

# Capstone Project – The Battle of the Neighborhoods

(Leveraging location data for problem solving)

Marlise Van Dyk

5 March 2020

## 1. Introduction

### 1.1. Background

Relocating for work can be intimidating – even more so when you have to relocate to another country which you have never been to before and now you have to take your partner or family into account as well. It's an overwhelming life event that makes you start thinking - **what should I do, where do I begin and how do I approach this?** The idea for this Capstone Project is to show how leveraging location data from FourSquare and other data sources can assist you with making decisions based on your individual needs when having to relocate.

The approach will be to follow and apply the systematic data science methodology to our scenario where I will:

1. Understand the problem and identify our approach
2. Identify the required data
3. Collect and understand the data
4. Prepare the data
5. Analyze the data
6. Model the data
7. Evaluate the model

### 1.2. Problem

In this scenario the individual needs to relocate to Boston, Massachusetts, United States where he will be working. He needs to take his partner (who is a chef) into consideration as she will be relocating with him and will also have to look for a new job. He also needs to take his child into consideration for a school.

He starts off by listing his and his family's basic wants and needs to identify how to approach the problem and to identify what data is required.

Basic Needs	Family Wants
Security	To live in a safe environment
Close to working locations	Must have loads of restaurants in the area, Close to Museum
Public Elementary school	Neighborhood close to school

The individual, who works at The Museum of Science, will have to identify a neighborhood to live in, taking into consideration the following key elements. It should be in a **safe environment, reasonably located from his work**, but also be in an **area where there are many restaurants** to provide a wide opportunity range for his partner to look for a job and have a **public Elementary school in the vicinity** for his kid. Our approach will be to predict the best neighborhood to live in based on the data gathered.

### 1.3. Interest

This solution is targeted to assist any individual who must relocate to another country, to identify ideal locations of which neighborhood to live in whilst taking more than one person i.e. the family's or partner needs into account.

## 2. Data

### 2.1. Data acquisition

From the basic wants and needs list in the scenario description the following data is required

1. [Crime data](#)
2. [School data](#)
3. FourSquare data on restaurants and work location

Open source data from [Analyze Boston](#) is obtained which is the City of Boston's open data hub. From here the crime data is obtained as well as school data. This will assist in identifying safe neighborhoods to live in as well as where the options are for Elementary schools.

The FourSquare API to query is used for geographical data on restaurants. It will be ideal to stay in the neighborhood or close to the neighborhood where there are many restaurants as a potential work option for the wife. FourSquare is also used to identify the location of the museum so that it can be taken into consideration when identifying possible neighborhoods to minimize extensive travel where possible.

For each dataset used in this study a similar approach is followed to first explain where the data comes from, what is contained in the data and how the data is prepared in order to start working with a clean dataset.

## **2.2. Data Cleaning and Feature Selection**

### **2.2.1. Boston Crime Data**

The Boston Crime dataset can be downloaded from the [Analyze Boston data portal](#). This is a dataset containing records for crime incident reports as provided by the Boston Police Department and includes types of incidents as well as when and where it occurred.

This data will allow to identify lower crime neighborhoods and to consider these as possible options for neighborhood selections. The libraries required to extract the data is installed and the data is read into a pandas dataframe.

From the dataframe there is information such as incident number, type of offense, district, date and location. Not all of these attributes in the dataset are required. Thus the following data was imported to create a new dataframe as we only need to know the type of offense and where and when it occurred. This data will allow to monitor the trend with regards to the crime in the neighborhoods.

- Offense Code Group
- Offense Description
- Year
- Month
- Street
- Occurred on date
- Lat
- Long

An example of the Crime Dataframe with only the required data can be seen in Figure 1 below after the unused columns have been removed.

	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	YEAR	MONTH	STREET	OCCURRED_ON_DATE	Lat	Long	Location
0	NaN	ASSAULT - AGGRAVATED	2019	10	RIVERVIEW DR	2019-10-16 00:00:00	NaN	NaN	(0.00000000, 0.00000000)
1	NaN	VERBAL DISPUTE	2019	12	DAY ST	2019-12-20 03:08:00	42.325122	-71.107779	(42.32512200, -71.10777900)
2	NaN	INVESTIGATE PERSON	2019	10	GIBSON ST	2019-10-23 00:00:00	42.297555	-71.059709	(42.29755500, -71.05970900)
3	NaN	WARRANT ARREST - OUTSIDE OF BOSTON WARRANT	2019	11	BROOKS ST	2019-11-22 07:50:00	42.355120	-71.162678	(42.35512000, -71.16267800)
4	NaN	SICK ASSIST	2019	11	WASHINGTON ST	2019-11-05 18:00:00	42.309718	-71.104294	(42.30971800, -71.10429400)
...	...	...	...	...	...	...	...	...	...
463072	NaN	INVESTIGATE PERSON	2020	1	HARRISON AVE	2020-01-05 00:00:00	42.339541	-71.069408	(42.33954100, -71.06940800)
463073	NaN	DISTURBING THE PEACE/ DISORDERLY CONDUCT/ GATH...	2020	1	HANOVER ST	2020-01-01 01:02:00	42.364167	-71.054070	(42.36416700, -71.05407000)
463074	NaN	WARRANT ARREST - OUTSIDE OF BOSTON WARRANT	2019	11	NaN	2019-11-30 21:00:00	42.360866	-71.061316	(42.36086600, -71.06131600)
463075	NaN	INVESTIGATE PERSON	2019	11	HYDE PARK AVE	2019-11-25 16:30:00	42.256215	-71.124019	(42.25621500, -71.12401900)
463076	NaN	THREATS TO DO BODILY HARM	2019	11	MORA ST	2019-11-12 12:00:00	42.282081	-71.073648	(42.28208100, -71.07364800)

463077 rows × 9 columns

Figure 1: Crime Dataframe

The dataframe currently consists of 9 columns and 463 077 entries. The data is then processed and all rows containing entries with no data is removed, reducing the entry amount to 396 320.

A list of the crime groups is then obtained and crimes not considered serious are removed from the dataset reducing the size to 124 851 entries. Offenses not considered as serious crimes are for example Investigate Person, Medical Assistance, Investigate Property, Warrant for arrest, Animal Incidents, Violations etc.

All related incidents were also renamed to form a single category i.e. Other Burglary and Residential Burglary will become just Burglary and Simple Assault will become Assault etc.

Now that the data cleaned up the Offense Description won't be required anymore and thus the column is removed. The Crime data consisting of 124 851 rows and 8 columns is now ready to be analyzed.

### 2.2.2. Boston School Data

The Boston School dataset can be downloaded from the [Analyze Boston data portal](#). This dataset gives general information about each school building, the type of schools as well as location data. This dataset contains information more pertaining to school building investments, but contains much more information than the standard public school dataset which might be required and thus is the dataset of choice. Knowing the type of school as well as the location will assist in deciding the neighborhood to live in.

The data is first extracted and read into a pandas dataframe. From the dataset, the information seems quite cryptic, but there are additional documents on the website that describes the data keys. This is used to assist in extracting the required data and renaming the columns to a better description.

From the information extracted there are 141 data entries and 251 columns. The following describes the data keys and associated information required:

- BPS\_School\_Name: School Name
- BPS\_Address: School Address
- BRA\_Neighborhood: School Neighborhood
- SMMA\_latitude: Latitude
- SMMA\_longitude: Longitude
- SMMA\_Typology: Type of School

Columns not relating to the aforementioned information is removed from the dataset and are given more meaningful headers. The data is then further filtered to only supply information with regards to Elementary Schools. From the 141 schools in the Boston area, there are only 48 Elementary schools. This dataset is now ready to be analyzed. A segment of the dataset can be seen in Figure 2.

The new dataframe size is: (48, 6)

	School_Name	Address	Neighborhood	Latitude	Longitude	Type
0	Adams, Samuel Elementary	165 Webster St East Boston, MA 02128	East Boston	42.365553	-71.034917	Elementary School
1	Alighieri, Dante Montessori School	37 Gove Street East Boston, MA 02128	East Boston	42.371565	-71.037608	Elementary School
4	Bates, Phineas Elementary	426 Beech St Roslindale, MA 02131	Roslindale	42.277663	-71.135353	Elementary School
5	Beethoven, Ludwig Van Elementary	5125 Washington St West Roxbury, MA 02132	West Roxbury	42.263520	-71.155824	Elementary School
6	Blackstone, William Elementary	380 Shawmut Ave Boston, MA 02118	South End	42.341012	-71.072056	Elementary School

*Figure 2: School Dataframe*

### 2.2.3. FourSquare Data

Now location data on work location and restaurants can be scraped from FourSquare which is an independent location data and technology platform. The FourSquare website is queried to also obtain additional data about the restaurants. The idea behind acquiring this data is to assist in identifying location and density of restaurants in each neighborhood and to consider the distance between working and school locations when mapped out.

Both the Museum location data and various restaurants' location data that's in the vicinity is required, as it would be ideal to have both partners' work location close to one another or close to the chosen neighborhood. The Museum's location data (i.e. coordinates) is obtained based on the known address.

Boston's size is approximately 232 km<sup>2</sup>, so to attempt to capture all possible restaurant options, a query is defined to search for restaurants within a 100km radius from the museum and to transform the data into a pandas dataframe. We will initially keep the range quite wide as not to limit options when only cleaning the data.

The dataframe is then cleaned to only include sensible information pertaining to restaurants and location. The category range remains quite wide initially, as not too limit potential options

for places for a chef too work at. Any city information not pertaining to Boston was also removed. An example of the cleaned and ready to be analyzed Restaurant dataset can be seen in Figure 3.

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	neighborhood	id
2	Q Restaurant	Hotpot Restaurant	660 Washington St	42.351707	-71.062715	[{"label": "display", "lat": 42.35170662953259...	1723	02111	US	Boston	MA	United States	[660 Washington St (at Beach St), Boston, MA ...	NaN	4c91546ab641236ab6a68079
3	Billy Tse's Restaurant	Sushi Restaurant	240 Commercial St	42.363832	-71.051115	[{"label": "display", "lat": 42.36383163608474...	1405	02109	US	Boston	MA	United States	[240 Commercial St, Boston, MA 02109, United States]	NaN	4b1af6bd9f64a52030cf623e3
4	Great Taste Bakery & Restaurant	Bakery	31 Beach St	42.351291	-71.060165	[{"label": "display", "lat": 42.35129067813932...	1828	02111	US	Boston	MA	United States	[31 Beach St, Boston, MA 02111, United States]	NaN	48e310cef964a520399021e3
6	Montien Boston - Thai Restaurant	Thai Restaurant	63 Stuart St	42.351094	-71.064498	[{"label": "display", "lat": 42.35109416020406...	1761	02116	US	Boston	MA	United States	[63 Stuart St (Tremont St), Boston, MA 02116, ...	NaN	4a04e5af9f64a5203e721fe3
7	Primo's Restaurant	Pizza Place	28 Myrtle St	42.359934	-71.065583	[{"label": "display", "lat": 42.35993273996034...	843	02114	US	Boston	MA	United States	[28 Myrtle St, Boston, MA 02114, United States]	NaN	4aa91a1af964a5203e5120e3
8	Thornton's Restaurant & Cafe	Diner	150 Huntington Ave	42.345288	-71.082010	[{"label": "display", "lat": 42.34528762816119...	2660	02115	US	Boston	MA	United States	[150 Huntington Ave (at W Newton St), Boston, ...	NaN	4aec58d2f964a52035c621e3
9	Pulcinella Mozzarella Bar and Restaurant	Italian Restaurant	78 Salem St	42.363693	-71.055872	[{"label": "display", "lat": 42.36369287757674...	1093	02113	US	Boston	MA	United States	[78 Salem St, Boston, MA 02113, United States]	NaN	502bd747e4b0802cd00820d
10	Last Corner Restaurant	Diner	49 High Street	42.376570	-71.063172	[{"label": "display", "lat": 42.37657, "lng": ...	1157	01867	US	Boston	MA	United States	[49 High Street (Chute St), Boston, MA 01867, ...	NaN	4b57b3679f64a520af93c28e3
11	New Jumbo Seafood Restaurant	Chinese Restaurant	5 Hudson St	42.350902	-71.059895	[{"label": "display", "lat": 42.35090215936083...	1876	02111	US	Boston	MA	United States	[5 Hudson St, Boston, MA 02111, United States]	NaN	4b08321cf964a520f05323e3
12	The Causeway Restaurant and Pub	BBQ Joint	65 Causeway St	42.364659	-71.062912	[{"label": "display", "lat": 42.3646590394404...	459	02114	US	Boston	MA	United States	[65 Causeway St (Lancaster), Boston, MA 02114, ...	NaN	53407e949e5cc70137182b
15	Moon Villa Restaurant	Chinese Restaurant	19 Edinboro St	42.351884	-71.095954	[{"label": "display", "lat": 42.35188354712937...	1784	02111	US	Boston	MA	United States	[19 Edinboro St (Near Beach St), Boston, MA 0...	NaN	40b21286f964a52094b63ce3
19	Toro Restaurant	Tapas Restaurant	1704 Washington Street	42.336988	-71.075924	[{"label": "display", "lat": 42.33698788, "lng": ...	3378	02118	US	Boston	MA	United States	[1704 Washington Street (at Massachusetts Ave)...	NaN	43e9e7ef9f64a520202f1fe3
21	International Restaurant & Pub	Restaurant	184 High St	42.357344	-71.052514	NaN	1631	02110	US	Boston	MA	United States	[184 High St, Boston, MA 02110, United States]	NaN	40b28c80f964a520bf771ee3
22	Carrie Nation Restaurant & Cocktail Club	American Restaurant	11 Beacon St	42.358316	-71.061458	[{"label": "display", "lat": 42.35831564666699...	1070	02108	US	Boston	MA	United States	[11 Beacon St (at Somerset St), Boston, MA 021...	NaN	518564e649e25e89f4c2679
23	Saus Restaurant	Belgian Restaurant	33 Union St	42.361076	-71.057054	[{"label": "display", "lat": 42.36107632380033...	1081	02108	US	Boston	MA	United States	[33 Union St (Marsh Lane), Boston, MA 02108, U...	NaN	4b0f18c43506e3bea63bd22
24	Amrheins Restaurant	Bar	80 W Broadway	42.341816	-71.055311	[{"label": "display", "lat": 42.34181624208448...	2956	02127	US	Boston	MA	United States	[80 W Broadway, South Boston, MA 02127, United...	NaN	487783e0f964a520b7e41fe3
28	Pearl Villa Restaurant	Chinese Restaurant	25 Tyler St	42.350690	-71.061010	[{"label": "display", "lat": 42.35069, "lng": ...	1869	02111	US	Boston	MA	United States	[25 Tyler St (at Kneeland St), Boston, MA 0211...	NaN	4af4e050f964a520a721e3
29	No Name Restaurant	Seafood Restaurant	15 Fish Pier St W	42.350384	-71.038321	[{"label": "display", "lat": 42.35038402915641...	3028	02210	US	Boston	MA	United States	[15 Fish Pier St W, Boston, MA 02210, United S...	NaN	4b9011b79f64a520a87333e3
31	Tosceno Restaurant	Italian Restaurant	47 Charles St	42.357360	-71.070026	[{"label": "display", "lat": 42.35736019165299...	1060	02114	US	Boston	MA	United States	[47 Charles St, Boston, MA 02114, United States]	NaN	3f6b6200f964a5204dec1ee3

Figure 3: Restaurant Dataframe

## 3. Methodology

### 3.1. Exploratory Data Analysis

Now that the data is cleaned up it's ready to be explored and analyzed in order to obtain insights. The data will be analyzed in the following order:

1. Crime data
2. School data
3. FourSquare data

A map of Boston (Figure 4) is created as a base map to superimpose other data and findings onto to visually assist in analyzing the data

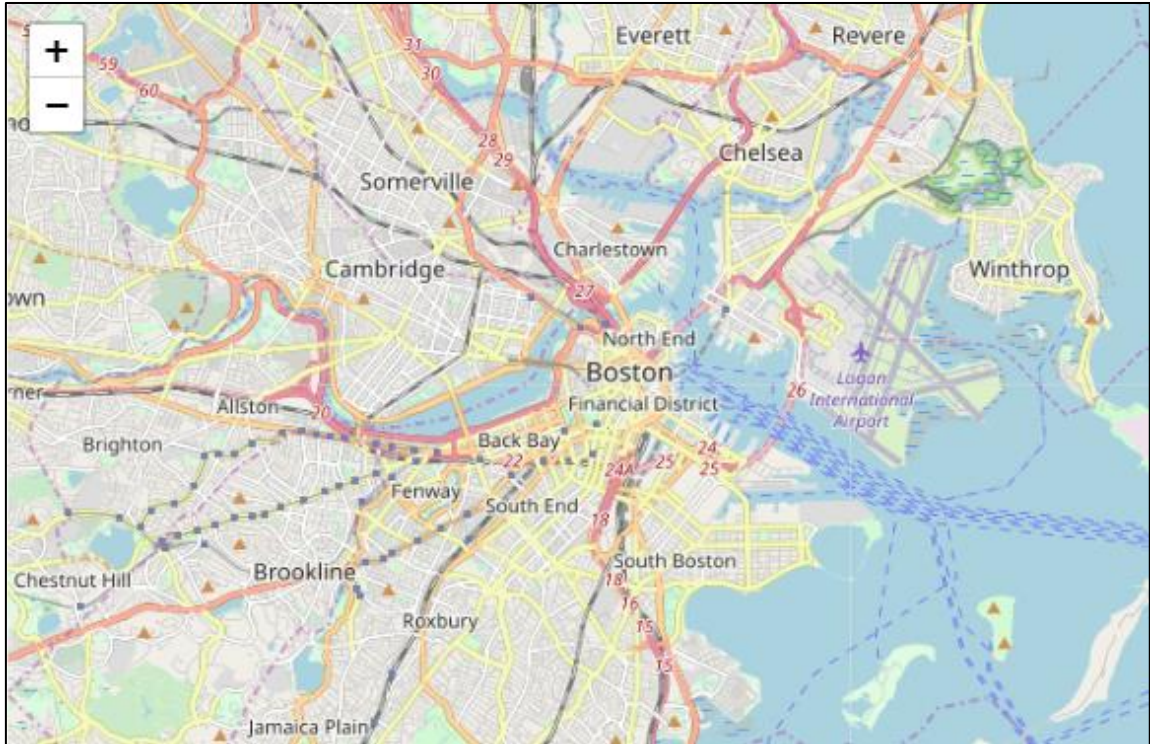


Figure 4: Boston base map

### 3.1.1. Crime Data

As there are so many entries of crime data, the entries were plotted on a scatter plot of latitude vs. longitude to see if there are any further outliers after cleaning the dataset (Figure 5).

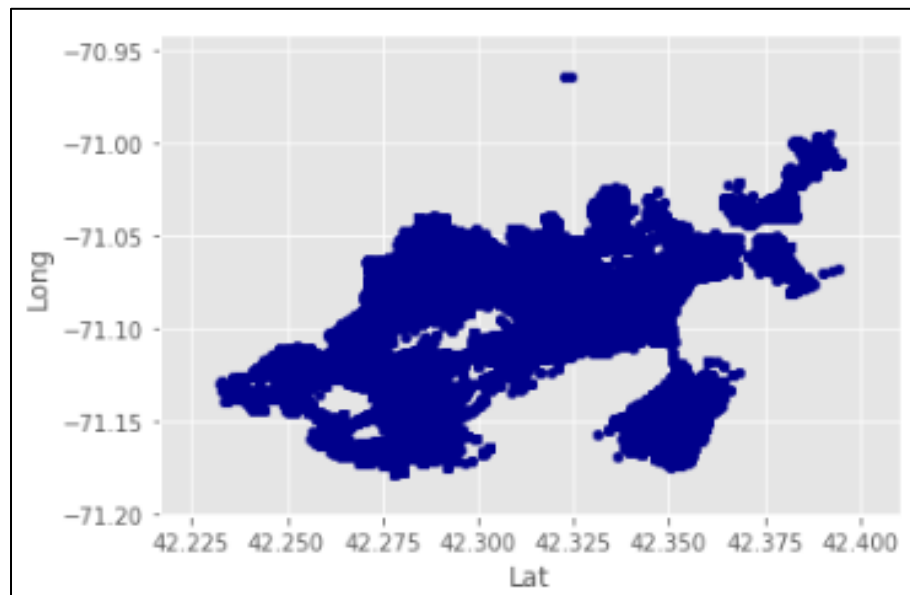


Figure 5: Scatter plot to locate outliers of crime data



The outlying coordinates (~42.3"N,70.96"W) we're identified to still be around the Boston area (more around the nearby islands) and thus remained in the dataset. The rest of the data look well grouped to the Boston location.

Crime trends over the supplied time period, from 2015 to 2019, was not provided and had to be calculated. The top crime streets were identified to be Washington Street followed by Boylston Street.

GroupByStreet	
STREET	
WASHINGTON ST	6182
BOYLSTON ST	4341
BLUE HILL AVE	2699
TREMONT ST	2238
DORCHESTER AVE	2170
MASSACHUSETTS AVE	1957
HARRISON AVE	1861
HUNTINGTON AVE	1808
COMMONWEALTH AVE	1780
NEWBURY ST	1653
CENTRE ST	1507
RIVER ST	1181
COLUMBIA RD	1173
HYDE PARK AVE	955
COLUMBUS AVE	943
WARREN ST	899
DUDLEY ST	890
BEACON ST	815
SUMMER ST	760
BOWDOIN ST	716

Figure 6: Top Boston crime streets



The data was then grouped by type of crime from where it became apparent that the top crime type in Boston is larceny, followed by assault and then vandalism.

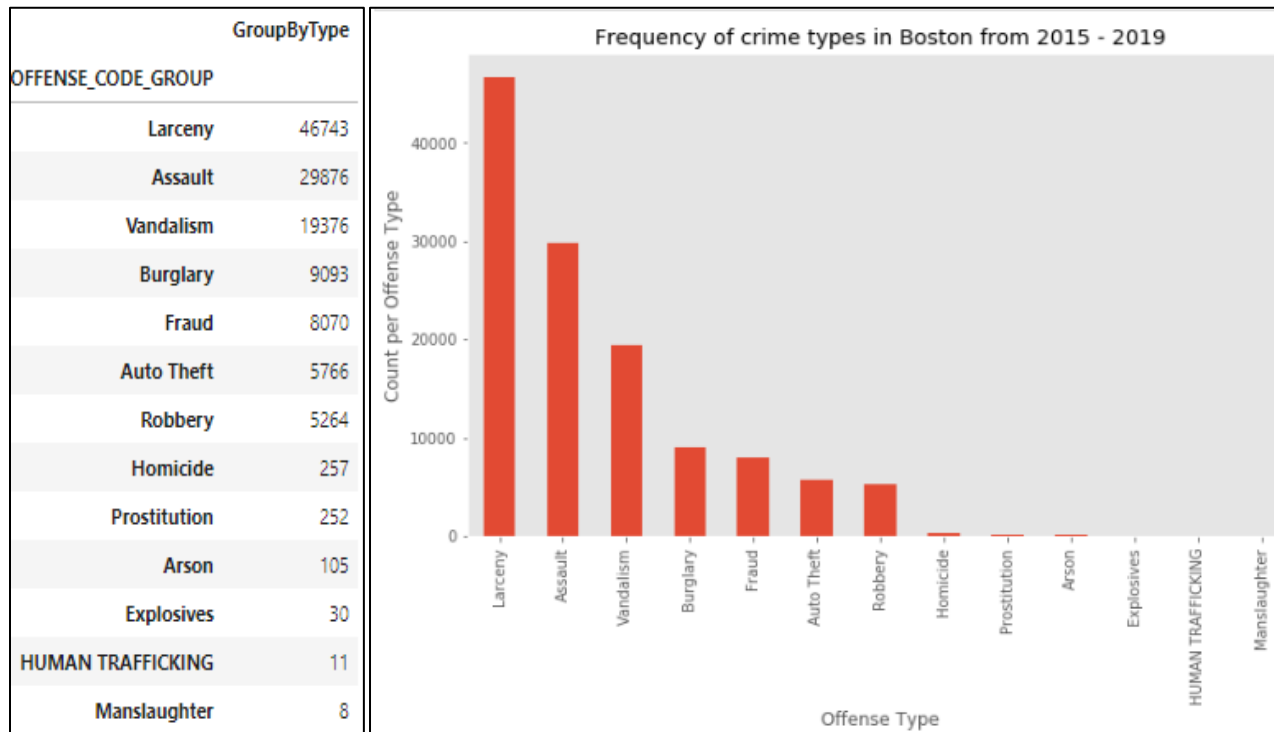


Figure 7: Boston top crime types

The number of crimes for each year was plotted to determine the crime trend.

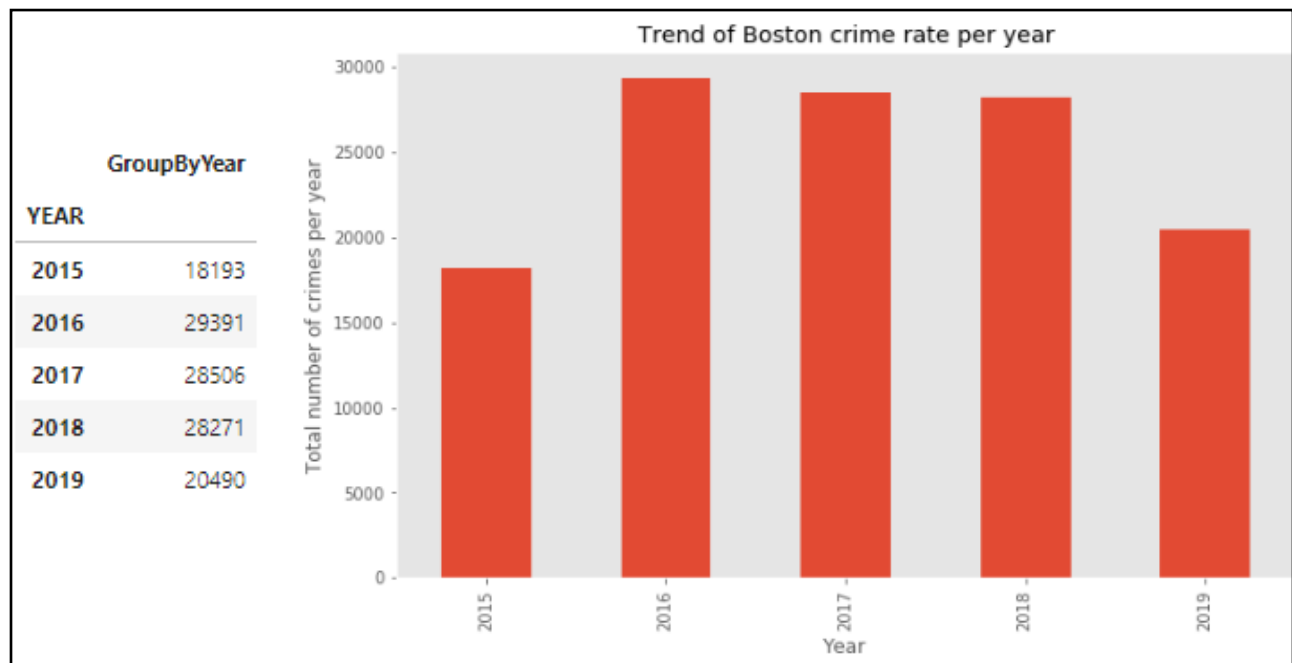


Figure 8: Boston crime trend from 2015 - 2019

From Figure 8 that shows the Boston crime trend, it can be seen that although there has been a significant increase in crime from 2015 to 2016, the crime rate has almost reverted back in 2019 to the rate seen in 2015 and the trend currently indicates a decrease in crime.

The top 7 crimes were then evaluated, and it was identified, as seen in Figure 9, that the top 3 contributing crimes (assault, larceny and vandalism) has been the main cause for the increase in crime rate for the period 2016-2018, but decreased in 2019. Comparing the lower crime rate years, 2015 and 2019, it is apparent that only burglary and robbery rates are lower in 2019 whereas all the other crime type rates have either remained the same or increased.

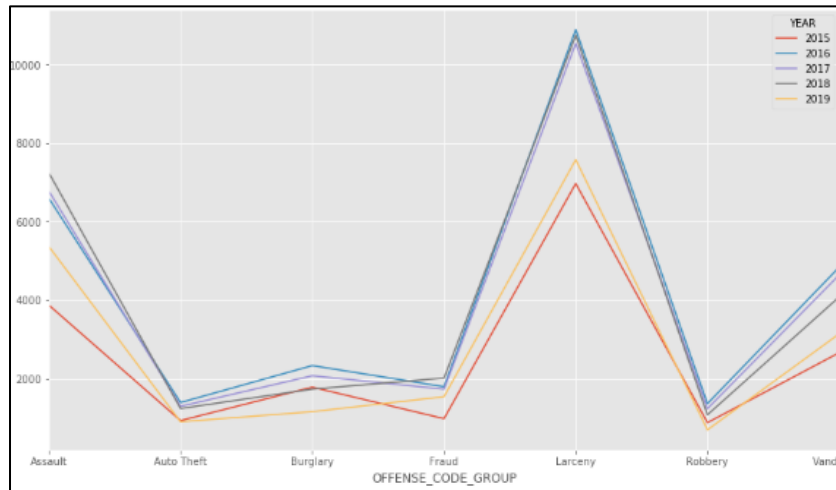


Figure 9: Top 7 crimes per year

The data was then further reduced indicate current or latest trends and only included data for the year 2019, this reduced the dataset to 20 490 entries. The data was then grouped per coordinates to indicate based on the most recent data the high crime streets in Boston. This data was then superimposed on the base Boston map to visually identify high crime areas.

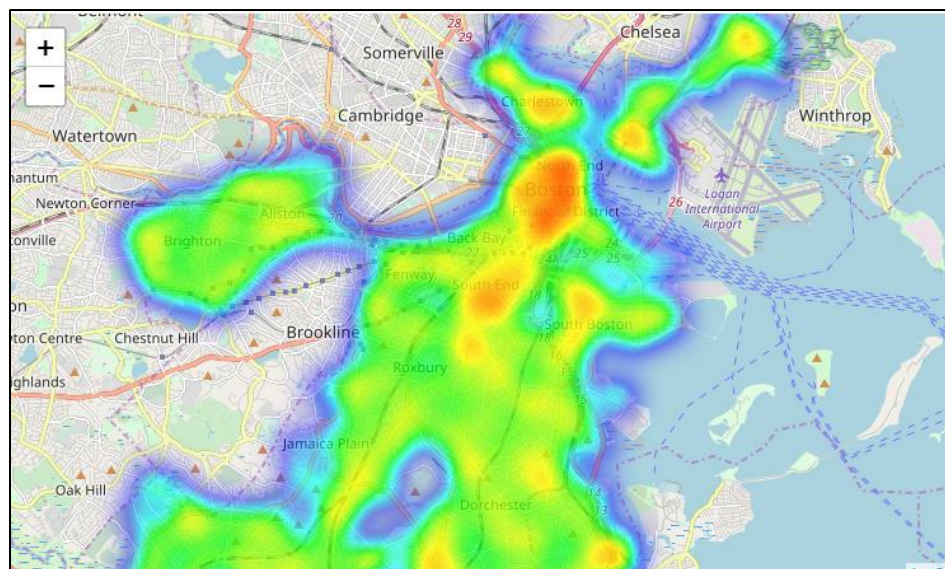


Figure 10: Heat map of 2019 Boston crimes

### 3.1.2. School Data

For the school data it was ideal to highlight which neighborhood contains more than one elementary school. The schools were thus grouped by neighborhood and to display the density of schools per location visually, the data was overlaid as purple markers on the Boston base map which now included the crime data as well.

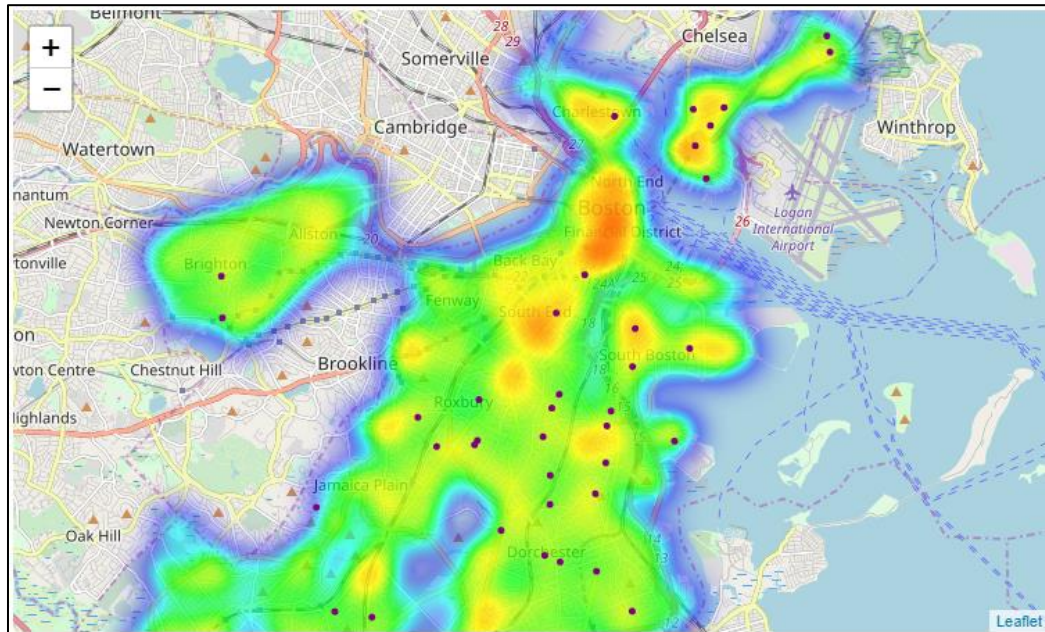


Figure 11: Boston elementary schools superimposed on the the crime heat map to display density

### 3.1.3. FourSquare Data

The Museum's location was then marked on the Boston base map as this will identify the center location to take into consideration when choosing a neighborhood to live in.

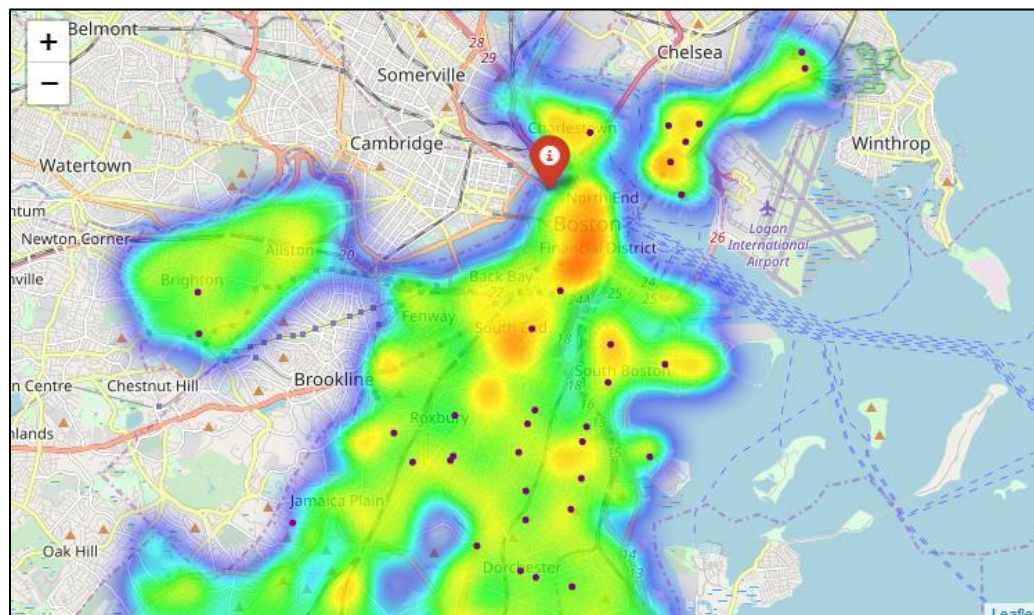


Figure 12: Red museum location marker added to the map

The data of the restaurants are then described below. It can be seen that there are 27 entries with the mean location at coordinates latitude = 42.356070, longitude = -71.058684 and that the general distance of the restaurants from the museum is approximately 1km. The furthest restaurant data entry is about 4.3km from the museum location.

	lat	lng	distance
count	27.000000	27.000000	27.000000
mean	42.356070	-71.058684	1915.740741
std	0.011424	0.013074	1000.221191
min	42.331051	-71.084968	459.000000
25%	42.350796	-71.063835	1082.500000
50%	42.357344	-71.060165	1726.000000
75%	42.362485	-71.055591	2654.000000
max	42.379929	-71.027055	4342.000000

Figure 13: Restaurant data description

The various types of restaurants in the data are then identified. It is ideal to see the top three restaurant type spread across Boston, as this will allow a chef to see the each location's desired cuisine allowing for an additional variable to take into consideration when choosing a neighborhood.

	name
categories	
Italian Restaurant	4
Chinese Restaurant	3
American Restaurant	2
Bar	2
Pizza Place	2

Figure 14: Restaurant types

The top 3 restaurant types were categorized (American – 1, Italian – 2 and Chines – 3) and the three types were plotted against each latitude and longitude.



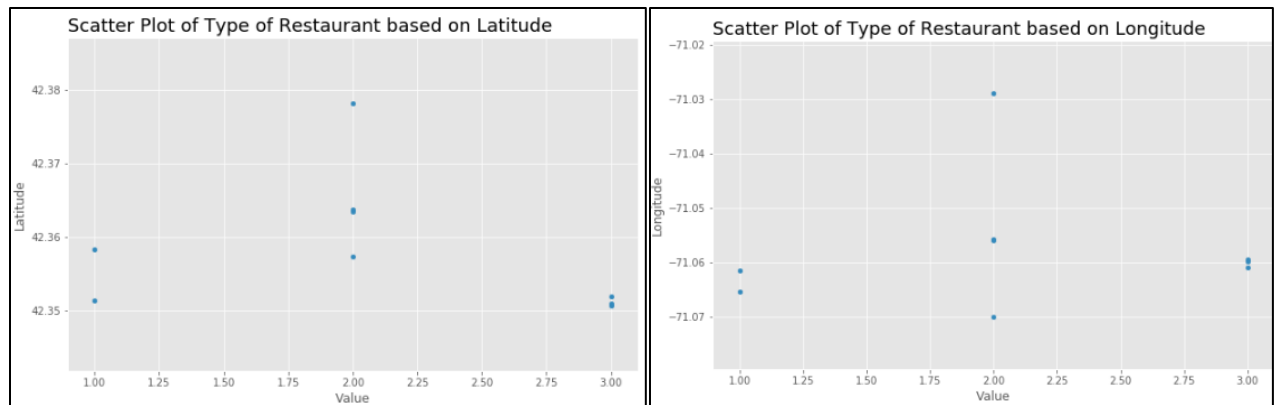


Figure 15: Restaurant types categorized by coordinates

From figure 15 it's visible that restaurant type 3, Chinese cuisine, is quite clustered together. This is confirmed when visualized on a map (figure 16 and figure 17).

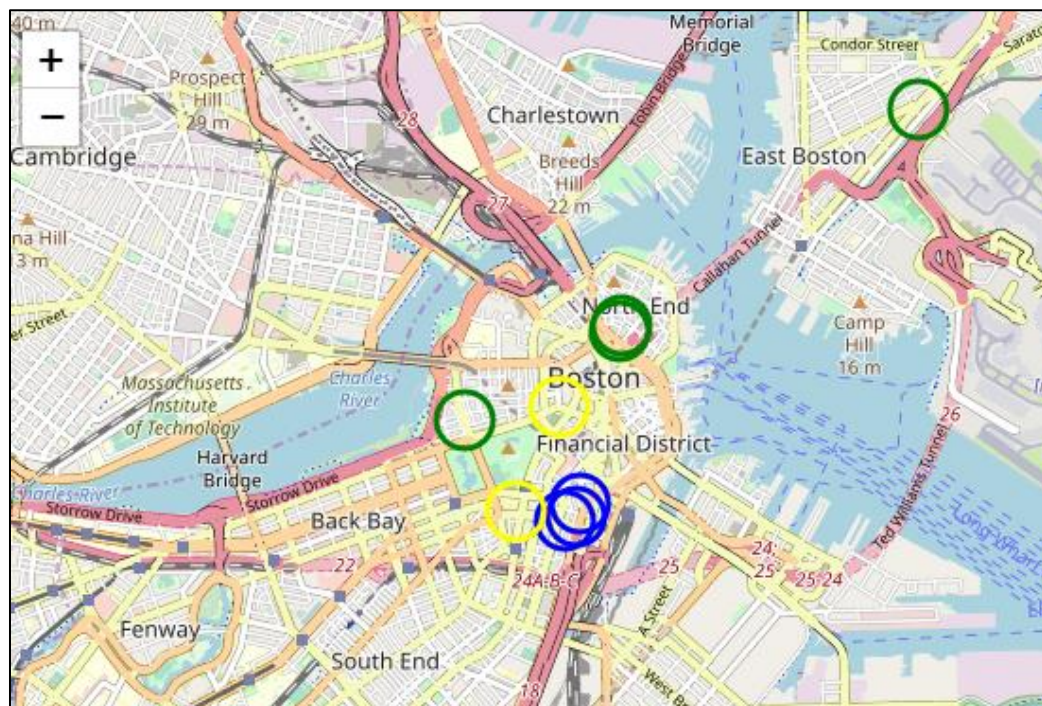


Figure 16: Boston Restaurant types: Yellow - American, Green - Italian, Blue - Chinese

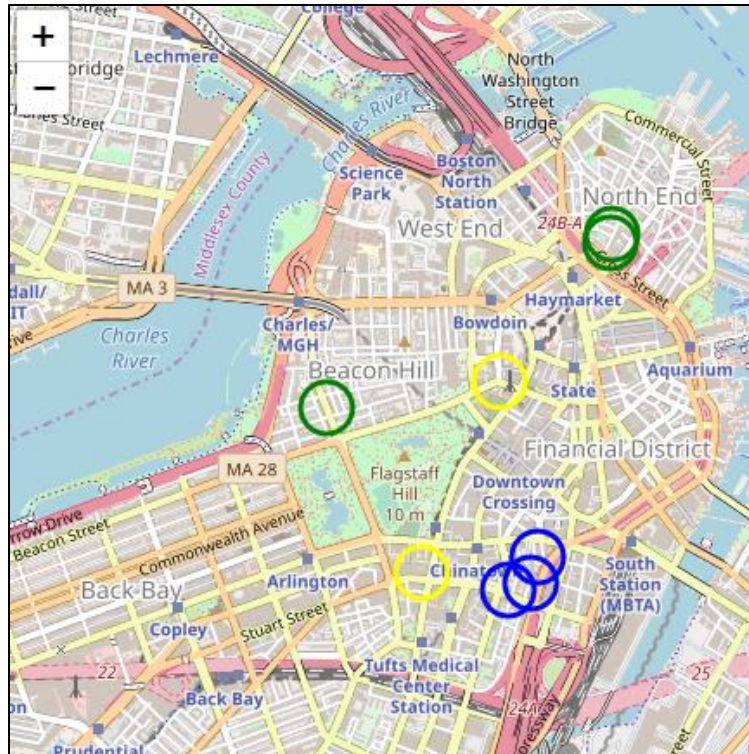


Figure 17: Zoomed in view of Boston restaurant types

Not wanting to be limited to restaurant cuisine type, all restaurant locations are overlaid on the Boston map figure 18.

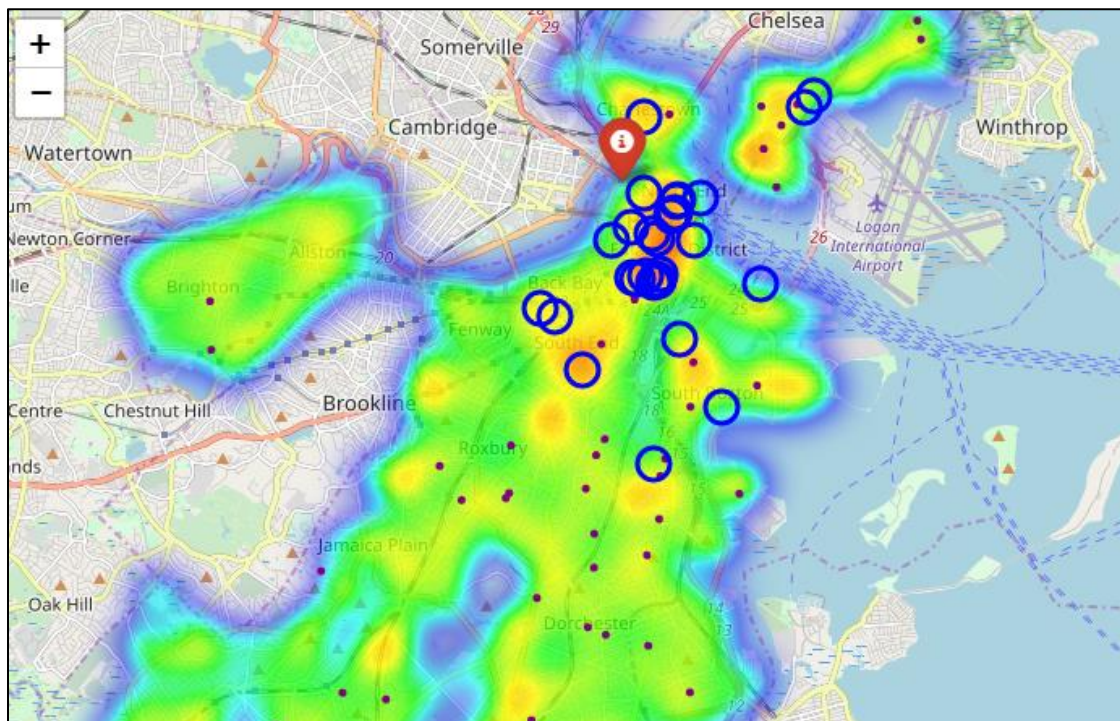


Figure 18: Restaurants mapped as circle markers

Looking at the Boston map containing all the required data (figure 18), it is visible that it's not that easy to identify an ideal neighborhood based on mapping all the preferences on a single map. Therefore, predictive modeling is required to assist in identifying an ideal neighborhood.

### 3.2. Predictive Modeling

The type of model that will assist in solving the problem needs to be identified. As the intent is to identify an area or neighborhood that features most of the location preferences, clustering is recognized as the ideal exploratory data analysis technique.

Clustering is the grouping of data based on shared characteristics. One type of partitioned based clustering algorithm is K-means Clustering. K-means is an iterative algorithm that uses distance-based measurements to determine similarity between data points. It's one of the simplest and most popular unsupervised machine learning algorithms.

The data to use for the model needs to be prepared to only include numerical data. K-means clustering will be used to identify ideal neighborhoods, so that indicates focusing on positive traits and thus only restaurant and school data will be used and the crime data will be excluded. The restaurant and school data is combined into a new dataframe and the dataframe is cleaned up to only include numerical data.

First the cluster centroids need to be determined. The Euclidean distance for each point is calculated from the centroid and an elbow curve is developed to visually show the ideal number of clusters to continue with based on the data. It can be seen from figure 19 that a good K to use in the model will be 4.



Figure 19: Determining the best K for k-means clustering



The K-means model is then developed, and the data fitted to the model. 4 Clusters are then obtained and visually marked on the Boston map with a crime heat map, the museum data and the school data in the background as seen in figure 20.

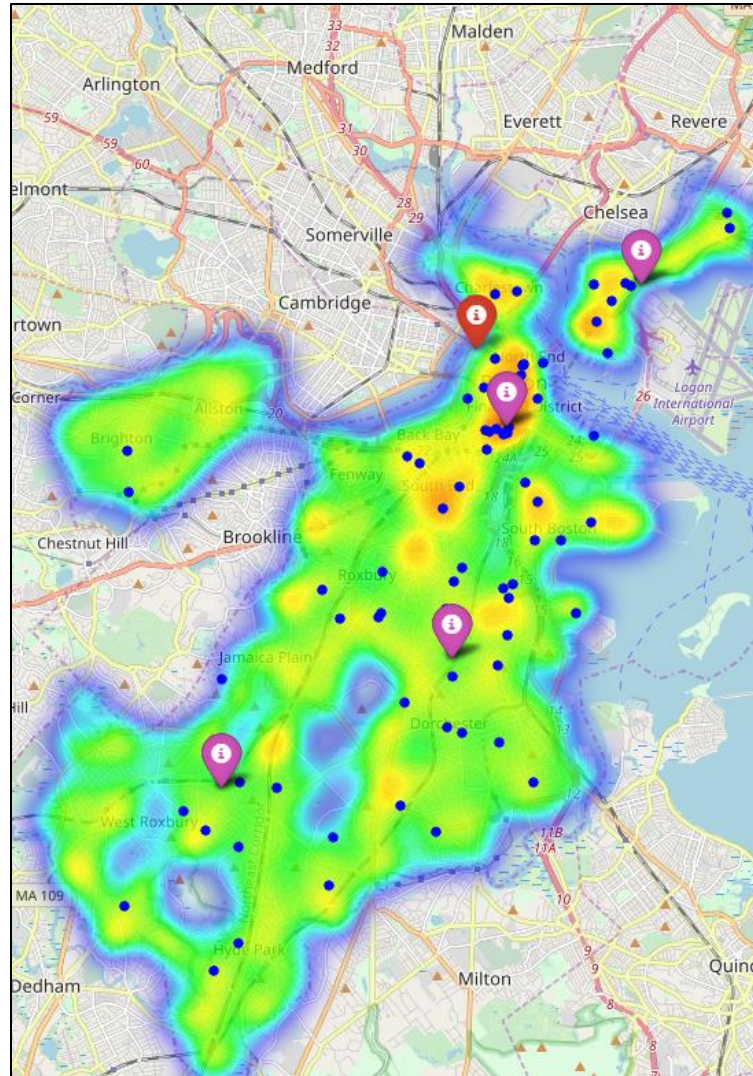


Figure 20: Ideal neighborhoods mapped out as purple markers

## 4. Results

From the results the highest crime area is located around the Financial District (downtown Boston) close by Chinatown. This would not be an ideally suited neighborhood to select.

Directing attention to the school data, it is beneficial if a neighborhood is chosen that provides multiple options, but it's important to take note that where there are more options the neighborhood might also be considerably bigger than other neighborhoods. From the school

results no conclusion can be made as yet on a preferred neighborhood as there are no immediate visible clusters.

From the restaurant data, apart from the Chinese cuisine type clustered in Chinatown, there seems to be quite a variety of restaurants with various cuisine types all over Boston allowing for a variety of restaurants for a chef to work at. This data on its own does not provide a concrete solution for a neighborhood.

From the clustering results of desirable traits (schools and restaurants) in a location, 4 neighborhoods were marked on a heat map of crime locations. These locations can then be further evaluated based on closer inspection or a zoomed in view of the map.

As a final result, the four ideal neighborhoods to live in was marked with the following coordinates and upon further investigation the neighborhoods were determined:

- [ 42.30965123, -71.07414251] – Roxbury,
- [ 42.28570345, -71.13181767] – Roslindale,
- [ 42.35243914, -71.06025271] – Chinatown,
- [ 42.37861659, -71.0266932] – East Boston.

These markers only considered the positive attributes of neighborhoods, so when looking at figure 20 and including the crime heat map, it can be seen that Chinatown is located in a high crime location and is thus eliminated as a possible option for a neighborhood to consider.

Final desirable neighborhoods to live in based on the individual needs and location from the museum ideal neighborhoods to consider are then East Boston, Roxbury and Roslindale in no specific order.

## **5. Discussion**

The recommended neighborhoods should be considered as a starting point for further, more detailed analysis where other preferences such as housing or traffic data can then be taken into consideration. Prioritization of needs can also be applied to assist in decision making.

Even though the crime rate has decreased, crime remains everywhere and thus when finally deciding on a neighborhood, it is important to look out for high-crime locations or streets in the chosen neighborhood and to also consider the type of crime in that area.

It would also be desirable to have more data pertaining to schools (where possible) as well as restaurants.

The analysis indicates that deciding on a neighborhood to fully adhere to the parameters of the individual's needs can be tricky but doable, especially when a machine learning algorithm is implemented to point out possible neighborhoods more clearly by means of clustering.

There is definitely room for improvement on this project and with more work this can be developed into a fully pledged application to assist individuals when relocating. Additional

data can be included such as traffic or property trends and school costs etc. and a wider area can also be considered. More detail can also be supplied with regards to the final results.

## **6. Conclusion**

The purpose of this project was met to leverage location data to assist an individual in his/her decision of which neighborhood to live in when relocating to an unfamiliar location. The model developed assists in narrowing down ideal neighborhoods to select from based on stakeholder needs or preferences. Through clustering preferences, the cluster locations created is an ideal starting point for final exploration by the stakeholders. Final decision on the chosen neighborhood will be made by each stakeholder taking in their own needs and priorities into consideration.