Landlord Mapping (Small Landlord Data Analysis)

Saisriram Gunturu, Stephanie Liu, Vineet Raju, Phillip Tran

Background and Motivation

In partnership with City Councilor At-large Ruthzee Louijeune, this project seeks to understand the current housing marketing and improve the supply of affordable housing in the city of Boston. Councilor Louijeune aims to create an avenue for owner occupied and small landlords to be added to a database of affordable rental units for tenants that qualify for such affordable housing programs.

Data Collection

The project did not require any original data collection. Instead, two datasets from the city of Boston were leveraged:

1. The Boston Income Restricted Housing Inventory (BIRHI) dataset contains all income-restricted housing units in the city of Boston. This helps us to understand the current supply of affordable housing available in the city.

	Project Name	Neighborhood	Zip Code	TtlProjUnits	RentUnits	OwnUnits	TtlMarket	MarketRent
0	Abbot Street/ Shawmut Ave	Roxbury	02119	16	16	0	0.0	0.0
1	Academy Homes I	Roxbury	02119	202	202	0	52.0	52.0
2	Academy Homes II	Roxbury	02119	236	236	0	0.0	0.0

2. The Property Assessment FY2022 (PA) dataset contains the property assessment data of every property in the city of Boston. This helps understand the entire housing market in the city.

	PID	CM_ID	GIS_ID	ST_NUM	ST_NAME	UNIT_NUM	CITY	ZIPCODE	BLDG_SEQ
0	100001000		100001000.0	104 A 104	PUTNAM ST	NaN	EAST BOSTON	02128	1.0
1	100002000		100002000.0	197	LEXINGTON ST	NaN	EAST BOSTON	02128	1.0
2	100003000		100003000.0	199	LEXINGTON ST	NaN	EAST BOSTON	02128	1.0

For our exploration, the team needed to match each property in the BIRHI dataset to a property in the PA dataset so that we could gain insights about properties enrolled in affordable housing programs as well as those that aren't. However, none of the properties in the BIRHI dataset contained an address, which is necessary for matching to a property in the PA dataset. In order to find the address of each property in the BIRHI dataset, the team utilized the Google Places API. By searching the project name and zip code using the API, we were able to find a street address for each of the properties in the BIRHI dataset and then match it with the appropriate property in the PA dataset.

Data Exploration, Data Visualization, and Questions Answered

We leveraged the following data science libraries to answer our key questions:

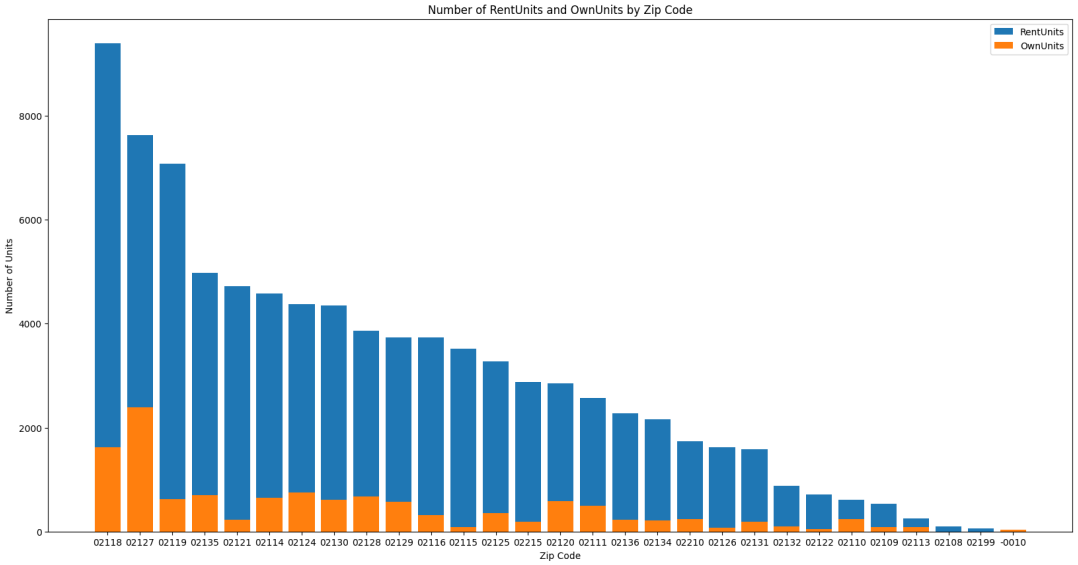
1. Pandas for loading and manipulating the datasets
2. numpy for performing arithmetic operations
3. matplotlib for generating data visualizations.

After exploring the data using the packages listed above, we were able to answer seven key questions about small landlords in Boston:

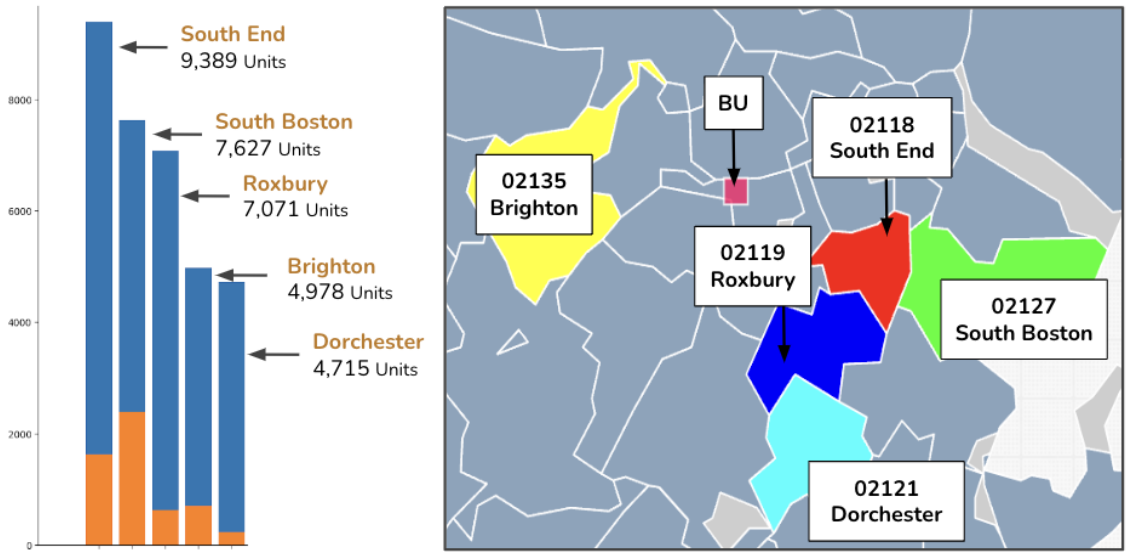
1. *What percentage of housing stock is owned by owner occupied landlords?*
   According to our analysis of the data, there are 78,242 owner occupied units. This represents 43.8% of total housing stock.

2. *What is the current number of affordable housing units (non-owner residential and residential) in current affordable housing programs?*
   Our analysis shows that there are 73,665 affordable housing units. Of these units, 62% are non-owner residential units while 48% are residential units.

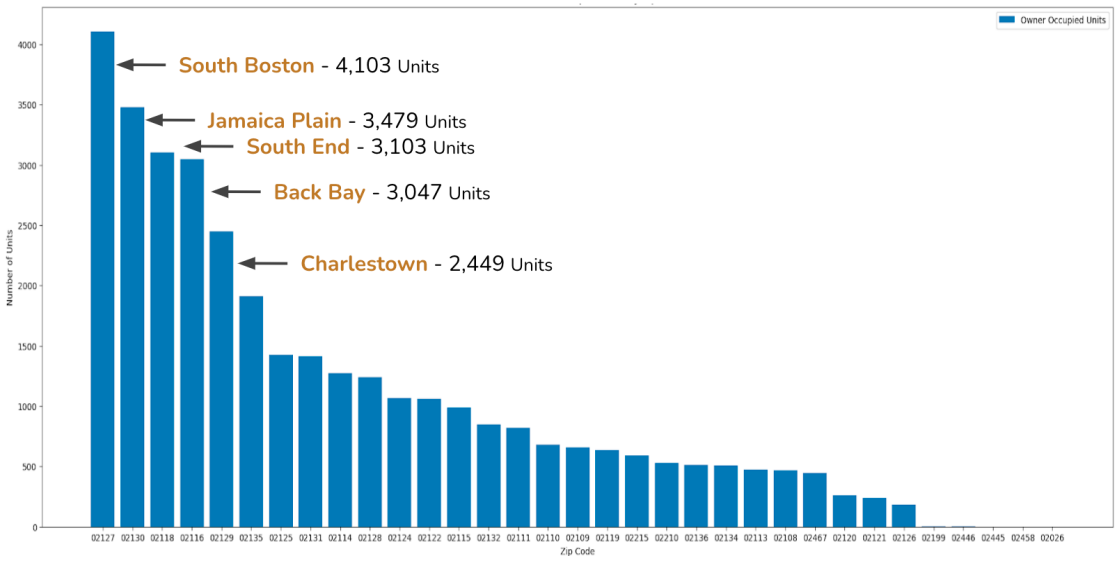
3. *What is the current distribution of landlords and housing listed in current affordable housing programs by zip code?*
   As seen in the bar graph below, the number of affordable rental units varies greatly from zip code to zip code. On the other hand, the distribution of affordable owner units is much less varied.



To gain further insight into the geographic distribution of affordable housing in Boston, we also decided to plot the top five locations on a map. After doing so, we found that four out of the top five zip codes – Roxbury, South End, South Boston, and Dorchester – were clustered together in the southeast side of the city. The remaining neighborhood, Brighton, is located much further from the other four in the northwest side of the city.

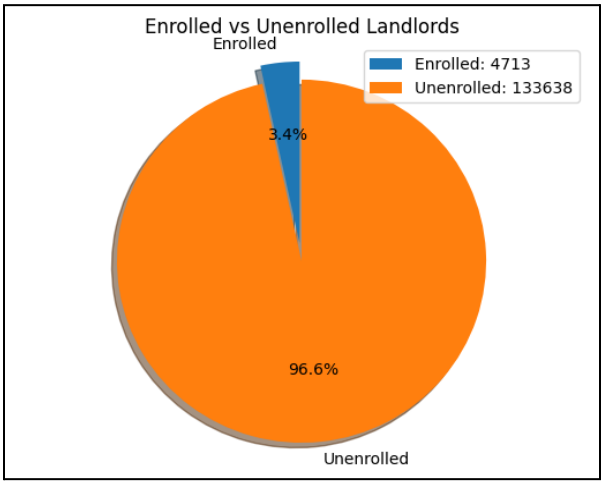


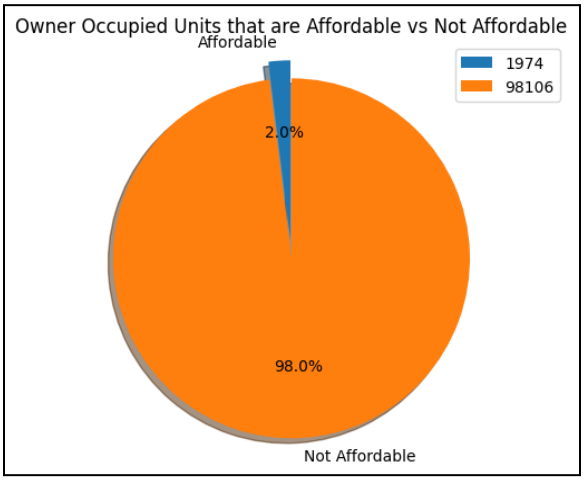
4. *What is the current distribution of owner occupied housing by zip code?*
   The top 5 zip codes for owner occupied units (South Boston, Jamaica Plain, South End, Back Bay, Charlestown) differ from the top 5 zip codes for affordable units (South End, South Boston, Roxbury, Brighton, Dorchester). However, South Boston is still represented on both visualizations. This shows that places with a lot of housing don't necessarily have more affordable housing options available.



5. *What percentage of owner-occupied housing stock is affordable-housing?*
   Through our analysis of the data, we have discovered that approximately 2% of all owner-occupied housing stock is affordable. This represents 1974 affordable owner-occupied units out of a total of 100,080 total owner-occupied units. This means that there is a large supply of owner occupied units that could potentially be added to affordable housing programs.

6. *What number of landlords are not currently enrolled in affordable housing programs?*
   We have discovered that a very small percentage of landlords are enrolled in an affordable housing program—3.4%. This represents 4,713 landlords out of a total of 13,3638 landlords that could potentially enroll to participate in an affordable housing program.





Limitations of Results and Challenges Faced

Although a great number of insights were able to be obtained through our exploration, we were limited by third-party data availability. Because we mostly relied on data collected by the city of Boston to gather insights about the housing market, our analysis was limited to the information provided in these datasets. For example, the BIRHI dataset does not contain the address of each property. This necessitated the use of the Google Places API to associate each property with an address. For the most part, these addresses seem correct; however, this process is imperfect and there is no guarantee that this third-party API produced a correct address for each property.

Another roadblock was the inability to access certain datasets. The team had also planned to use the Boston Planning and Development Agency dataset. This was preferred over the BIRHI dataset because it also contained the street addresses for properties enrolled in affordable housing programs, eliminating the need to guess project addresses with the Google Places API; however, it was nearly impossible to export or otherwise extract the data from this dataset. As such, it ultimately had to be abandoned in favor of the BIRHI dataset.

Suggestions for the Future of the Project

As an extension to the given project, the team wishes to match the owner and address of each owner occupied unit to the Massachusetts voter file. This would allow us to see how many of these owners are voters which, in turn, allows the project group to gain an understanding of the demographics and concentration of owners that would be participating in Counselor Loujeune's affordable housing program. In order to do so, we hope to find what zip codes these owner-voters are concentrated in as well as the demographics and party affiliations of these owners.

To achieve this goal, we will utilize two data sources:

1. The MA Voter File will be used to match each owner to a voter on file.
2. The Census Neighborhood Data will be used to contextualize the demographics of the neighborhoods in which these owners live.

Using these data sources, we will hope to answer three key questions about property owners in Boston:

1. *How does the distribution of registered voter-owners differ from the distribution of non-registered voter-owners by zip code?*
2. *Is there a difference in the number of properties owned by registered voters and those owned by non-registered voters?*
3. *What are the demographics and party affiliations of voter-owners?*

Question 1:

In order to perform our analysis, we first had to overlay the MA Voter File with the PA dataset. This was done by matching the information in the two datasets using the name of the property owner and their home address. All property owners that had a match in the MA Voter File were considered registered, and the rest were considered unregistered. For the scope of this project, this is sufficient for merging the two datasets; however, we would like to note that the name and address formatting could be more thorough, and differences in how addresses were entered into the datasets as well as any inaccuracies in the Google Places API could inflate the number of unregistered voters (unmatched owners) in our results.

After creating this new dataset, we were now able to look at the distribution of registered and non-registered voter-owners per zip code. This information is summarized in the three bar charts above. Observing the bar charts, we can see that there is a greater number of owners that were not registered to vote in each zip code (on average, about 76.30% of owners are not registered voters). Moreover, we can see that this proportion is fairly consistent across all zip codes.

Ranking	Total	Registered Only	Unregistered Only
1	02132 (West Roxbury)	02132 (West Roxbury)	02127 (Telegraph Hill)
2	02130 (Jamaica Plain)	02136 (Stony Brook)	02132 (West Roxbury)
3	02127 (Telegraph Hill)	02131 (Metropolitan Hill)	02130 (Jamaica Plain)
4	02124 (Codman Sq)	02130 (Jamaica Plain)	02124 (Codman Sq)
5	02136 (Stony Brook)	02124 (Codman Sq)	02131 (Metropolitan Hill)

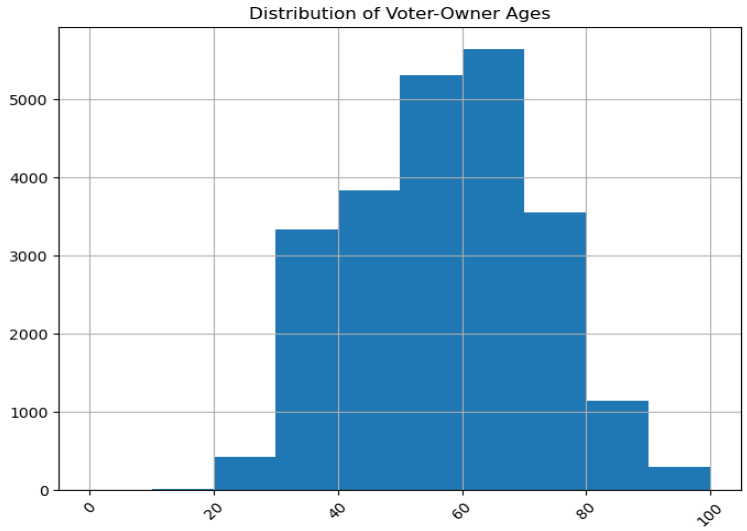
The table above shows the top five zip codes for each of the groups (total owners, registered owners only, and unregistered only). The neighborhoods are very similar across these three groups: West Roxbury, Codman Square, and Jamaica Plain are in the top five in all groups. The differences are as follows: Telegraph Hill is not in the top five for unregistered property owners and Metropolitan Hill, which is in the top five for the two separated groups, is not in the overall top five. Stony Brook is only in the top five for the overall distribution and for registered only. Because of the similarities, it seems that there aren't any obvious differences in the distribution of voter-owners across Boston.

Question 2:

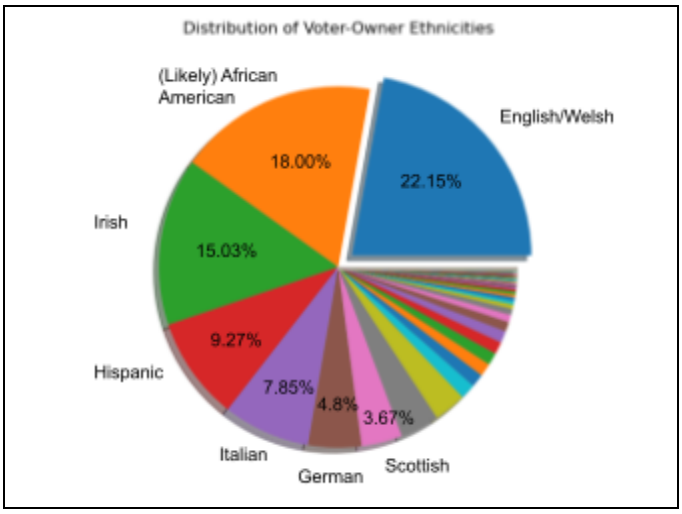
Using the overlaid data from the previous question, we were able to calculate the percentage of owner-occupied properties that were owned by registered/unregistered voters. We found that 29.98% of owners were registered and, therefore, that 70.02% of owners were not. Using the RES_UNITS field, we were also able to do the same calculations, but looking at the proportion of residential units and not just the proportion of properties. After doing so, we found that 26.20% of all residential units were owned by voter-owners and 73.80% were owned by non-voter-owners.

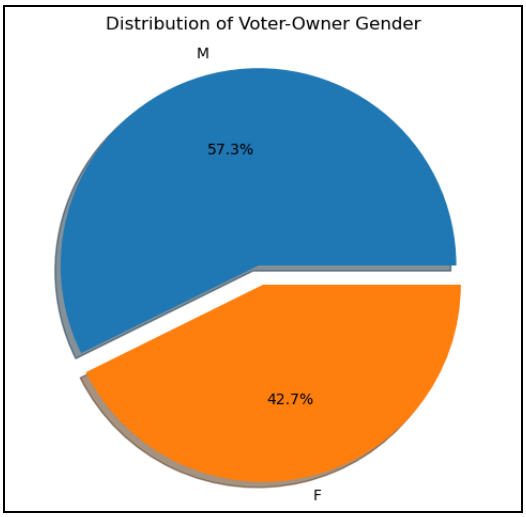
Question 3:

The next step was to look at the demographic information of owner-occupied property owners. The first aspect we looked at was age which is summarized in the histogram below. As seen in the histogram, the ages of voter-owners are approximately normally distributed. The mean age is 57 years old, the minimum age is 18 years, and the maximum age is 99 years.

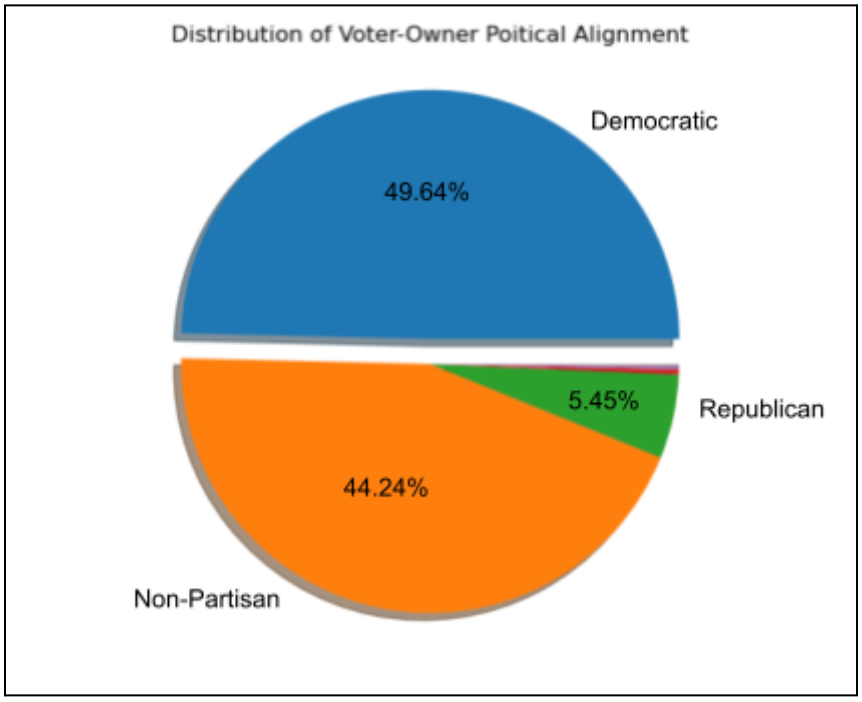


Next, we looked at the distribution of ethnicities and genders of voter owners. For the sake of readability, only the top seven ethnicities are labeled on the pie chart below. As seen on the pie chart on the below (left) , the most represented ethnicity is English/Welsh, comprising 22.15% of all voter-owners.





Next we took a look at the distribution of genders for voter owners. As shown in the pie chart above (right), the majority of voter-owners identify as male (57.3%) and the remainder identify as female. While there is a slightly greater percentage of male voter-owners than female, the population of voter-owners is pretty evenly divided between the two observed genders.



Finally, we looked at the distribution of voter-owner party affiliations. The majority of voter-owners were registered Democrats (49.64%) or Non-Partisan (44.24%). A small minority were registered Republican (5.45%) while the remaining 0.67% were registered under third-parties. This is visualized in the pie chart shown above.