

CS 506

Building Violations

Early Insights Report

Team C

Contents

1 Data Cleaning and Preprocessing	3
2 Preliminary Analysis of the Data	3
2.1 Top 20 Streets with the Highest Number of Violations	3
2.2 Streets with Repeated Violations	4
2.3 Heatmap of Violation Types	4
2.4 Data analysis results applied to the map	5
2.5 Landlords / Management with Most Violations	6
2.6 Most Affected Neighborhoods	7
3 Key Questions Answered	8
3.1 Are there certain landlords/ management companies that have repeated violations?	8
3.2 What neighborhoods / communities are affected most?	8
3.3 What kinds of building complaints are people making around the city? (partial)	8
4 Future Work	9
5 Repo Link	9

1 Data Cleaning and Preprocessing

We found that there are a large number of missing rows and missing columns in the provided data set. The existence of these noise information affects the accuracy of data processing.

1. Remove columns with substantial missing data to reduce interference from irrelevant data.
2. Delete rows with missing information to ensure all data are valid.
3. Sort the dataset to make it more structured and organized.

2 Preliminary Analysis of the Data

2.1 Top 20 Streets with the Highest Number of Violations

By counting the occurrences of `violation_street` entries, this chart ranks the top 10 streets with the highest number of reported violations. This visualization helps to pinpoint hotspots of non-compliance within the area, shedding light on where resources might be best allocated to address these issues.

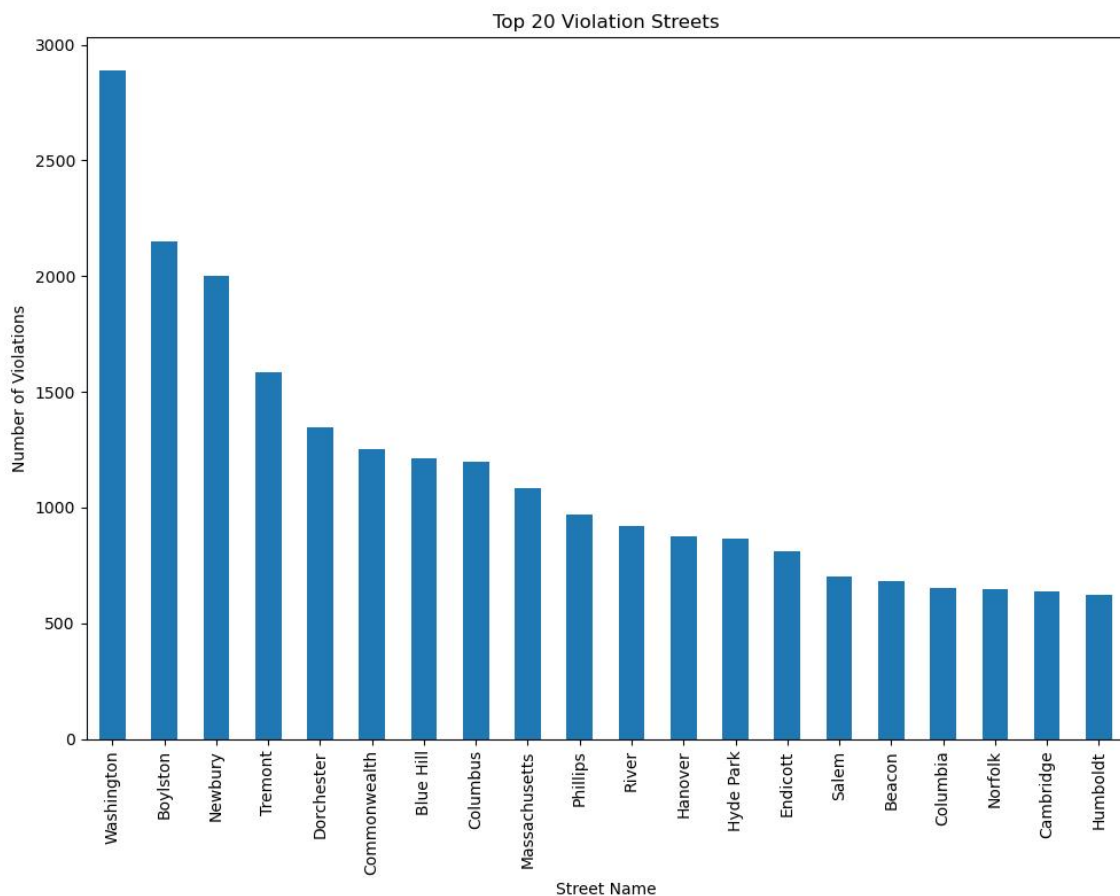


Figure 1: Top 20 Violation Streets

2.2 Streets with Repeated Violations

This visualization identifies streets with recurring violations by aggregating records based on the `sam_id`, a unique identifier for properties. It filters for `sam_id` occurrences greater than two and then counts how frequently each corresponding street is mentioned. This graph highlights streets prone to repeated violations, suggesting areas that might benefit from targeted enforcement or preventative measures.

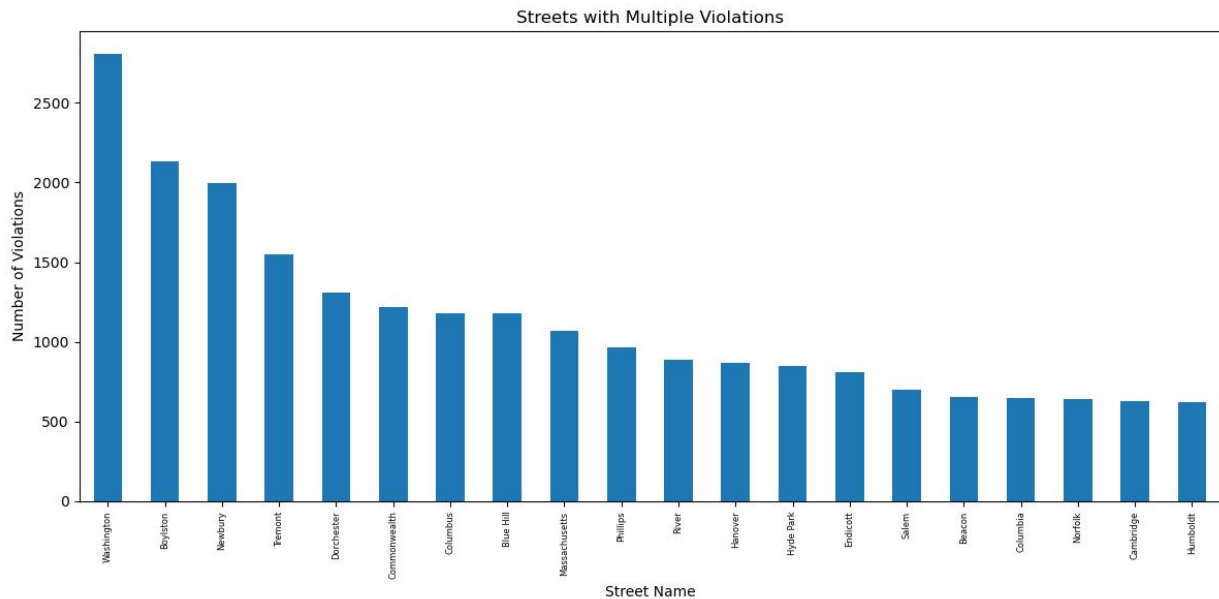
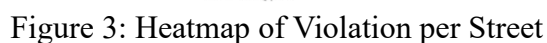


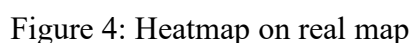
Figure 2: Streets with Repeated Violations

2.3 Heatmap of Violation Types

Utilizing the `description` field, this heat map displays the frequency of each type of violation across different streets. It reveals patterns in the prevalence of specific violations per street, providing insights into common compliance issues in certain areas. This information could guide targeted interventions or public awareness campaigns.



Based on the latitude and longitude of each record, this heatmap visualizes the concentration of violations in specific geographic areas. Areas with a higher density of violations appear more prominently, allowing for a quick visual assessment of which neighborhoods might be experiencing higher levels of non-compliance or other issues.



This map marks the exact locations of recorded violations using the geographic coordinates provided in the dataset. It offers a detailed view of where violations are occurring within the city, enabling a granular analysis of problem areas. This visualization can be particularly useful for local authorities or urban planners looking to address specific issues or improve citywide compliance strategies.

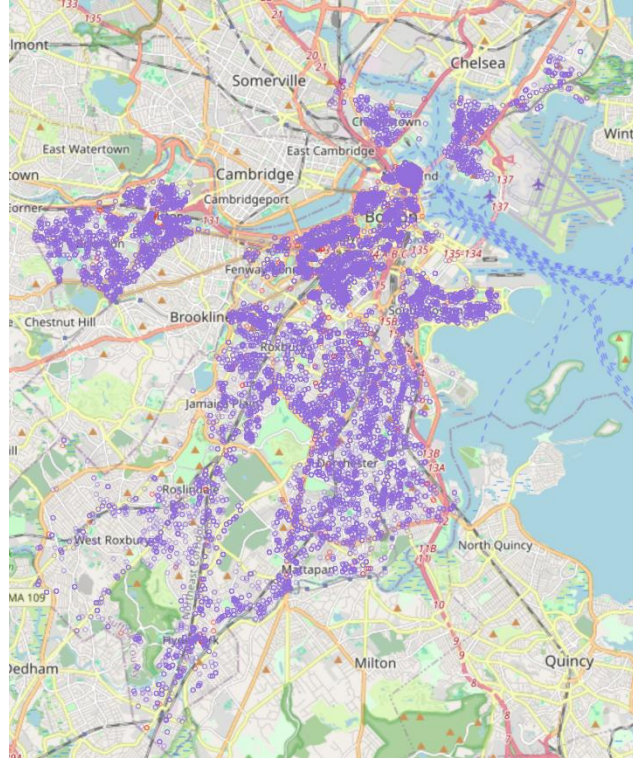


Figure 5: Boston Violations Map

2.5 Landlords / Management with Most Violations

By merging the ST_NUM and ST_NAME columns from the Property Assessment dataset, along with the violation_stno, violation_street, and violation_suffix from the Building and Property Violations dataset, we have created full_address columns in both tables. Subsequently, by matching these datasets on full_address and tallying the OWNER column in the resulting matches, we are able to determine the number of violations attributed to each landlord/management entity. This analysis is instrumental in identifying and safeguarding against landlords with a history of violations, thereby protecting potential tenants from undesirable rental situations.

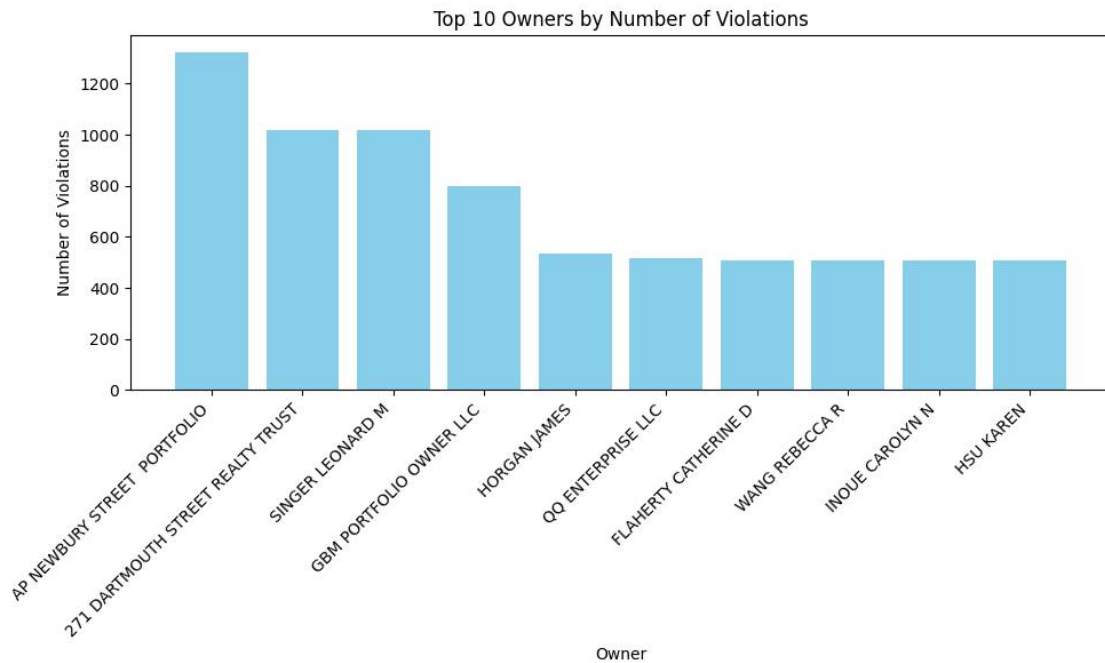


Figure 6: Owners by Number of Violations

2.6 Most Affected Neighborhoods

These two charts count the ‘neighborhood’ column in the data from 2020 onwards for the ‘Building and Property Violations’ and ‘Public Work Violations’ datasets, respectively. The goal is to identify which neighborhoods have been most severely affected by violations in recent years.

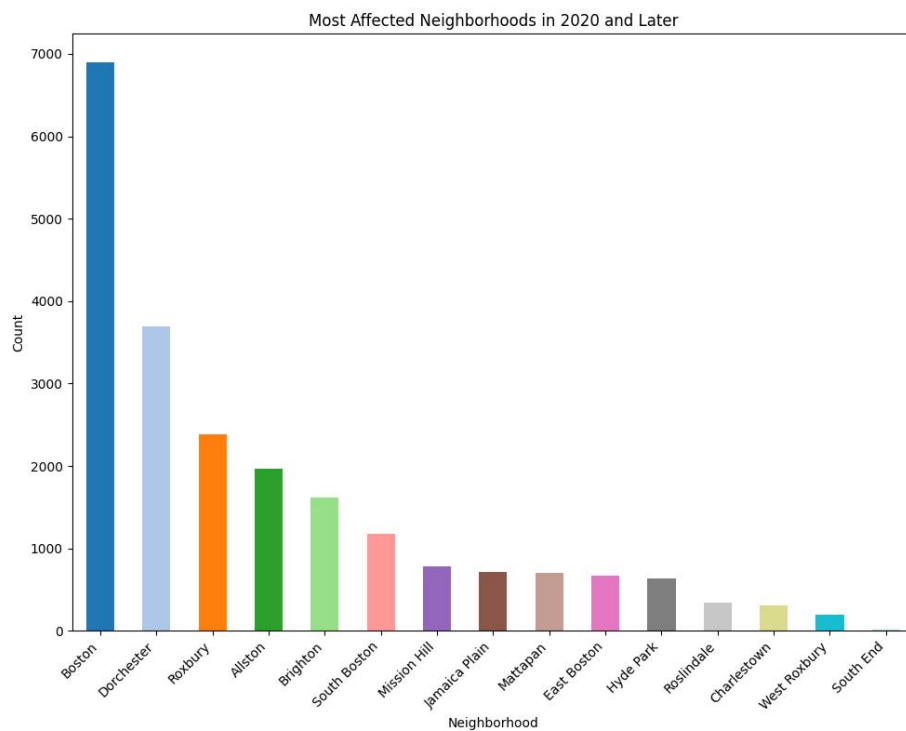


Figure 7: Most Affected Neighborhoods based on Building and Property Violations

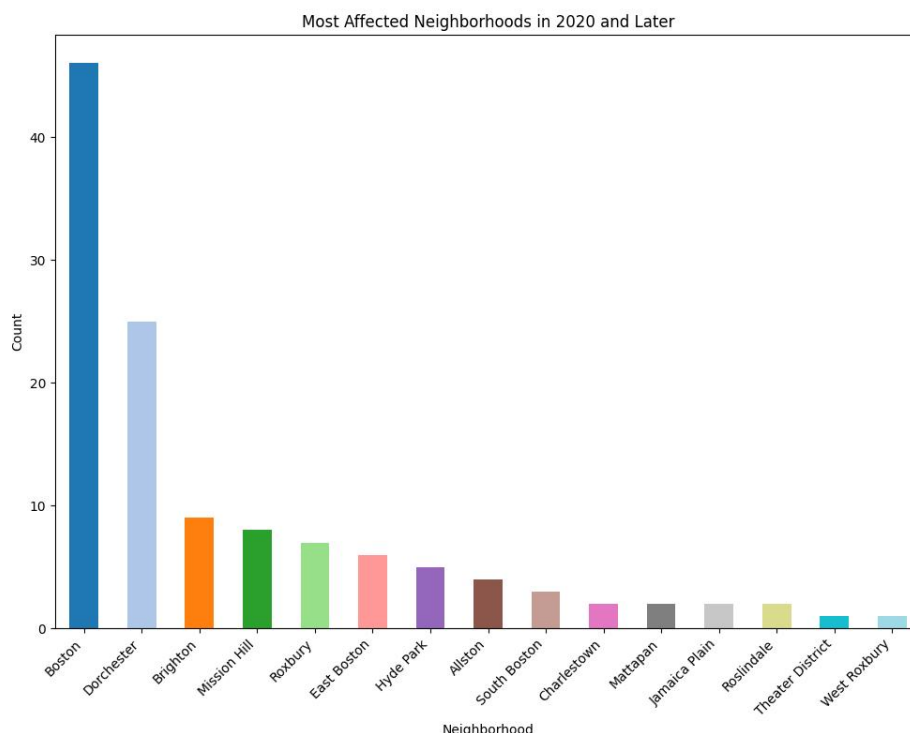


Figure 8: Most Affected Neighborhoods based on Public Work Violations

3 Key Questions Answered

3.1 Are there certain landlords/ management companies that have repeated violations?

As we can see in 2.5, there are many landlords / management companies that have repeated violations. For instance, Ap Newbury Street Portfolio, 271 Dartmouth Street Reality Trust and Singer Leonard M have significant repeated violations.

After our analysis, we believe that the number of buildings occupied by these landlords/ management companies will also significantly affect this value. I hope it can be handled reasonably in future work.

3.2 What neighborhoods / communities are affected most?

As we can see in 2.1, 2.2, 2.4 and especially 2.6, these neighborhoods / communities are affected seriously. In addition to the dendrogram, we also plotted heat maps to visualize the distribution of building violations in Boston.

3.3 What kinds of building complaints are people making around the city? (partial)

We can see the density of different types of building violations on different streets and be able to work out which violations are most prevalent in a particular area. Additionally, this report shows an abbreviated version because the full version of the heat map is too large.

4 Future Work

As of now, our project has not yet utilized the 311 datasets. However, we are poised to conduct further analysis and refinement on these datasets in the future, aiming to reduce any inherent biases and seamlessly integrate them into our ongoing work.

In the future, we will stay in touch with our client and collaborate as a team to answer additional key questions and improve existing answers.

5 Repo Link

<https://github.com/BU-Spark/ds-boston-building-violations/tree/team-c>