

Councilor Louijeune Small Landlord Project Report

Yuesi Liu
Haoran Yang
Hanlin Zou
Shun Yao

Background

The affordable housing Program in the Boston area is designed to help people with limited incomes find affordable housing. There are many types of housing in the Boston area, and housing prices are not evenly distributed. We hope to use some of the techniques of data science to help the project find more suitable homes for people with limited incomes. In addition, we intend to conduct a summary of complaints about homes in the Boston area.

Motivation

We chose this project because we want to help a certain group of people with low income to find and live in suitable and affordable houses or apartments. Based on our analysis, we hope it would be possible to create a path between affordable housing and qualified tenants. We hope to make some contribution to this project through our efforts.

For the extension part, our group members all have experience renting in the Boston area. We have encountered some problems in our life, so we wanted to look into the landlord complaints of houses in the Boston area and the main reasons behind the data.

Compilation Instruction

We have two folders, which are Spark Landlord and extension. The Spark Landlord folder contains all the data used to answer the base questions and the graph generated. The extension folder contains all the data used in the extension question and the graphs generated.

Code should be run in Jupyter Notebook. Run all the code again to reproduce all the results.

Datasets Navigation

Base:

1. Boston Assessor Added Income Restricted.csv

Using google API to get the detailed address for the unit in the income-restricted dataset and then separate the address to get some new attributes. Through the same address to merge BostonAssessorsDataCleaned and income-restricted-inventory-2021

2. boston-neighbourhood-data.csv

Recorded the distribution of human races in each city in the Boston area.

3. BostonAssessorsDataCleaned.csv

Dataset specifies owner occupied and number of units in a location.

4. city_lat.csv

The geographic location of each city within the Boston area.

5. income-restricted-inventory-2021.csv

Income restricted housing detailed information.

6. income-restricted-inventory-address.csv

Income restricted housing detailed information and their address.

7. lat.csv

The geographic location of income restricted housing.

Extension:

1. RentSmart.csv

Recorded housing violation type, location, and other detailed information.

2. Boston Assessor Added Income Restricted.csv

Using google API to get the detailed address for the unit in the income-restricted dataset and then separate the address to get some new attributes. Through the same address to merge BostonAssessorsDataCleaned and income-restricted-inventory-2021

3. Building and Properties Violations.csv

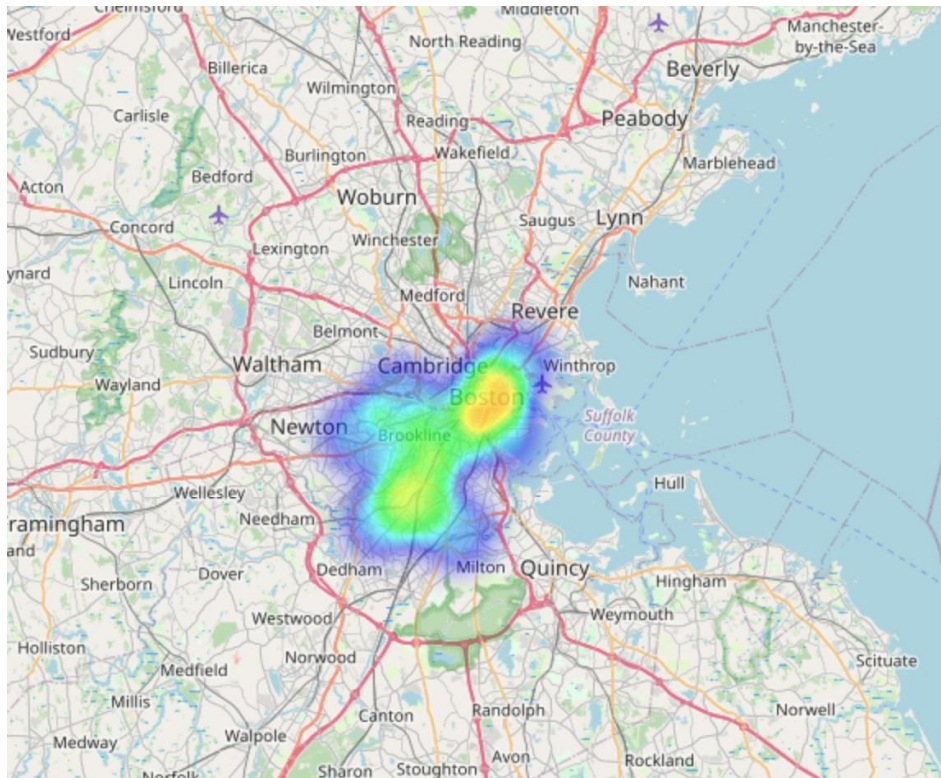
Recorded detailed housing violation description and housing location.

Data Visualization and Exploration

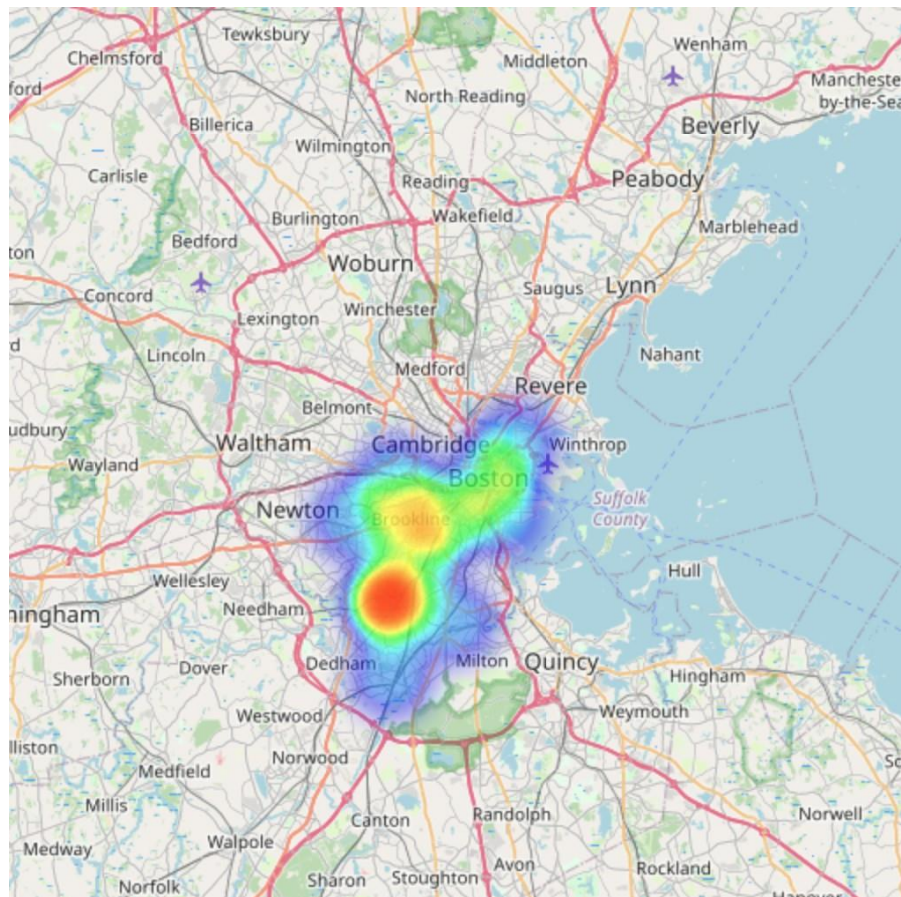
Base:

	OWNER1	UNITS
0	MILLENNIUM TOWER TRUST	442
1	HARBOR TOWERS 1 CONDO TR	308
2	SOUTH END 10 LLC	273
3	MILLENIUM AVERY CONDOMINIUM TRUST	256
4	ONE CHARLES CONDOMINIUM	233
...
32391	FRANCOIS CHRISTIAN	0
32392	FRANCOIS CAMILLE	0
32393	FRANCOIS ACHILLE S	0
32394	FRANCO NICHOLAS R	0
32395	ZZI REALTY TRUST	0

Chart of the number of housing landlords in affordable housing in the Boston area



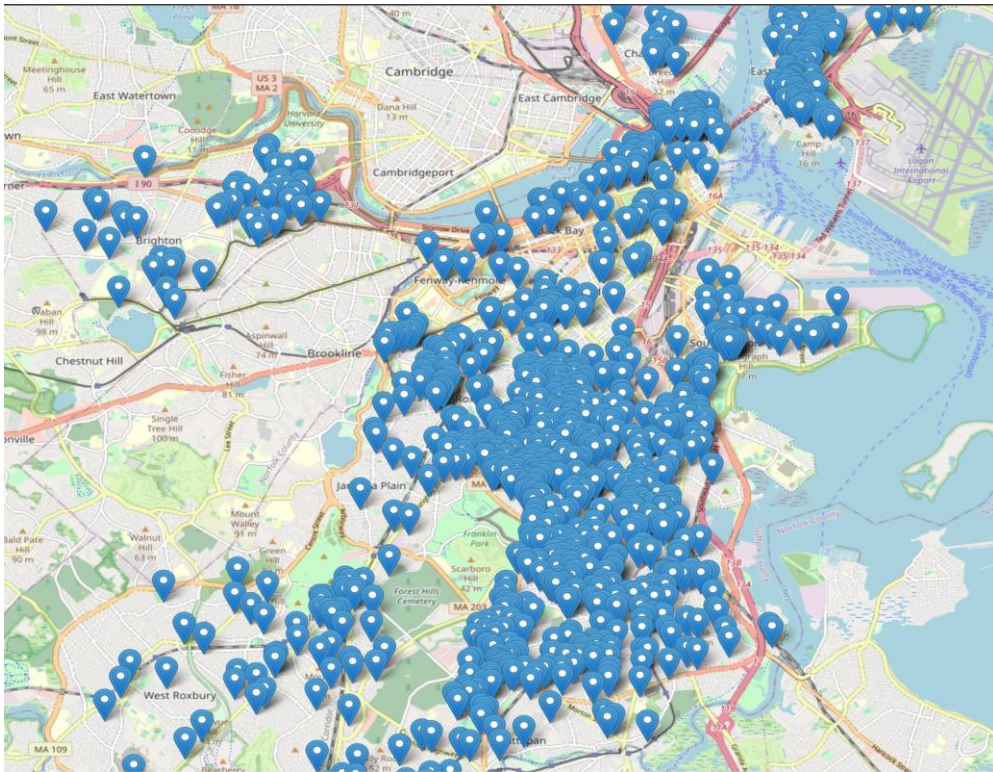
Distribution of affordable housing heatmap



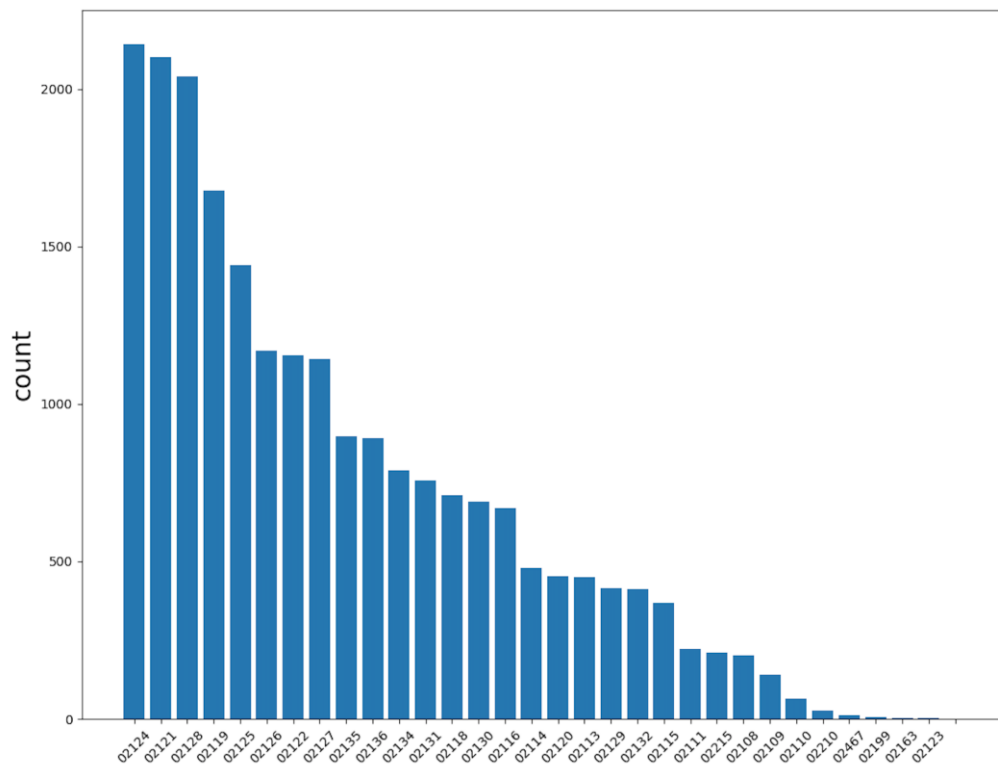
Distribution of non-affordable housing heatmap

The above two heat maps show the distribution of affordable and non-affordable housing in the Boston area.

Extension:



Houses with complaints/violations on the map through the house address



The number of complaints per housing by Zip Code

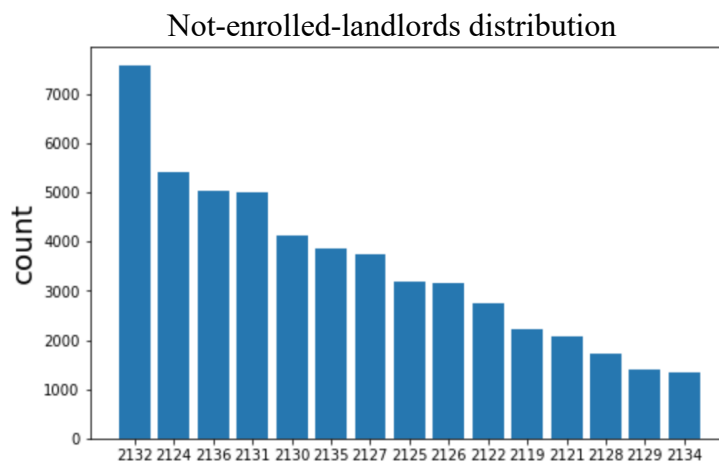
Results obtained / questions answered

Base Questions:

1. What is the current distribution of landlords NOT currently enrolled in different affordable housing programs? # of units?

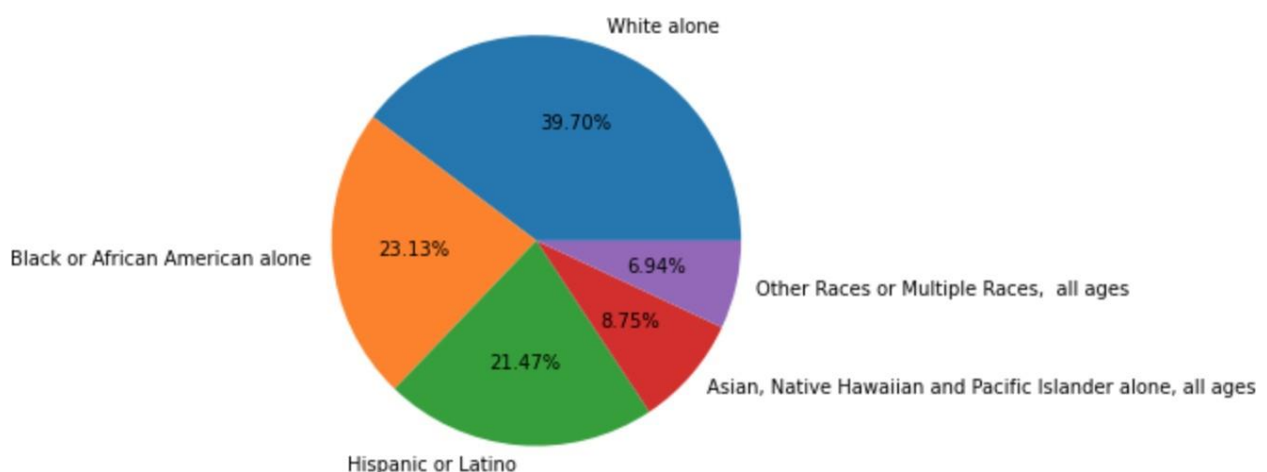
19304 landlords are not currently enrolled in different affordable housing programs.

2. Geographic distribution (by zip code)?

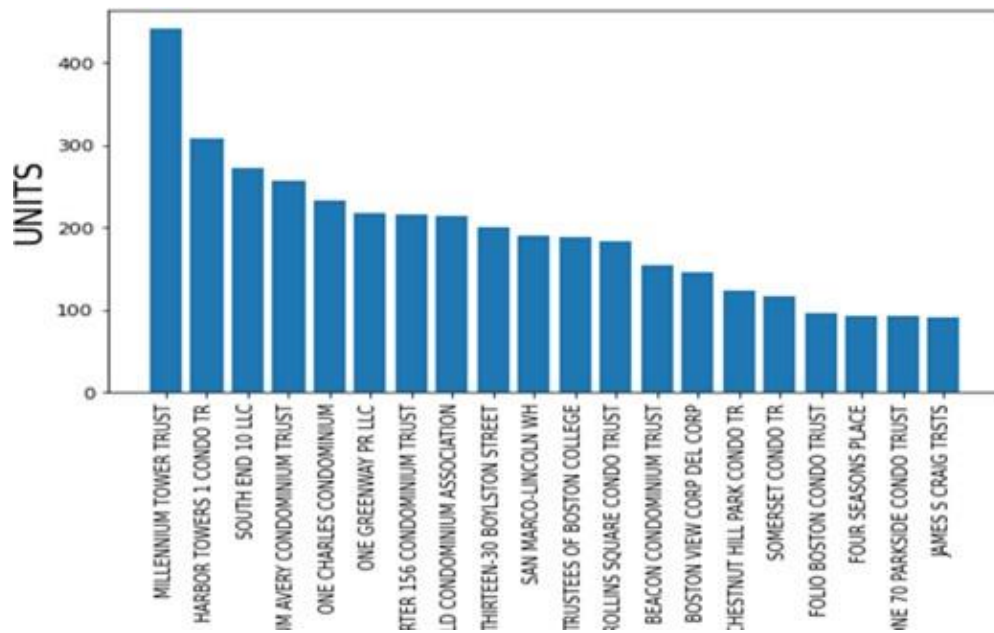


3. Demographic profile of census block group (majority race, ethnicity, income)

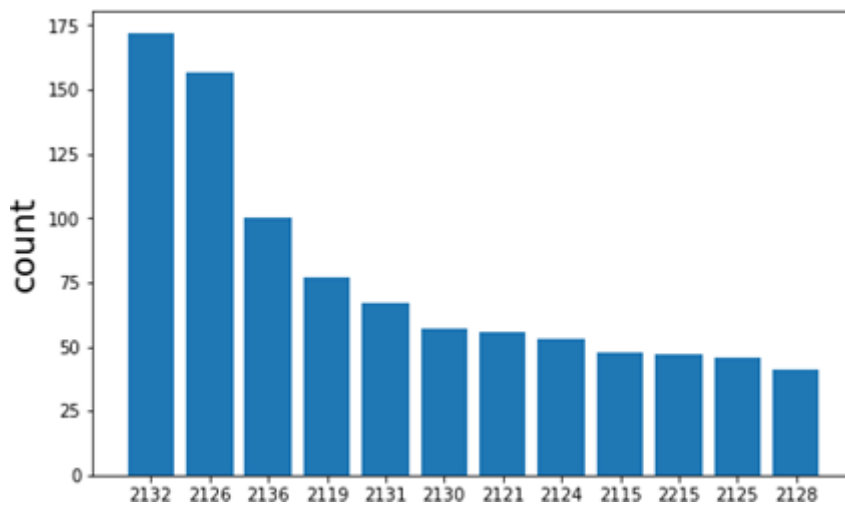
Distribution of races in Boston



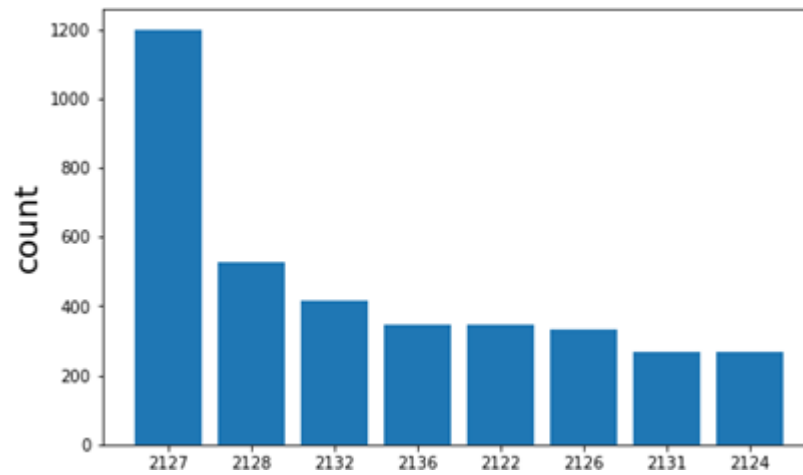
4. What is the current distribution of landlords and housing listed in current affordable housing programs?



current distribution of landlords



Zip code 02132 has the most landlords.



Zip code 02127 has the most housing.

5. What is the geographic distribution of these landlords by city council district?

	CITY	count
0	dorchester	9486
1	boston	6322
2	east boston	4710
3	hyde park	3045
4	south boston	2977
5	roxbury	2585
6	jamaica plain	2187
7	rosindale	1994
8	west roxbury	1955
9	charlestown	1654
10	mattapan	1366
11	brighton	1342
12	allston	886
13	roxbury crossing	708
14	chestnut hill	22

Affordable housing distribution.

	CITY	count
0	dorchester	13310
1	west roxbury	7599
2	hyde park	5087
3	rosindale	4965
4	boston	4290
5	jamaica plain	4196
6	brighton	3876
7	south boston	3393
8	mattapan	3111
9	roxbury	2229
10	east boston	1813
11	charlestown	1416
12	allston	1330
13	roxbury crossing	654
14	chestnut hill	154
15	brookline	21
16	dedham	7
17	readville	1

Non-affordable housing distribution

6. What percentage of housing stock is owned by owner occupied and small landlords, and at what % affordable

Owner-occupied and small landlords: **75.91%**

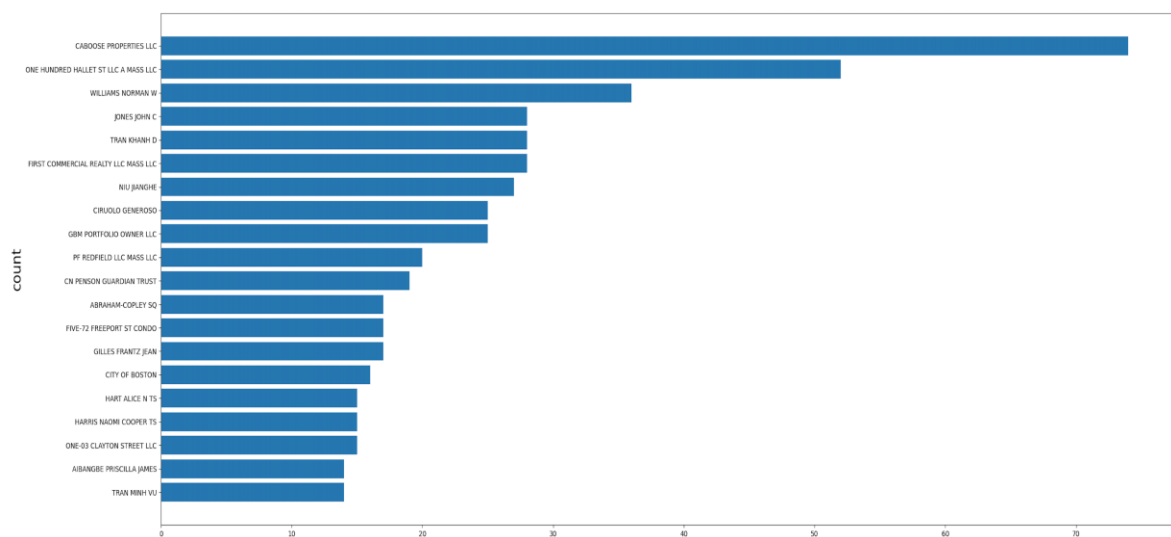
Affordable Owner occupied and small landlords: **36.27%**

Extension:

1. Which landlords have the most housing complaints?

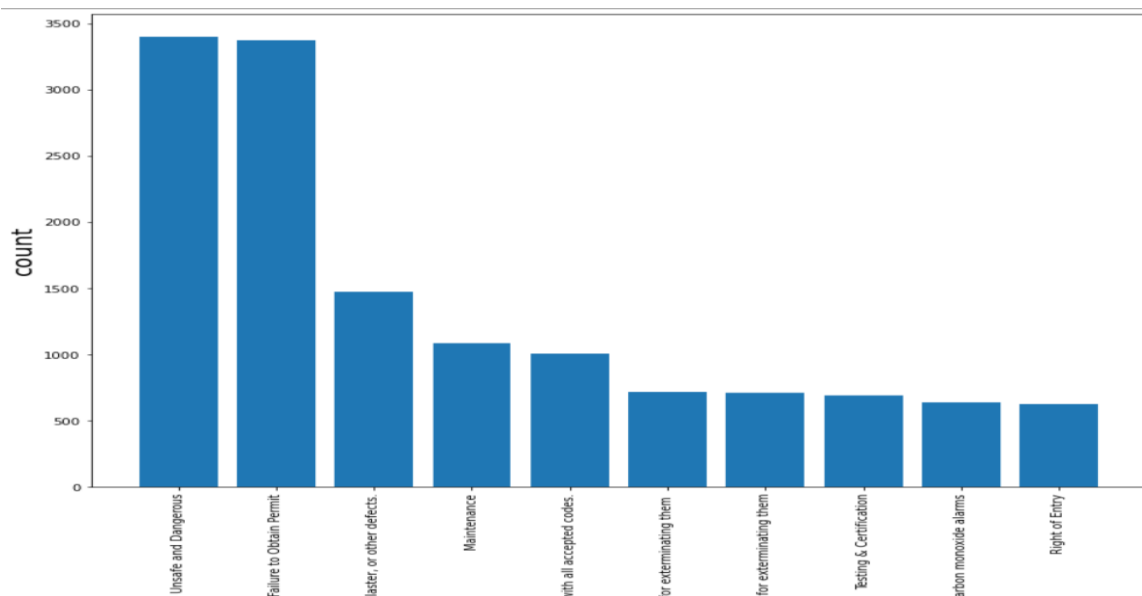
The top three companies with the most complaints are:

1. CABOOSE PROPERTIES LLC
2. ONE HUNDREND HALLEST ST LLC A MASS LLC
3. WILLIAMS NORMAN W



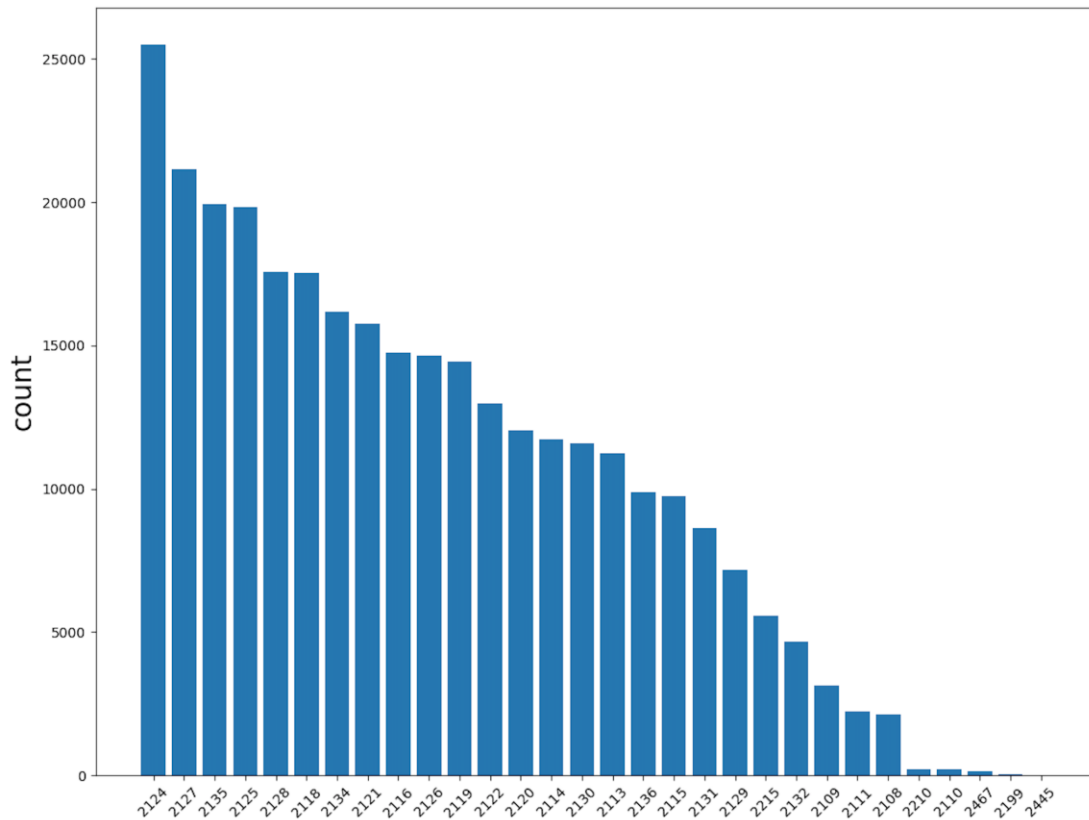
2. What are the most frequently complained reasons?

Unsafe and dangerous, and failure to obtain permits are the two problems most being complained about.



3. Which city has the highest number of complaints?

The city which zip code is 2124 has the most complaints. According to the basic question c, maybe Dorchester city has the most housings so it has the most complaints.



Limitations Encountered

1. We can't get a breakdown of races for each city in the Boston area, so we can't get a breakdown of the races of people who own housing in each city.
2. We suspect that the addresses we get from google API are not only located in Boston but also some other cities like NYC share the same street addresses. This may lead to a larger calculation of the number of homes owned by the landlord.
3. Some data in the combined data set is incomplete, which may affect the results of the final analysis.
4. The complainants may complain about the same problem several times at different times because the problem was not solved in time, resulting in inaccurate analysis results.

Challenges Encountered

1. Not familiar with the zip code and street names in the Boston area. So we don't have a very intuitive sense of the data.
2. In the process of getting a detailed address through google API, the process of splitting

the address is very complicated.

3. For the two datasets in the extension project, we are unable to find a column to merge the dataset. We stacked the data according to the time of complaints. This makes our task very complicated.

Suggestions for the future of the project

1. The data obtained through the Google API will be further cleaned to reduce the bias brought by the data.
2. Obtain more data on Dorchester City to further study the distribution of housing in the city.
3. The density of each category in housing in each zip code are different. For example, most housing in 02124 is rental. Most housing in another zip code is small landlords or self-occupied. So, in the future we might want to do normalization of data. Then the analysis for neighborhood would be more precise and reliable.