

Safest Place to Start Restaurant in Boston, MA

Harold Okai

October 8, 2020

1. Introduction

1.1 Background

The city of Boston is a vibrant city with plenty of diversity and plenty of fun things to do. Despite the outbreak of the current COVID-19 pandemic, people are still enjoying their favorite foods from their favorite restaurants. The restaurant industry is in its most trying time ever, perhaps since the great depression. It is for this reason now, more than ever that business owners and investors are given the proper information that will save them time and money.

1.2 Problem

One of the leading causes of restaurant failure, according to Forbes is that restaurant menus are sometimes self-serving and do not reflect the tastes of the community in which they are established. New businesses, according to Forbes struggle to find new capital during their initial phases of operation, which leads to a lot of restaurants folding. A high crime area can lead to burglaries, theft, dining and dashing and even employee theft, which lead to high insurance premiums and unexpected costs.

1.3 Interest

It is therefore necessary to find better locations and cut cost when opening a new restaurant. The safety of the location is important, which is why I added an analysis of crime trends in each neighborhood of Boston. This information will ensure that future investors, restaurateurs etc. Make better business decisions.

2 Data Sets and Data Wrangling

2.1 Data Sets

I was able to get some data from Kaggle, which included crime types, year crime was committed and the location of the crime. The data set was in a csv format. I was also able to utilize the Foursquare REST API to gather the venues that were in each neighborhood. In order to show the data analysis in the map, I also had to use some of the Geojson files uploaded to GitHub. It was a convenient and fast way to access the Geojson through a simple URL link. The Geojson file was for the Boston area was separated by neighborhood boundary lines. That was the right file to visual the neighborhood since my analysis was at the neighborhood level only.

2.2 Data Wrangling

Some of the data sets were to not full. As in, they did not sometimes provide all the information need in each file. I had to create a loop to fill some of the neighborhood information regarding venues via Nominatim. The geocode API gave me the longitudes and the latitudes of each neighborhood. I found in my data analysis that the neighborhood and the venues often had different longitude and latitude coordinates. As can be seen below the longitude and latitudes are mismatched. The Foursquare API, however provided the categories or the type of establishment, the venue and neighborhood the restaurants were situated.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roslindale	42.291209	-71.124497	Peters Hill	42.293617	-71.128063	Scenic Lookout
1	Roslindale	42.291209	-71.124497	Roslindale House Of Pizza	42.287989	-71.126549	Pizza Place
2	Roslindale	42.291209	-71.124497	Guira Y Tambora	42.291845	-71.122254	Cuban Restaurant
3	Roslindale	42.291209	-71.124497	Arboretum Grill	42.291813	-71.122762	American Restaurant
4	Roslindale	42.291209	-71.124497	Target	42.288204	-71.126659	Big Box Store

3 Methodology

3.1

The first approach I took was to visualize how much crime was in each neighborhood. This I believe gives a quick glance of crime in each neighborhood. From the bar chart in Fig1, it is very easy to see the which neighborhood has the highest crimes and which has the lowest.

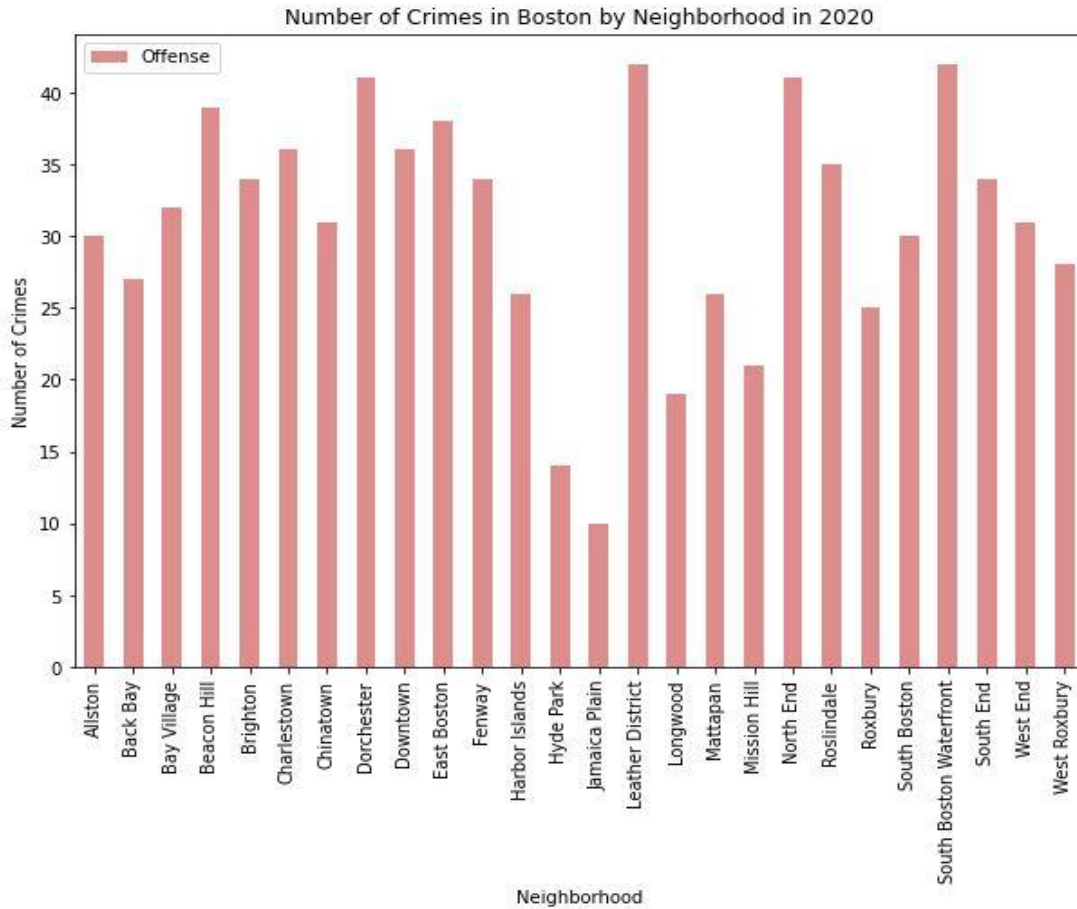


Fig2

Now, let's look at a different kind of visualization which can help us to see the intensity of crime in each neighborhood. For this kind of approach, I used a choropleth map with the help of the Geojson file, which shows each neighborhood and how much intensity of crime can be found there. The amount of crime increases the darker the blue color gets. With this, you can see that some of the neighborhoods on the map appear quite lightly colored, indicating far less crime, thus the safest in the Boston area.

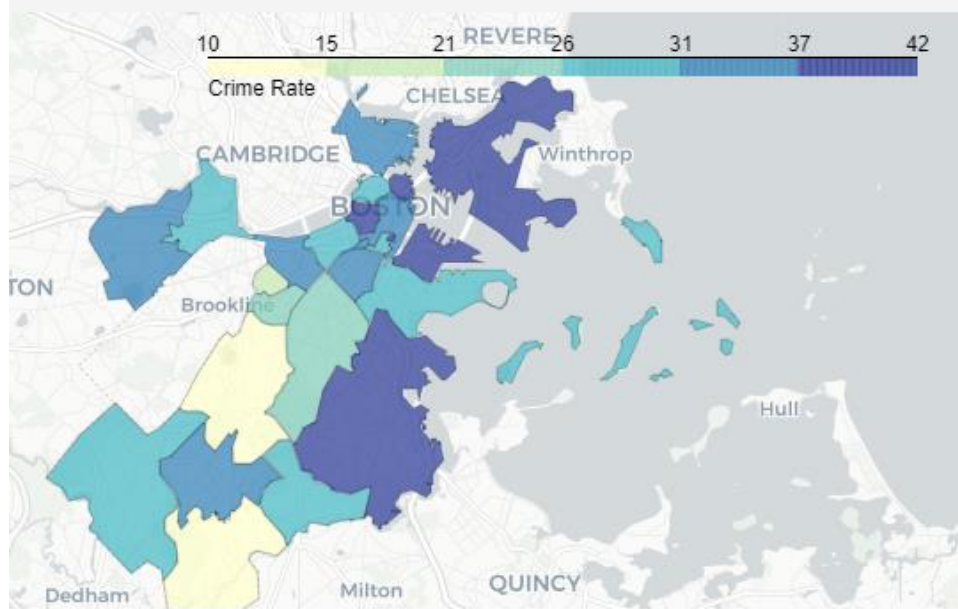


Fig3

This json file below is the file returned when a call is made to the Foursquare API. It provides a dictionary of venues and categories, as well as their latitude and longitude coordinates.

```
results = requests.get(url).json()
results
```

```
Out[46]: {'meta': {'code': 200, 'requestId': '5fa96d2ea6646a5e76ff8245'},
  'response': {'suggestedFilters': {'header': 'Tap to show:',
    'filters': [{'name': 'Open now', 'key': 'openNow'},
      {'name': '$-$$$$', 'key': 'price'}]},
    'headerLocation': 'Boston',
    'headerFullLocation': 'Boston',
    'headerLocationGranularity': 'city',
    'totalResults': 240,
    'suggestedBounds': {'ne': {'lat': 42.67525371500031,
      'lng': -70.63279056994504},
      'sw': {'lat': 42.045253084999686, 'lng': -71.48379183005495}},
    'groups': [{'type': 'Recommended Places',
      'name': 'recommended',
      'items': [{'reasons': {'count': 0,
        'items': [{'summary': 'This spot is popular',
          'type': 'general',
          'reasonName': 'globalInteractionReason'}]},
        'venue': {'id': '4a8e279ff964a5201f1220e3',
          'name': 'The Rose Kennedy Greenway',
          'location': {'lat': 42.3505, 'lng': -71.0515}}]}
```

There was however a few problems when retrieving data from some of the json files as some of them were missing some pieces of data. The solution was to create to run a known list of Boston neighborhoods against the Nominatim geocode API to successfully retrieve the Neighborhoods data. The figures show a side by side comparison of the actual list of neighborhoods and the retrieved location coordinates for each neighborhood.

```

0      Roslindale      [['Roslindale', 42.2912093, -71.1244966],
1      Jamaica Plain  [['Jamaica Plain', 42.3098201, -71.1203299],
2      Mission Hill   [['Mission Hill', 42.33255965, -71.10360773640765],
3      Longwood       [['Longwood', 42.3417067, -71.1099518],
4      Bay Village    [['Bay Village', 42.35001105, -71.0669477958571],
5      Leather District [['Leather District', -33.6976001, 19.0061312],
6      Chinatown      [['Chinatown', 42.3494053, -71.0635823927835],
7      North End      [['North End', 42.3650974, -71.0544954],
8      Roxbury        [['Roxbury', 42.3248426, -71.0950158],
9      South End       [['South End', 42.34131, -71.0772298],
10     Back Bay        [['Back Bay', 42.3507067, -71.0797297],
11     East Boston     [['East Boston', 42.3750973, -71.0392173],
12     Charlestown     [['Charlestown', 42.3778749, -71.0619957],
13     West End        [['West End', 42.3639186, -71.0638993],
14     Beacon Hill     [['Beacon Hill', 42.3587085, -71.067829],
15     Downtown        [['Downtown', 42.2630688, -71.8020015],
16     Fenway          [['Fenway', 42.3451868, -71.1045987],
17     Brighton        [['Brighton', 42.3500971, -71.1564423],
18     West Roxbury    [['West Roxbury', 42.2792649, -71.1494972],
19     Hyde Park       [['Hyde Park', 42.2556543, -71.1244963],
20     Mattapan        [['Mattapan', 42.2675657, -71.0924273],
21     Dorchester      [['Dorchester', 42.2973205, -71.0744952],
22     South Boston Waterfront [['South Boston Waterfront', 42.3334312, -71.0494949],
23     South Boston    [['South Boston', 42.3334312, -71.0494949],
24     Allston

```

With this I was able to get the list of venues and their categories in each neighborhood. I run I against the Foursquare API to the full list of venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roslindale	42.291209	-71.124497	Peters Hill	42.293617	-71.128063	Scenic Lookout
1	Roslindale	42.291209	-71.124497	Roslindale House Of Pizza	42.287989	-71.126549	Pizza Place
2	Roslindale	42.291209	-71.124497	Guira Y Tambora	42.291845	-71.122254	Cuban Restaurant
3	Roslindale	42.291209	-71.124497	Arboretum Grill	42.291813	-71.122762	American Restaurant
4	Roslindale	42.291209	-71.124497	Target	42.288204	-71.126659	Big Box Store
5	Roslindale	42.291209	-71.124497	BCYF- Flaherty Pool	42.288133	-71.122913	Pool
6	Roslindale	42.291209	-71.124497	Enterprise Rent-A-Car	42.292576	-71.122067	Rental Car Location
7	Roslindale	42.291209	-71.124497	Dunkin'	42.287912	-71.126991	Donut Shop
8	Jamaica Plain	42.309820	-71.120330	Arnold Arboretum: Hunnewell Building	42.307112	-71.120159	Trail
9	Jamaica Plain	42.309820	-71.120330	Footlight Club	42.310394	-71.115957	Theater
10	Jamaica Plain	42.309820	-71.120330	Vee Vee	42.310210	-71.115143	American Restaurant
11	Jamaica Plain	42.309820	-71.120330	Espresso Yourself	42.310199	-71.115422	Coffee Shop
12	Jamaica Plain	42.309820	-71.120330	Salmagundi	42.310184	-71.115145	Accessories Store

This I needed in order to be able cluster the venues to which neighborhoods would be favorable for each type of establishment. The table shows the venues and the categories in each neighborhood. Each neighborhood had multiple venues in different categories, but that didn't necessary show some form favorability. The solution was to cluster the neighborhoods based which types of venues they most favored. Each venue in each neighborhood was tallied in order to separate each neighborhood with a list of its own values. The algorithm I used was the K means clustering algorithm. It sorted the common neighborhoods in each neighborhood.

4 Results

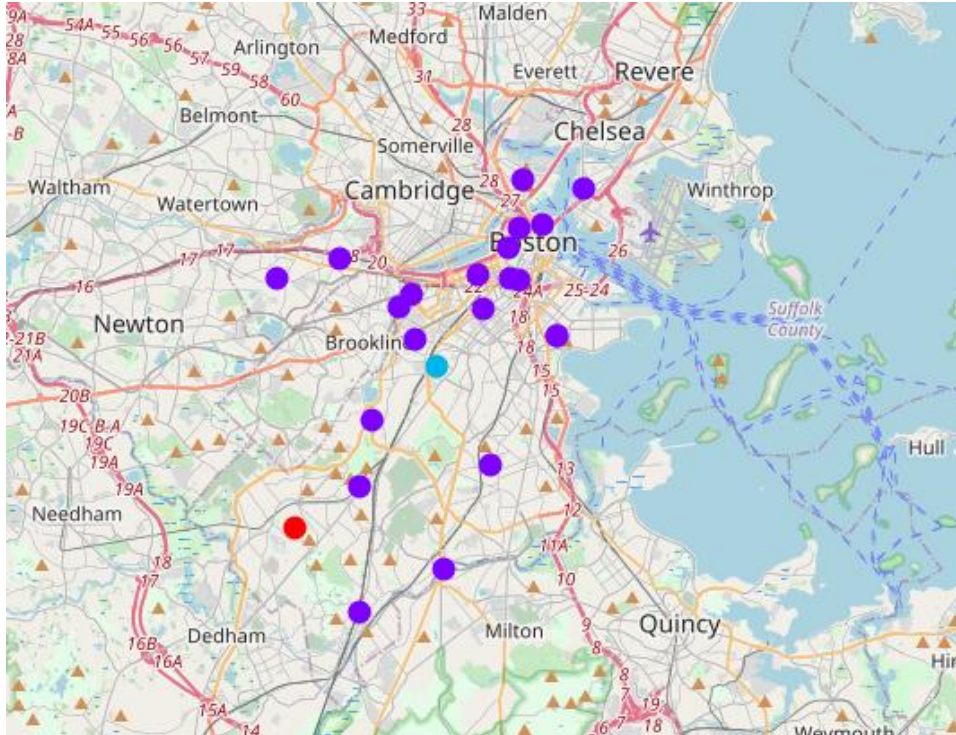
The final sorting of the algorithm shows us below, most common venues in each neighborhood. Each neighborhood as its own tastes and what seems to be common in that neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allston	Korean Restaurant	Chinese Restaurant	Pizza Place	Bakery	Thai Restaurant
1	Back Bay	Coffee Shop	Italian Restaurant	American Restaurant	Gym / Fitness Center	Clothing Store
2	Bay Village	Hotel	Bakery	Italian Restaurant	Sandwich Place	Spa
3	Beacon Hill	Hotel Bar	Pizza Place	Italian Restaurant	Plaza	French Restaurant
4	Brighton	Pizza Place	Bank	Bakery	Dry Cleaner	Coffee Shop
5	Charlestown	Pizza Place	Coffee Shop	Gastropub	Liquor Store	Pub
6	Chinatown	Chinese Restaurant	Bakery	Asian Restaurant	Seafood Restaurant	Bubble Tea Shop
7	Dorchester	Convenience Store	Gym	Plaza	Café	Shoe Store

The top three safest neighborhoods from my analysis were.

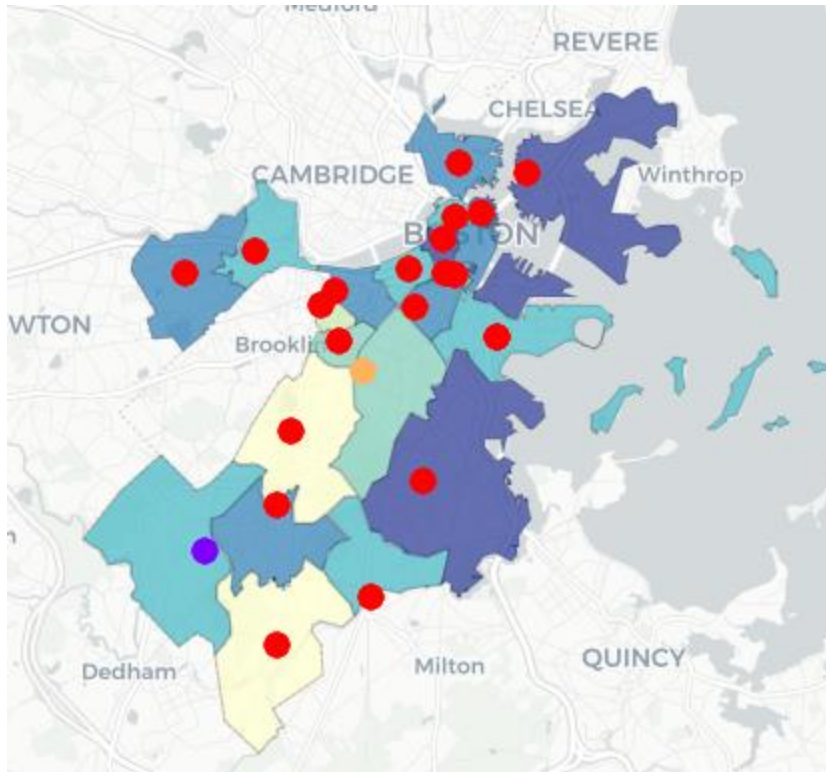
1. The Jamica plains
2. Hyde Park
3. Longwood

Not all of the neighborhoods favored the establishment of a restaurant, but some of them had a high favorability for them. We know how each neighborhood was different from the other. The map below shows us that.



5 Discussion

Even though certain neighborhoods tended to favor certain types of venues, the effect of crime in the neighborhood cannot completely be glossed over. The high crime area may not necessarily be unfavorable to a restaurant based on patronage. The decision lies risk and reward analysis for each investor. The models and the analysis could have been improved with more data, which leaves this analysis for further improvement and updating.



The neighborhood of Roslindale will be a good place to setup a pizza place, but it is also one of the neighborhoods with highest crime rates in the city. The Jamica plain neighborhood is has far less crime but also not favorable for a new restaurant.

6 Conclusion

The city of Boston holds great prospects for the future in terms of new businesses. The analysis proves that. Many thriving restaurants in many other places. As already stated, the model can be further and better improved with more data. The city over all is great place to start a restaurant, but the overall decision lies with investor. Many more restaurants have a great future in the Boston area, and many more will be established even as we socialize from a social distance in this pandemic.