2135 (Allston)
 - Total: 35,395, 5.85%
- 2118 (South End)
 - Total: 35,021, 5.78%
- 2116 (Back Bay)
 - Total: 33,308, 5.50%



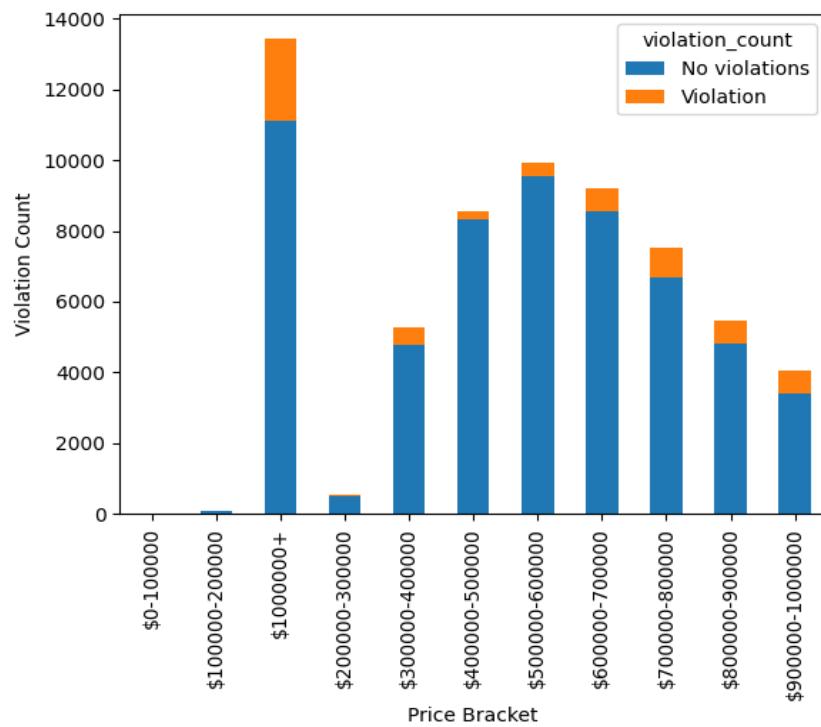
2. Whether affordability affects the amount of property violations?



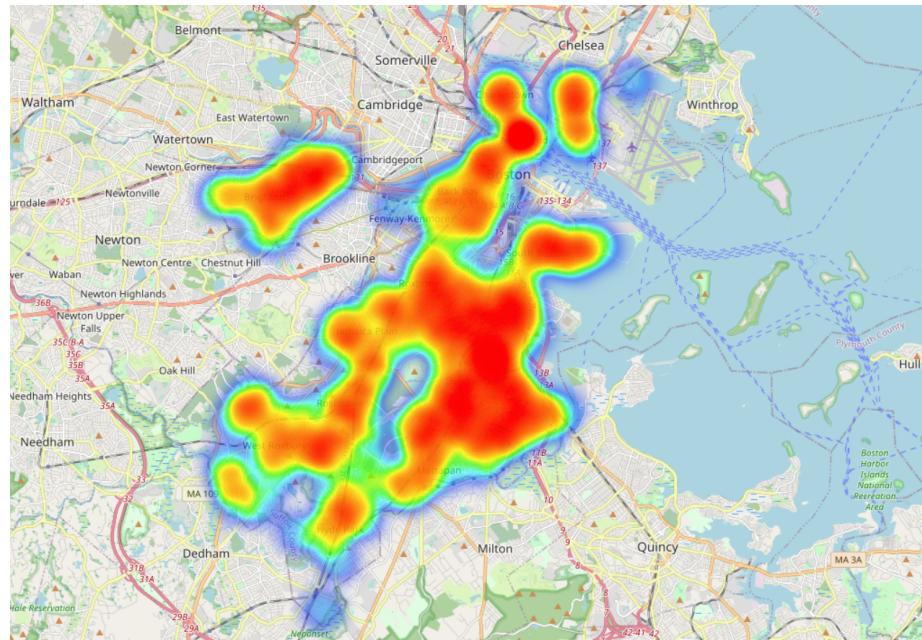
- Determining whether affordability has any correlation with the amount of property violations was a bit tricky considering we were not given any data related to property values at the start. However, using the Harvard dataverse, we were able to grab 2021 property values, and thus, able to estimate rent using the 1% rule stating that the rent of a property is usually around 1% of its property value. Furthermore, the data was also filtered for only residential properties since the focus of this project is on bad landlords. As a result, we were able to plot the number of residential properties that fall between different price brackets, and their projected rent, as seen below:



- This, albeit not too surprising considering that Boston is a high cost of living city, shows how expensive properties are in Boston. Now relating this to the amount of violations per property type which is shown below, we note that properties priced at more than a million dollars dominate the number of violations amongst all other price brackets. Although maybe not surprising considering it contains the most properties, it is surprising in the sense that you would typically expect that lower priced properties are more worn down which would create a greater number of violations. However, this does not seem to be the case.



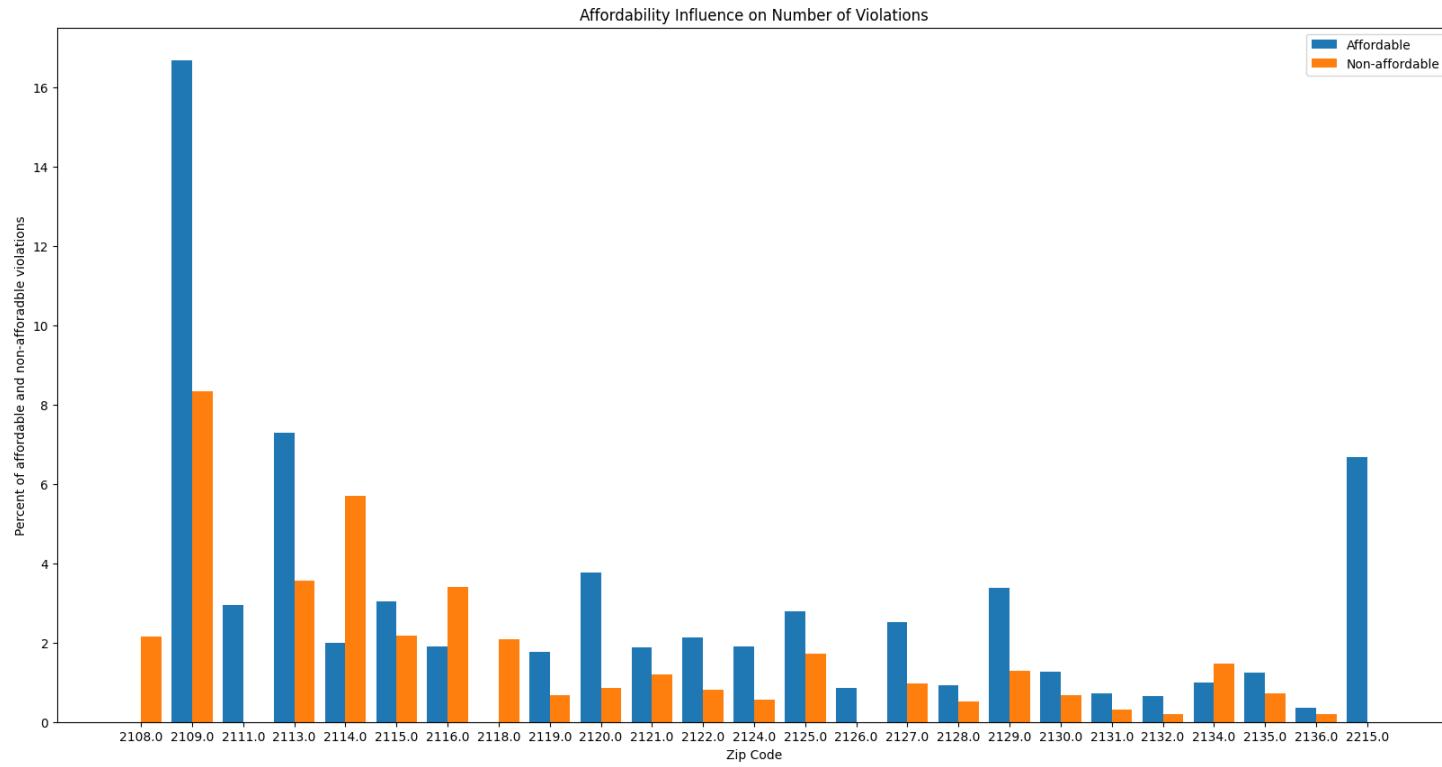
- With these two graphs above, we plotted a heat map displaying the areas with the highest correlation of number of violations as well as price. It shows very similar results to that of the zip code map answered from the first question.



We then analyzed how affordability affects number of violations. In order to do that we had to first distinguish between properties that would be considered affordable vs non-affordable. In order to do that we compared monthly rent calculated in the previous steps to the “fair rent prices”. If a monthly rent price was lower than the “fair rent price” than we considered such property affordable and non-affordable otherwise.

When we looked at properties from our dataset some of them were built for more than one family. So, we had to calculate monthly rent for each family. Besides, we approximated that each family in the US consists of 3 people based on census.gov data. And that meant that we would look at the “fair rent prices” of 3 bedroom apartments because this is an average recommended apartment size for a family of 3 as mentioned by various realtor sources.

After identifying affordable vs non-affordable properties we matched them with the properties from Harvard dataset of housing violations using exact geographic coordinates. This left us with properties that had violations and were already classified as affordable or non-affordable. We analyzed properties by most common zipcodes and our analysis revealed the following graph.



The graph demonstrates the percent of affordable and non affordable properties that had violations and were sharing common zipcode. If we look at the found data we can see that higher percentage of affordable properties would have violations under their name. This finding confirms our intuitive idea that more affordable properties would have more violations associated with them. However, we can notice that for some zipcodes this was not necessarily the case. One way to explain that would be that those zipcode areas would have way less number of affordable properties which would lead to such difference.

The challenge of the given analysis was that it was hard to find many matches from original dataset and the violations dataset. It would be nice to further repeat the analysis but on a larger dataset. Besides, there were various approximations that we had to use during the process (1 percent rule, for example), however, the analysis still gives a good idea that affordability does affect the number of violations.

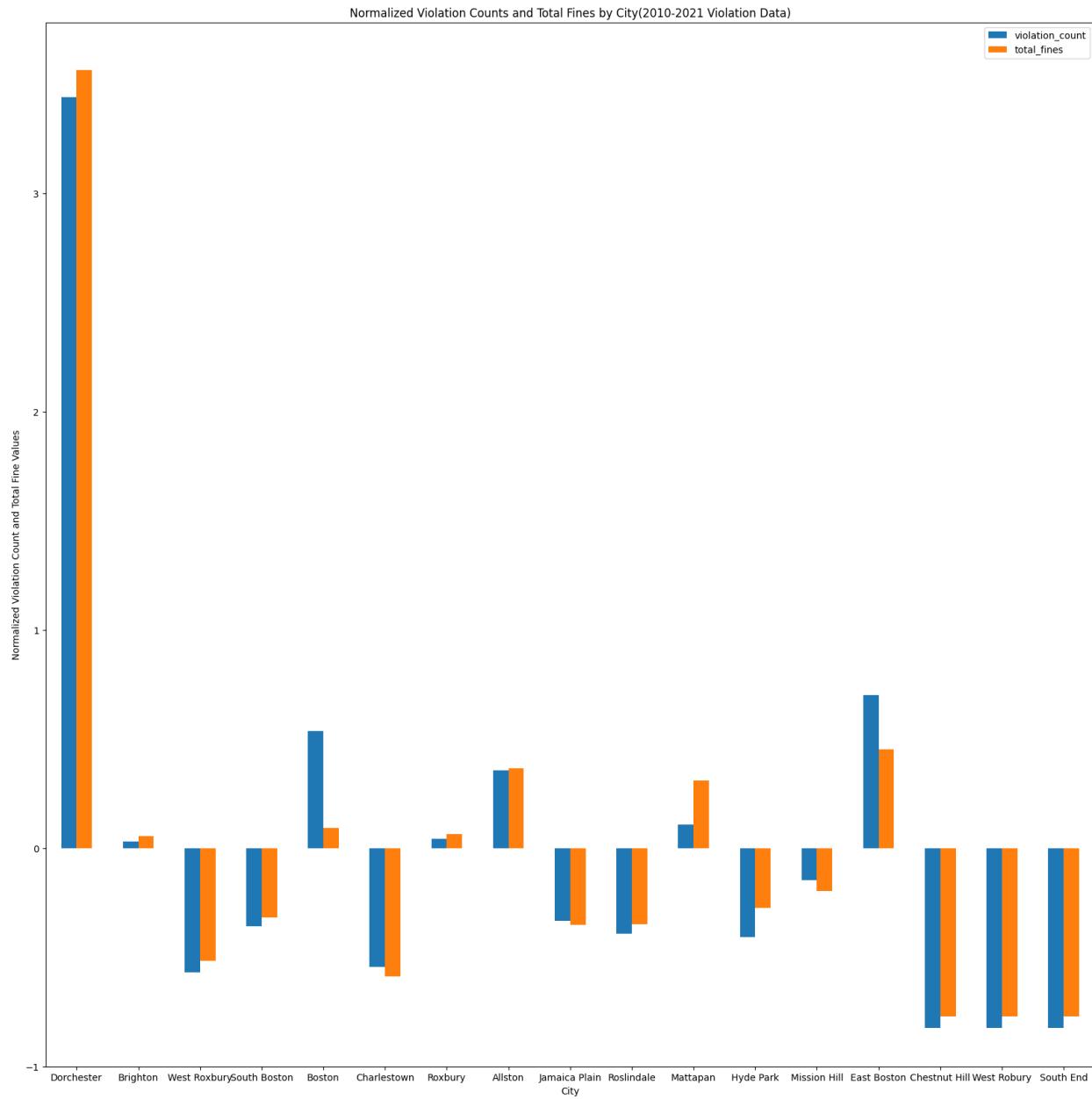
Extension Project:

1. Big Picture: Do a higher number of bad landlords, or do a specific type of violation have long term impacts on people moving into certain neighborhoods of Boston?
 - For our extension question we wanted to look at how the behavior of landlords affect people choosing certain types/areas of housing. There is a lot of data available for New York City where top 50 bad landlords are listed from all over the city along with all the violations they did and the

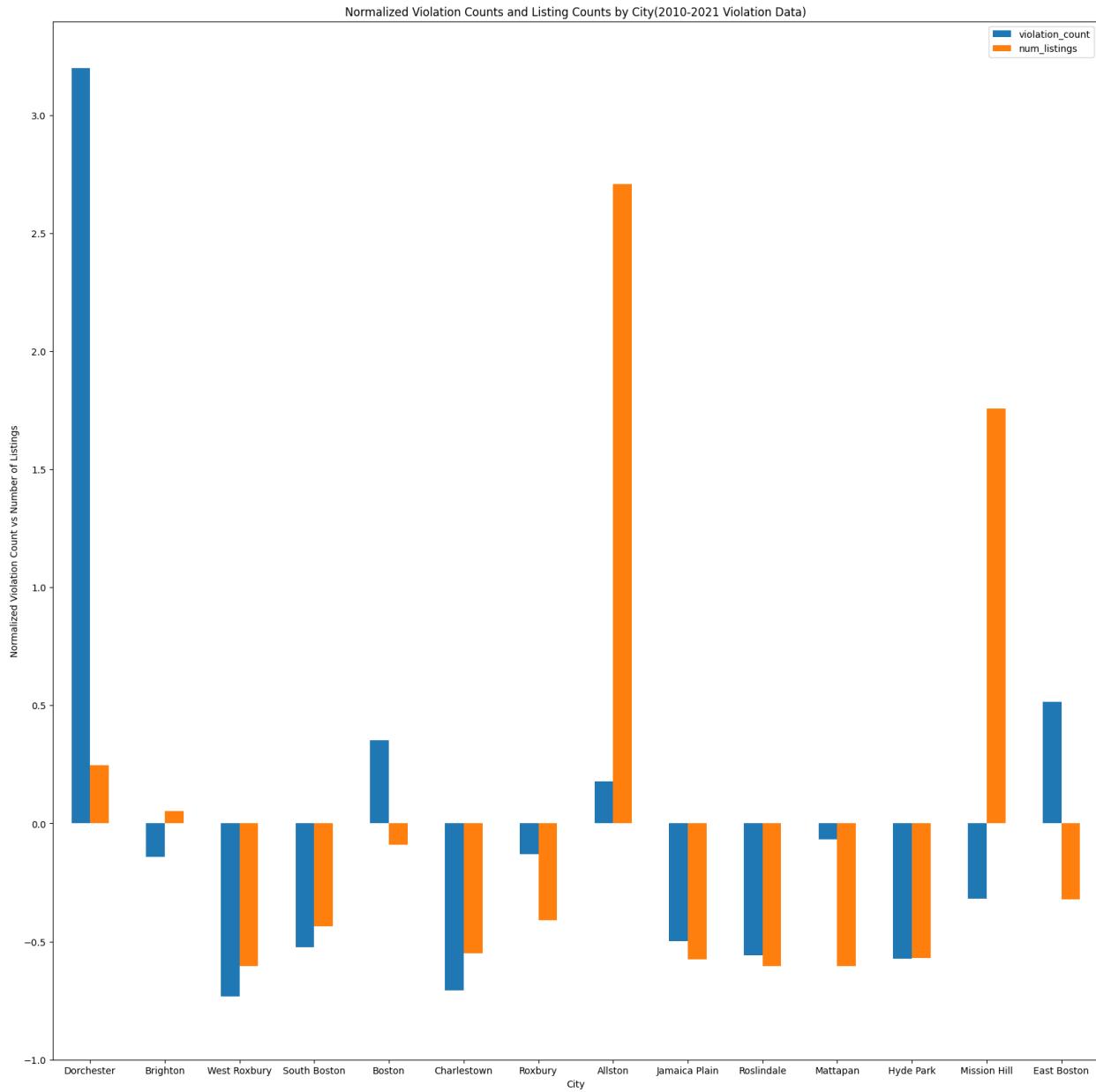
number of units they rent out. But no such data is available for Boston and nearby regions. We wanted to find suitable data so we can prove our assumption of whether people prefer not to choose bad landlords or it has no effect on their preferences. By bad landlord we typically mean that if landlord A has committed 2 minor violations whereas landlord B committed 1 major violation, it might be reasonable to conclude that landlord B is the worse landlord among the two. Although, we did find the data that shows the number of violations and the rental units in areas which have the most violations. It is an uneven distribution but we do see that the number of units rented decreases as the violation goes up in the following table. “But a review of data provided by Boston’s Inspectional Services Division, which inspects rental housing and enforces the city’s housing code, shows that while most of the scores of residential rental buildings have few or no housing violations, more than a dozen buildings have racked up hundreds in the last 18 months.”(<https://www.nbcboston.com/news/local/despite-progress-boston-tenant-vs-landlord-conflicts-abound/136310/>) Studies do show that people suffer from long term impacts from the violations committed while renting. Some of these include people getting long term respiratory infections, asthma, lead poisoning, injuries, and mental health.

Area	Number of violations	Number of rental units
58-80A Cedar St	582	24
731 River St	478	21
720 River St	370	15
134-136 Everett St	264	6
3 McCormack Square	256	7
119 Peterborough St	195	31
461 Park Dr	180	21
11 Cummings Rd	160	6
53 Hancock St	168	9
62 Clarendon Sr	166	8

- We were also able to determine the violation counts and fines for each Boston suburb as seen in the graph below, normalized by their respective columns:



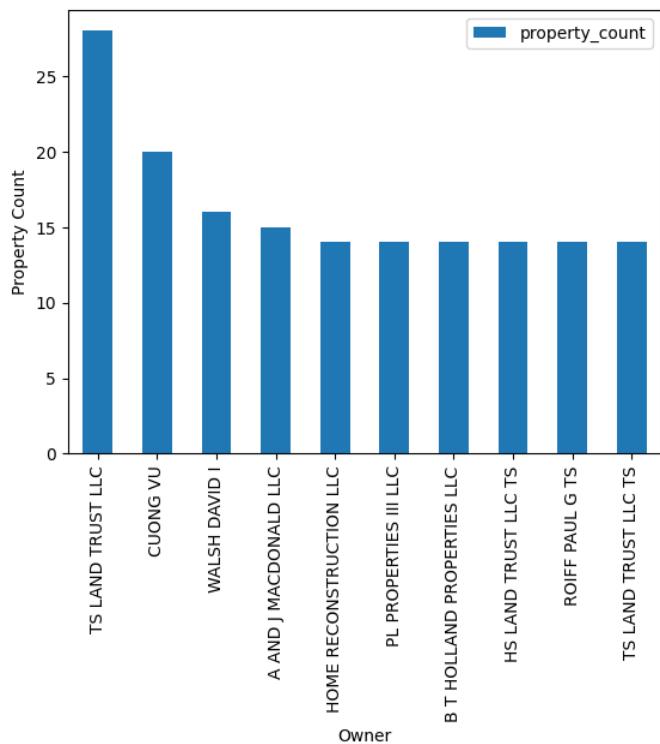
- As seen, Dorchester leads the way by a heavy margin in both categories, racking up a large amount of fines while also maintaining the most amount of violations. Outside of that, Boston, Roxbury, Allston, Mattapan, Brighton, and East Boston are all those who have positive normalized data for these two categories. As a result, these are the Boston suburbs that need the most amount of attention when it comes to cracking down on bad landlords.
- Similarly, we can look at this data normalized by also looking at the number of listings based off data we have web scraped from Zillow in the graph below:



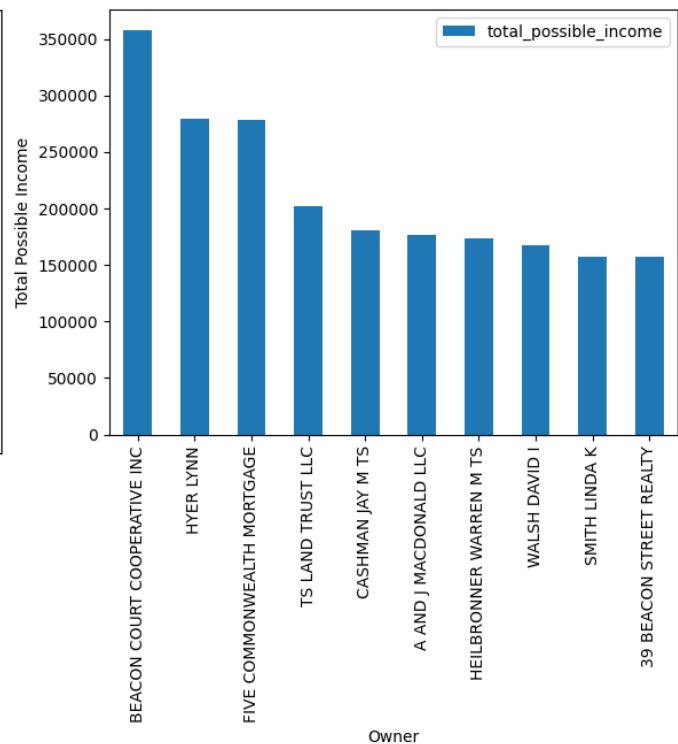
- What we can gather here is that there is not a huge correlation between the amount of violations that occur in a city and the number of listings that are available. Allston for example, has many listings currently available on Zillow, however, relative to Dorchester, East Boston, and Boston commits significantly less violations. On the contrary, the fact that West Roxbury, South Boston, and many others have negative normalized data for these two categories suggest that they are possibly more desired locations while also committing less violations. This gives us a good idea of the amount of violations occurring in different areas of Boston while seeing if that matches with there being a lot of properties available in that particular region.

- Smaller Scope: Do individual landlords see a decrease in revenue/number of properties they are able to rent out based on how bad of a landlord they are/number of violations they commit?
 - To start, we analyze which landlords have the most properties. To do this, we used data from the Boston Buildings Inventory Dataset and matched it with the Harvard Dataverse dataset in order to obtain the most amount of properties possible. From there, using the 1% rule that was previously mentioned, the probable rent value was calculated for each property. From this, we generate two metrics:
 - Landlords with the most properties
 - Landlords with the most possible income(based on the 1% rule).

Landlords with Most Properties:

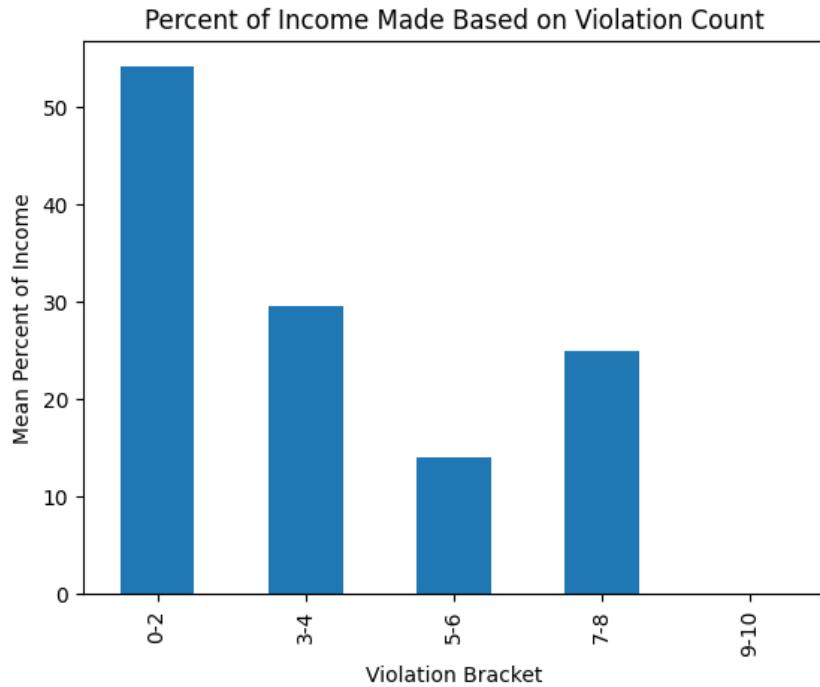


Landlords with highest possible income:



- As we can see from the graphs above, there does not seem to be a huge correlation between having a large property collection and total possible income as only 3 of those with the most properties appear in the highest possible income graph(TS Land Trust LLC, Walsh David I, A and J MacDonald LLC).
- With this information in hand, we were able to web scrape data off of Zillow which included current listings for properties in the Boston area. By comparing the available listings on Zillow at their current rent price, we can compare to see how much money each landlord is not only missing out on, but also, how many properties they are able to lease at a time. In the graph below, we are able to see

whether the amount of violations committed by a landlord affects the percent of their total possible income they are able to make:



- As seen from the following bar chart, it does seem to convey some correlation between good landlords being able to retain a majority of their total income. Landlords who commit greater than 9 violations do not seem to have a large amount of their properties leased while those who are relatively clean(0-2) violations are able to make greater than 50% of their total income. This chart is made from our web-scraped Zillow data and the Rentsmart dataset giving us insight into violations committed by each landlord.

3. Looking at whether based on the answers above, the number of violations/bad landlords in vulnerable groups/areas is getting better/worse over years compared to general Boston?

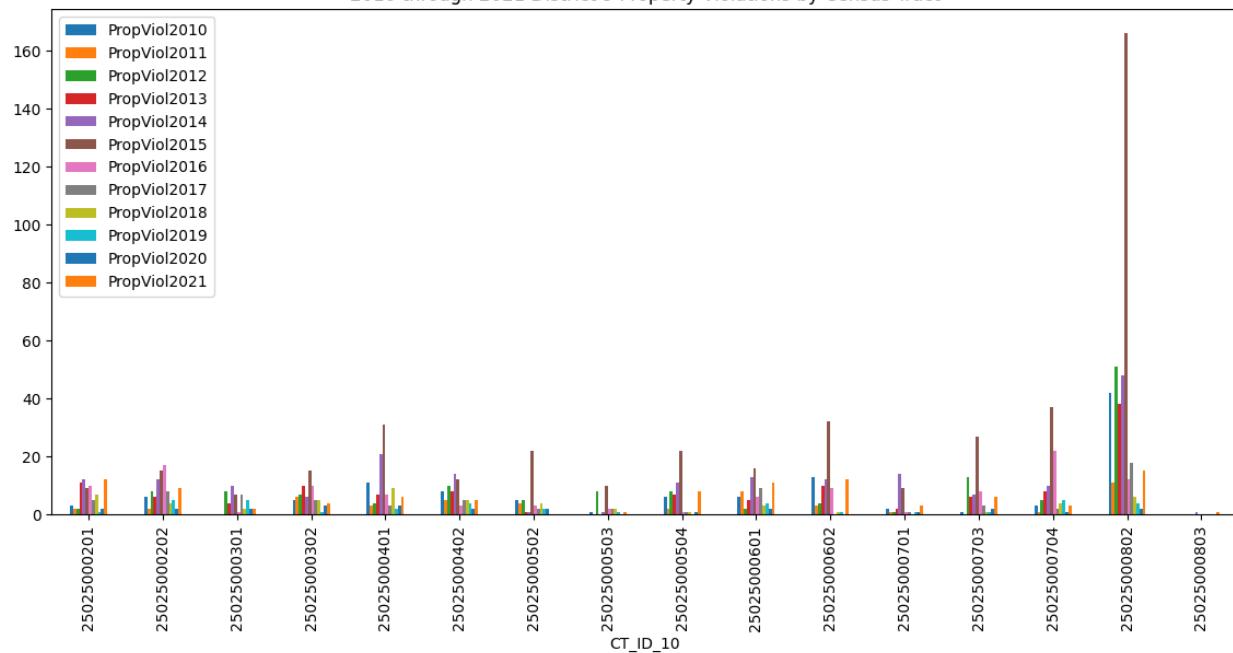
- In order to determine whether or not the effects of violations or unhelpful landlords on vulnerable groups is changing over time, we must compare the amount of violations over time in a certain area. We must also utilize a set definition for vulnerable groups: "Social vulnerability is defined as the disproportionate susceptibility of some social groups to the impacts of hazards, including death, injury, loss, or disruption of livelihood." This extension project will help us learn about trends related to vulnerability and affordability in the city of Boston. People and communities who are impacted the most under our definition of vulnerability would likely be those who fall on the lower end of the income scale.

- Data collection involved utilizing the Boston Area Research Initiative's (BARI) Longitudinal Property Assessment dataset from the years 2010 through 2021. This dataset documents housing ownership and value ownership which is useful for charting trends and analyzing changes over time. The BARI Code and Property Violations dataset containing property and code violations from 2010 through 2021 was also used to aggregate the amount of violations. Additionally, we used the Census Reporter's Household Income in the Past 12 Months dataset to retrieve the amount of individuals who fell in certain categories. This was used to help us identify socioeconomic status. Finally, a map of 2010 Census Tracts and Neighborhoods from the Boston Planning and Development Agency was also used to determine which properties corresponded to which neighborhoods. We discovered that there are 16 census tracts in the neighborhoods of Brighton and Allston, so we tailored our analysis to these 16 tracts to make it more relevant for District 9.
- We had to aggregate data from all datasets and find a way to map data regarding the census tracts to their corresponding ID in the property violations dataset. We then analyzed the amount of violations by tract, and then split the dataset into Brighton and Allston specific dataframes. We then analyzed the amount of violations again to get a narrower focus of which neighborhoods (areas) are prone to the most violations. Finally, we extracted data from the Boston Planning and Development Agency and retrieved the amount of census responses corresponding to each income level based on census tract. All income information is taken from the 2020 American Community Survey, ACS, and is recorded in inflation adjusted dollar amounts.
- Our analysis of property violations overall revealed that Census Tract ID 25025000802 had the greatest amount of violations by far, especially in 2015. It appeared that 2015 tended to be a year where most census tracts experienced spikes in the amount of property violations. Generally speaking, 2010 and 2019 appeared to have the least property violations across the board. Looking at the detailed bar graphs over time for Brighton and Allston property violations separately, out of the five tract ID's assigned to Allston, only 25025000802 experienced large amounts of violations. 25025000803 experienced almost negligible amounts of violations. The data appears skewed left for Allston, and almost a bit of a bell curve distribution for Brighton. The ACS graph reveals that the income also appears to be skewed left. An important note is that the Census tract ID can be determined by multiplying by 100 and concatenating it to 25025000. 25025000201 appears to be the lowest income neighborhood given the large amount of ACS responses corresponding to an income below \$10,000. Looking at the property violations for this tract, it does seem to be like violations are more constant over time compared to other neighborhoods which have spikes at different points in time. 2502500040, a Brighton neighborhood, has the greatest proportion of responses in the \$60,000 - \$74,999, and appears to

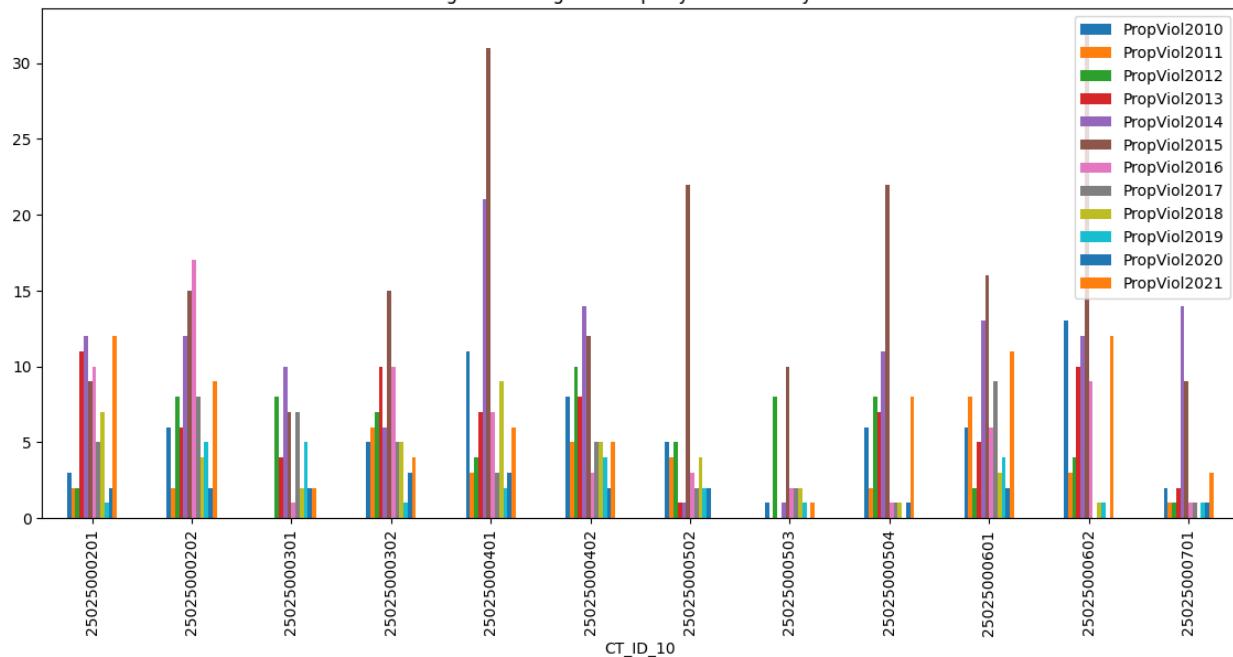
be more wealthy. It also experiences the most violations on average, especially in 2015. The data does not seem to be conclusive about whether or not violations (as a sign of a bad landlord) directly affect vulnerable groups.

- To delve more deeply into socially vulnerable groups, we plotted the population of Little to No Income individuals by ACS 2008-2012 census tracts. We found that 25025000802 consists of the greatest population of low income households by far. The graph is included below. 25025000802 has just under 3500 individuals, and tracts 2502500302 and 25025000504 have right under 2000 qualifying individuals. Tract 25025000802 also has the most property violations by far compared to any other tract.
- We also plotted various factors informing the distribution of social vulnerability across various areas in District 9. We found that 2502500401 has the highest number of older adults, defined as individuals over 65. This same tract also has the highest number of children in households. Children are expensive, and require a variety of resources. Parents and primary caregivers may have to miss work to take care of their children which can limit the total family income. This tract also had the highest number of individuals making between \$60,000 and \$74,999, as evidenced by the bar graph below. This is the sixth highest income bracket, so it can be inferred that this tract may represent a middle-class family. 2502500802 had the highest number of people of color. People of color are statistically more likely to experience poverty, gain a lower income, and have limited English proficiency. The areas with the most socially vulnerable groups appear to be this same tract according to the graph. Individuals with medical illnesses and limited English proficiency seemed to be concentrated in this tract as well. Overall, census tract 2502500802 appears to have the most impacted groups and property violations.

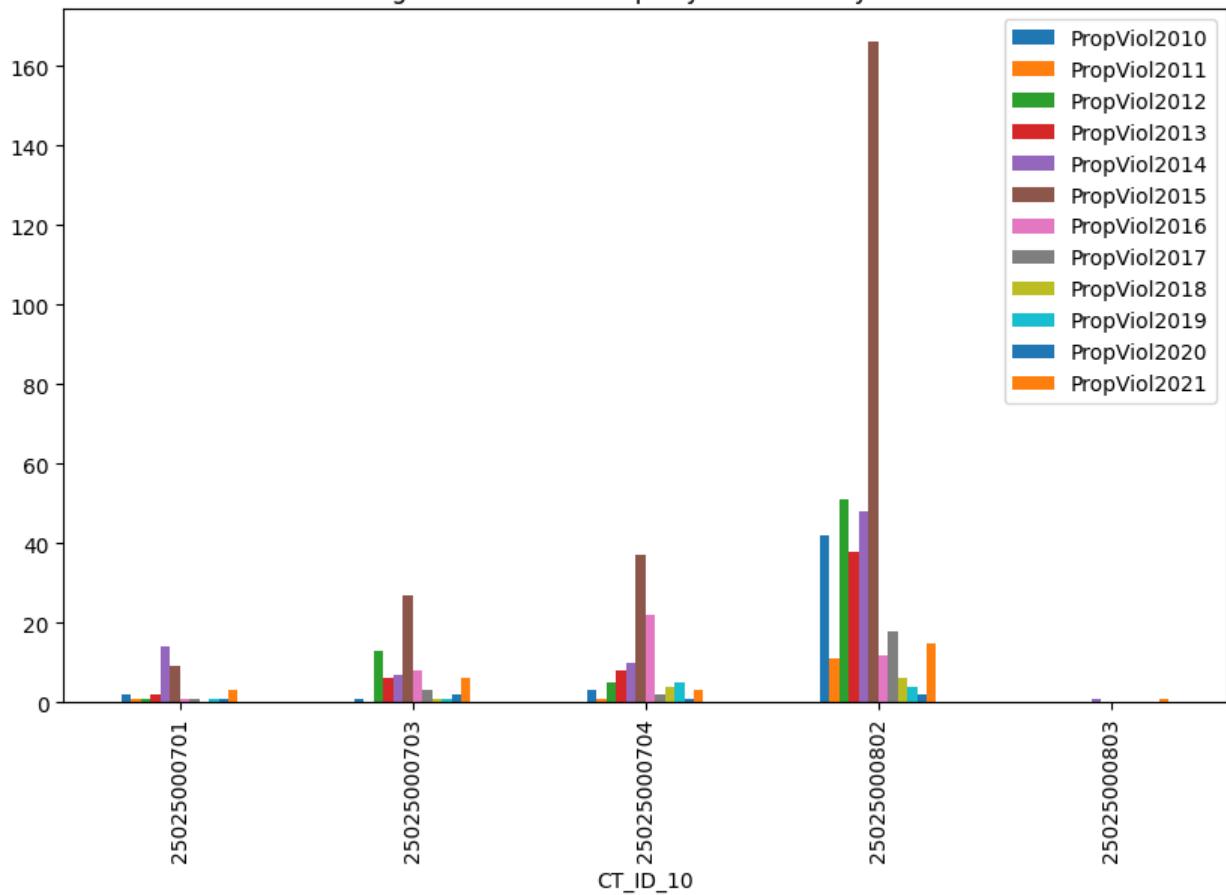
2010 through 2021 District 9 Property Violations by Census Tract



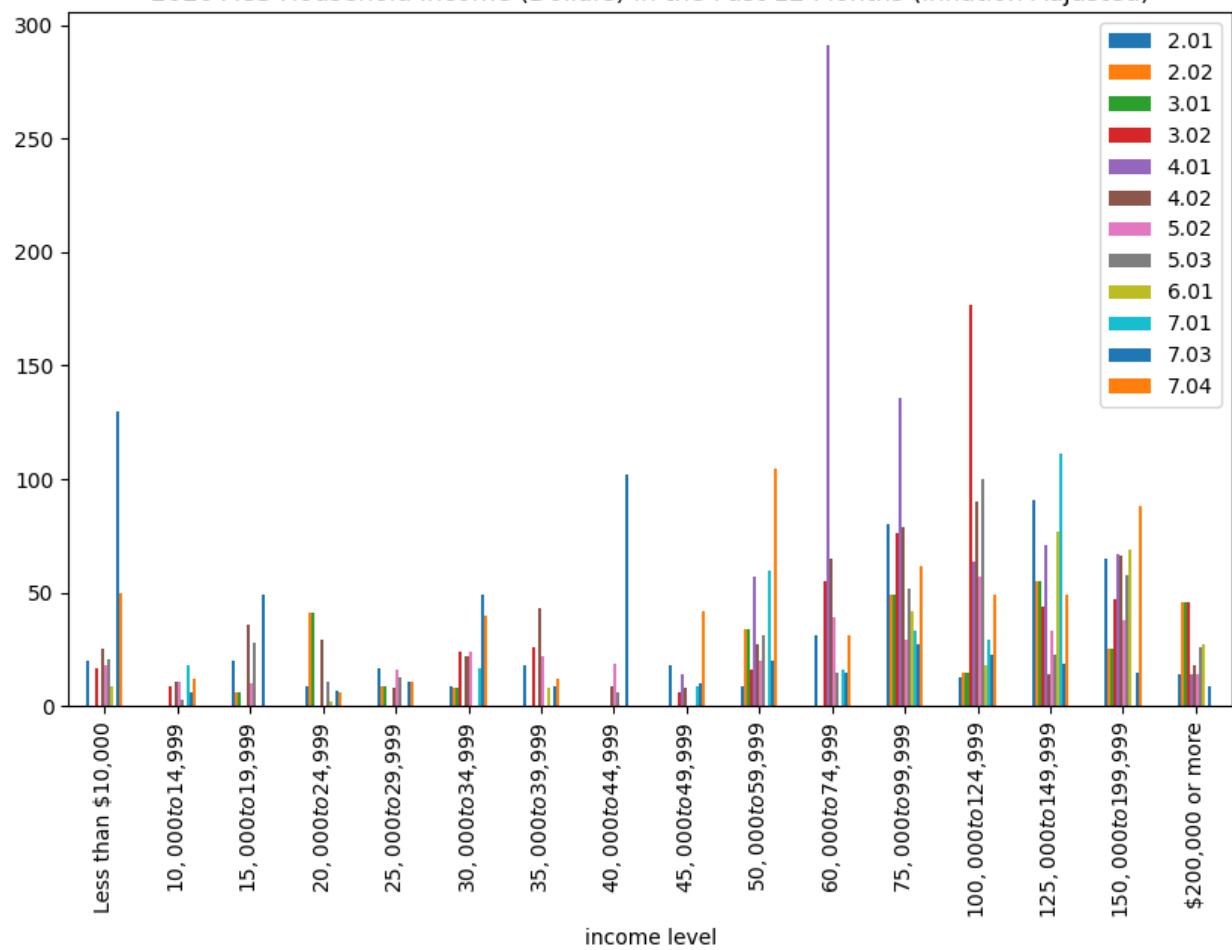
2010 through 2021 Brighton Property Violations by Census Tract

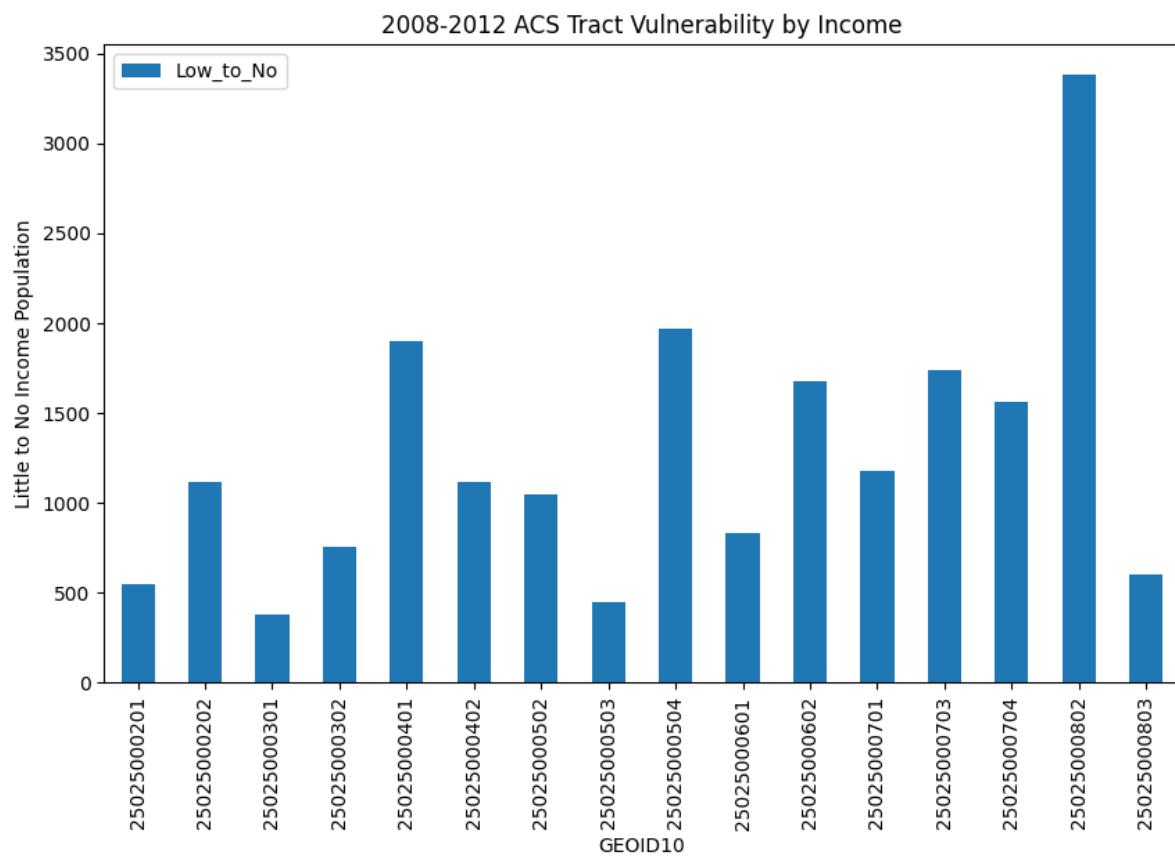


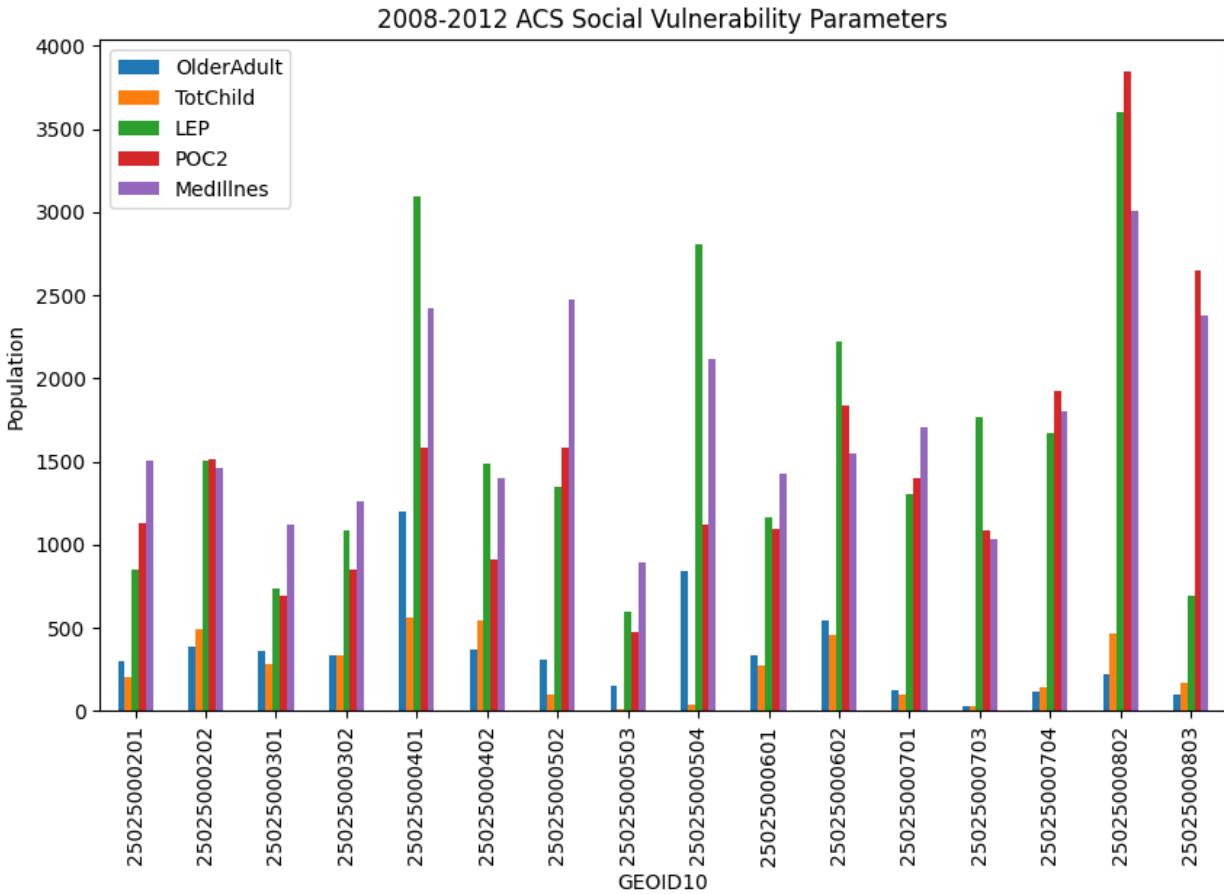
2010 through 2021 Allston Property Violations by Census Tract



2020 ACS Household Income (Dollars) in the Past 12 Months (Inflation Adjusted)







Challenges Faced:

1. Finding data related to exact rent prices. Although we had data related to property values, we could not find exact rent prices, leaving us to use
2. Defining AMI brackets for specific areas of Boston based on Zip code and being able to link this data to the property values.
3. Some of the provided data has missing information, which possibly skewed some of the results of our analysis.
4. Finding data related to property sales, and when they specifically occur.
5. Limitations involve that data had to be correlated, it was quite difficult to find datasets including information regarding property value, number of landlord violations, and zip code / tract ID. This made it extremely difficult to make any detailed conclusions and analysis, especially since data had to be pulled from multiple different sources. The time and duration also did not match up with the datasets. The graphs of property violations are over time from 2010 through 2021, whereas the median household income is from 2020. It would have been better to find a dataset with all three key pieces of information, and until then, it is hard to make a concrete conclusion. Additionally, quite a few tracts of

land had missing information for the household income in the ACS dataset, namely two neighborhoods in Allston and two in Brighton.

6. Another limitation was the lack of data available regarding only landlord violations. A large portion of data revolved around property violations, as evidenced by the ACS Census data.
7. In the future, we hope to find a more detailed dataset with key information that would be extremely useful. Data collected from the same source is often subject to the same method of retrieval, analysis, and timeframe.