

---

# Deliverable 4: Small Landlord Analysis

---

**Ciaran Hikaru Ueda Fitzgerald**  
Boston University  
cueda@bu.edu

**Haoran Hu**  
Boston University  
hhran@bu.edu

**Daniel Scrivener**  
Boston University  
dscriv@bu.edu

**Julie Le**  
Boston University  
tjle@bu.edu

## 1 Introduction

### 1.1 Motivation

Compiling and interpreting information on small landlords who have opted into Boston's affordable housing program provides renters with a valuable source of information regarding where to seek housing. Our project makes use of available municipal data concerning rented properties in the Greater Boston area. We have organized this data so that owner-occupied properties — as well as properties belonging to non-commercial entities — are properly accounted for in an overall assessment of Boston's affordable housing stock.

The importance of this project is evident in terms of its impact on the broader community: the city of Boston seeks to widen the array of affordable housing options available to tenants. This proves beneficial not only to tenants, but to landlords as well in terms of their ability to effectively advertise available units.

### 1.2 Key Questions Addressed

1. What is the current distribution of non-commercial landlords *not* currently enrolled in different affordable housing programs?
  - (a) # of units
  - (b) Geographic distribution by neighborhood
2. What is the current distribution of non-commercial landlords currently enrolled in different affordable housing programs?
3. What percentage of the total housing stock is rented out by small landlords (non-commercial), and what percentage of affordable properties?

### 1.3 Extension Project: Problem Properties

1. What is the density of property violations by neighborhood?
2. Which non-commercial units that are *not* in affordable housing programs have recorded property violations?

## 2 Methods

### 2.1 Generating Distributions

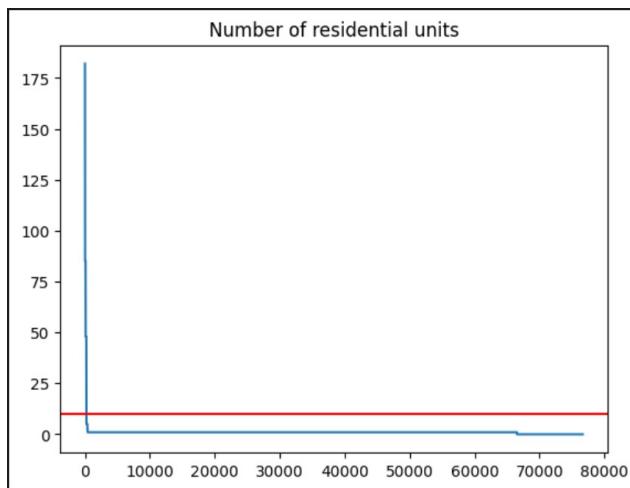
From our meetings, we decided that grouping properties by neighborhood is the best way to visualize the distribution. It's also very important to consider only the properties which are "non-commercial", so we needed to establish what qualifies as a non-commercial property.

#### 2.1.1 Filtering Out Commercial Units

The following analysis explains how we came to define "non-commercial" to mean **properties with a non-commercial tenure, also consisting of 10 or fewer units**.

In order to get an accurate picture of non-commercial units, we filter out large-scale rental properties with many units. To decide where to make the cutoff, we examined the distribution of the number of units in each dataset.

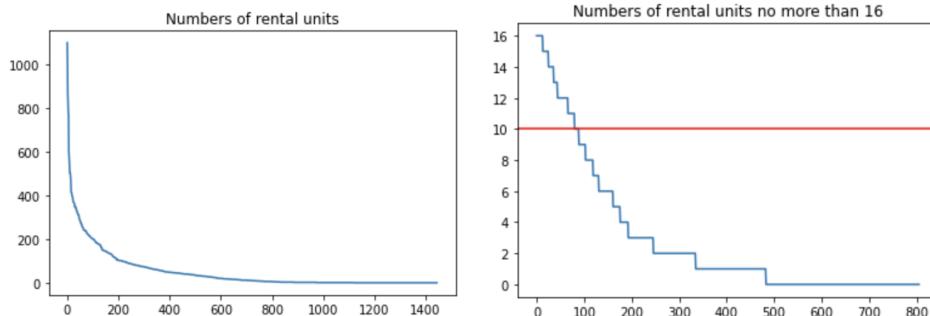
Figure 1: Plot of the number of rental units in each property. Taken from the Analyze Boston Property Assessment FY2022.



As the figure above illustrates, properties with more than 10 units are likely to be large-scale commercial properties. Reassuringly, this finding aligns with the cutoff instituted by Boston's Inclusionary Development Policy, which requires properties with 11 or more units to designate a certain percentage of their units as affordable housing; 10 units serves as a natural threshold.

The following figures show a similar trend in the income-restricted housing:

Figure 2: Plots of the number of rental units in each property. Right plot is a filtered version of the first. Taken from the Analyze Boston Income-restricted Housing Inventory 2021.



As such, we excluded all properties with 0 rental units, and more than 10 rental units from both datasets.

Our only concern is that this rule may not be strict enough. We discussed that there could still be some large-scale landlords which consist of multiple, small properties which were not dropped by our definition. Future work may want to look into identifying these landlords and excluding them as well.

### 2.1.2 Geocoding Properties from the Main Dataset

All properties within Boston’s main housing inventory were assigned coordinates using geocoding functions available within the Google Cloud suite of APIs. Only properties that remained after filtering by non-commercial units were geocoded (66293 total). Google cloud was successful in obtaining coordinates for the vast majority of these properties — approximately 20 addresses could not be located.

Geocoded properties from the main dataset were then merged with any corresponding property violation records. Coordinates were used as the join attribute (with slightly reduced precision to account for slight discrepancies between different geocoding strategies).

## 3 Results

### 3.1 Non-commercial housing percentages

This addresses **key question 3** of the base project.

Of the entire housing stock,  $\frac{66293}{178598} = 37\%$  of units are rented out by non-commercial landlords.

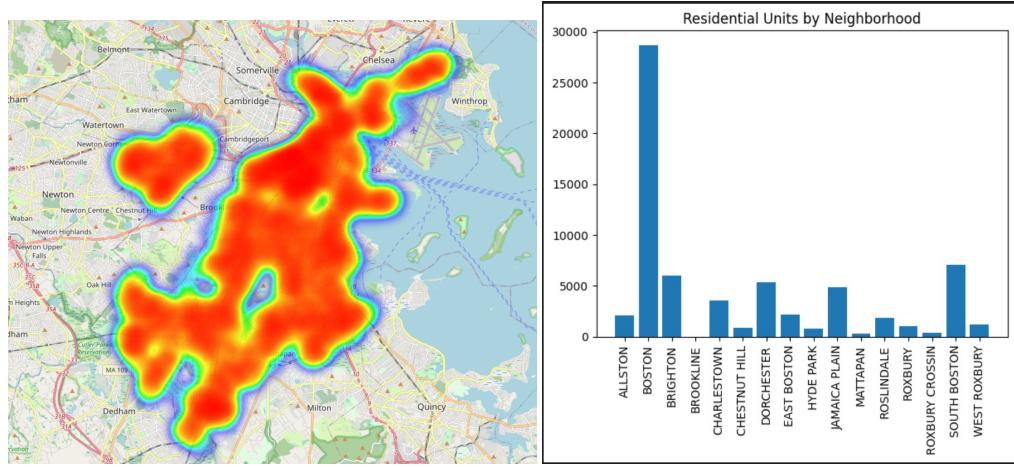
Of the affordable properties,  $\frac{403}{1445} = 28\%$  of units are rented out by non-commercial landlords.

### 3.2 Distribution of non-commercial rental units

This addresses **key question 1** of the base project.

We plotted the distribution of non-commercial units by neighborhood:

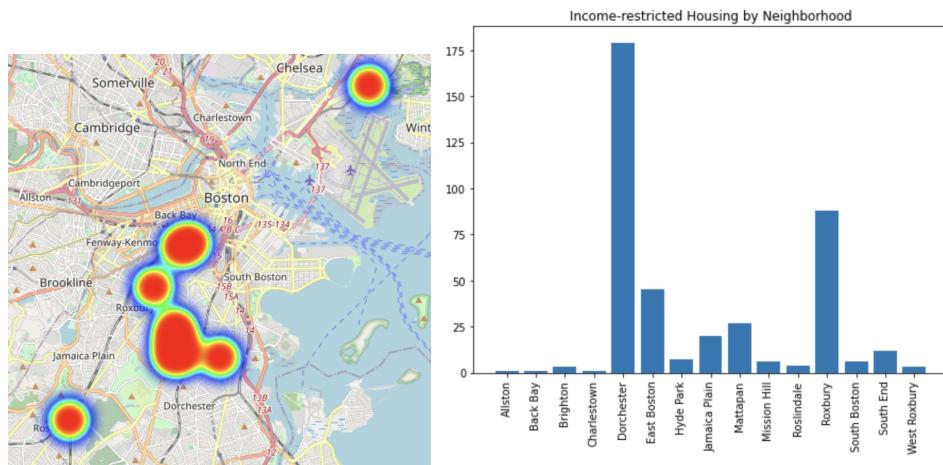
Figure 3: Distribution of all non-commercial rental units. Taken from the Analyze Boston Property Assessment FY2022.



### 3.3 Distribution of income-restricted units

This addresses **key question 2** of the base project.

Figure 4: Distribution of non-commercial, income-restricted rental units. Taken from the Analyze Boston Income-restricted Housing Inventory 2021.

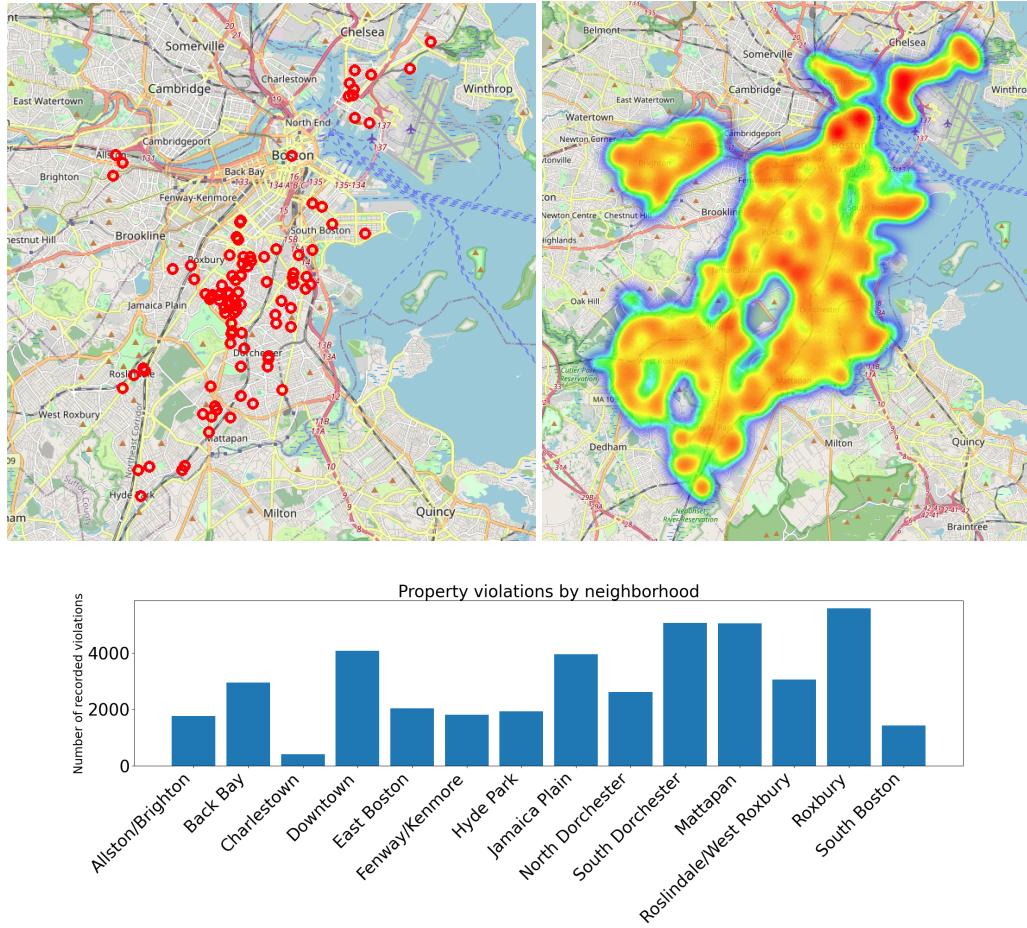


<sup>1</sup>Note: not all properties were geocoded properly for the income-restricted housing heatmap

### 3.4 Property violations

This addresses **key question 1** of the extension project.

Figure 5: Distribution of "problem" properties with past violations. From left to right: Top 100 properties with the most repeat violations, heatmap (density) of violations, bar chart of violations per neighborhood. Taken from Analyze Boston Building and Property Violations.



In terms of property violations, many of the top repeat offenders are located in Roxbury; others are scattered throughout Dorchester, East Boston, Hyde Park, and South Boston. This aligns with the general overview provided by our graph of violations per neighborhood, in which Roxbury, Mattapan, and South Dorchester appear to be the largest contributors of property violations.

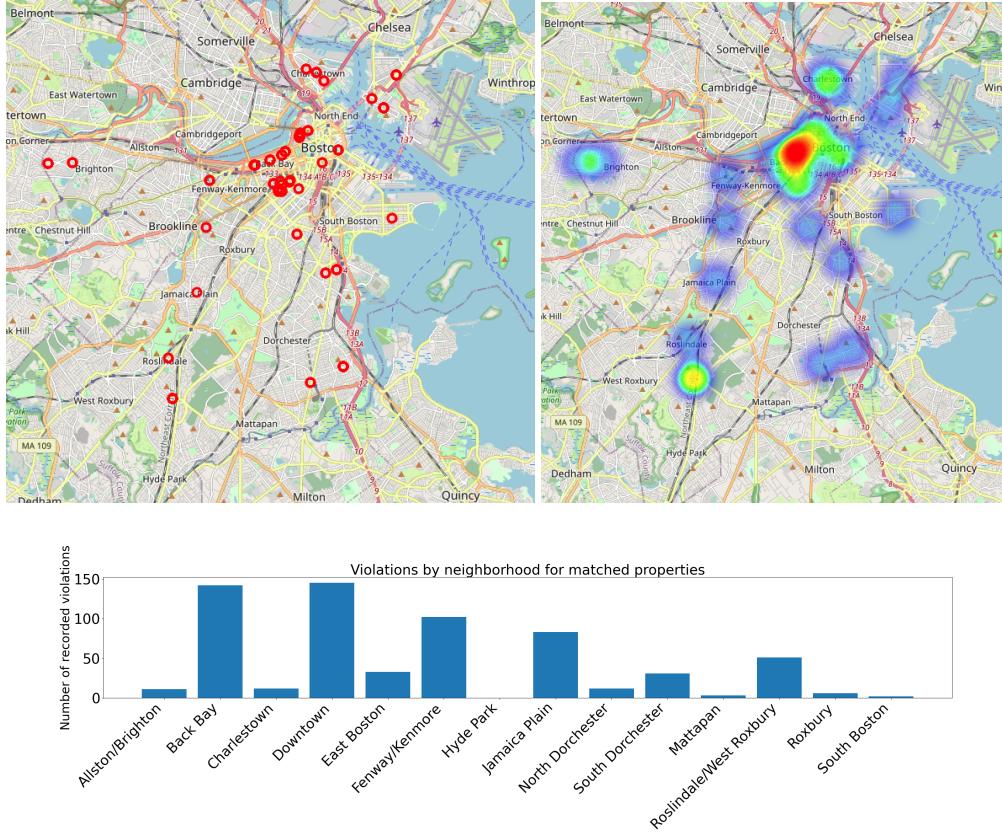
Interestingly, although the relative majority of property violations occur in Roxbury, the densest parts on the heatmap are Downtown Boston and East Boston. It stands to reason that these might be considered potential "hot spots" of problematic landlords.

### 3.5 Non-commercial property violations

This addresses **key question 2** of the extension project.

Of the 66293 non-commercial properties, 296 also appeared in the property violations dataset.

Figure 6: Distribution of **non-commercial** "problem" properties. From left to right: exact location, heatmap (density), count of non-commercial properties with violations by neighborhood. Taken from Analyze Boston Building and Property Violations.



The heatmap displays a significant hotspot spanning downtown and Back Bay. Smaller clusters are located in surrounding areas such as Fenway or Boston's South End.

### 3.6 Next steps

Further analysis should focus on the extent to which units from the main dataset can be identified as strong candidates for Boston's affordable housing program — it might be useful to isolate the characteristics of units that were matched to the dataset of property violations.

It's also worth verifying that the owners of these units don't live in them themselves.

## 4 Challenges

The data with which we are working is liable to change constantly as landlords continue to make daily decisions regarding their housing stock. Much of the available data is collected in bulk and captures the state of the housing market at particular moments in time. Additionally, guarantees about the quality of the datasets have been uncertain; the data collection practices themselves may be subject to bias. For example, we imagine that small landlords are likely to be less rigorous in their reporting practices than larger commercial entities, further complicating data collection.

In particular, incomplete data (units missing geographic coordinates) has proved to be the biggest obstacle thus far.