

Finding best neighbourhood to open American restaurant in Seattle

DEVANG JAGDALE

JULY 12, 2021

1. Introduction

1.1 Background

Seattle is a seaport city on the West Coast of the United States. It is the seat of King County, Washington. With a 2019 population of 753,675, it is the largest city in both the state of Washington and the Pacific Northwest region of North America. The Seattle metropolitan area's population is 3.98 million, making it the 15th-largest in the United States. Even though Seattle is such a big city there are very few American Restaurants present there. This lack of authentic American restaurants makes it great market place for a person who wants to open an American style Restaurant in Seattle. Choosing a right location to open a restaurant in such a big city might be a herculean task for any new restaurant owner. With this goal in my mind this capstone project will be helping new entrepreneur to find best Neighbourhood in Seattle to open an American Restaurant.

1.2 Business Problem

Entrepreneur who wants to find best Neighbourhood in Seattle to open an American Restaurant. We can compare all venues present in different neighbourhoods in various boroughs, find out number of restaurants and other competitor places and find out best place by analysing this data.

2. Data acquisition and cleaning

2.1 Requirements

- a. We will require all the Borough and Neighbourhoods in it.
- b. We will require Latitude and Longitude of each Neighbourhood.
- c. We will require all nearby places details in that Neighbourhood.

2.2 Data Resources

- a. Wikipedia: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle use this to solve our first requirement.
- b. GeoPy: We will use this package to solve our second requirement.
- c. Foursquare API: We will use this API to solve our third requirement.

2.3 Data Cleaning

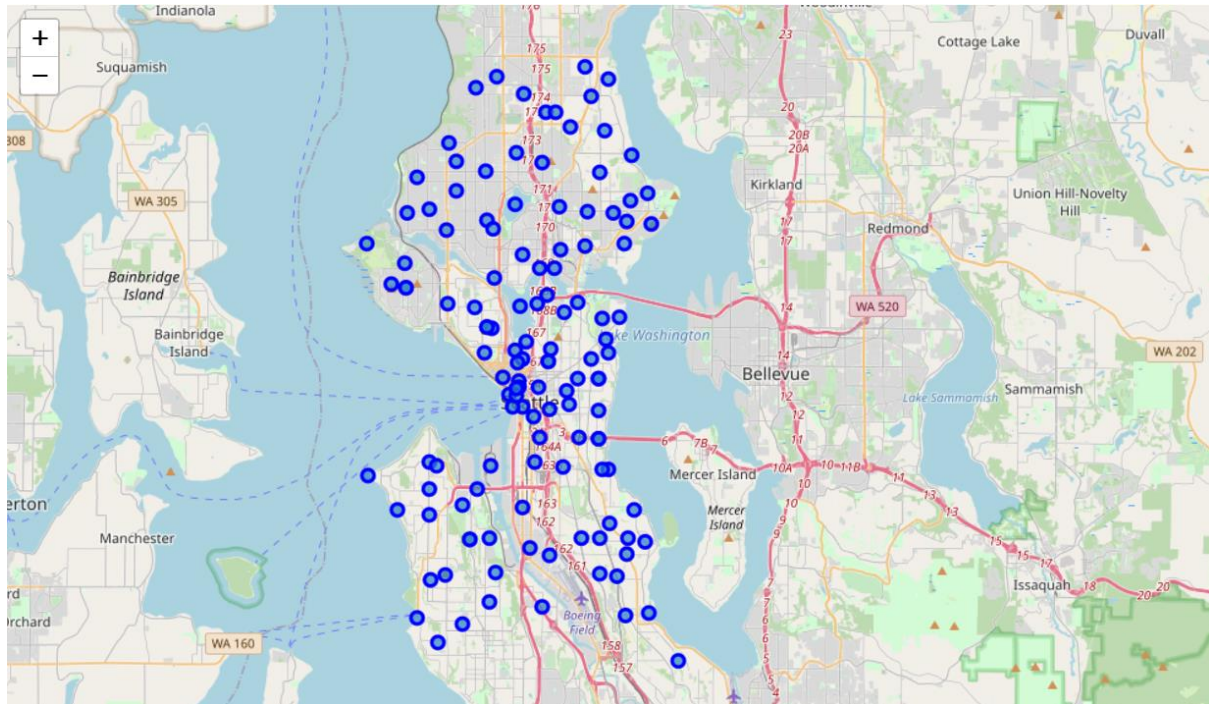
The data obtained and scrapped from various sources has to be collected and combined into one single table. Also above data might be having unwanted entries like nan, none, etc, so data cleaning must be done.

Data we obtained from Wikipedia had many unwanted things like square brackets inside which some numbers are written like [12], then instead of comma they used slash to separate two areas, then after each entry \n was there for new line added as data in our table, etc. All such data needed to be cleaned before being used. So, we used regex package to remove all such unwanted data and obtained 127 rows of data. After applying geopy we obtained latitude and longitude for all neighbourhoods but out of 127 there were 7 rows for which we obtained none as their result. So, we applied two approaches out of which first was adding data manually (this method was adopted for further operations) and second was to remove these rows and have 120 number of rows.

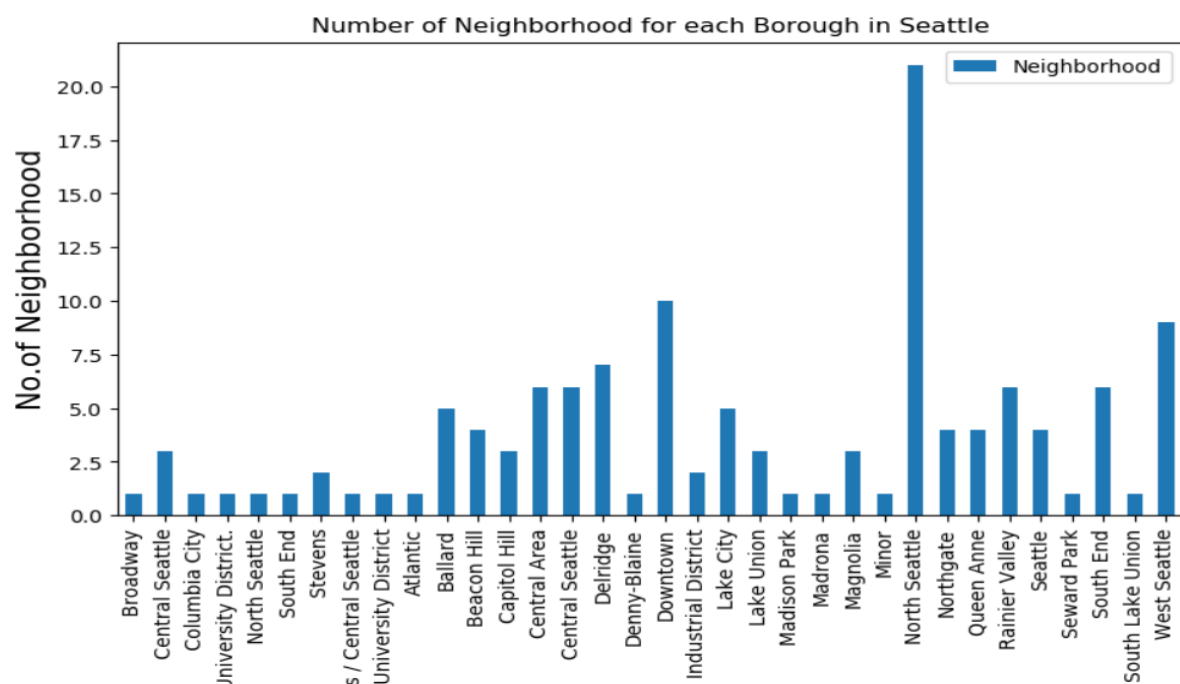
3. Exploratory Data Analysis

3.1 Exploring neighbourhoods

After obtaining Seattle neighbourhood data we use geopy to get latitude and longitude of Seattle and map all neighbourhoods on map using folium library.



Now we try to get number of different neighbourhoods in each borough.

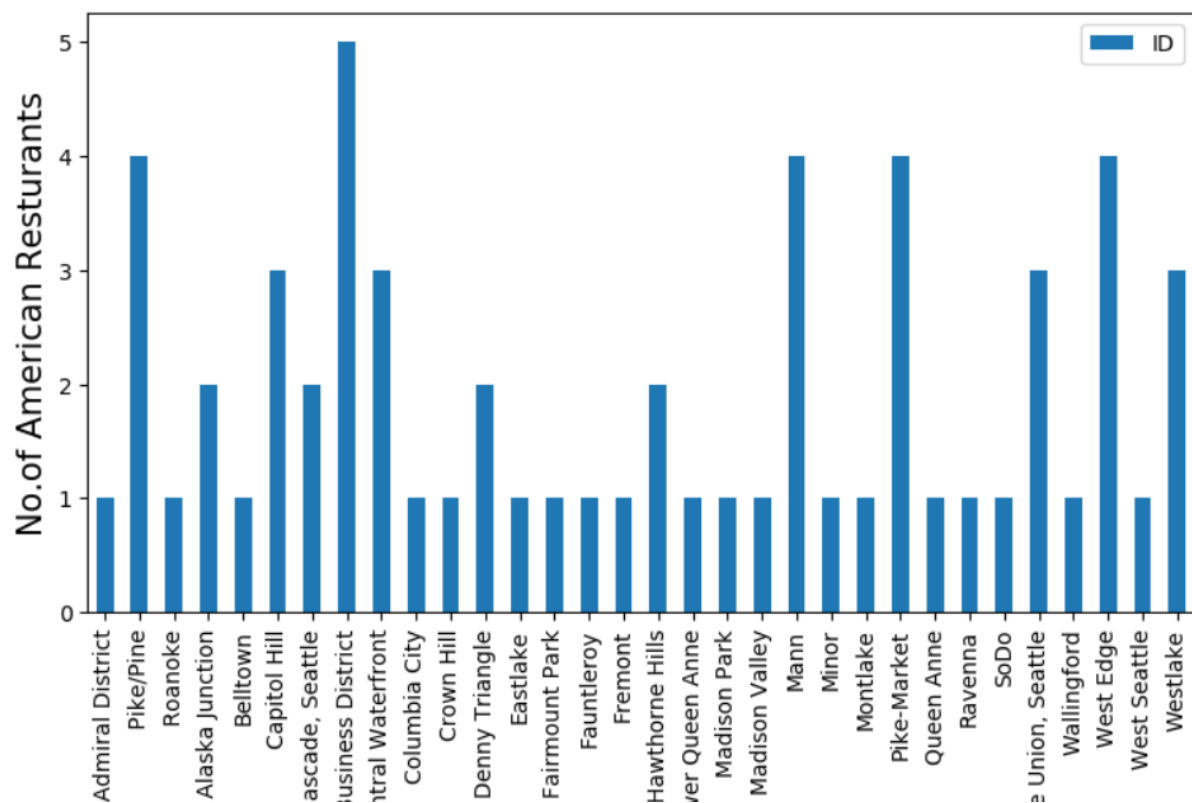


From above graph we see that North Seattle has highest number of neighbourhoods having 21 neighbourhoods.

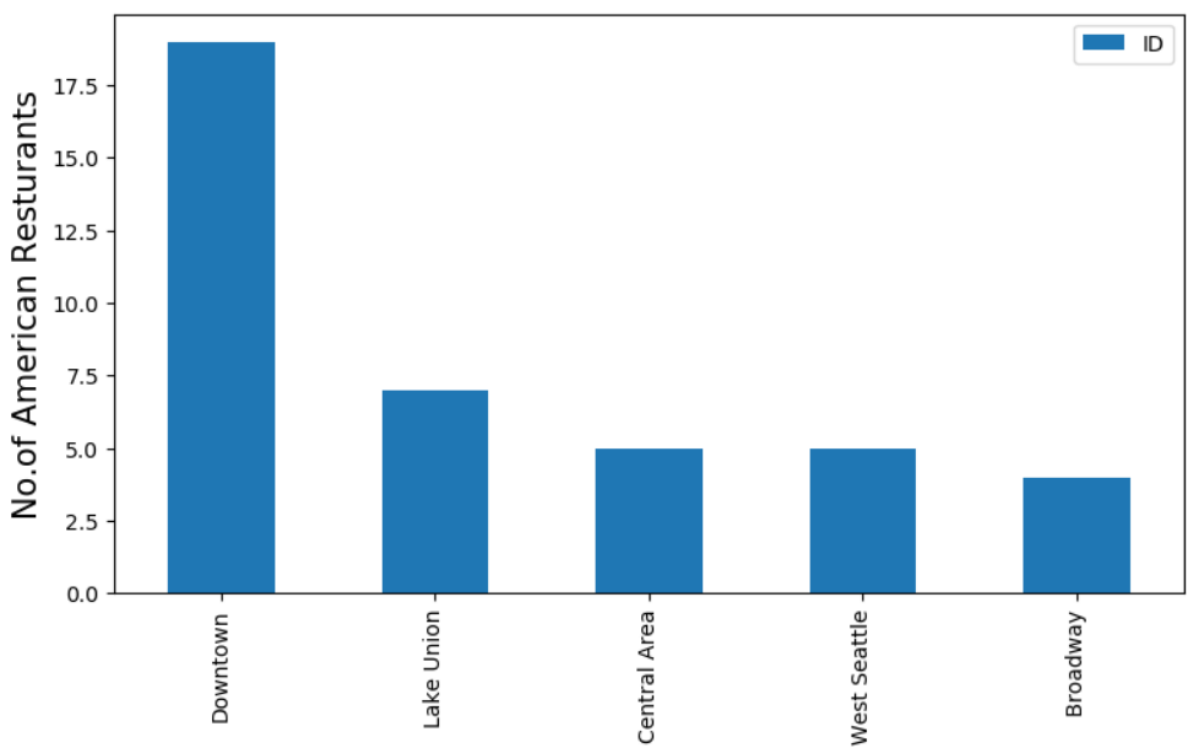
Now we try to get nearby venues for each neighbourhood. For this we use Four Square API free tier version. We find top 100 venues for each neighbourhood within a radius of 500 meters. After this we create a new data frame where we collect American Restaurants from each neighbourhood.

```
( 66 / 127 ) American Resturants in Lake Union, Eastlake:1
( 67 / 127 ) American Resturants in Central Seattle, Downtown:0
( 68 / 127 ) American Resturants in Downtown, Denny Triangle:2
( 69 / 127 ) American Resturants in Downtown, Belltown:1
( 70 / 127 ) American Resturants in Downtown, Pike-Market:4
( 71 / 127 ) American Resturants in Downtown, Central Business District:5
( 72 / 127 ) American Resturants in Downtown, First Hill:0
( 73 / 127 ) American Resturants in Downtown, Pioneer Square:0
( 74 / 127 ) American Resturants in Downtown, International District :0
( 75 / 127 ) American Resturants in Downtown, Yesler Terrace:0
( 76 / 127 ) American Resturants in Downtown, Central Waterfront:3
( 77 / 127 ) American Resturants in Downtown, West Edge:4
( 78 / 127 ) American Resturants in Central Seattle, Central District :0
( 79 / 127 ) American Resturants in Central Area, Mann:4
( 80 / 127 ) American Resturants in Central Area, Minor:1
( 81 / 127 ) American Resturants in Minor, Cherry Hill & Squire Park:0
( 82 / 127 ) American Resturants in Central Area, Atlantic:0
( 83 / 127 ) American Resturants in Atlantic, Judkins Park:0
( 84 / 127 ) American Resturants in Central Area, Madrona:0
( 85 / 127 ) American Resturants in Madrona, Madrona Valley:0
```

Plotting Number of American restaurants for each neighbourhood in Seattle.



Plotting Number of American Restaurants for each Borough in Seattle.



Now we perform one hot encoding venue category for each neighbourhood. Now we get top 5 common venues for each neighbourhood.

```
---- Mee-Kwa-Mooks----
      venue  freq
0      Gym  0.33
1      Park  0.33
2  Scenic Lookout  0.33
3      Zoo Exhibit  0.00
4  Other Repair Shop  0.00

---- NewHolly----
      venue  freq
0      Park  0.5
1  Playground  0.5
2      Zoo Exhibit  0.0
3      Pet Store  0.0
4  Performing Arts Venue  0.0
```

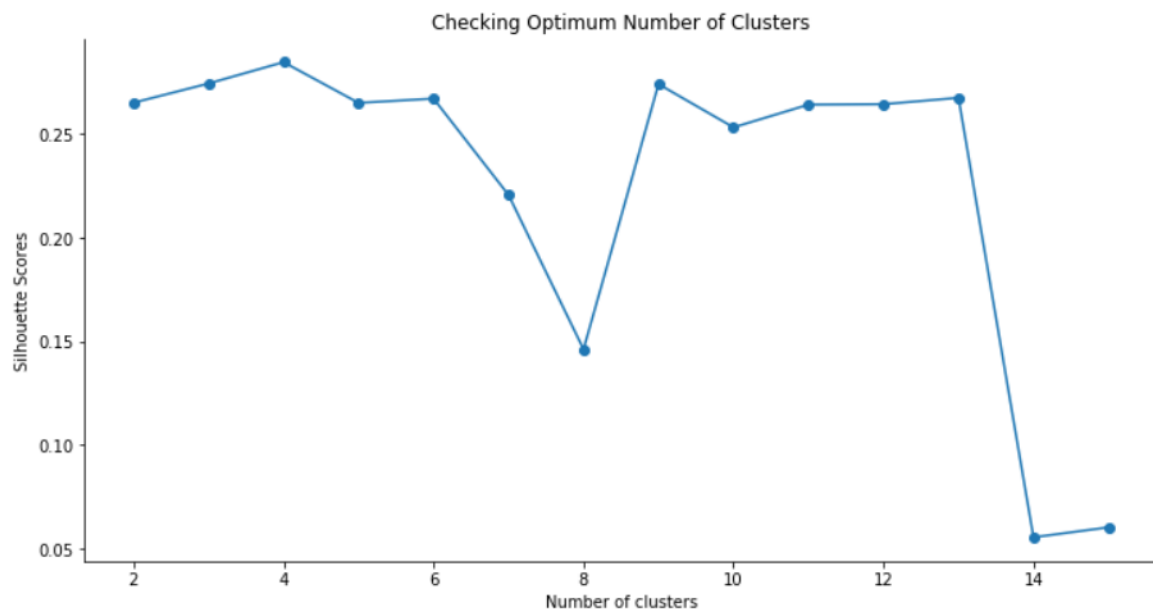
Now we sort venues in descending order and print top 10 venues for each neighbourhood.

Neighborhood	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 Admiral District	Coffee Shop	Theater	Thai Restaurant	Pub	Frozen Yogurt Shop	Middle Eastern Restaurant	Bar	Gas Station	Supermarket	Market
1 Atlantic City Beach	Mexican Restaurant	Garden	Grocery Store	Pharmacy	Gym / Fitness Center	Coffee Shop	Pier	Marijuana Dispensary	Pizza Place	Breakfast Spot
2 Blue Ridge	Park	Garden Center	Café	Pool	Zoo	Fish & Chips Shop	Farm	Farmers Market	Fast Food Restaurant	Field
3 Central District	Coffee Shop	Performing Arts Venue	Ethiopian Restaurant	Vietnamese Restaurant	Bus Stop	Fish & Chips Shop	Chinese Restaurant	Liquor Store	BBQ Joint	Gym
4 Denny-Blaine	Park	Playground	Beach	Surf Spot	Monument / Landmark	Food Truck	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant

4. Clustering

4.1 Clustering neighbourhoods

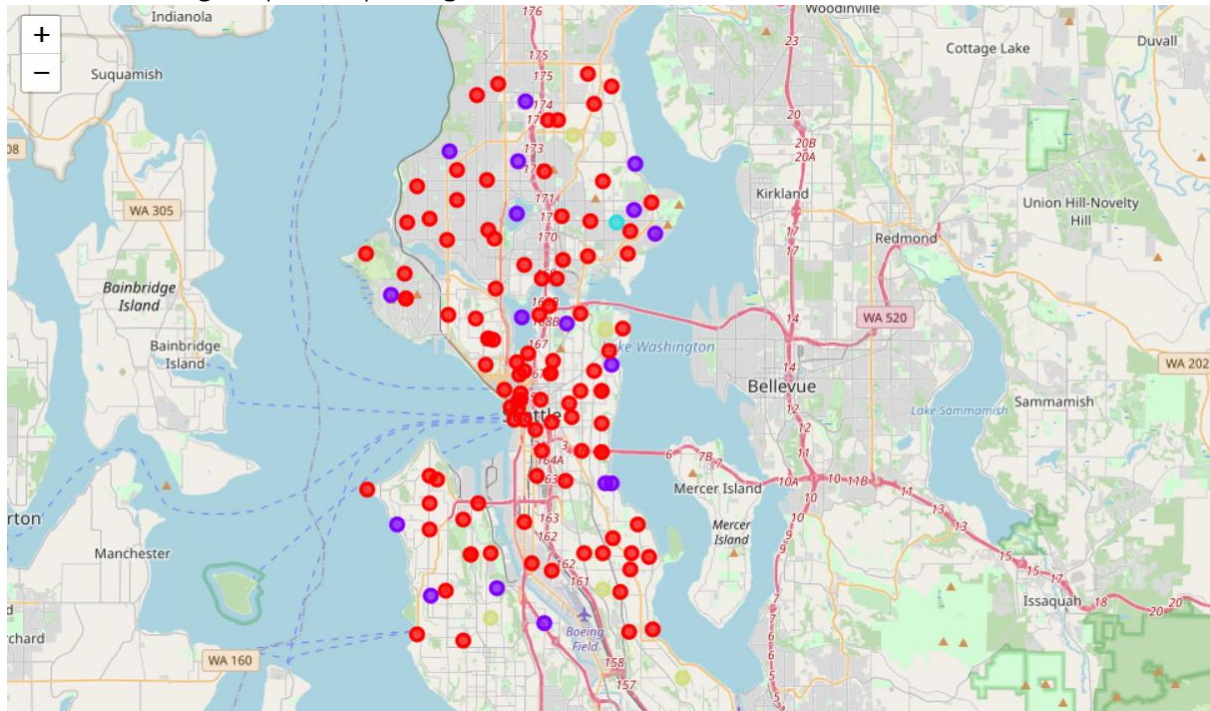
Now we can use K-Means clustering method to cluster the neighbourhoods. First, we need to determine how many clusters to use. This will be done using the Silhouette Score. We will define a function to plot the Silhouette Score that will be calculated using different number of clusters. Plotting graph for score vs cluster gives:



Above graph shows that $k=4$ gives the highest score so we cluster our neighbourhoods at $k=4$. We perform clustering and labels to that data frame. We drop nan rows and convert cluster labels column from float to int.

65	Eastlake	Lake Union	47.640825	-122.325723	0	Sandwich Place	Coffee Shop	Italian Restaurant	Playground	Bus Stop	Pizza Place	General Travel
66	Downtown	Central Seattle	47.598122	-122.328053	0	Chinese Restaurant	Coffee Shop	Sports Bar	General Entertainment	Bubble Tea Shop	Japanese Restaurant	Lounge
67	Denny Triangle	Downtown	47.618565	-122.337059	0	Coffee Shop	Hotel	Food Truck	Café	Mexican Restaurant	Bar	Italian Restaurant
68	Belltown	Downtown	47.613231	-122.345361	0	Coffee Shop	Bar	Hotel	Sushi Restaurant	Bakery	Seafood Restaurant	Breakfast Spot
69	Pike-Market	Downtown	47.619002	-122.319851	0	Bar	Coffee Shop	Mexican Restaurant	American Restaurant	Salon / Barbershop	Smoke Shop	Pizza Place
70	Central Business District	Downtown	47.609532	-122.336290	0	Coffee Shop	Hotel	American Restaurant	Breakfast Spot	Cocktail Bar	Seafood Restaurant	Clothing Store
71	First Hill	Downtown	47.609305	-122.324984	0	Sandwich Place	Coffee Shop	Hotel	Asian Restaurant	Pharmacy	Pizza Place	Food & Drink Shop

After clustering we plot map and get result as follows:



Now, you can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster.

5. Results and Discussion

After scrapping data from Wikipedia, we get 127 neighbourhoods and 34 Borough in Seattle. From above analysis shows that there are many restaurants present in Seattle out of which only 60 restaurants are American restaurants. Out of which most of the American restaurants in Neighbourhoods are present in Downtown, Lake Union, Central Area, West Seattle and Broadway. Highest number of American restaurants in Borough are in Downtown and in Neighbourhoods are in Central Business District. After performing k-means clustering for $k=4$ we get 4 different clusters out of which cluster 1 is the largest followed by cluster 2, cluster 4 and then cluster 3.

6. Conclusion

If we take careful look, we see that Downtown has highest number of American restaurants and all the downtown neighbourhoods are present in cluster 1. In Cluster 1 there is a lot of competition but there is an upside that is people might be familiar with American restaurants taste and if new restaurant has best taste it might work. Cluster 2 has different types of restaurants and cafe in its top 10 venues which shows that people are open to different tastes and as there is no American restaurant in there top 10 venues new American restaurants might have a chance. Cluster 3 and 4 is cluster which has liking of Fast-Food Filipino and Falafel Restaurants so opening new American restaurant there might be a challenging task.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhood's and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.