Deliverable 1: Preliminary Report

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

GitHub Repo: https://github.com/BU-Spark/ds-councilor-louijeune-small-landlord

1 Key question

After pre-processing the data, we performed a preliminary analysis by answering the first 2 key questions regarding the current distribution of landlords. We plotted the units with are/are not in our list of affordable housing as a heatmap, and also plotted a histogram of these units by zip code.

1.1 Distribution of all landlords

Figure 1: Partial plot of registered properties in the Boston area. Taken from the Analyze Boston Property Assessment FY2022.

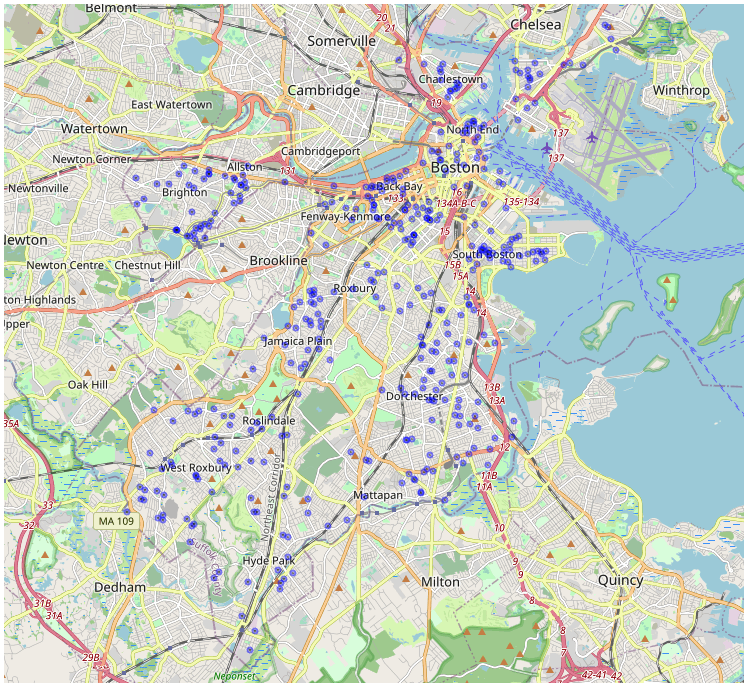
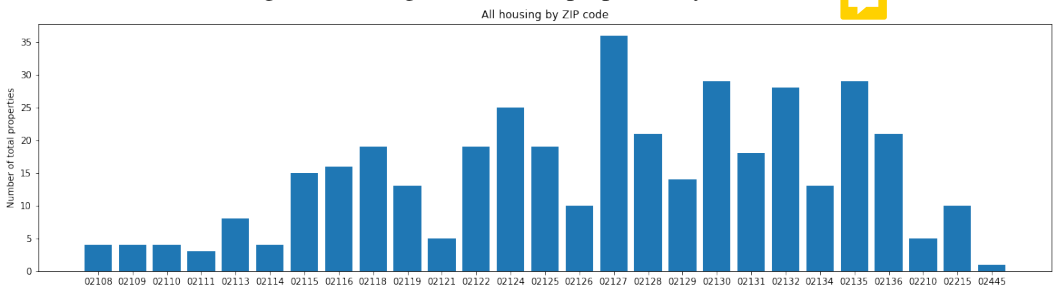


Figure 2: Histogram of rented properties by ZIP code.



1.2 Distribution of landlords in affordable housing

Figure 3: Partial plot of affordable properties in the Boston area. Taken from the Boston Planning & Development Agency.

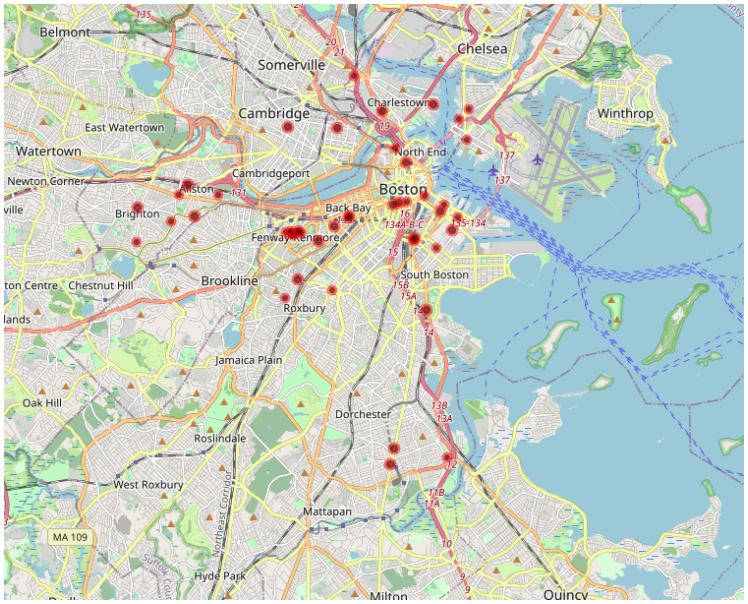
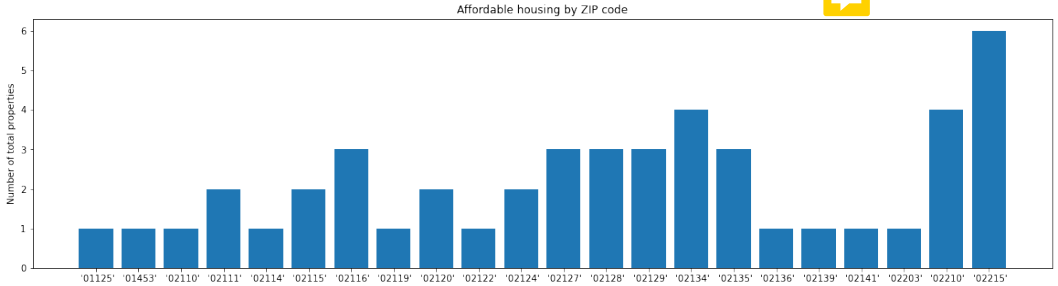


Figure 4: Histogram of affordable properties by ZIP code



2 Refined project scope

2.1 Limitations of data

As mentioned in the original project assessment, the housing market is constantly changing: this impacts the accuracy of our data in reflecting the current state of affordable housing in Boston. Additionally, the provided Affordable Housing Stock Data appears incomplete, and so we will need to extrapolate features of this dataset in order to make broader claims regarding the geographic distribution of affordable housing.

For the purpose of plotting properties on a map, we discovered when working with these datasets that some entries are missing relevant address information. As an example, one set of data did not include each property's ZIP code, a crucial feature for delineating geographic regions of interest. Furthermore, some addresses could not be converted to coordinates with an external Python library (geopy, which queries OpenStreetMaps), and we were therefore unable to represent them in our results. We will need to explore using other mapping APIs (such as Google Cloud) as a fallback for our default library in order to maximize the amount of usable data. Lastly, it is difficult to verify the accuracy of the addresses obtained with geopy without manually cross-referencing the results with the original dataset: we have implemented some simple error-checking to this end.

2.2 Potential risks of achieving goal

We originally intended to provide an accessible user interface for browsing properties; however, it would appear that this goal lies in conflict with the formatting & accuracy issues described above. As such, it may be necessary to take a more interpretive approach to working with this data: rather than facing the logistical challenges of maintaining completely accurate, updated information on every property, we will instead try to analyze features and trends in the data that would be provide a more general sense of how to understand affordable housing in Boston.

As we continue to work with additional datasets (such as MAPC and the Analyze Boston income-restricted properties) we will be able to further calibrate our goals to the features of the data. Ideally, we will be able to leverage new APIs to collect additional information on each property listed — particular in terms of landlord history.