Battle of the Neighborhoods

IBM Applied Data Science Specialization - Capstone

By Antony Pradeep Clement

Description of the Problem

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 sightsee, 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 be used:

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
2) Data of different Neighborhoods in Boston
    a) Source:https://data.boston.gov/dataset/boston-neighborhoods
3) Venue data to different venues near the locations
    a) Source: Foursquare Developer Access to Various Venues (https://foursquare.com/).
4) Geospatial Data of Boston's Neighborhoods (geojson file)
    a) Source:https://data.boston.gov/dataset/boston-neighborhoods

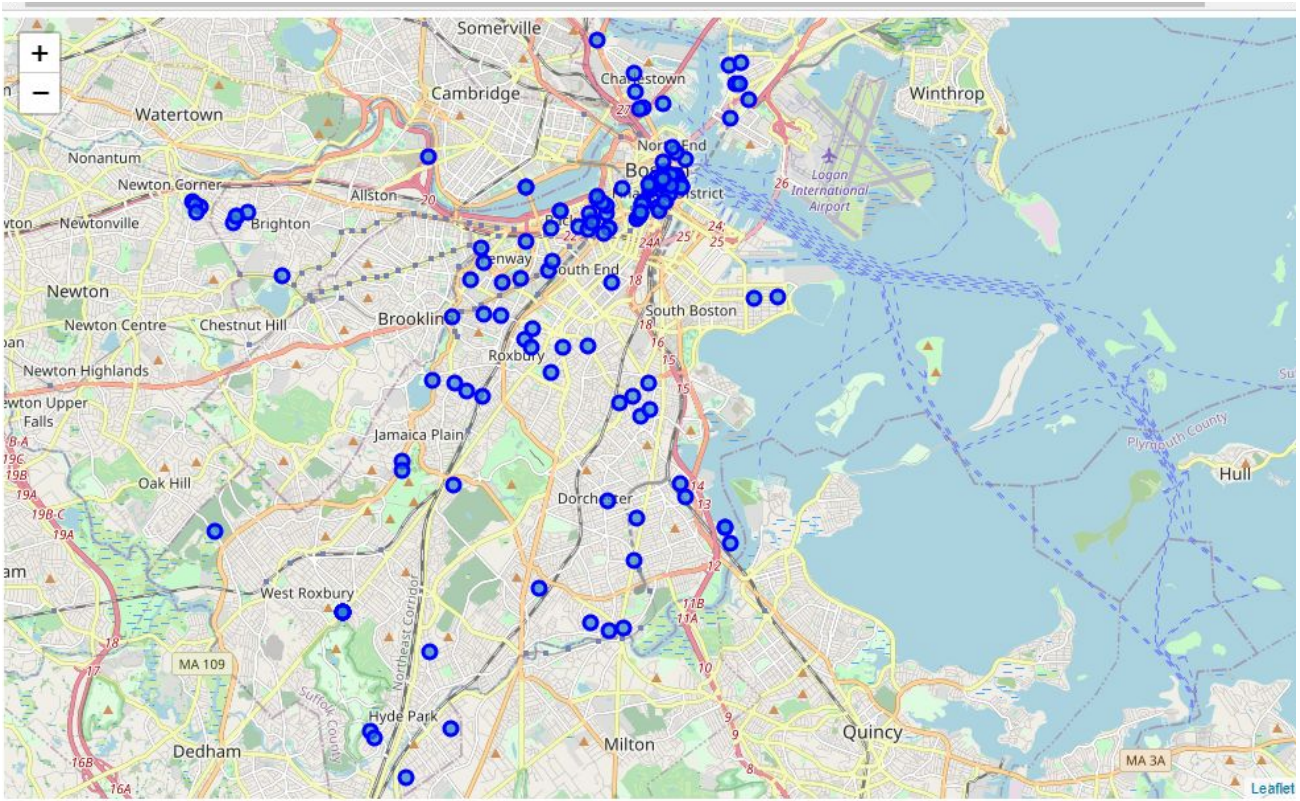
Methodology and Steps followed

Methodology :

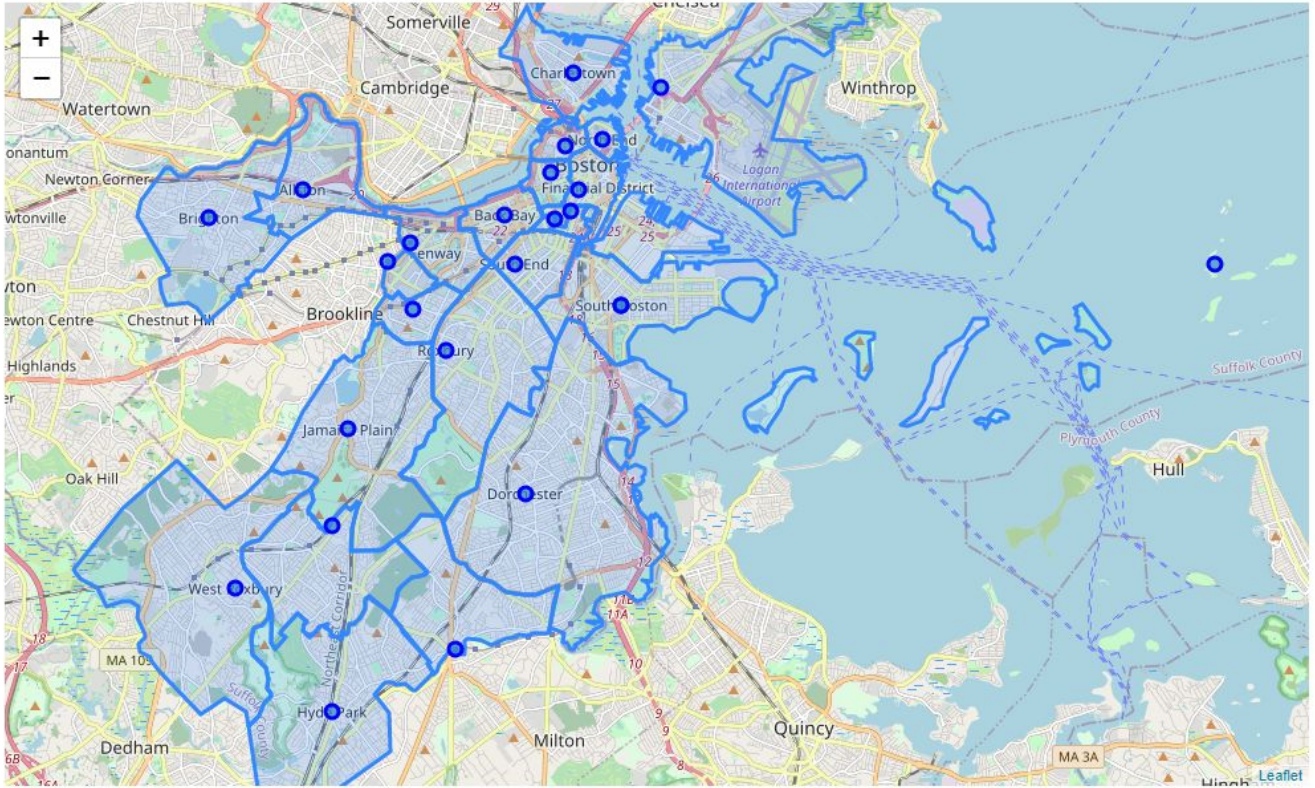
The methodology followed for landmarks and neighborhoods basically revolve around:

1. Loading and preprocessing the required dataset.
2. Obtain the location data for each Landmark/Neighborhood and plot the on a map using folium.
3. Use the Foursquare API to obtain the nearby venues for the places of interest.
4. Cluster the places of interest based on the common venues.
5. Plot the clusters on the map

Map of Landmarks in Boston



Map of Neighborhoods in Boston

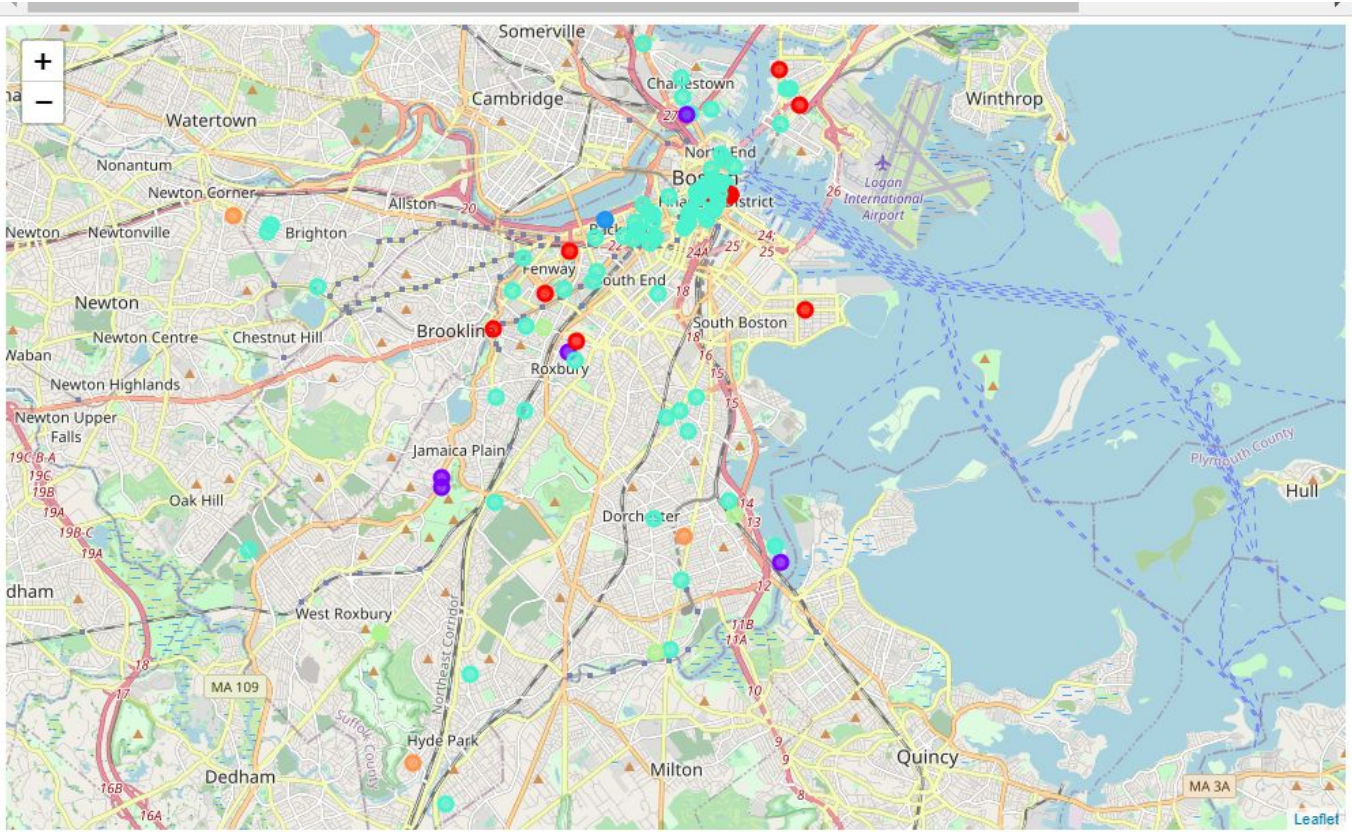


Results

Clustering results for Landmarks

CLUSTER	COUNT	NOTABLE COMMON VENUES
0	1	Sculpture gardens and fast food outlets
1	5	Parks, Yoga Studios, Fried Chicken Joints, Food Trucks, Food courts and Flower Shops.
2	2	Venues in Beacon Street
3	96	Various Restaurants, Cafes, Coffee Shops and Stores
4	9	Pizza Places and Pharmacies, Yoga studios, Farmer Markets, Liquor Stores and Donut Shops
5	3	Furniture/Home Stores, French Restaurants, Fried chicken joints and Food trucks, Bars and Yoga Studios
6	12	Sandwich places, Parks, Donut Shops and Food trucks, Beer Gardens, Pubs, Fried Chicken joints and Offices

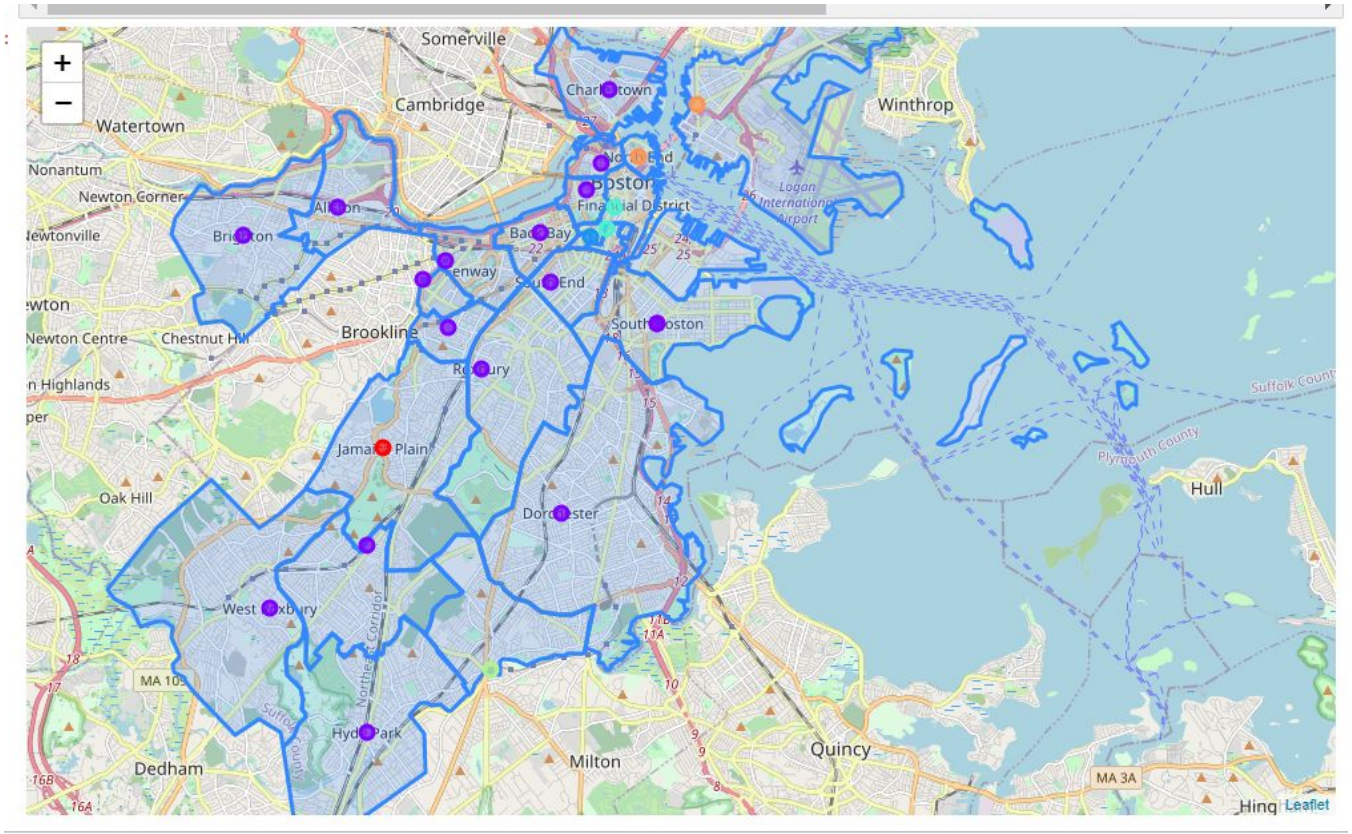
Map of Clustered Landmarks



Clustering results for Neighborhoods

CLUSTERS	COUNT	NOTABLE COMMON VENUES
0	1	Parks, American Restaurants, Coffee shops, Breweries and Bakeries. (Jamaica Plain only)
1	17	Pizza Places, Coffee Shops, Cafes, Bars and American Restaurants.
2	1	Seafood Restaurants, Theaters, Spas, Rental Car Outlets and Sandwich Places. (Bay Village only)
3	2	Chinese Restaurants, Theaters and Asian Restaurants.
4	9	Pizza Places, Pharmacies, Donut Shops, Liquor stores and Electronics stores, as well as Farmers markets and Yoga studios.
5	2	Parks, Italian Restaurants, Pizza Places and Seafood Restaurants.

Map of Clustered Neighborhoods



Conclusion

Discussion of results:

From the results obtained we can see that while most of the clusters seem accurate with regards to its classification, certain lusters gave a slightly vague classifications.

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.

Concluding Remarks

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.

Thank you!