Businesses & Venues Data Analysis of Boston City

I. Introduction

Boston is one of the most historically abundant cities in the United States. It is the capital of the Massachusetts, and it has about 26 neighborhoods in total which cover an area of 49 square miles. The city has an estimated population of 692,600 in 2019 making it the most populated city in New England [1].

Boston is placed among the top 30 most economically powerful cities in the world and is highly innovative because of presence of academia, access to venture capital, and the presence of many high-tech companies. It also has several historic sites and museums, social activities , dinning services and outdoors activities .If a contractor is trying to start their own business whether it is a food store/retail stores, then where they locate their business depends in part on the location of the target market/customers, business partners, and their personal preferences. Here, we create a map and cluster the neighborhood venues in Boston to provide venue density and provide information that will guide investors where to best locate their businesses e.g. service/consulting company, retail store, restaurant etc.

II. Data Description

To consider the problem:

Boston neighborhoods dataset was used from Analyze Boston, which is Boston City's open data hub. The .csv file has the Neighborhood boundaries data layer which is a grouping of zoning neighborhood boundaries, zip code boundaries and 2010 Census tract boundaries. This dataset was used for visualization purposes for zoning exploration.

Foursquare API was used to help provide the information needed to segment and cluster the neighborhoods in the city of Boston. As well as, the most common venue categories in each neighborhood in Boston was used to create a map of the venue density in Boston City.

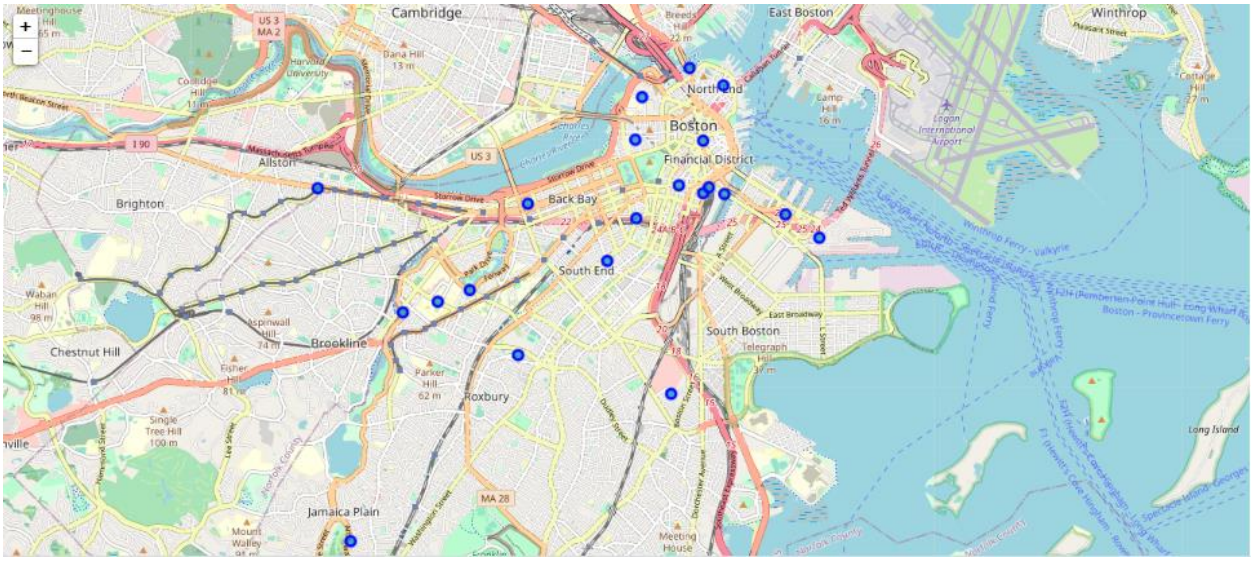
Certified business directory dataset was used from Analyze Boston to determine which neighborhood has the most small/local business in Boston. This data set has minority-owned and woman-owned businesses, small and small local businesses, and veteran and service disabled veteran-owned small businesses.

Food establishment inspection dataset was used from Analyze Boston to find the all the active food establishment found in the Boston City. This was used to analyze the most neighborhoods with the most cuisines and restaurants

Community center dataset was used from Analyze Boston to find the all the community centers/social venues found in the Boston City. This was used to analyze the most neighborhoods with the most social venues and activities.

III. Methodology

Geopy library was used to retrieve the latitude and longitude values of Boston City and its neighborhoods. Folium library was later used to visualize the neighborhoods in Boston City. A map of Boston was produced with its neighborhoods superimposed on top, as shown below.



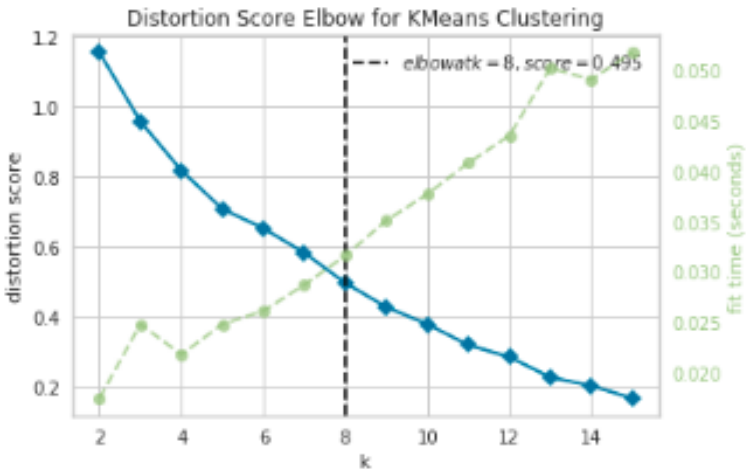
Foursquare API was utilized to explore the Boston neighborhoods and segment them. The top 100 venues we obtained by Foursquare API, for each neighborhood in Boston within a radius of 500 meters. Afterwards, the total number of venues found in each neighborhood was found and it was found that there are 206 unique venue categories of venues in Boston.

One hot-encoding was used to analyze the venue in each neighborhood and later the neighborhoods where grouped and the mean of the frequency of occurrence of each category of venue was found. Each neighborhood along with the top 5 most common venues was tabulated as shown below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allston	Metro Station	Pharmacy	Market	Park	Fried Chicken Joint
1	Back Bay	Clothing Store	Coffee Shop	Hotel	Women's Store	Accessories Store
2	Bay Village	Theater	Sandwich Place	Seafood Restaurant	Bakery	French Restaurant
3	Beacon Hill	Pizza Place	French Restaurant	Italian Restaurant	Hotel	Music Venue
4	Brighton	Coffee Shop	Yoga Studio	Pizza Place	Gym / Fitness Center	Thai Restaurant
5	Charlestown	Park	Pizza Place	Donut Shop	Sports Bar	Bar
6	Chinatown	Chinese Restaurant	Bakery	Asian Restaurant	Coffee Shop	Sandwich Place
7	Dorchester	Coffee Shop	Italian Restaurant	Bakery	Sandwich Place	American Restaurant
8	Downtown	Hotel	Sandwich Place	American Restaurant	Historic Site	Park
9	East Boston	Bakery	Chinese Restaurant	American Restaurant	Sandwich Place	Coffee Shop
10	Fenway	Art Museum	Café	Art Gallery	Park	Coffee Shop
11	Harbor Islands	Seafood Restaurant	Coffee Shop	American Restaurant	Boat or Ferry	Donut Shop
12	Hyde Park	Liquor Store	Supermarket	Pizza Place	Wings Joint	Pharmacy
13	Jamaica Plain	Park	Pet Store	Bakery	Plaza	Record Shop
14	Leather District	American Restaurant	Mobile Phone Shop	Coffee Shop	Gym	Department Store
15	Longwood	Coffee Shop	Café	Donut Shop	Falafel Restaurant	Sandwich Place
16	Mattapan	Park	Construction & Landscaping	Event Service	Pizza Place	Ice Cream Shop
17	Mission Hill	Sandwich Place	Café	Coffee Shop	American Restaurant	Pharmacy
18	North End	Italian Restaurant	Pizza Place	Park	Bakery	Seafood Restaurant
19	Roslindale	Bookstore	Pet Store	Italian Restaurant	Bus Line	Latin American Restaurant
20	Roxbury	Pharmacy	Rental Car Location	Pizza Place	Donut Shop	Bus Station
21	South Boston	Chinese Restaurant	Bakery	Asian Restaurant	Sandwich Place	Bubble Tea Shop
22	South Boston Waterfront	Seafood Restaurant	Music Venue	American Restaurant	Café	Steakhouse
23	South End	Wine Bar	American Restaurant	Park	Mexican Restaurant	Wine Shop
24	West End	Hotel	Donut Shop	Pizza Place	Bar	Museum
25	West Roxbury	Home Service	Pizza Place	Train Station	Bagel Shop	Food

Since we have some common venue categories in the neighborhoods, therefore unsupervised learning K-means clustering was used to cluster the neighborhoods. The elbow method was used select the optimal number of clusters by fitting the model with a range of values for K. For each value/data point, we are calculating the sum of squared distances from each point to its assigned center (distortions). The silhouette score calculates the mean Silhouette Coefficient of all samples, while the calinski harabasz score computes the ratio of dispersion between and

within clusters. The “elbow” (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. From the graph below, we can see that K=8 gives the optimal clusters.

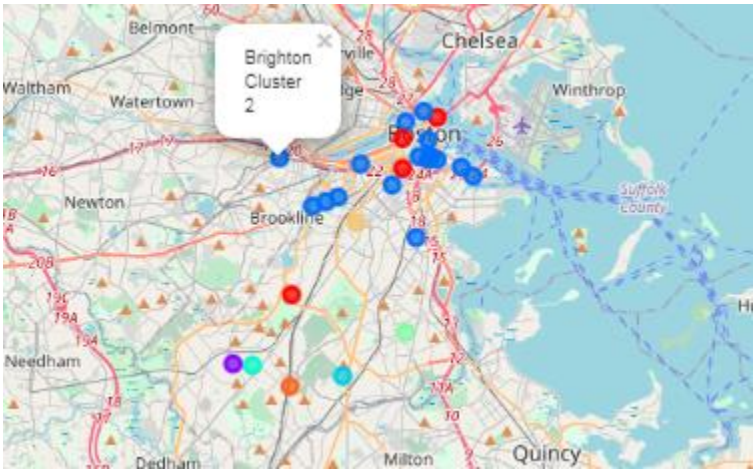


After the number of number of clusters was determined K-mean clustering was performed on the venues of each neighborhood and the clusters were plotted on the map to show their distribution across Boston City.

The clusters were examined to determine the labelling of the neighborhood and clusters in terms of the venue category.

IV. Results

A map with the clusters of the Boston neighborhoods was created with the name and cluster as point markers superimposed on top as shown below.



Cluster 1: Malls, Historic sites, Hotels and Dinning

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Jamaica Plain	0	Park	Bakery	Pet Store	Art Gallery	Library
Bay Village	0	Theater	Seafood Restaurant	Sandwich Place	Hotel	Bakery
North End	0	Italian Restaurant	Pizza Place	Park	Bakery	Seafood Restaurant
Beacon Hill	0	Pizza Place	Italian Restaurant	French Restaurant	Outdoor Sculpture	Kids Store

Cluster 2: Multiple social venues (Park, Malls, Historic sites, Nightclubs) & dining services

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Mission Hill	2	Sandwich Place	Pharmacy	American Restaurant	Coffee Shop	Café
Longwood	2	Café	Coffee Shop	Falafel Restaurant	Sushi Restaurant	Sandwich Place
Leather District	2	American Restaurant	Department Store	Coffee Shop	Mobile Phone Shop	Chinese Restaurant
Chinatown	2	Chinese Restaurant	Bakery	Asian Restaurant	Coffee Shop	Sandwich Place
South End	2	Park	Wine Bar	American Restaurant	Gift Shop	Theater
Back Bay	2	Clothing Store	Coffee Shop	Hotel	Bookstore	Ice Cream Shop
East Boston	2	Bakery	Chinese Restaurant	Coffee Shop	American Restaurant	Sandwich Place
Charlestown	2	Pizza Place	Sports Bar	Park	Hotel	Bar
West End	2	Hotel	Donut Shop	Pizza Place	Bar	Café
Downtown	2	Historic Site	Sandwich Place	Hotel	American Restaurant	Park
Fenway	2	Art Museum	Café	Park	Restaurant	Gym
Brighton	2	Coffee Shop	Pizza Place	Yoga Studio	Taco Place	Nightclub
Dorchester	2	Coffee Shop	Sandwich Place	Italian Restaurant	Bakery	Hotel
South Boston Waterfront	2	Seafood Restaurant	Restaurant	American Restaurant	Steakhouse	Donut Shop
South Boston	2	Chinese Restaurant	Bakery	Asian Restaurant	Sandwich Place	Bubble Tea Shop
Harbor Islands	2	Seafood Restaurant	Coffee Shop	Donut Shop	American Restaurant	Boat or Ferry

Cluster 3: Service venues

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Mattapan	3	Park	Ice Cream Shop	Pizza Place	Event Service	Business Service

Cluster 4: Cultural dining venues

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Roslindale	4	Latin American Restaurant	Sandwich Place	Pet Store	Italian Restaurant	Chinese Restaurant

Cluster 5: Accommodation (bank, market, pharmacy) venues

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Allston	5	Bank	Pharmacy	Fried Chicken Joint	Fast Food Restaurant	Market

Cluster 6: Diverse food restaurants

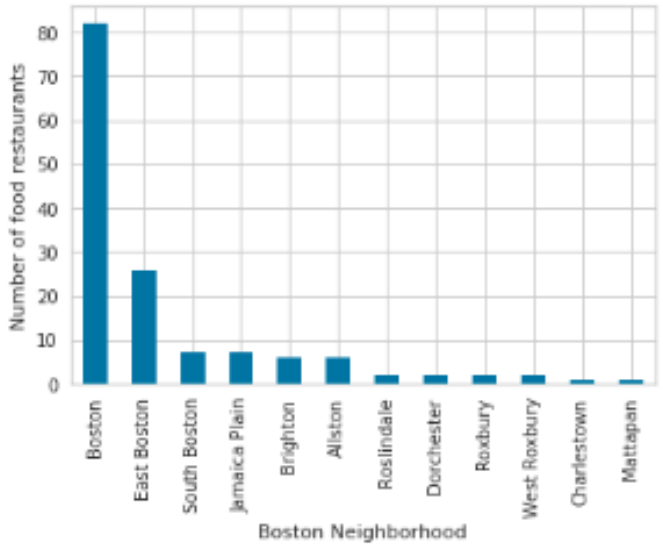
Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Roxbury	6	Donut Shop	BBQ Joint	Pharmacy	African Restaurant	American Restaurant

Cluster 7: Drink/Food venues

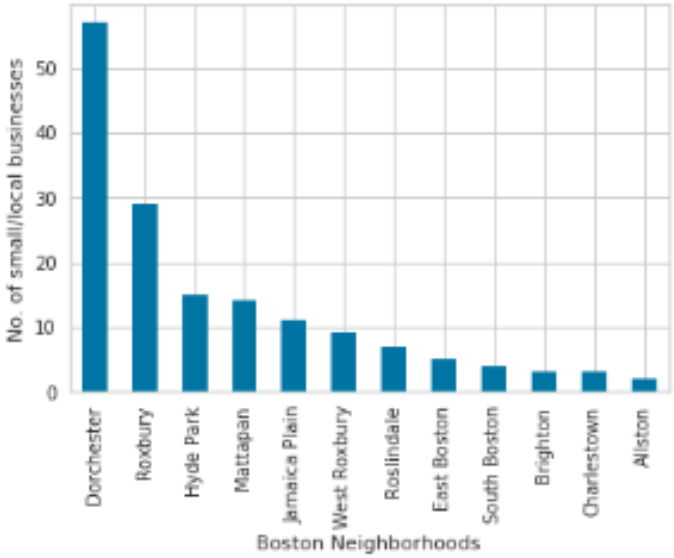
Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Hyde Park	7	Liquor Store	Pizza Place	Gym / Fitness Center	Mexican Restaurant	Buffet

The community center dataset of Boston was used to highlight where social activities and services are provided in Boston. After cleaning the dataset and extracting the necessary parameter as graph of the most community centers in Boston was plotted. From the graph below, Roxbury and Dorchester have the most community centers and this supports the clusters above as these are public locations where members of a community tend to gather for group activities, social support, public information, and other purposes such as dining.

The food establishment dataset of Boston was used to highlight where food restaurants and cuisines are provided in Boston the most. After cleaning the dataset and extracting the necessary parameters a graph of the most food restaurants which are active in Boston was plotted. From the graph below, the center of Boston and East Boston have the most dining and food restaurants in the Boston area.



The Certified business directory dataset was used to determine which neighborhood has the most minority-owned and woman-owned businesses, small and small local businesses, and veteran and service disabled veteran-owned small businesses. After cleaning the dataset and extracting the necessary parameters a graph of the most businesses, was plotted. From the graph below, Dorchester and Roxbury have the most minority-owned and woman-owned businesses, small and small local businesses, and veteran and service disabled veteran-owned small businesses in the Boston area.



IV. Discussion

Boston is a populous city with lots of innovations with a lot of access to venture capital, and the presence of many companies and businesses. It is has many historic sites, social activities and dining services provided for its population. There are about 26 neighborhoods in Boston and each is diverse and populous with different venues. Here, K-means clustering method was used to classify the neighborhoods in Boston to determine which areas have which category of venues. It was found that a cluster of 8 provides the best fit for the model. The cluster of the neighborhoods where plotted on the Boston map.

Each cluster had neighborhoods which had food restaurants as part of their top 5 venues and thus this made the clustering distinction and labelling challenging. Cluster 1 is for social activities for residents as it has restaurants, parks, theaters and kids store. Cluster 2 is for touristic attraction types of services as it includes Malls, Historic sites, Hotels and Dinning venues. Cluster 3 provides services mostly, yet is also have restaurants and food shops. Cluster 4 has culturally diverse cuisine venues. Cluster 5 is for residential living as it has market, banks, pharmacy and restaurants. Cluster 6 is a food venue mostly American food and finally cluster 7 is mostly a drink and food venue.

Data analysis was also performed on food establishment, small/local businesses, and community centers in Boston area to provide more information about where abundance of which venues belongs to which neighborhood. It was found the Dorchester and Roxbury have the most community centers for social activities and gathering, as well as, small/local business. Additionally, as mentioned previously food restaurants are present all the neighborhoods, but there is a higher concentration of such venues in the center of Boston and East Boston.

V. Conclusion

Modelling and data analysis performed in the project provide contractors information to know where best to locate their business. This platform provides information about the density of venues in Boston City and highlights the neighborhoods with high concentration in dining, social activities and local/small business. With the use of the map visualization and the data visualization methods, better decisions can be made by contractors looking to opening a business e.g. retail, food or service venue in the Boston area.

VI. References

[1] https://en.wikipedia.org/wiki/Boston

[2] “Food Establishment Inspections.” *Analyze Boston*, data.boston.gov/dataset/food-establishment-inspections/resource/4582bec6-2b4f-4f9e-bc55-cbaa73117f4c

[3] “Certified Business Directory.” *Analyze Boston*, data.boston.gov/dataset/certified-business-directory

[4] “Community centers.” *Analyze Boston*, data.boston.gov/dataset/community-centers