CS506 Final Project

PlaceMe & Housing in Great Boston

PlaceMe

Kunlin Cai, Yating Yang, Yihao Shi

In this final report, we found four topics that connect with Boston housing. In order to relate the information that can help PlaceMe Living, we provide detailed data, analysis and recommendation associated with each topic.

Table of Contents

Part 1: Housing Price V.S. Housing Market

1.1 Introduction

The first part is about the housing market of Boston. We found two useful rates to describe house sources in Boston. Through availability and vacancy rate, we hope that we can understand which areas of Boston are easier to rent and which areas are worth signing a contract with the local owner. At the same time, house prices in Boston are relatively high compared to other cities. If the company signs a contract with someone but does not rent the house out, it will definitely cause big losses to the company. Therefore, we want to dig into the relationship between vacancy rate and availability rate in order to gain a deeper understanding of the company's focus on the regions and the impact of the two rates toward the prices.

1.2 Experiment

We collect data from Boston pads. From the data we collected, we find out the relationship between Real Time Availability Rate, Real Time Vacancy Rate and House Prices.

Key Word:
Real Time Availability Rate：(Total Apartments Currently Vacant (Not in the market) + Apartments Set to Become Available on a Later Day) / total apartment in database
Real Time Vacancy Rate：(Total apartments currently available to rent (In the market) + all of those becoming available in the future) / total number of apartments

Neighborhood	Real Time Availability Rate	Real Time Vacancy Rate	Price
Quincy	1.02%	0.59%	$1,573
Charlestown	1.59%		$2,173
South Boston	1.78%		$2,130
South End	1.97%	0.33%	$2,444
All Areas	3.62%	0.43%	$2,121
City Of Boston	4.19%	0.54%	$2,153

Outside Boston	2.81%	0.98%	$2,046
Fenway	8.27%	0.42%	$2,370
Symphony	6.40%	0.52%	$2,644
Roxbury	6.23%	1.39%	$1,655
Mission Hill	5.80%	0.33%	$1,894
East Boston	4.75%	0.94%	$1,788
Brighton	4.50%	0.43%	$1,872
Allston	4.36%	0.43%	$1,857
Back Bay	2.69%	0.55%	$2,694
North End	4.00%	0.43%	$2,289
Brookline	2.93%	0.71%	$2,218
Cambridge	3.05%	0.96%	$2,345
Newton	4.01%	1.28%	$2,319
Medford	2.65%	1.01%	$1,561
Malden	4.11%	1.25%	$1,670
Dorchester	2.52%	0.90%	$1,724

Correlation between 1b1b price and Real Time Availability Rate: **0.1065869552**
Correlation between 1b1b price and Real Time Vacancy Rate: **-0.4392974535**

1.3 Explanation

Correlation Coefficient

Related Level	Range:
Low	**-0.1 < R < 0.1**
Middle	**-0.6 < R < -0.1 or 0.1 < R < 0.6**
High	**-1.0 < R < -0.6 or 0.6 < R < 1.0**

By the definition above, we can see the correlation between price and real time vacancy rate: -0.4392974535. It's a middle rate. **The correlation means that when the vacancy rate decreases, the price of the house will increase**. However, we can see that the correlation of 1b1b price and real time availability rate is just 0.1065869552 which is low related. **This happens because the availability rate is not related to houses in the market currently.**

1.4 Recommendation for PlaceMe

PlaceMe could rent the house in the area with a high current Vacancy Rate + low current Availability Rate, because the rent in this area may increase in the future.

1.5 Extension and Future Improvement

- Try correlation analysis on other data.
- Try other ways to find relationships(e.g:Spearman Correlation Coefficient)

Part 2: Keywords Frequency Analysis to Reviews

2.1 Introduction

Our second task is to do a sentiment analysis on customer reviews of PlaceMe and their competitors: Bedly and Bungalow. Thus, we plan to analyze the advantages of PlaceMe and the competitors from the reviews and want PlaceMe to keep their advantage and learn from other companies, and then we will analyze the negative reviews so that we hope they can make some corrections.

We collected tenant reviews of PlaceMe and its competitors, Bedly and Bungalow, from Google, Facebook and Yelp. We then learned that PlaceMe has the highest star rating and we decided to analyze the advantages of PlaceMe through vocabulary analysis in the review. Because there are few PlaceMe negative reviews, we then focus on collecting and reviewing negative reviews in its competitors, hoping that Placeme can be vigilant to these contents.

Note*: We want to do a sentiment analysis through building a model originally. However, since there are too few data entries and most of the reviews are positive. So the data is not balanced and we decided to do a vocab frequency analysis to see what PlaceMe can improve.

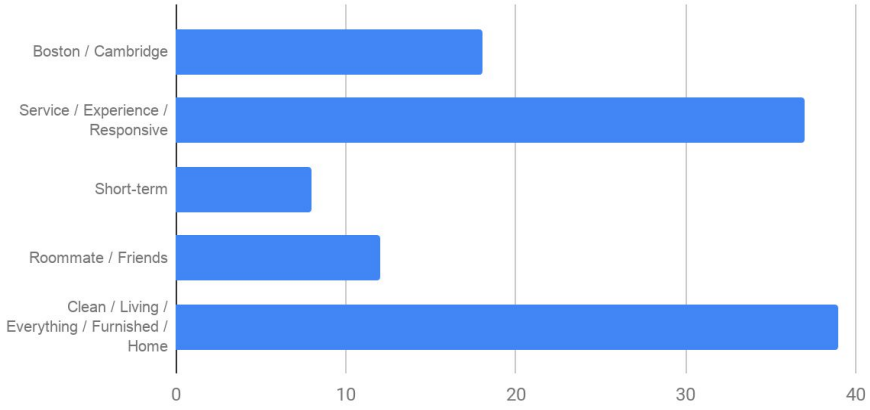
2.2 Experiment

For all the reviews, we clean the punctuations and lowercase all the characters. Then, we write a python program to transfer the reviews we collected into words and count their frequencies.

2.3 Explanation

The vocabulary above are some meaningful keywords which have high frequency in the collected reviews. So we believe that they are useful for identifying customers' preferences.

PlaceMe Positive Reviews

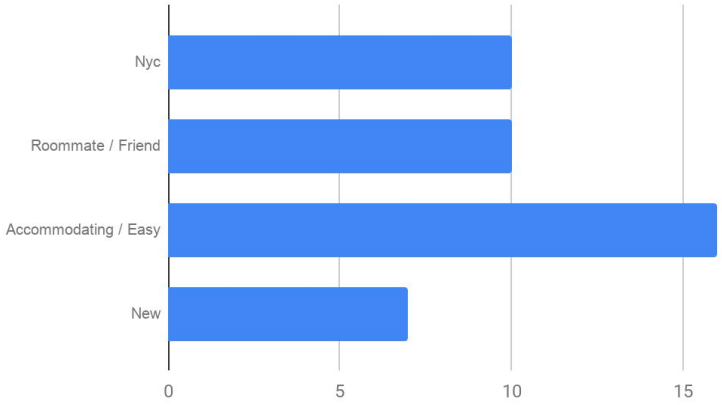


There are 43 positive reviews we found for PlaceMe. From all these reviews, here are some vocabs with high frequency worth notifying.

- "Boston/Cambridge": 18 -- PlaceMe have their major services in Boston, especially in Cambridge.
- "Service/Experience/responsive": 37 -- Good pre/after rent services
- "Short-term": 8 -- Allow short-term rent
- "Roommate/friends":12 -- PlaceMe provide matched Roommate/Friendships

- “Clean/ Living/Everything/furnished/home”:39 -- PlaceMe provides all the things for customers and makes them feel like home.

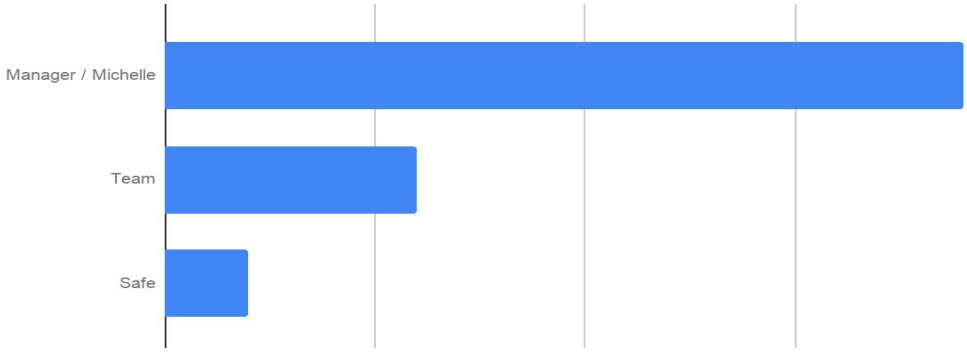
Bedly Positive Reviews



There are 25 positive reviews we found for Bedly. From all these reviews, here are some keywords with high frequency worth notifying.

- “Nyc”: 10 -- Bedly have their major services in New York City
- “Roommate/Friend”: 10 -- Bedly provide matched Roommate/Friendships
- “accommodating/easy”: 16 -- Quick/humanized services
- “New”: 7

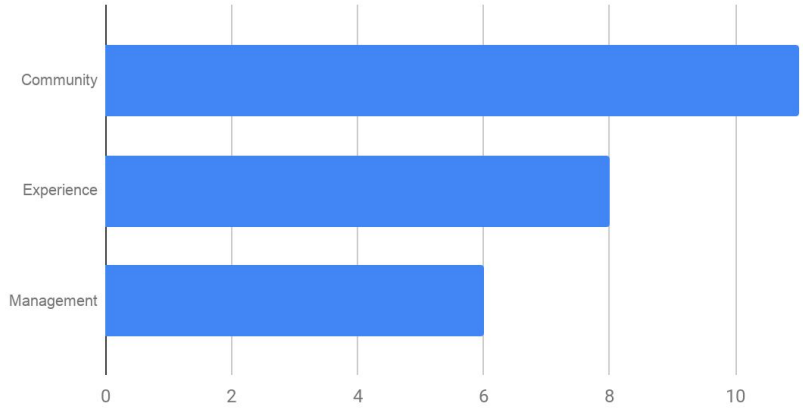
Bungalow Positive Reviews:



There are 11 positive reviews we found for Bungalow. From all these reviews, here are some keywords with high frequency worth notifying.

- “Manager/Michelle”: 19 -- Good managers(Especially a manager called Michelle)
- “Team”: 6 -- Good response and nice teams
- “Safe”: 2 -- It just mensions twice but we did not see this field in other companies’ reviews. We believe safety is also important.

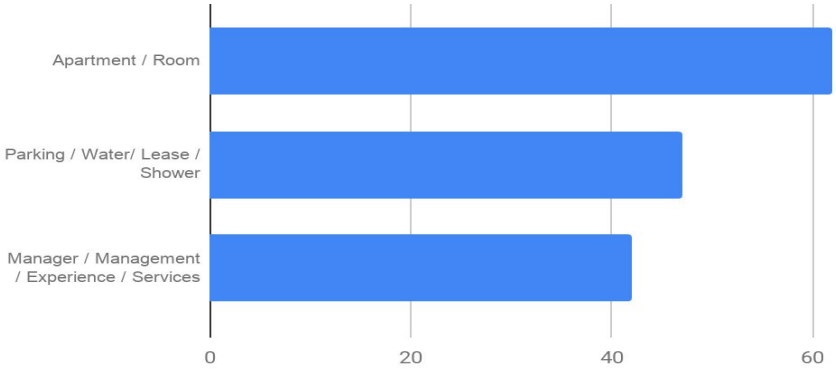
Common Positive Reviews



There are 38 positive reviews we found for Bungalow. From all these reviews, here are some keywords with high frequency worth notifying.

- “Community”: 11 -- Co-living company give tenants chances to make more friends and build communities when they are living
- “Experience”: 8 -- Having great experience
- “Management”: 6 -- Providing good services from managers

All Negative Reviews



There are 70 negative reviews we found for three companies. From all these reviews, here are some keywords with high frequency worth notifying.

- “apartment”: 62 -- Some low quantity apartment
- “Parking/Water/Lease/Shower”: 45 -- Apartment facilities are important

- “Manager/ Management/ Experience/ Service”: 42 -- Need to provide good service

2.4 Recommendation for PlaceMe

- Keep the good things.
- Learn from the good reviews of other companies.
- Avoid the bad things in the negative reviews.

2.5 Extension and Future Improvement

- Collect more reviews and maybe conduct a survey to get feedback from customers directly.
- Try other methods to analyze the data.

Part 3: Potential Customer and Their Expected Prices

3.1 Introduction

If we can help PlaceMe Living select the appropriate target and consumer, it will greatly reduce the company’s expenditure on advertisement. Therefore, we want to find customers who match the ideal price of the corresponding house from the analysis of the dataset and PlaceMe could advertise them.

We select variables and create a dataset from ipums. We choose 22 variables that we think might correlate to their willingness to pay for rent. We had a readme file explaining labels and associated numbers we used in this dataset.

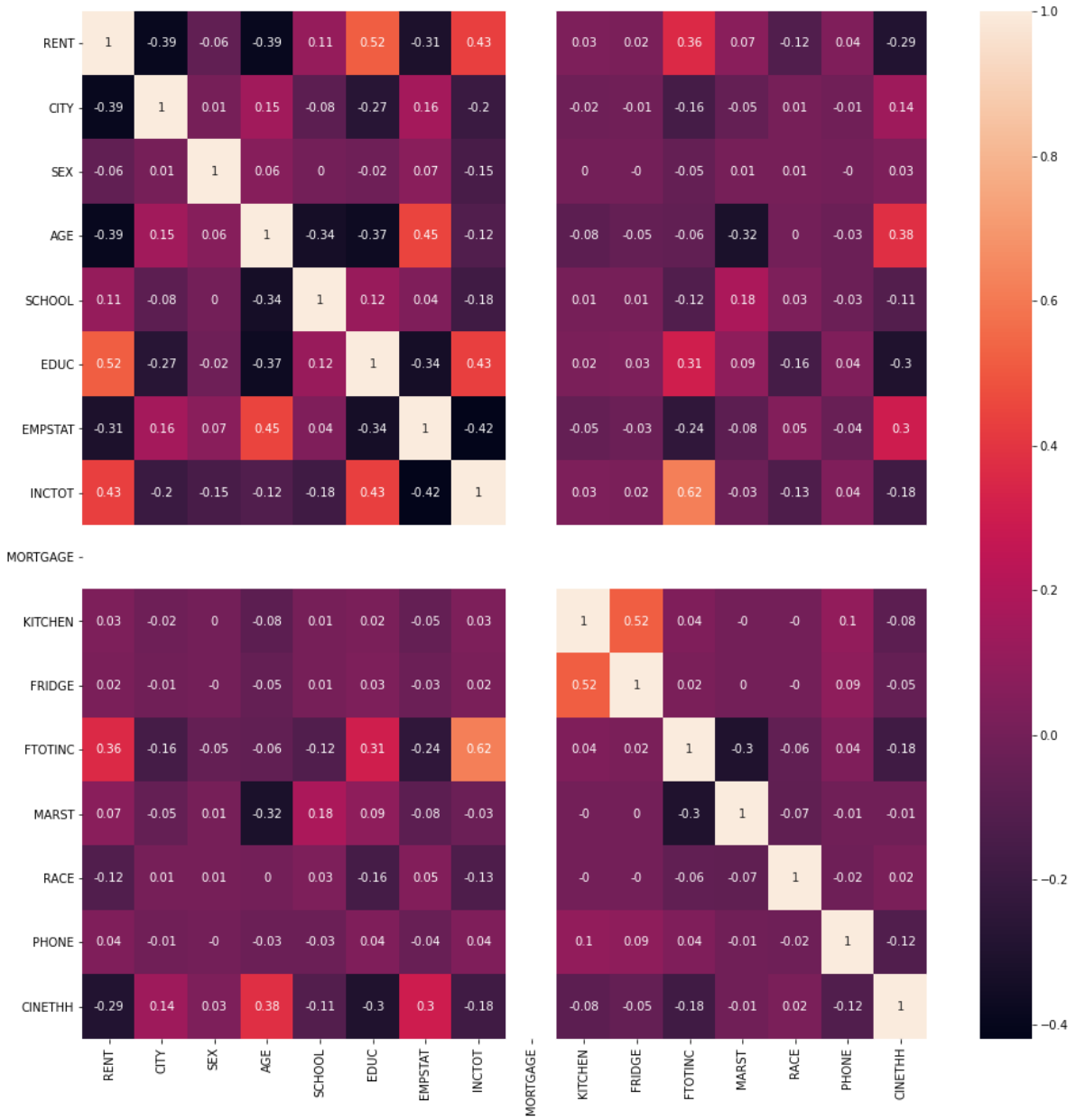
3.2 Experiment

Choosing Data features:

Since we are predicting the rent, rent will be the label and we are using the rest of the data as features.

We start by finding useful columns and their relationship to rent. So we think correlation is a good tool to use.

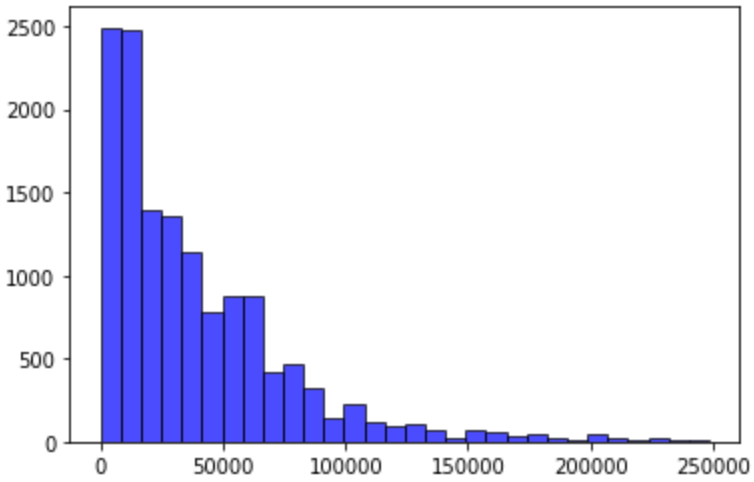
Correlation of Each Field



From the graph above we can see that “RENT” have the highest correlation efficiency with “EDUC”(0.52), “INCTOT”(0.43), “AGE”(0.39), So these three fields are probably the most important fields for us to predict “RENT”.

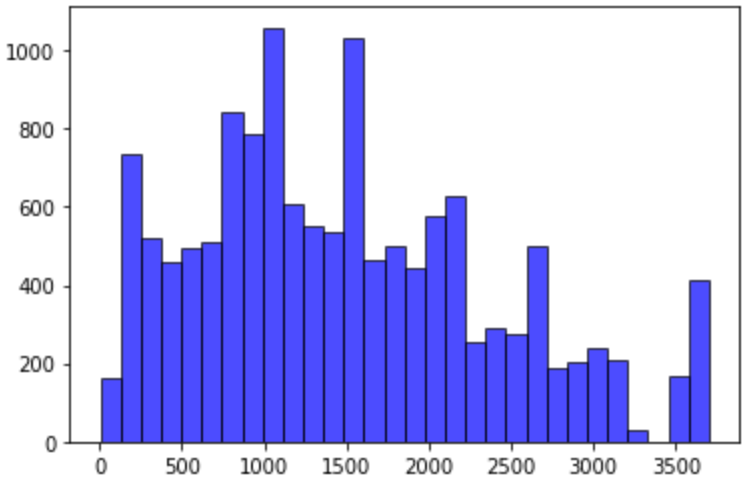
Now we draw the distribution of “INCTOT” and “RENT” to understand the real situation of people's incomes and the money they spent on rents:

Income



We can see that most of the people in the dataset are in the Low income range.
Only very little of them have High Income > $90000
< $20000 (Low Income)
$50000 < $90000 (Middle Income)
> $90000 (High Income)

Rent



We can see the distribution of Rents above.

- Most of the people are renting the houses between $100 - $1300
- Then it’s $1300 - $1900

- There are much less people have 1900 < rent < $3000
- There are just a little people whose rent is > $3000

The next thing to do is to process data:

We only use data in MA and cities we choose are Boston, Cambridge and Brookline. Secondly, we drop data who has rent 0 because they have their own house and do not want to rent. Now we end up having a dataset with 13844 data entries and 15 columns. Each of the entries is a person with 15 variables to describe their state.

We use the rents as labels. According to the rent distribution, we decided to put them into 4 classes:

0: rent = $0 - $1300
1: rent = $1300 - $1900
2: rent = $1900 - $3000
3: rent > $3000

And we are going to predict which class a person is in with the other 14 columns(We have tried to only use columns with higher correlation, however, when we use all the columns the results are better).

3.3 Results

Methods:	**Accuracy:**
Logistic Regression	0.5960555149744339
Random Forest	0.5558802045288532
MLP((20,15))	0.5924032140248356
SVM (C = 1.5)	0.5989773557341125
LinearSVM(C = 3)	0.5953250547845143
KNN(K = 47)	0.5938641344046749
SVM + Logistic Regression	0.6194945848375452
SVM + Random Forest	0.5949458483754513
SVM + SVM	0.6180505415162455

SVM + LinearSVM	0.6151624548736462
LogisticRegression + LogisticRegression	0.6209386281588448
LogisticRegression + SVM	**0.6274368231046932**
LogisticRegression + Random Forest	0.5891696750902528
LogisticRegression + LinearSVM	0.6166064981949458

3.4 Explanation

The best result is given by the combination of 2 models:

LogisticRegression + SVM

We first use Logistic Regression to train on the ["INCTOT","AGE"] columns of the data since they are different based on people. We then use SVM to train based on the rest data columns with just multiple classes because they are on the same scale. The data columns including: ['CITY', 'SEX', 'SCHOOL', 'EDUC', 'GRADEATT', 'EMPSTAT', 'FAMSIZE', 'MORTGAGE', 'KITCHEN', 'FRIDGE', 'MARST', 'RACE', 'PHONE', 'CINETHH']

3.5 Recommendation for PlaceMe

- Advertisement on school housing websites for certain house sources.
- Recommend/email certain house sources for people moving to boston.

3.6 Extension and Future Improvement

- Find data entries with higher correlation to improve dataset quality.
- Try other more complex machine learning models.

Part 4: Visualization of Short-Term-Rent Houses in the Boston Area

4.1 Introduction

From the last meeting with the client after Deliverable 3, PlaceMe asked us to make use of the Short-Term-Rent Houses data from Boston government website and Cambridge government website.

The Boston dataset have the following fields:

['sam_address_id' 'issued_registration' 'sam_address' 'home-share eligible' 'limited-share eligible' 'owner-adjacent eligible' 'owners_current_license_types' 'income restricted' 'problem property' 'problem property owner' 'open violation count' 'violations in the last 6 months' 'legally restricted' 'unit owner-occupied' 'building owner-occupied' 'units in building' 'building single owner']

The Cambridge dataset have the following fields:

['ID' 'Issue Date' 'Status' 'Location' 'Latitude' 'Longitude' 'Short Term Rental Type' 'Property Type' 'Property Type Additional Description' 'Condo Association' 'Total Bedrooms' 'Rented Bedrooms' 'Maximum Renter Capacity' 'Kitchen' 'Bathrooms' 'Airbnb' 'HomeAway' 'FlipKey' 'VRBO' 'Craigslist' 'Couch Surfing' 'Boston Rentals' 'Other' 'Other Platform Descriptions' 'All Rental Services']

Unfortunately, the data above do not have many useful features to build a model/Statistic analysis because most of the values are “None” in the form. So we decided to build a visualization tool for drawing the distribution of houses on the map.

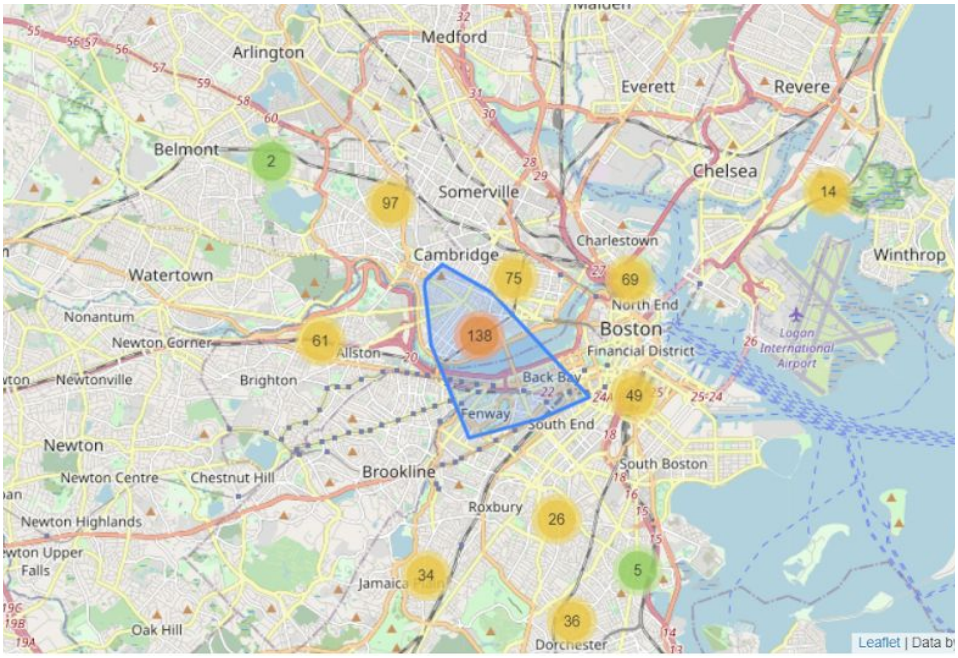
From the data fields we can see that the Boston dataset does not have ['Latitude' 'Longitude'] that we need for drawing on the map. So we used a library -- **“geocoder”** that uses google map api to translate all the addresses in the Boston dataset into ['Latitude' 'Longitude'].

Now we can use the location information to draw on the map with the python package **“folium”**.

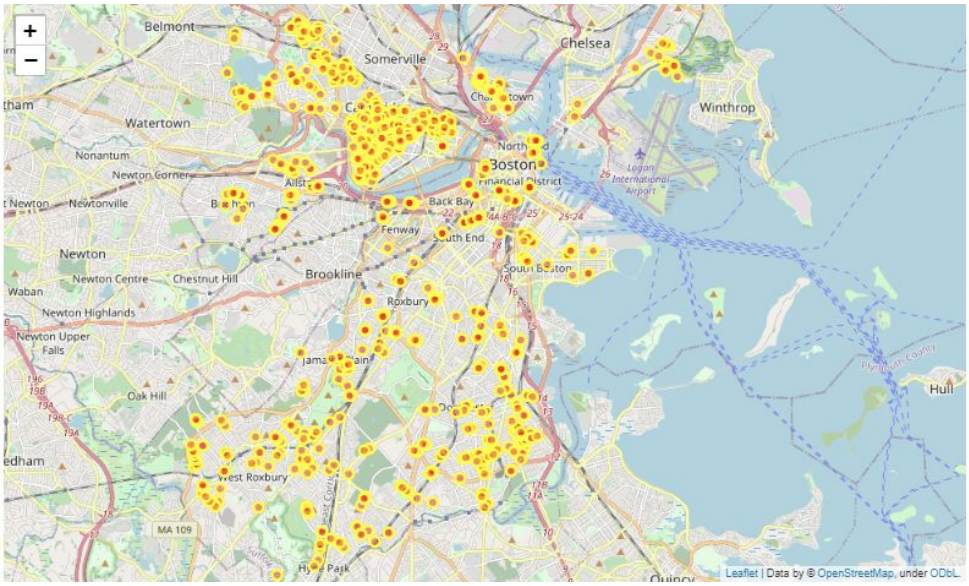
4.2 Experiment

We are using the first 500 entries in the Boston dataset (there are a total 397948 entries and it takes too long to run) and all 281 entries in the Cambridge dataset to create the map.

MarkerCluster



CircleMarker



4.3 Explanation

This is basically a visualization tool which can transfer address inputs (e.g: 700 Commonwealth Ave, Boston, MA) into coordinators (['Latitude' 'Longitude']) and draw the marks on the real map. From the graph, we can be informed of the distribution of available housing and quantity provided.

4.4 Recommendation

- Take advantage of the tool to show available apartments.
- Offer apartments in the area that has less apartments availability from other companies

4.5 Extension and Future Improvement

Add a link to each point on the map to show the information about the apartment (require data about apartment).

Reference

- https://bostonpads.com/2019-boston-apartment-rental-market-report/
- https://usa.ipums.org/usa
- https://data.cambridgema.gov/Inspectional-Services/Short-Term-Rentals/wxgv-w968/data
- https://www.facebook.com/pg/PlaceMeLife/reviews/?ref=page_internal