

# Deliverable 1

## 1. Goals

For this deliverable, our team is working on **Task 2: Rank buildings based on hazardous criteria (mold, heat, hot water, lead, asbestos etc.)** as listed in the [Bad Landlords Project Description](#).

Since we have not yet had a client meeting, we chose our own questions related to our task to guide our preliminary analysis that we felt still addressed their questions as we understood them:

- a. Are certain buildings and neighborhoods more susceptible to the specified hazardous criteria?
- b. Do the buildings and neighborhoods that have a higher frequency of hazardous conditions also have a higher frequency of other violations?

Once we are able to discuss with the client we will expand this preliminary analysis to better address their specific questions.

## 2. Methods and Data Preprocessing

- We were provided with many datasets in the project proposal and so for this task we did not collect our own data. We chose to use the [Building and Property Violations Dataset](#) to determine the frequency of violations occurring at each building and in each city.
- There are many violations listed in the data set so we consulted the [Violation Codes Dictionary](#) to determine which should be classified as “hazardous criteria” and therefore included in our analysis.
  - This is one area in particular where we will request clarification from the client. We may need to expand our definition of “hazardous conditions” or examine other datasets.
  - Following the examples given and based on the information in the dictionary we identified 6 violations that represented hazardous conditions: **Hot Water, Potable Water, Heating Facilities Required, Temperature Requirements, Asbestos Material, and Use of Lead Paint Prohibited.**
  - We then hard coded these 6 conditions and their corresponding violation codes into a dictionary so that we could easily identify them when searching the dataset.
- We were looking at the code column to count the frequency of the pre-selected violations.
- We then looked at the const\_address1 and city columns separately to count the frequency of violations by building and by city.
  - These columns were pre-processed to change all letters to lowercase and to remove trailing or leading whitespace in order to prevent duplicates of the same location.
  - We searched for each violation separately because we had to ignore other letters that enclosed the code

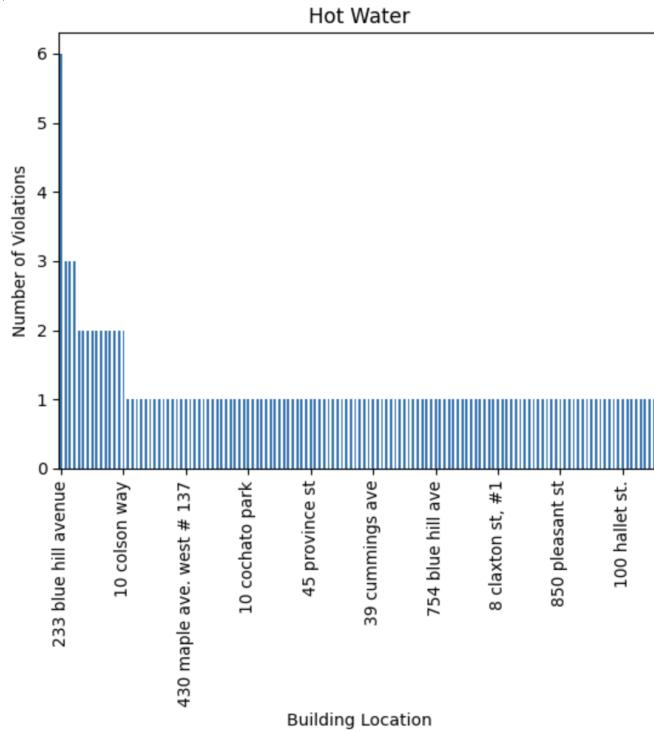
- Unfortunately we had to ignore NaN values since there is not way to fill an address without potentially skewing the data
- The counts of each violation by address were then put into a data frame with a column for each violation, location, and finally a Total for each row (representing the total number of hazardous conditions violations for that location)
  - This was repeated with the counts of violation by city (instead of street address).

	Location	Hot Water	Potable Water	Heating Facilities Required	Temperature Requirements	Asbestos Material	Use of Lead Paint Prohibited	Total
143	546 east broadway	0.0	0.0	4.0	4.0	0.0	0.0	8.0
0	233 blue hill avenue	6.0	0.0	0.0	0.0	0.0	0.0	6.0
142	20 reverand burke street	0.0	0.0	6.0	0.0	0.0	0.0	6.0
7	6 rockland av	2.0	0.0	3.0	1.0	0.0	0.0	6.0
75	40 upton st	1.0	0.0	2.0	2.0	0.0	0.0	5.0
3	320 washington st	3.0	0.0	1.0	1.0	0.0	0.0	5.0
79	30 bickford st	1.0	0.0	2.0	2.0	0.0	0.0	5.0
144	270 roxbury st	0.0	0.0	3.0	2.0	0.0	0.0	5.0
300	10 causeway st	0.0	0.0	0.0	5.0	0.0	0.0	5.0
157	53 georgia street	0.0	0.0	2.0	2.0	0.0	0.0	4.0

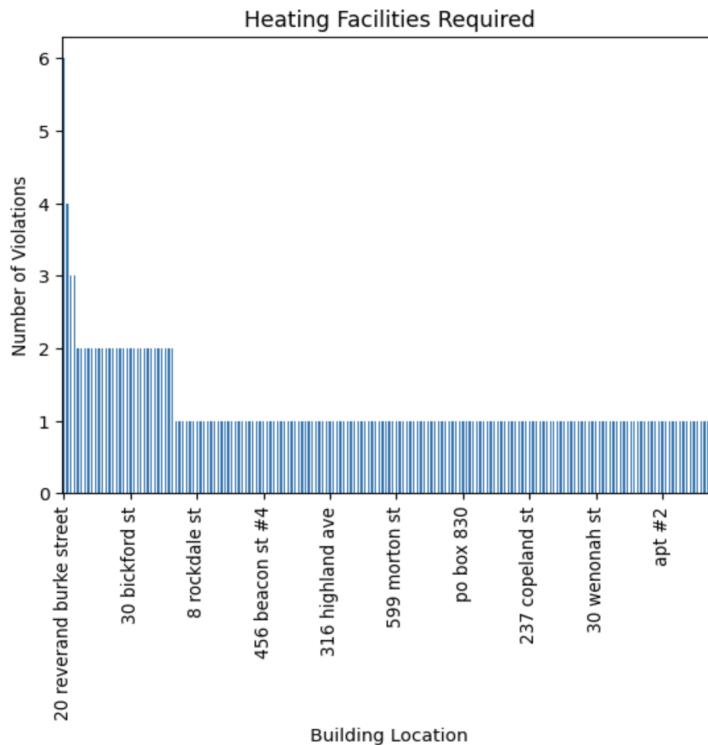
Figure 2.1: Compiled Data for Buildings

### **3. Analysis**

**Rank buildings by the number of violations based on each hazardous criteria.**



*Figure 3.1:* Number of Hot Water Violations per building



*Figure 3.2:* Number of Heating Facilities Required Violations per Building

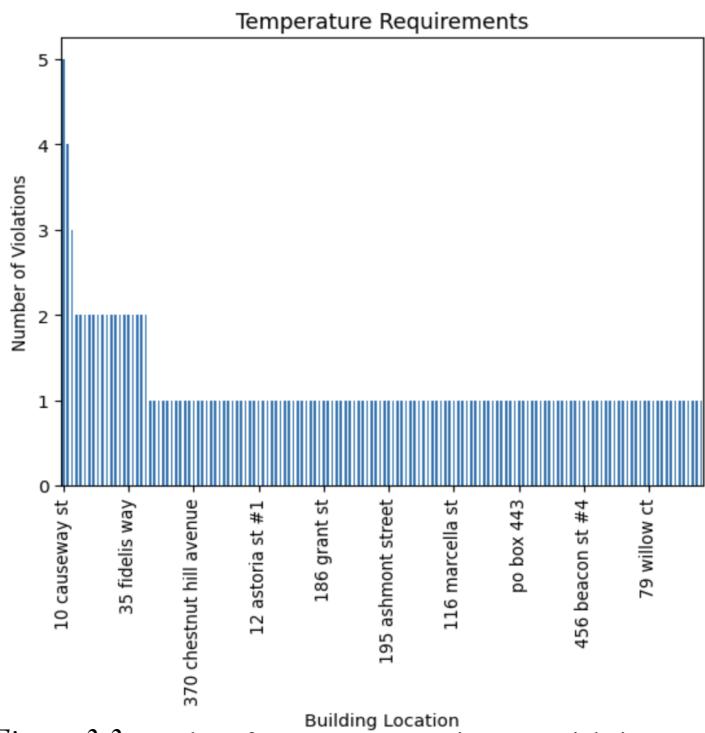


Figure 3.3: Number of Temperature Requirements Violations per Building

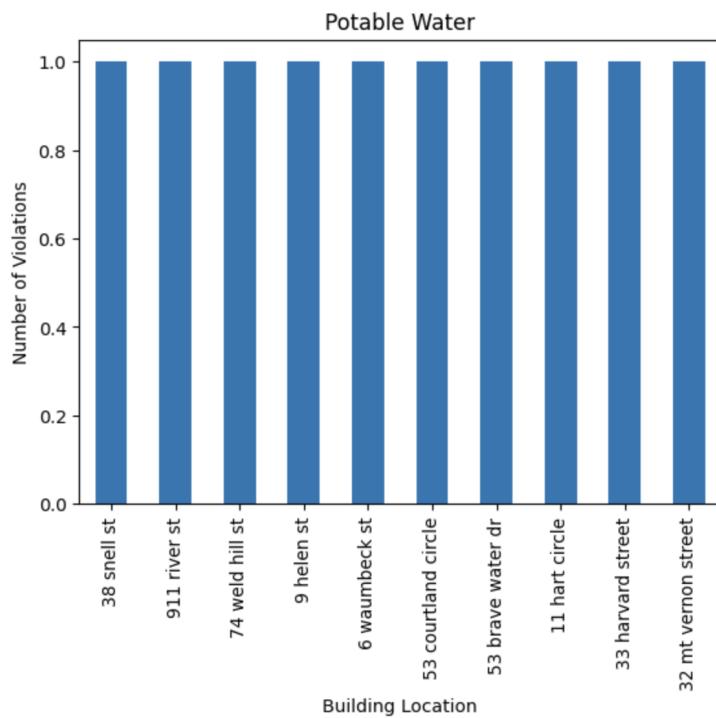
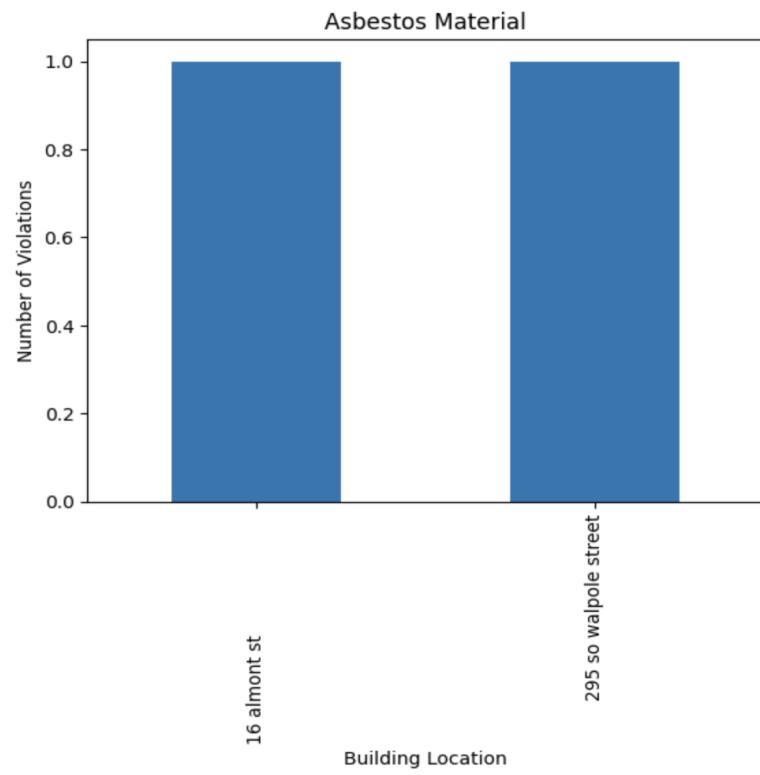
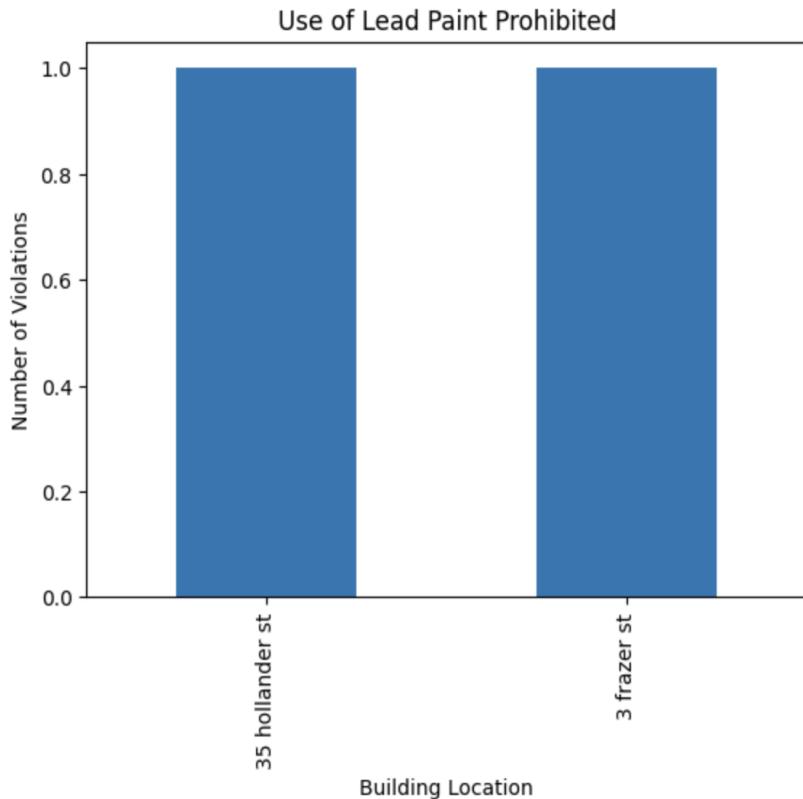


Figure 3.4: Number of Potable Water Violations per Building

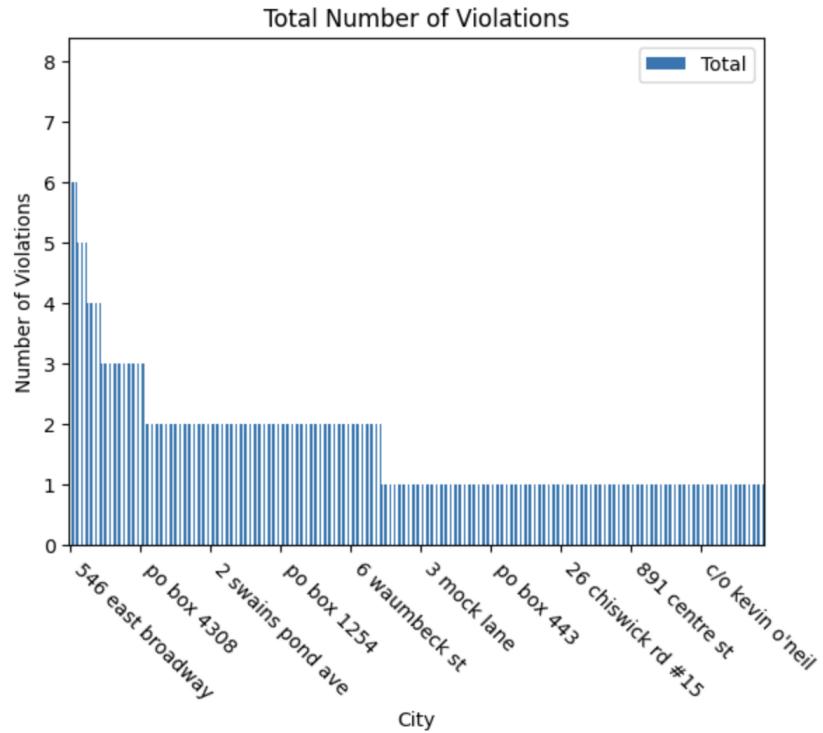


*Figure 3.5:* Number of Asbestos Material Violations per Building

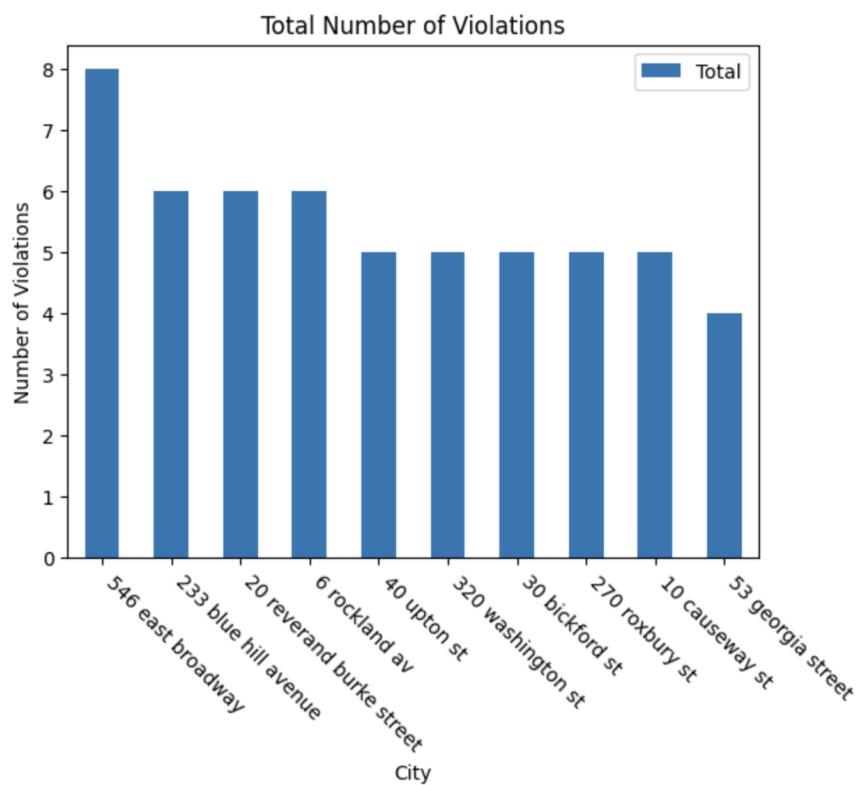


*Figure 3.6:* Number of Use of Lead Paint Prohibited Violations per Building

### **Rank Buildings by the Total Number of Hazardous Criteria Violations**



*Figure 3.7:* Total Number of Violations per Building



*Figure 3.8: Top Ten Buildings with Greatest Number of Violations*

### Rank Cities by the Total Number of Hazardous Criteria Violations

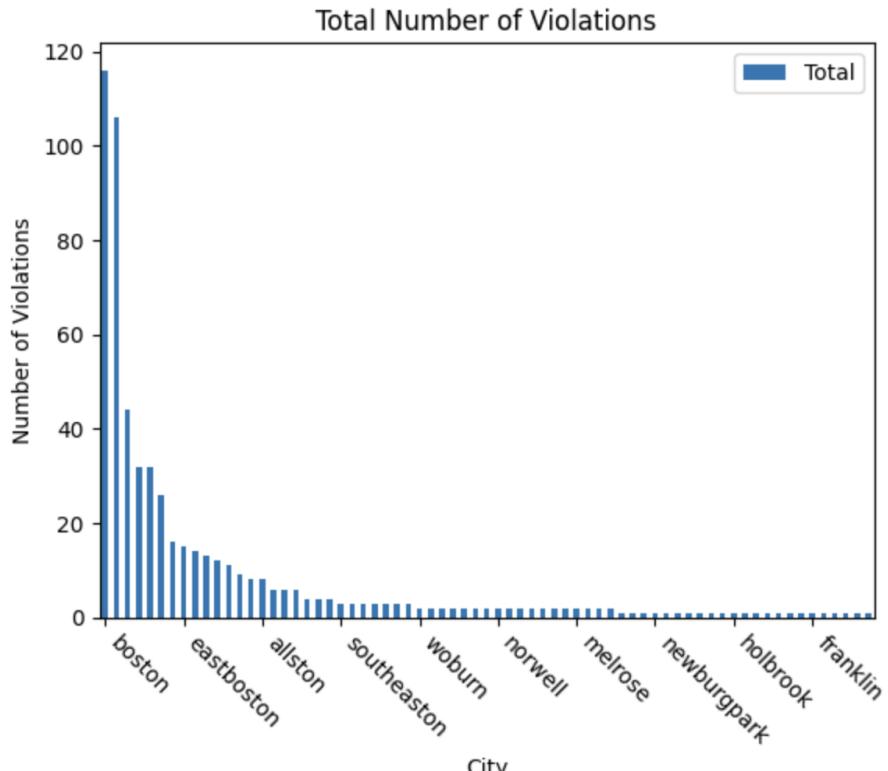


Figure 3.9: Total Number of Violations per City

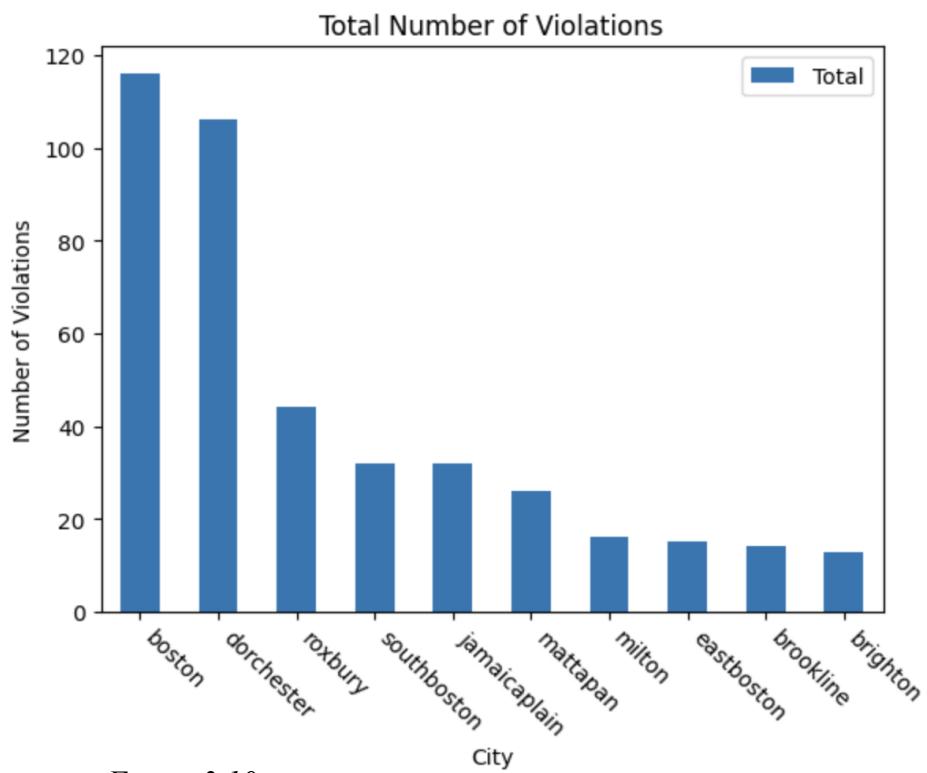
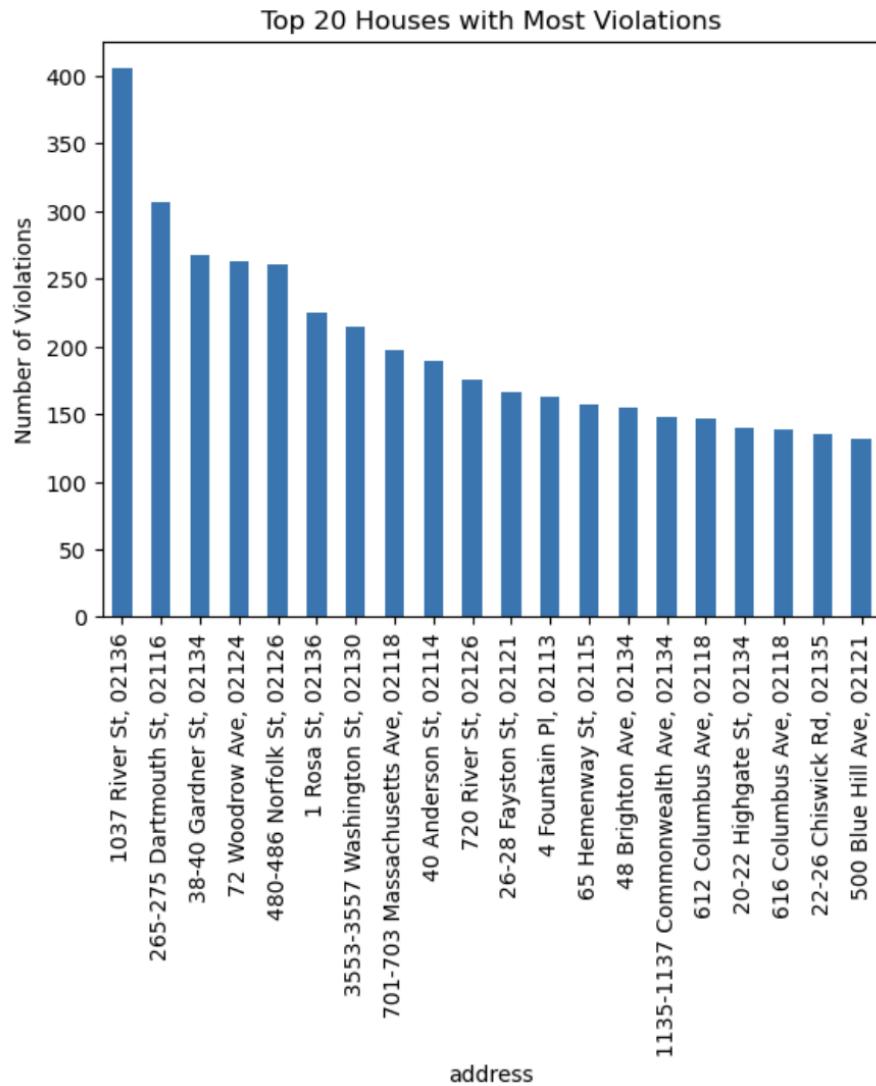


Figure 3.10: Top Ten Cities with Greatest Number of Violations

## Rank Buildings by the Total Number of All Violations (based on all types of violations)

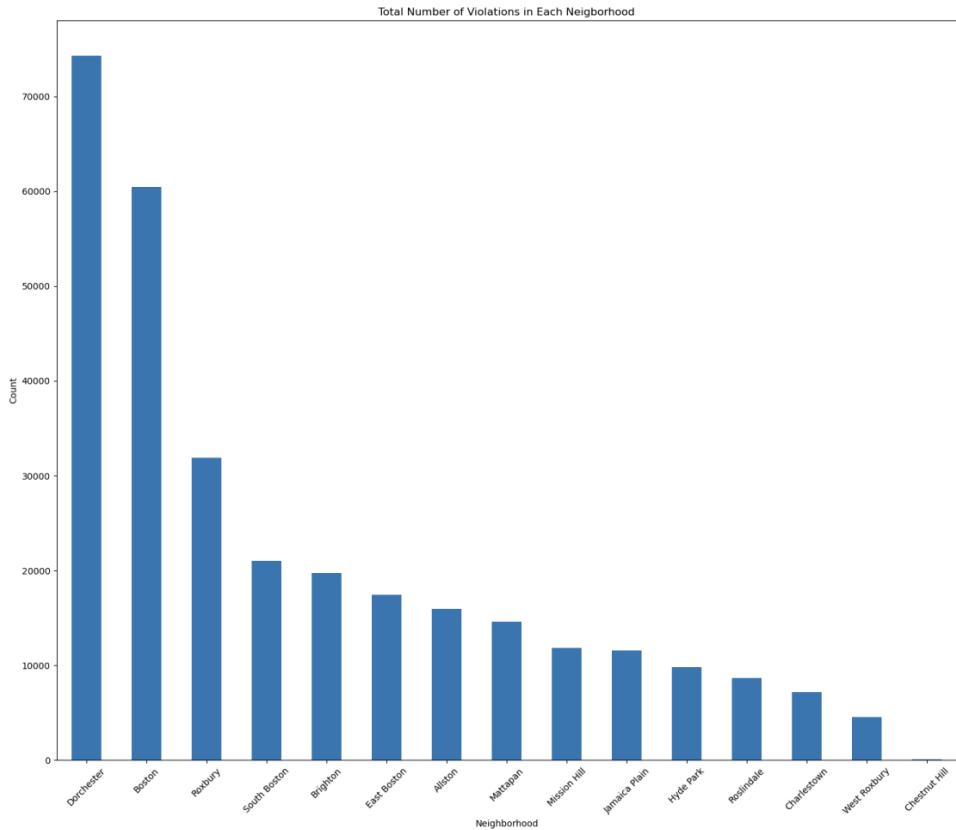
Source: Fall 2022 Team 2 Deliverable 4



*Figure 3.11: Top Twenty Buildings with Greatest Number of ALL Violations*

## Rank Cities by the Total Number of All Violations (based on all types of violations)

Source: Fall 2022 Team 2 Deliverable 4



*Figure 3.12: Total Number of Violations by City*

## Geographical distribution of all violation types in the Boston Metropolitan Area

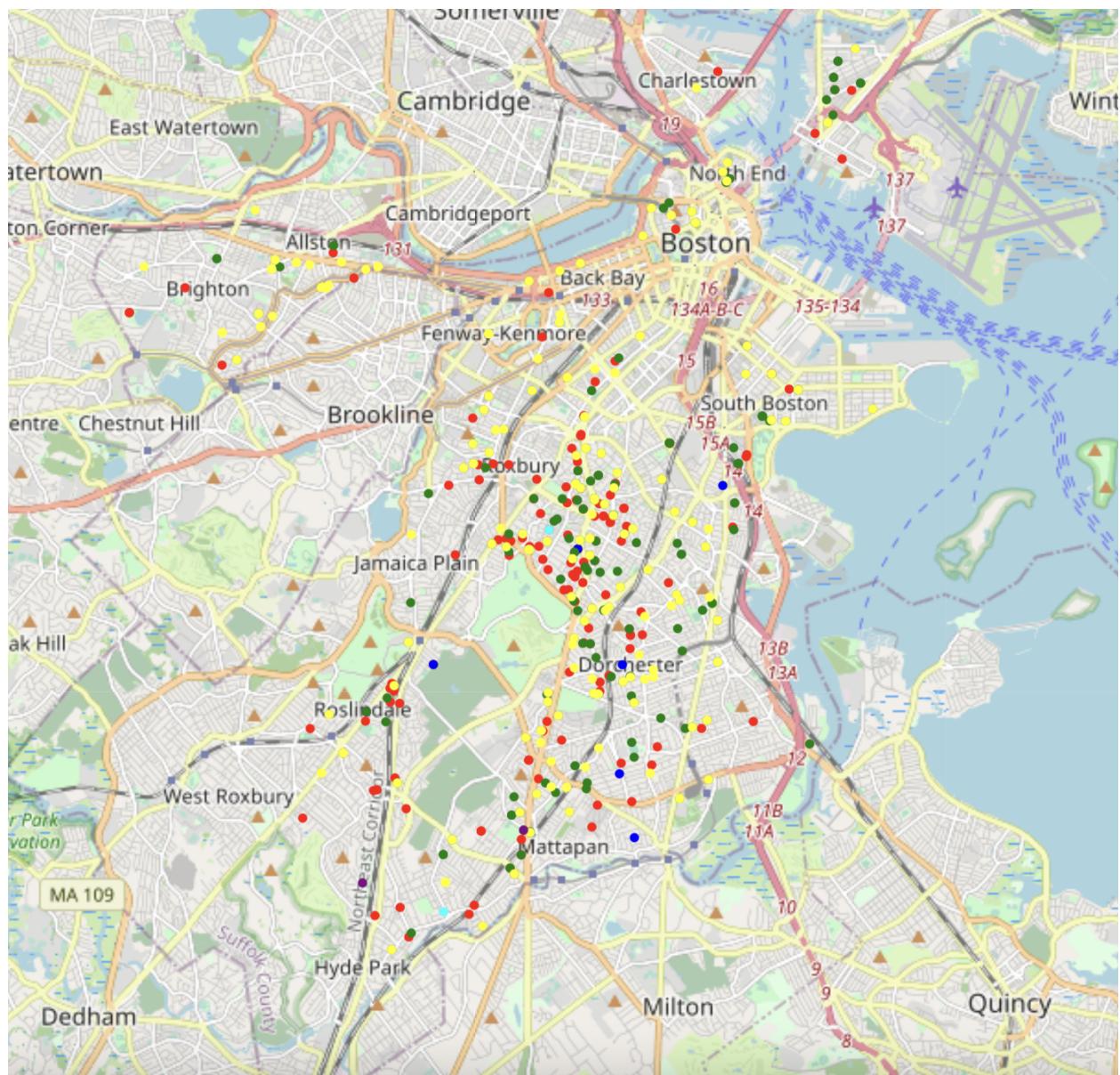


Figure 3.13: Hot Water: red; Potable Water: blue; Heating: green; Temperature: yellow; Asbestos: purple; Lead: cyan.

#### **4. Answering Questions**

##### **Are certain buildings and neighborhoods more susceptible to the specified hazardous criteria?**

From the graphs we created to rank buildings by the number of violations based on each hazardous criterion, we can see that all the buildings have at most one violation per building of potable water (*figure 3.4*), asbestos material (*figure 3.5*), and use of lead paint prohibited (*figure 3.6*). However, there are up to six violations per building of hot water (*figure 3.1*), heating facility required (*figure 3.2*), and temperature required (*figure 3.3*). It seems that several buildings are more susceptible to the hazardous criteria, including 233 Blue Hill Avenue with six "hot water" violations, 20 Reverend Burke Street with six "heat facilities required" violations, and 10 Causeway St with five "temperature requirement" violations.

We have created two more graphs to rank buildings by the number of violations based on all the hazardous criteria together (*figures 3.7 and 3.8*). As you can see from the first graph, there are numerous buildings with more than one violation, with the highest number of violations of eight at 546 East Broadway. In our second graph, we have plotted the top 10 buildings with the most hazardous violations, and we can see that there are numerous buildings with more than four violations. These buildings are definitely more susceptible to the hazardous criteria.

Moreover, we grouped all the buildings by the cities and ranked the cities by the number of violations based on all the hazardous criteria together (*figures 3.9 and 3.10*). As we can see, Boston has the most number of violations of almost 120, followed by Dorchester of almost 110. Then the number of violations per city drops off significantly with other cities in Massachusetts. It could be true that Boston and Dorchester are more susceptible to the hazardous criteria, but it could also be true that there are more buildings in Boston and Dorchester. Our group will do some further analysis to find the number of violations per city and normalize it by the number of buildings in each city.

Lastly, we created a scatter map visualizing all violation types in the spatial context (*figure 3.13*). The map result aligns closely with our bar graphs and tables which all show dense violations in Boston, Dorchester, Roxbury and their closely surrounded cities. The violation data points are more sparsely populated in the rest of the study area. The most frequent violation types in these areas are hot water (red dots), temperature requirements (yellow dots), and heating facilities (green dots). Cases of lead paint, asbestos, and potable water violation are much

less frequent. For instance, there are only two lead paint violations in the entire area, with one in Dorchester and the other in Hyde Park.

### **Do the buildings and neighborhoods that have a higher frequency of hazardous conditions also have a higher frequency of other violations?**

We don't really see a correlation between the number of hazardous conditions and the frequency of other violations by building. In Fall 2022 Team 2 Deliverable 4 (*figure 3.11*), they plotted buildings with the most violations while including all types of violations. None of the buildings on our top 10 list for hazardous violations are in the top 20 houses with the most violations.

However, we do see a correlation between the number of hazardous conditions and the frequency of other violations by city. In Fall 2022 Team 2 Deliverable 4 (*figure 3.12*), they plotted cities with the most violations while including all types of violations, and the top two cities with the most violations are Dorchester and Boston, respectively. Additionally, the third city with the most violations is Roxbury, which has significantly fewer violations than Dorchester and Boston.

## **5. Refining Project Scope/ Looking Forward**

We faced some challenges in this deliverable since we had a team member drop the class and have not yet had a client meeting. Once we are able to meet with the client we will better understand how to refine our analysis moving forward.

At this time we see several area for further exploration:

- Expanding (and refining) our definition of hazardous conditions so we can collect more data
- For cities, looking at the proportion of hazardous criteria violations so that we can better compare which cities have more violations
- Determining which criteria are most egregious (at this time, we know which ones are most common)
- Examining other data and possible cross-referencing with the analysis we have already done