

Deliverable 1

Our team's task for this deliverable was to review the most common types of violations from the fall project (Task 1 from the [Bad Landlords Project Description](#)).

Since we have yet to meet with the client, we broke the task down into the following key questions:

1. What are the top 3-5 most common violation types from the fall project?
2. Intuitively, what are some factors that affect violation frequency/type?
3. Based on available data, what are some factors that affect violation frequency/type?

Preliminary Analysis of the Data

We have selected 4 datasets to begin our data analysis, with each dataset assigned to a team member:

Maria - [RentSmart](#)

Francesco - [311 Complaints](#)

Aaron - [Building and Property Violations](#)

Eddie - [Income-Restricted Housing Inventory](#)

Answering Key Questions

1. What are the top 3-5 most common violation types from the fall project?

After consulting some of the deliverables from the Fall semester, we concluded that the most common violation types they outlined were Enforcement violations, followed by Housing complaints and Sanitization requests, with most violations classified in the 'low severity' category. (Source: [team 2 deliverable 4](#), [team 4 deliverable 5](#))

2. Intuitively, what are some factors that affect violation frequency/type?

It makes sense that the most common violations would be of low severity, since they are easier to report and occur more often than major issues/violations.

Enforcement violations (most common type) include the following subcategories:

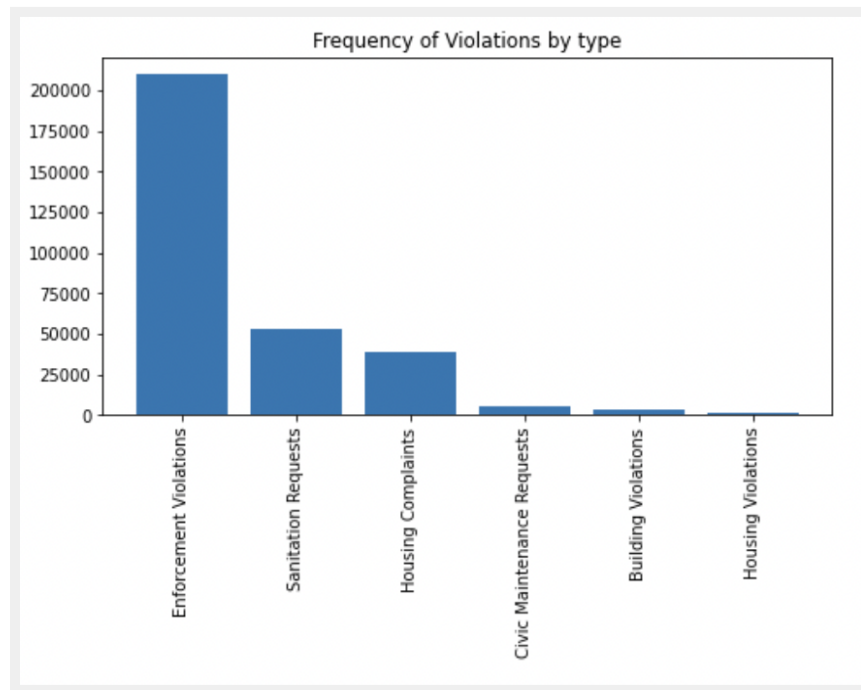
- Improper Storage Trash
- Unregistered Motor Vehicles
- Failure to Obtain Inspection
- Failed Multiple Rental Inspections
- Overgrown Weeds on the Property
- No Number on the Building
- Overfilling of a Dumpster

These violations follow laws enforced by the city, thus are more heavily regulated, so it makes sense for them to be more common than violations filed by renters.

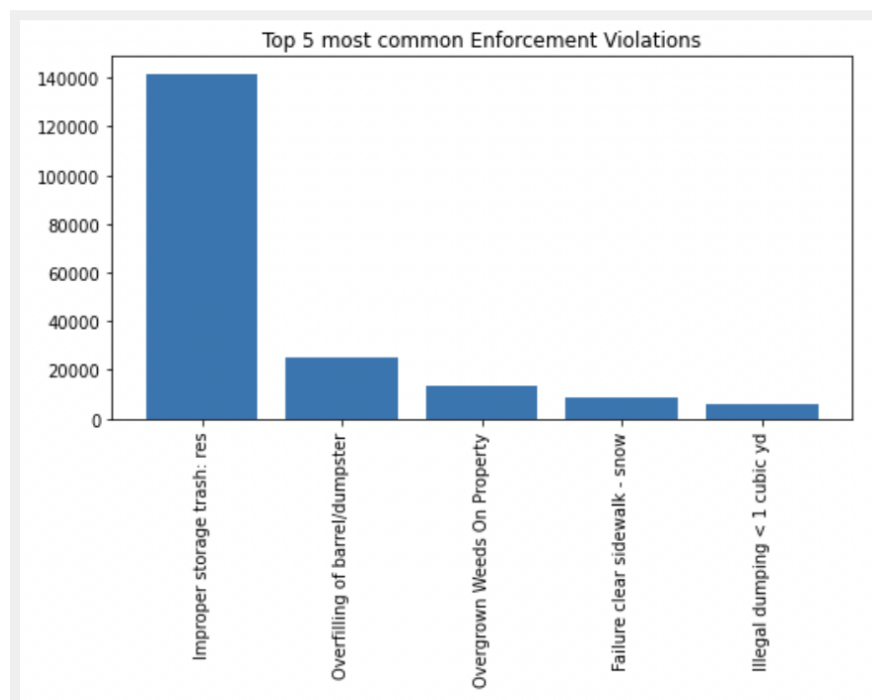
3. Based on available data, what are some factors that affect violation frequency/type?

Based on our preliminary analysis, we were able to draw the following conclusions:

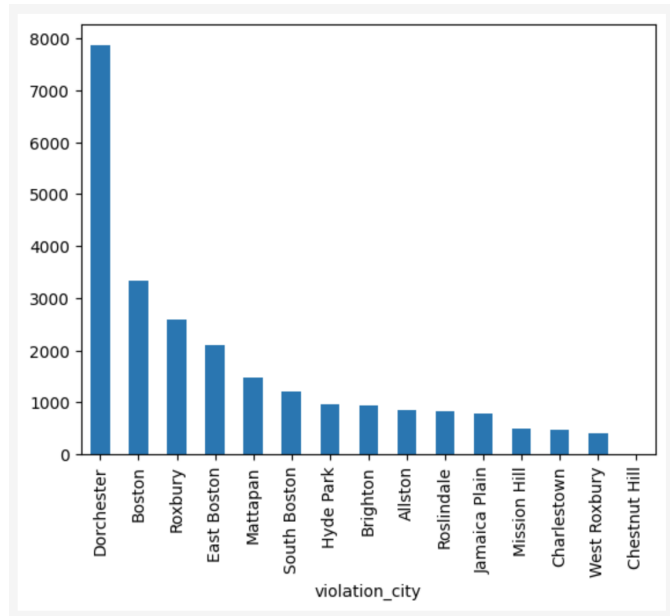
Our findings from the [RentSmart](#) dataset analysis agree with those of team 4 (deliverable 5). We found Enforcement Violations to be the most common type of violation, followed by Sanitation requests and Housing Complaints.



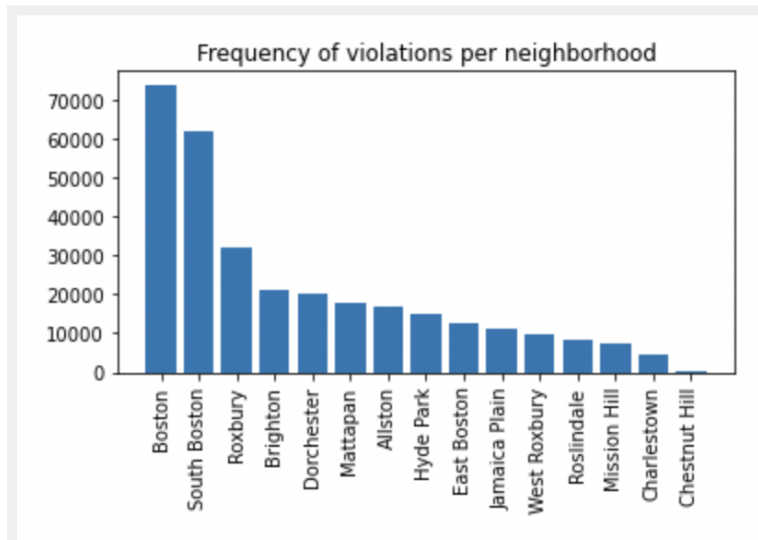
Within the Enforcement Violations, the most common violation is for improper trash storage.



We can confirm with [Team 1's findings](#) from Fall 2022 that Dorchester is the city with the highest number of violations from the [Building and Property Violations](#) dataset. Others seem to include Roxbury, East Boston, etc; however, their counts were much lower than Dorchester.

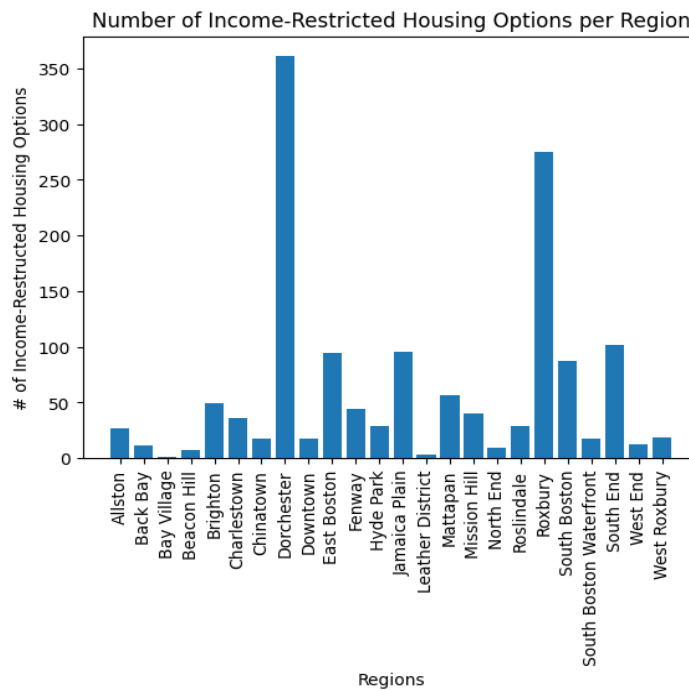


The results from the [RentSmart](#) dataset, however, were very different, with Boston having the most violations, followed by South Boston and Roxbury. There are several reasons that this may be the case, first off, these datasets explore different types of violations from one another (with RentSmart geared more specifically towards rental properties), and second, the RentSmart dataset used had more recent data (starting from 2016).



Income-Restricted Housing

According to the [Income-Restricted Housing Inventory](#) dataset, which tracks all housing in Boston reserved for households earning at or below a certain income, the majority of income-restricted housing in Boston is found in Dorchester and Roxbury, with Dorchester taking the lead. This correlates with the findings found above from the Building and Property Violations dataset, which lists Dorchester as the region where the most violations occurred. Judging by this, is possible that income-restricted housing is a factor that affects violation frequency. This could be due to “bad landlords” finding such households easy targets to slap violations on, though further analysis of the data will be required before any conclusions can be certainly drawn.



Common Violations

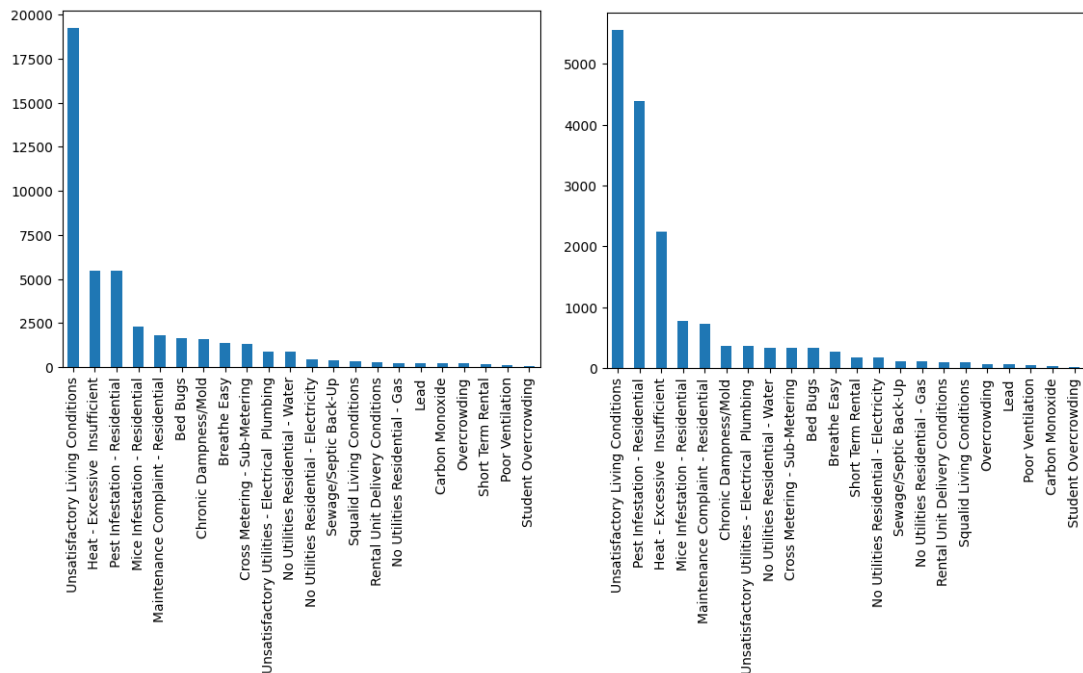
According to the [Building and Property Violations](#) dataset, the most common violation for fitness of “human habitation” by the Department of Public Health is violation CMR410.500, “Owner's Responsibility to Maintain Structural Elements;” it is followed by CMR410.351 “Owner's Installation and Maintenance Responsibilities.” Other notable ones are CMR410.550 and CMR410.482, which are “Extermination of Insects, Rodents and Skunks” and “Smoke Detectors and Carbon Monoxide Alarms” respectively. Because the dataset focuses on buildings and property, it makes sense that maintenance takes up a large number of violations, especially in lower-income communities.

	code	case_count
58	CMR410.500	2472
30	CMR410.351 A	1379
70	CMR410.550 B	935
53	CMR410.482 A	794
52	CMR410.481	263
9	CMR410.200 A	214
82	CMR410.602 D	190
11	CMR410.201	174
8	CMR410.190	157
59	CMR410.501 A	149
45	CMR410.451	131
79	CMR410.602 A	125
31	CMR410.351 B	117
69	CMR410.550 A	113
47	CMR410.480 A	109
71	CMR410.551	106
60	CMR410.501 B	87
63	CMR410.503 A	77
5	CMR410.150 D	57
72	CMR410.552	55

311

From the 311 service requests' dataset it is possible to obtain two slightly different ranks of most common complaints, if fact it depends on how strict is the filter applied to the data; for each entry is provided a "closure_reason" field in which there is sometimes inserted interesting information, such as the actual presence of a violation or the reason for which that entry was invalid. On the other hand often this field is empty. Trying to reconduct this field to a reasonable set of closure status the data can filter out all the non-violations.

But over 55% of the entries have no reason, so we hereby compare the results including (on the left) or excluding (on the right) the unknown closure reasons.



The main difference is on the absolute numbers, since the top 5 violations of both the ranks are the same, with different order.

Affordable renting

We tried to compare the data from the 311 requests' dataset with the locations of the affordable renting units to see if there were any obvious correlations between them but from the mere map it doesn't seem so.

In the following map the violations (as markers) and the units position (as heatmap).

