CAPSTONE PROJECT - APPLIED DATA SCIENCE SPECIALIZATION

THE BATTLE OF THE NEIGHBORHOODS (WEEK-2)

INTRODUCTION:

Boston is one of the most popular cities in the United States of America, for its cultural history and significance. It covers 127 square km, and is the 21st most populous city in the US, with about 692,600 inhabitants, as of 2019. Being the capital and largest city in the Commonwealth of Massachusetts, Boston is a diverse and heritage-rich city, consisting of various locales with storied histories.

Being as well-known as it is, Boston attracts visitors, tourists and immigrants. With a lot of different destinations to visit and sight-see, it would be useful to analyse the different places of interest and landmarks, as well as the major neighbourhoods, with regard to the amenities, such as parks, hotels, restaurants and cafes, available within their vicinity.

In this analysis we wish to achieve the following objectives:

1. List and visualize major landmarks in Boston.
2. List and visualize and neighborhoods in Boston.
3. To determine the most common and widely available venues within the vicinity of a certain location.

4. To cluster different Landmarks on the basis of the venues available.
5. To cluster different Neighbourhoods on the basis of the venues available.

DATA:

To accomplish the described objectives, we need the following information:

1) Data of different Landmarks in Boston
    a) Source:https://data.boston.gov/dataset/boston-landmarks-commission-blc-landmarks
    b) Description: The dataset describes different Landmarks in the city of Boston, according to the Boston Government. This data will be used to explore Landmarks in Boston.
2) Data of different Neighborhoods in Boston
    a) Source:https://data.boston.gov/dataset/boston-neighborhoods
    b) Description: The dataset describes different Neighborhoods in the city of Boston, according to the Boston Government. This data will be used to explore the neighborhoods in Boston.
3) Venue data to different venues near the locations
    a) Source: Foursquare Developer Access to Various Venues (https://foursquare.com/).
    b) Description: With a Foursquare Developer Account, we can access the data of venues around a location, within a certain radius.
4) Geospatial Data of Boston's Neighborhoods
    a) Source:https://data.boston.gov/dataset/boston-neighborhoods
    b) Description: We use the geojson file containing the geospatial data of Boston's data to visualize the Neighborhoods.

METHODOLOGY:

Part 1 - Analysing Boston Landmarks:

1. We load the csv file found at: https://data.boston.gov/dataset/boston-landmarks-commission-blc-landmarks

and observe the data.

```
df_poi=pd.read_csv('D://ML_IET//Boston.csv', encoding='ISO-8859-1')
df_poi
```

	OBJECTID_1	OBJECTID	WARD	PARCEL	PID_LONG	WPD	BATCH	Petition	Name_of_Pr	Areas_Desi	Address	Neighborho	Date_Desig	PID
0	401	0	0	1227	0501227000	05-06-060	0	15	Arlington Street Church	Exterior only (Parcel 1227)	355 Boylston Street	Back Bay	4/25/1978	0501227000
1	402	0	0	2940	0702940000	07-09-020	0	20	James Blake House	Exterior & Interior (parcel 2942)	210 East Cottage Street	Dorchester	4/25/1978	0702940000
2	403	0	0	4662	0304662000	03-06-635	0	6	International Trust Company Building	Exterior only (parcel 4662)	45 Milk Street	Boston	4/25/1978	0304662000
3	404	0	0	3686	0203686000	02-01-135	0	59	Austin Block	Exterior only (Parcel 3686)	90-92 Main Street	Charlestown	1/31/1981	0203686000
4	405	0	0	6485	2006485000	20-18-030	0	73	Theodore Parker Unitarian Church	Exterior and windows	1851 Centre Street	West Roxbury	4/9/1985	2006485000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
164	565	0	0	3679	0903679000	09-05-115	0	259.17	Richard Bond House		88 Lambert Avenue	Roxbury		0903679000
165	566	0	0	3871	0103871000	01-03-040	0	260.17	144-146 Maverick Street		144-146 Maverick Street	East Boston		0103871000
166	567	0	0	163	1100163000	11-02-010	0	262.18	Saint James African Orthodox Church		50 Cedar Street	Roxbury		1100163000
167	568	0	0				0	229.07	Charles River Esplanade	See original application on file		Beacon Hill/Back Bay	6/23/2009	
168	569	0	0				0		Calf Pasture Pumping Station	Calf Pasture Pumping Station only		Dorchester		

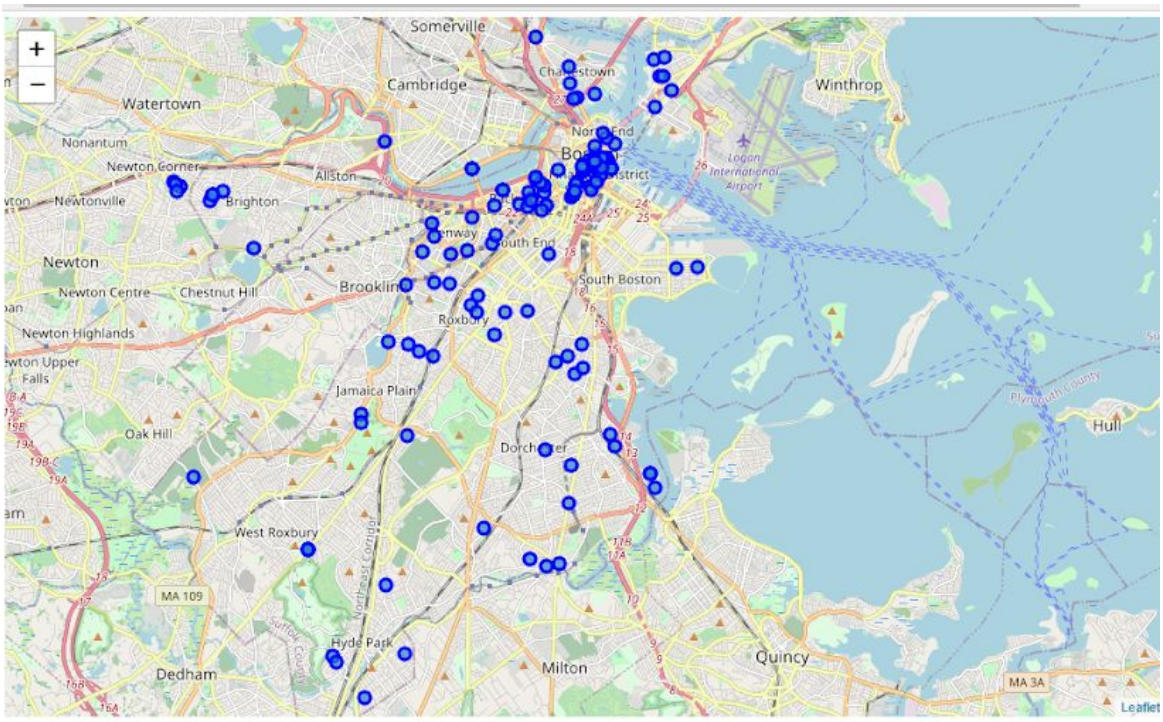
169 rows × 18 columns

2. We drop certain columns and rows that are unnecessary for analysis and contain NaN values respectively.

	OBJECTID_1	Name_of_Pr	Address	Neighborho
0	401	Arlington Street Church	355 Boylston Street	Back Bay
1	402	James Blake House	210 East Cottage Street	Dorchester
2	403	International Trust Company Building	45 Milk Street	Boston
3	404	Austin Block	90-92 Main Street	Charlestown
4	405	Theodore Parker Unitarian Church	1851 Centre Street	West Roxbury
...	...	...	...	...
162	563	Quaker Lane	Quaker Lane	Boston
163	564	Sidewalk Clock - 9 Chelsea Street	9 Chelsea Street	
164	565	Richard Bond House	88 Lambert Avenue	Roxbury
165	566	144-146 Maverick Street	144-146 Maverick Street	East Boston
166	567	Saint James African Orthodox Church	50 Cedar Street	Roxbury

167 rows × 4 columns

3. We obtain the location data for each Landmark using geopy and visualize them using folium.

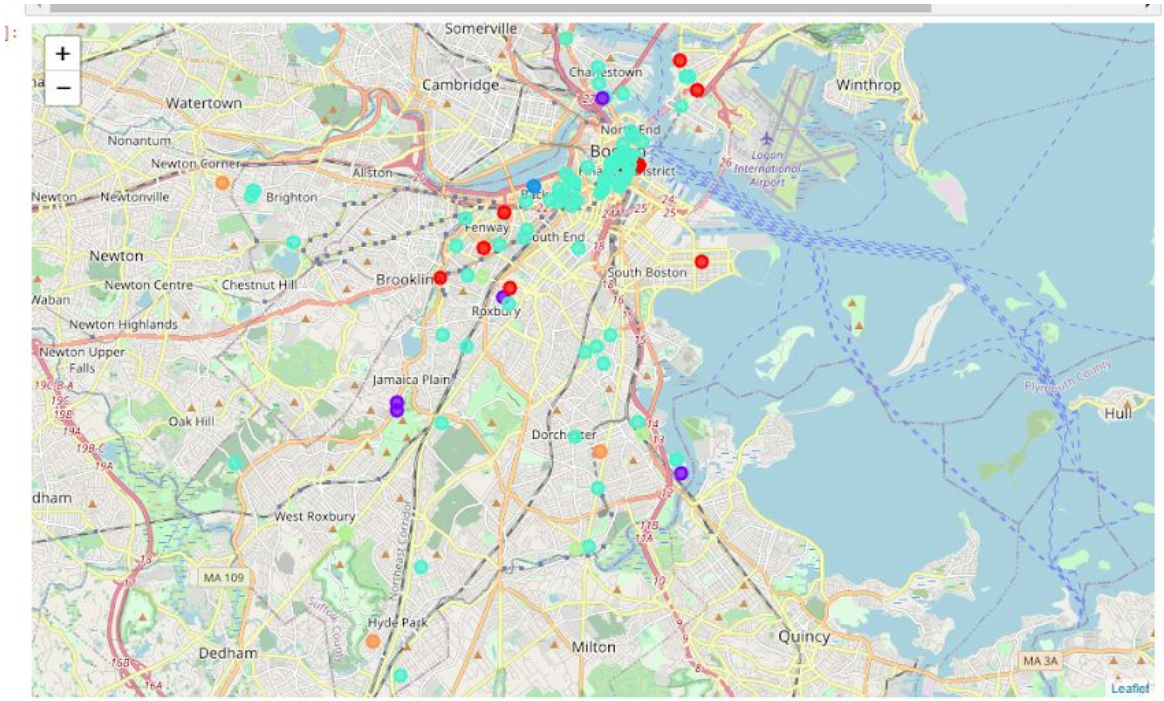


4. The Foursquare API is used to get the nearby venues within a radius of 150 meters for each Landmark.
5. We encode the Venue categories and find the top 10 venues for each Landmark.

5]:

	Place_of_Interest	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	10 Post Office Square	Italian Restaurant	Hotel	Mediterranean Restaurant	Park	Burger Joint	Donut Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Court
1	102 Broad Street	Sandwich Place	Pub	Donut Shop	Food Truck	Beer Garden	Coffee Shop	Park	Electronics Store	Dive Bar	Distribution Center
2	109-113 Broad Street	Sandwich Place	Park	Beer Garden	Food Truck	Pub	Donut Shop	Office	Coffee Shop	Food Court	Flower Shop
3	115-119 Broad Street	Sandwich Place	Park	Beer Garden	Food Truck	Pub	Donut Shop	Office	Coffee Shop	Food Court	Flower Shop
4	1203-5 Adams Street	American Restaurant	Burger Joint	Ice Cream Shop	Pizza Place	Donut Shop	Bakery	Mexican Restaurant	Coffee Shop	Thai Restaurant	Fast Food Restaurant

6. The Landmarks are clustered into 7 clusters based on the Venues in their vicinity.



Part 2 - Analysing Boston Neighborhoods:

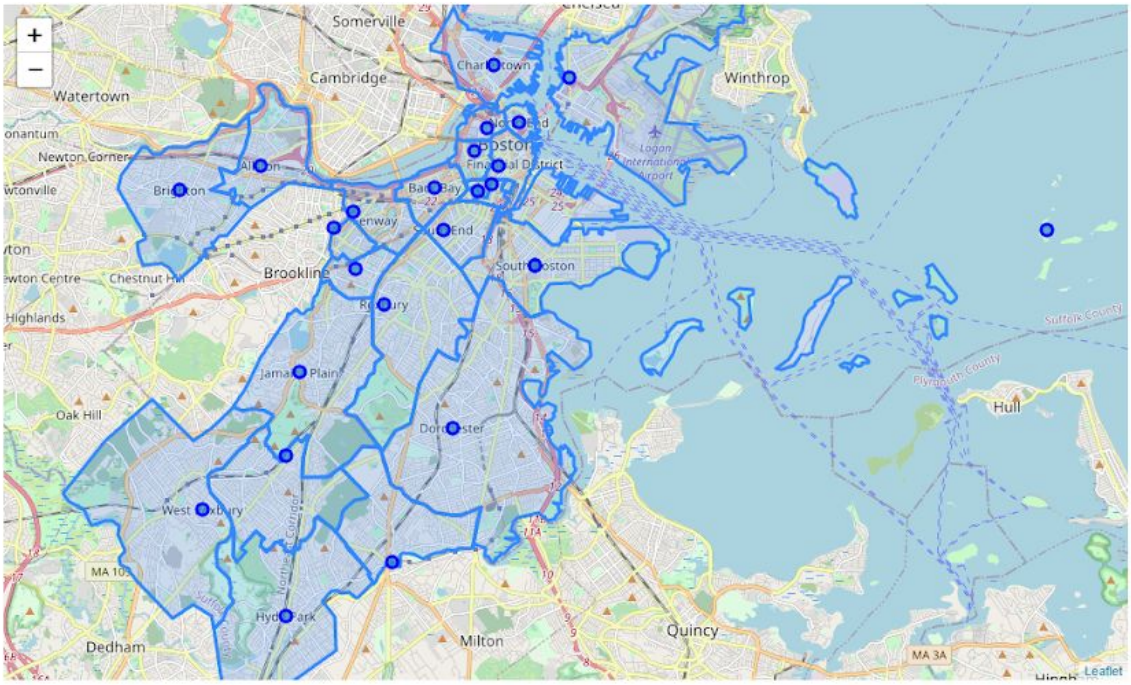
7. We load the csv file found at: https://data.boston.gov/dataset/boston-neighborhoods and observe the data.

	OBJECTID	Name	Acres	Neighborhood_ID	SqMiles	ShapeSTArea	ShapeSTLength
0	27	Roslindale	1605.568237	15	2.51	6.993827e+07	53563.912597
1	28	Jamaica Plain	2519.245394	11	3.94	1.097379e+08	56349.937161
2	29	Mission Hill	350.853564	13	0.55	1.528312e+07	17918.724113
3	30	Longwood	188.611947	28	0.29	8.215904e+06	11908.757148
4	31	Bay Village	26.539839	33	0.04	1.156071e+06	4650.635493
5	32	Leather District	15.639908	27	0.02	6.812717e+05	3237.140537
6	33	Chinatown	76.324410	26	0.12	3.324678e+06	9736.590413
7	34	North End	126.910439	14	0.20	5.527506e+06	16177.826815
8	35	Roxbury	2108.469072	16	3.29	9.184455e+07	49488.800485
9	36	South End	471.535356	32	0.74	2.054000e+07	17912.333569
10	37	Back Bay	399.314411	2	0.62	1.739407e+07	19455.671146

8. We drop certain columns and rows that are unnecessary for analysis and contain NaN values respectively.

	OBJECTID	Name	SqMiles
0	27	Roslindale	2.51
1	28	Jamaica Plain	3.94
2	29	Mission Hill	0.55
3	30	Longwood	0.29
4	31	Bay Village	0.04
5	32	Leather District	0.02
6	33	Chinatown	0.12
7	34	North End	0.20
8	35	Roxbury	3.29
9	36	South End	0.74
10	37	Back Bay	0.62

9. We obtain the location data for each Neighborhood using geopy and visualize them using folium, along with the geospatial data using the geojson file.

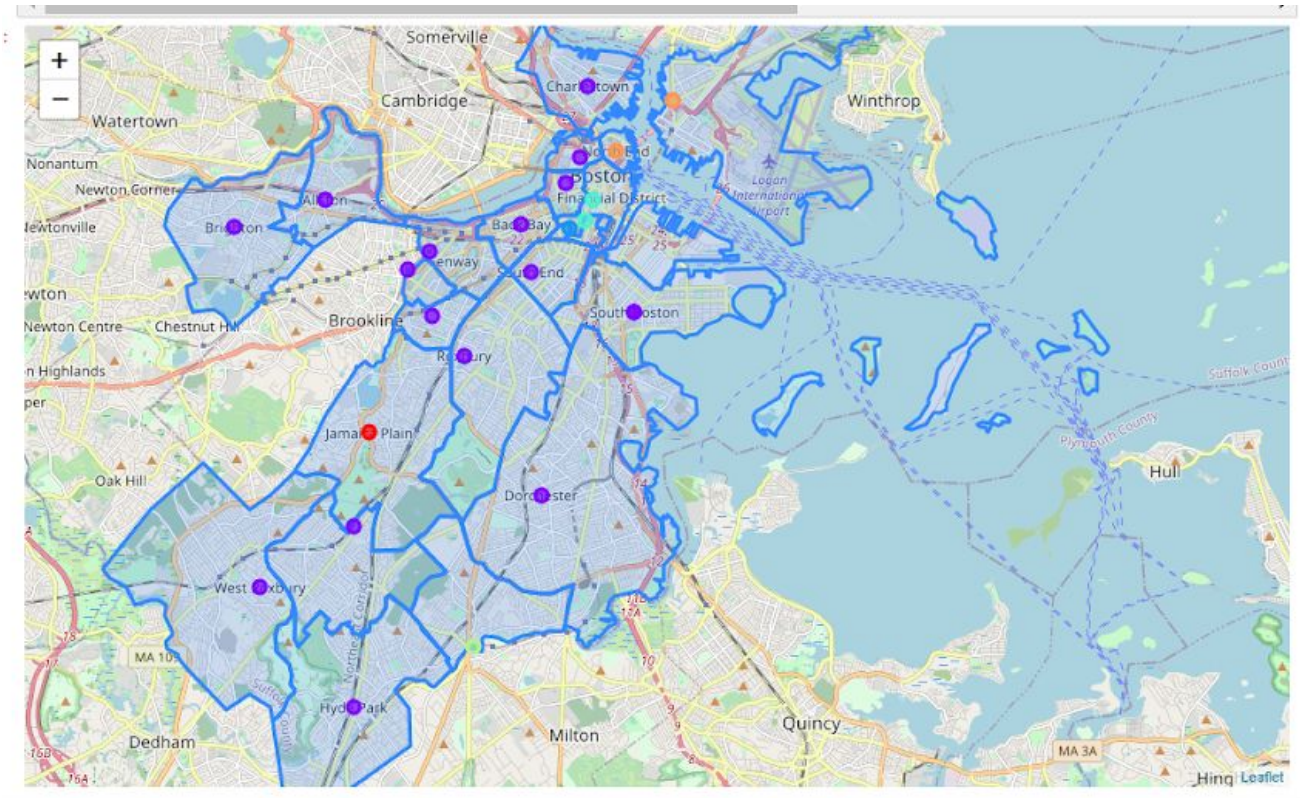


10. The Foursquare API is used to get the nearby venues within a variable radius, determined using the Area of each neighborhood.

11. We encode the Venue categories and find the top 10 venues for each Neighborhood.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allston	Pizza Place	Bar	Rock Club	Liquor Store	Thai Restaurant	Korean Restaurant	Yoga Studio	Seafood Restaurant	Café	Bakery
1	Back Bay	Clothing Store	Hotel	Seafood Restaurant	Coffee Shop	Spa	Bakery	Ice Cream Shop	Gym	Wine Shop	Italian Restaurant
2	Bay Village	Seafood Restaurant	Theater	Spa	Rental Car Location	Sandwich Place	Performing Arts Venue	Ethiopian Restaurant	Dry Cleaner	Electronics Store	Empanada Restaurant
3	Beacon Hill	Italian Restaurant	Pizza Place	French Restaurant	Gourmet Shop	Hotel	Korean Restaurant	Bar	Park	Kids Store	Sushi Restaurant
4	Brighton	Pizza Place	Café	Bakery	Sushi Restaurant	Gym	Mexican Restaurant	Chinese Restaurant	Grocery Store	Coffee Shop	Bar
5	Charlestown	Donut Shop	Pizza Place	Café	Gastropub	Park	Convenience Store	History Museum	Scenic Lookout	Grocery Store	Pub
6	Chinatown	Chinese Restaurant	Bakery	Asian Restaurant	Theater	Sushi Restaurant	Hotel Bar	Performing Arts Venue	Coffee Shop	Pizza Place	Hotpot Restaurant
7	Dorchester	Vietnamese Restaurant	Caribbean Restaurant	Pharmacy	Café	American Restaurant	Pizza Place	Pub	Restaurant	Bank	Breakfast Spot
8	Downtown	Chinese Restaurant	Coffee Shop	Italian Restaurant	Historic Site	Bakery	Hotel	Sushi Restaurant	Café	Asian Restaurant	Theater
9	East Boston	Park	Italian Restaurant	Pizza Place	Seafood Restaurant	Café	Mexican Restaurant	History Museum	Sandwich Place	Brewery	Gastropub
10	Fenway	Coffee Shop	American Restaurant	Park	Café	Donut Shop	Lounge	Baseball Field	Hotel	Mexican Restaurant	Bakery
11	Hyde Park	Pizza Place	Donut Shop	Pharmacy	Grocery Store	American Restaurant	Discount Store	Supermarket	Gym	Park	Optical Shop
12	Jamaica Plain	Park	American	Coffee Shop	Brewery	Trail	Bakery	Bar	Yoga Studio	Ice Cream Shop	Skating Rink

12. The Neighborhoods are clustered into 6 clusters based on the Venues in their vicinity.



RESULTS:

Note: The clustering was experimented with cluster values ranging from 5 - 10 each for landmarks and neighborhoods before deciding on the seemingly more accurate values chosen.

1. **Clustering of Landmarks:**
    a. **Cluster 0** - This cluster contains only 1 element and can be inferred to be the one near Sculpture gardens and fast food outlets.
    b. **Cluster 1** - This cluster contains 5 elements and can be inferred to be landmarks primarily near Parks, Yoga Studios, Fried Chicken Joints, Food Trucks, Food courts and Flower Shops.
    c. **Cluster 2** - This cluster contains 2 elements and can be inferred as the Landmarks in Beacon Street
    d. **Cluster 3** - This cluster contains 96 elements and is more tedious to generalize than the rest as it is mostly a mixed bag of sorts. It can be deduced as the Landmarks within the vicinity of Various Restaurants, Cafes, Coffee Shops and Stores.
    e. **Cluster 4** - This cluster contains 9 elements and can be inferred to be the Landmarks primarily near Pizza Places and Pharmacies, with a degree of prevalence of Yoga studios, Farmer Markets, Liquor Stores and Donut Shops as well.
    f. **Cluster 5** - This cluster contains 3 landmarks, primarily near Furniture/Home Stores, French Restaurants, Fried chicken joints and Food trucks with some prevalence of Bars and Yoga Studios as well.
    g. **Cluster 6** - This cluster contains 12 landmarks, primarily near Sandwich places, Parks, Donut Shops and Food trucks, with some prevalence of Beer Gardens, Pubs, Fried Chicken joints and Offices.

2. **Clustering of Neighborhoods:**
    a. **Cluster 0** - This cluster contains only 1 neighborhood, which is Jamaica Plain, primarily near Parks, American Restaurants, Coffee shops, Breweries and Bakeries.

b. **Cluster 1** - This cluster contains 17 neighborhoods, primarily near Pizza Places, Coffee Shops, Cafes, Bars and American Restaurants.
c. **Cluster 2** - This cluster contains only 1 neighborhood, which is Bay Village, primarily near Seafood Restaurants, Theaters, Spas, Rental Car Outlets and Sandwich Places.
d. **Cluster 3** - This cluster has 2 neighborhoods, namely Chinatown and Downtown, primarily with Chinese Restaurants, Theaters and Asian Restaurants.
e. **Cluster 4** - This cluster contains 9 neighborhoods, primarily near Pizza Places, Pharmacies, Donut Shops, Liquor stores and Electronics stores, as well as Farmers markets and Yoga studios.
f. **Cluster 5** - This cluster has 2 neighborhoods, namely North End and East Boston, primarily with Parks, Italian Restaurants, Pizza Places and Seafood Restaurants.

DISCUSSION:

From the results obtained we can see that while most of the clusters seem accurate with regards to its classification, certain lusters, particularly the cluster 3 for landmarks, which, regardless of cluster values chosen from 5 - 10, gave a slightly vague classification. The accuracy of the clustering might see an improvement on experimenting on higher cluster values or different radii for finding nearby venues. However the results have been promising and can go on to act as a guide for tourists and hopeful immigrants in the future.

CONCLUSION:

Thus, we can obtain a better understanding of the landmarks and neighborhoods in Boston, based on the amenities and facilities available in and near them.
We might obtain a more comprehensive clustering if we adjust the hyperparameters or use a more detailed dataset. This can be improved upon in the future and there is always room for improvement.

- By ANTONY PRADEEP CLEMENT