Capstone project IBM

Hazem Alaa Shams

Introduction

- 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.

Problem

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.

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

Data

- 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.

Boston Crime Data

	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

Boston School Data

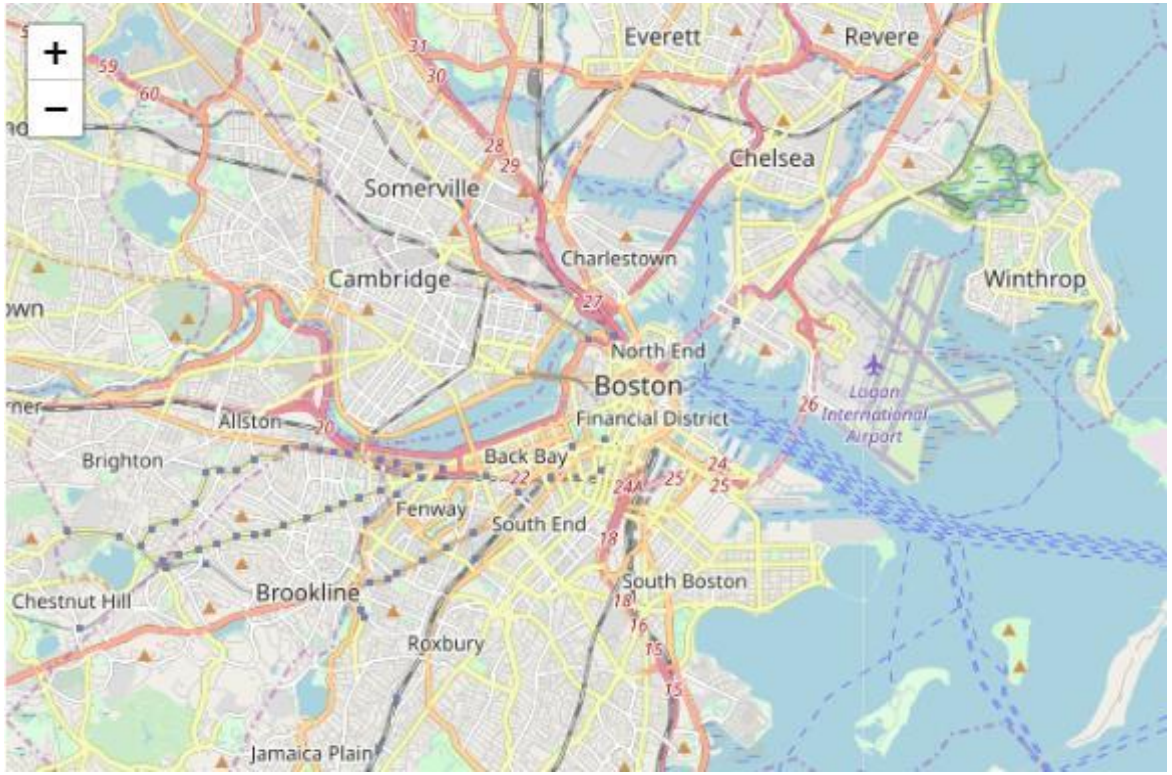
```
The new dataframe size is: (48, 6)
```

	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

FourSquare data

	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 S...	NaN	4b1afebdf964a5200cf623e3
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	4ae310cef964a520399021e3
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	4a04e5aff964a5203e721fe3
7	Primo's Restaurant	Pizza Place	28 Myrtle St	42.359324	-71.065583	[{'label': 'display', 'lat': 42.35932373996034...	843	02114	US	Boston	MA	United States	[28 Myrtle St, Boston, MA 02114, United States]	NaN	4aa91a1af964a520fe5120e3
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...	1033	02113	US	Boston	MA	United States	[78 Salem St, Boston, MA 02113, United States]	NaN	502bd747e4b082dc2d00820d
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	4b57b367f964a520af3c28e3
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	4b08321cf964a520ff0523e3
12	The Causeway Restaurant and Pub	BBQ Joint	65 Causeway St	42.364659	-71.062912	[{'label': 'display', 'lat': 42.3646590394484,...	459	02114	US	Boston	MA	United States	[65 Causeway St (Lancaster), Boston, MA 02114,...	NaN	534f07e9498e5cc70137182b
15	Moon Villa Restaurant	Chinese Restaurant	19 Edinboro St	42.351884	-71.059554	[{'label': 'display', 'lat': 42.35188354712937...	1784	02111	US	Boston	MA	United States	[19 Edinboro St (Near Beach St.), Boston, MA 0...	NaN	4bb21286f964a52094b63ce3
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	43e9e7eff964a520202f1fe3
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	40b28c80f964a520bff71ee3
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	518564e8498e25e8f94c2679
23	Saus Restaurant	Belgian Restaurant	33 Union St	42.361076	-71.057054	[{'label': 'display', 'lat': 42.36107632380333...	1081	02108	US	Boston	MA	United States	[33 Union St (Marsh Lane), Boston, MA 02108, U...	NaN	4bf18c843506ef3bea63bd22
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	4a7783edf964a520b7e41fe3
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	4af4e050f964a5202af721e3
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	4b9011b7f964a520a87333e3
31	Toscano 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	3fd66200f964a5204dec1ee3

Methodology



Boston's map

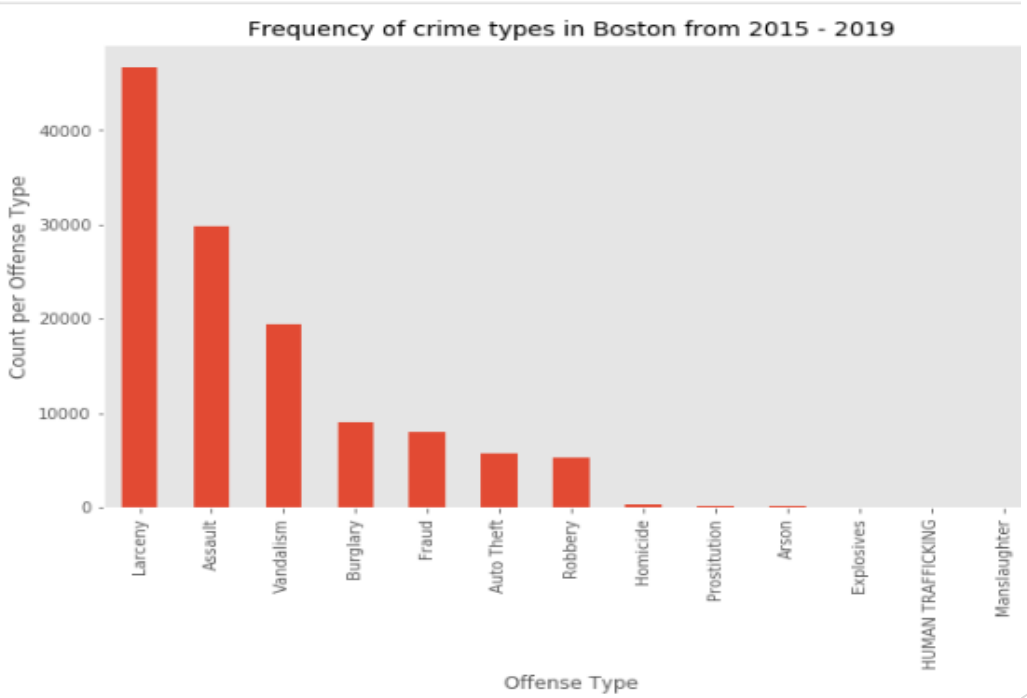
Data Analysis Now that the data is cleaned up it's ready to be explored and analyzed in order to obtain insights. The data will be analyzed in the following order:

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

Crime data

- 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.

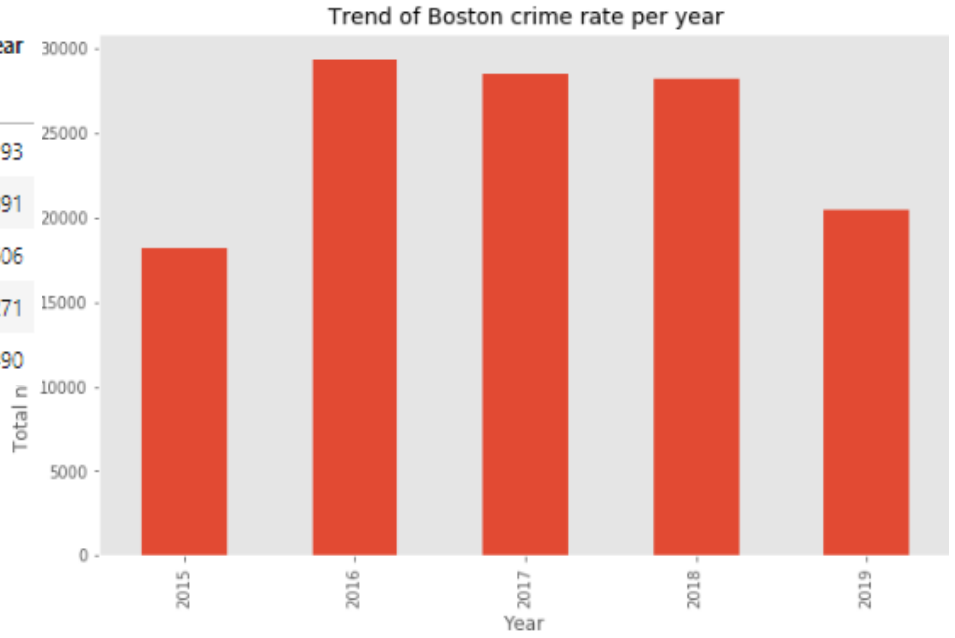
OFFENSE_CODE_GROUP	GroupByType
Larceny	46743
Assault	29876
Vandalism	19376
Burglary	9093
Fraud	8070
Auto Theft	5766
Robbery	5264
Homicide	257
Prostitution	252
Arson	105
Explosives	30
HUMAN TRAFFICKING	11
Manslaughter	8



Crime data

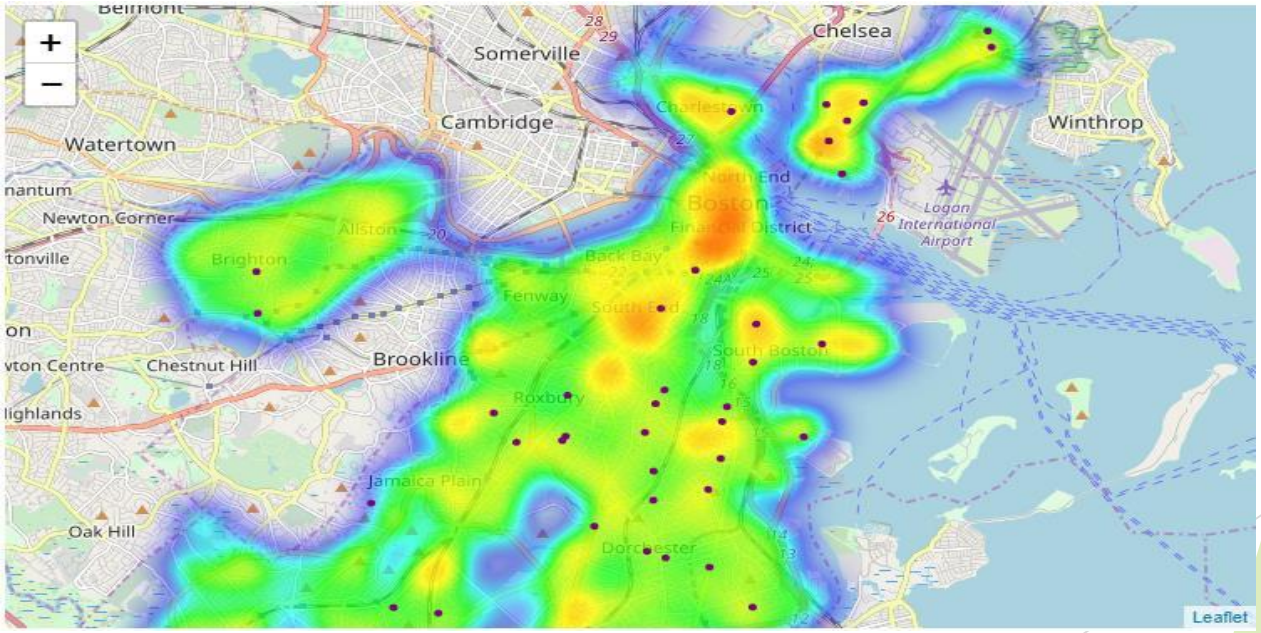
- the Boston crime trend, it can be seen that although there has been a significant increase in crime from 2015 to 2016, the crime rate has almost reverted back in 2019 to the rate seen in 2015 and the trend currently indicates a decrease in crime.

YEAR	GroupByYear
2015	18193
2016	29391
2017	28506
2018	28271
2019	20490



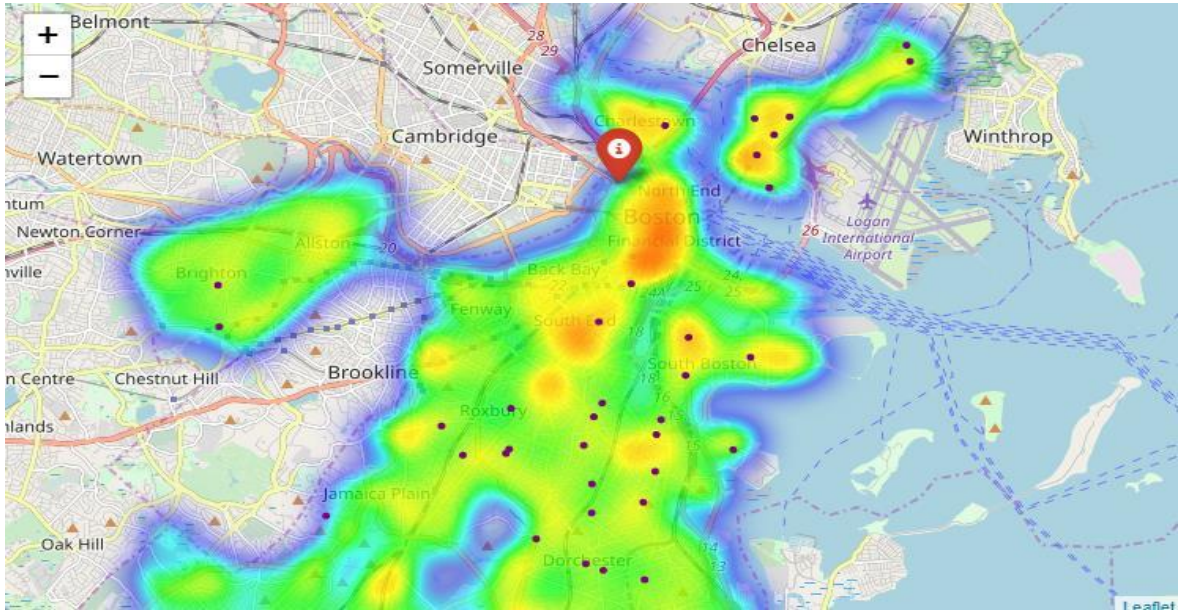
School data

- 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.



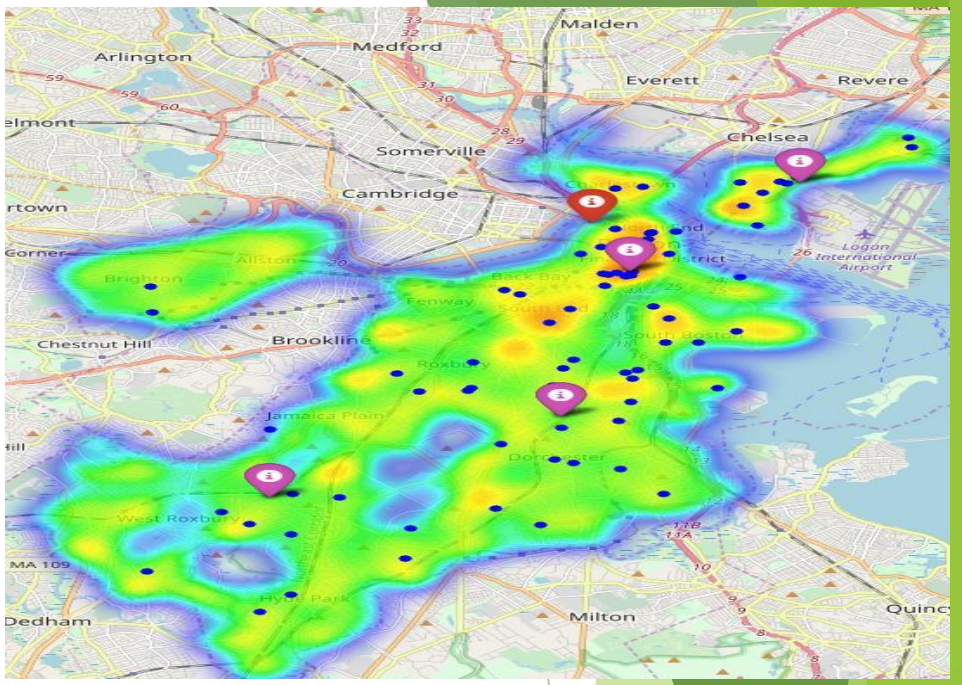
Foursquare data

- 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.



Building clusters using K-mean

- 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.
- 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 the map pic.



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 induvial needs and location from the museum ideal neighborhoods to consider are then East Boston, Roxbury and Roslindale in no specific order.

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.