Clustering Restaurant Cuisines in Central Toronto

I. Introduction

Putting up a restaurant may require some research or market study. For one, when setting up a restaurant, it is essential to determine the current set-up of the area or the city. It may help to see the available restaurants in the area. Look into the trends, and what are the available cuisines in the city. Setting up a restaurant that introduces a new cuisine can be good. However, this can be risky at times. People of a community may have certain preferences, and this is that this project will aim to see.

The aim of this project is to look into the available cuisines, or restaurant in the area of **Central Toronto** . This will help determine the business person know what type of restaurants are available in a certain community. Normally, if a certain community have a lot of Fast Food Restaurants, it may mean that there is a great demand for it. People may prefer it. So for this project, we will see if **people from Central Toronto prefer Chinese, French, Italian, Japanese or others? Which among the neighborhood (of Central Toronto) prefers sushi restaurants, for example.**

II. Data

To solve the problem, it will rely on the available data of the Central Toronto. Provided below are some of the references.

1. For the available Neighborhoods in Central Toronto, we will be utilizing the Wikipedia page. Below is a sample snapshot.

M1A *Not assigned*	M2A *Not assigned*	M3A North York (Parkwoods)	M4A North York (Victoria Village)	M5A Downtown Toronto (Regent Park / Harbourfront)	M6A North York (Lawrence Manor / Lawrence Heights)	M7A Queen's Park (Ontario Provincial Government)	M8A *Not assigned*	M9A Etobicoke (Islington Avenue)
M1B Scarborough (Malvern / Rouge)	M2B *Not assigned*	M3B North York (Don Mills) North	M4B East York (Parkview Hill / Woodbine Gardens)	M5B Downtown Toronto (Garden District, Ryerson)	M6B North York (Glencairn)	M7B *Not assigned*	M8B *Not assigned*	M9B Etobicoke (West Deane Park / Princess Gardens / Martin Grove / Islington / Cloverdale)
M1C Scarborough (Rouge Hill / Port Union / Highland Creek)	M2C *Not assigned*	M3C North York (Don Mills) South (Flemingdon Park)	M4C East York (Woodbine Heights)	M5C Downtown Toronto (St. James Town)	M6C York (Humewood-Cedarvale)	M7C *Not assigned*	M8C *Not assigned*	M9C Etobicoke (Eringate / Bloordale Gardens / Old Burnhamthorpe / Markland Wood)
M1E Scarborough (Guildwood / Morningside / West	M2E *Not assigned*	M3E *Not assigned*	M4E East Toronto (The Beaches)	M5E Downtown Toronto (Berczy Park)	M6E York (Caledonia-Fairbanks)	M7E *Not assigned*	M8E *Not assigned*	M9E *Not assigned*

2. Since there are not much available information on the Latitude and Longitude to plot each neighborhood, we will use the data from a CSV file. Below is a screenshot of the available data:

Postal Code	Latitude	Longitude
M1B	43.80669	-79.1944
M1C	43.78454	-79.1605
M1E	43.76357	-79.1887
M1G	43.77099	-79.2169
M1H	43.77314	-79.2395
M1J	43.74473	-79.2395
M1K	43.72793	-79.262
M1L	43.71111	-79.2846
M1M	43.71632	-79.2395
M1N	43.69266	-79.2648
M1P	43.75741	-79.2733
M1R	43.75007	-79.2958
M1S	43.7942	-79.262
M1T	43.78164	-79.3043
M1V	43.81525	-79.2846
M1W	43.79953	-79.3184

3. Foursquare Data. To determine the actual restaurants available in Central Toronto, the study will utilize information from Foursquare!

III. Methodology

Here are the available steps to answer the problem stated in the introduction.

Since of the available data are in the internet (Wikipedia page), it is essential to clean and obtain data from the webpage for easier analysis. Below is the result after cleaning the data from Wikipedia.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

Since there are not much information on the Latitude and Longitude values, we will use the available csv file to integrate with the data obtained from Wikipedia. Thus, the captured value will be:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

Since we are only concern with Central Toronto, we will limit the Boroughs to this. Thus it can be seen that we will have 9 Postal Codes that are of concern.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
1	M5N	Central Toronto	Roselawn	43.711695	-79.416936
2	M4P	Central Toronto	Davisville North	43.712751	-79.390197
3	M5P	Central Toronto	Forest Hill North & West	43.696948	-79.411307
4	M4R	Central Toronto	North Toronto West	43.715383	-79.405678
5	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678
6	M4S	Central Toronto	Davisville	43.704324	-79.388790
7	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160
8	M4V	Central Toronto	Summerhill West, Rathnelly, South Hill, Forest...	43.686412	-79.400049

It can be plot using Folium to better visualize.

Next, we will access information from Foursqaure! to get information on the restaurants in each neighborhood.

```
CLIENT_ID = '' # your Foursquare ID
CLIENT_SECRET = '' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

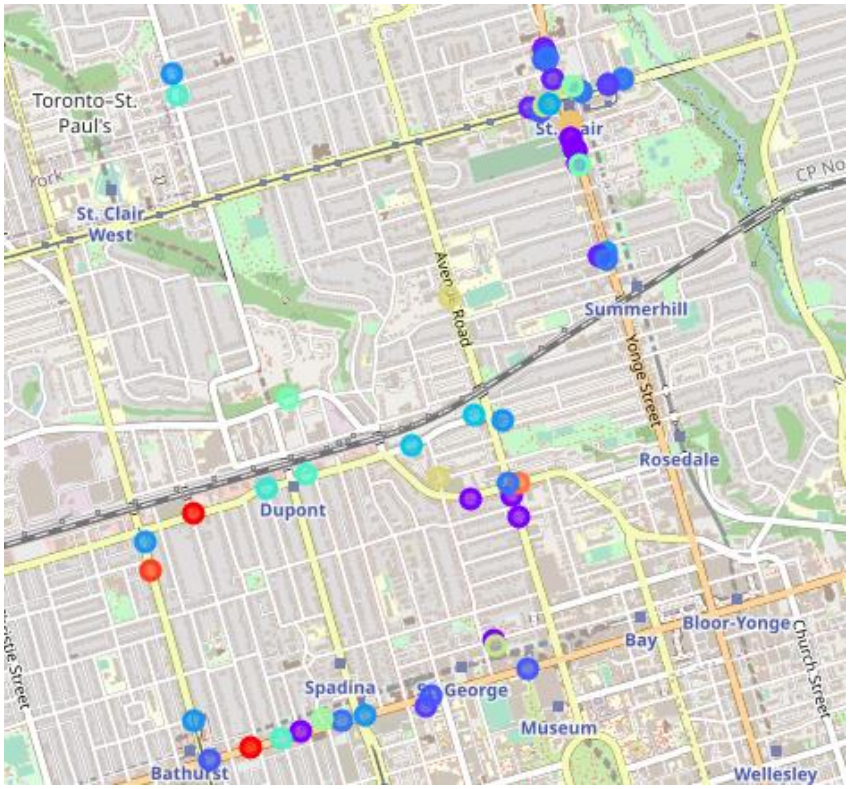
Below are the captured restaurant information in the Central Toronto Area.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roselawn	43.711695	-79.416936	7 Numbers	43.703630	-79.413724	Italian Restaurant
1	Roselawn	43.711695	-79.416936	Tokyo Sushi	43.704146	-79.410631	Sushi Restaurant
2	Roselawn	43.711695	-79.416936	Ferraro	43.703655	-79.413167	Italian Restaurant
3	Roselawn	43.711695	-79.416936	EDO	43.703754	-79.412802	Japanese Restaurant
4	Roselawn	43.711695	-79.416936	Kimono	43.704241	-79.410085	Sushi Restaurant
5	Roselawn	43.711695	-79.416936	Zen Sushi	43.703717	-79.414113	Sushi Restaurant
6	Davisville North	43.712751	-79.390197	Bar Buca	43.706961	-79.394808	Italian Restaurant
7	Davisville North	43.712751	-79.390197	La Vecchia Ristorante	43.710167	-79.399086	Italian Restaurant
8	Davisville North	43.712751	-79.390197	Kinton Ramen	43.707302	-79.395854	Ramen Restaurant
9	Davisville North	43.712751	-79.390197	Grazie Ristorante	43.709329	-79.398823	Italian Restaurant

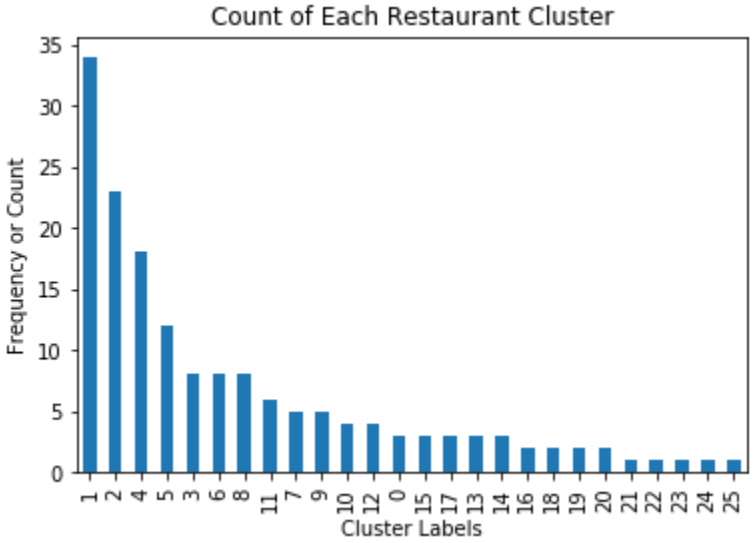
Once the data are properly formatted. These will be fitted in a kNN Clustering which will be further discussed in the Results section. The total number of clusters used is 26 which is equal to the number of unique Restaurant Categories.

IV. Results

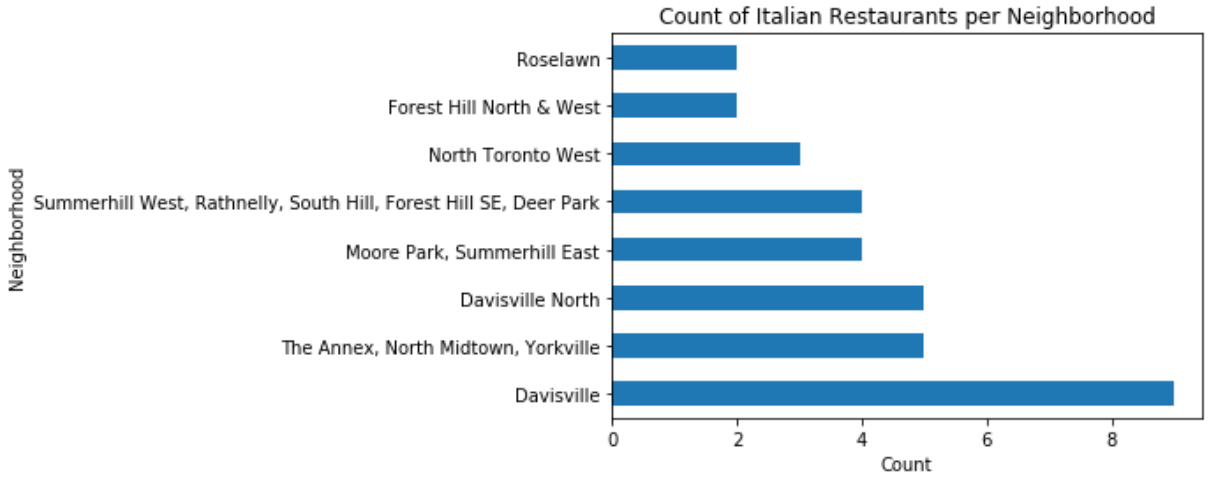
After the kNN clustering, the map below depicts the Clusters of Each restaurant. After careful observation, it can be said that each type of restaurant belong to the same cluster. For example, all Japanese Restaurants are in the same cluster.



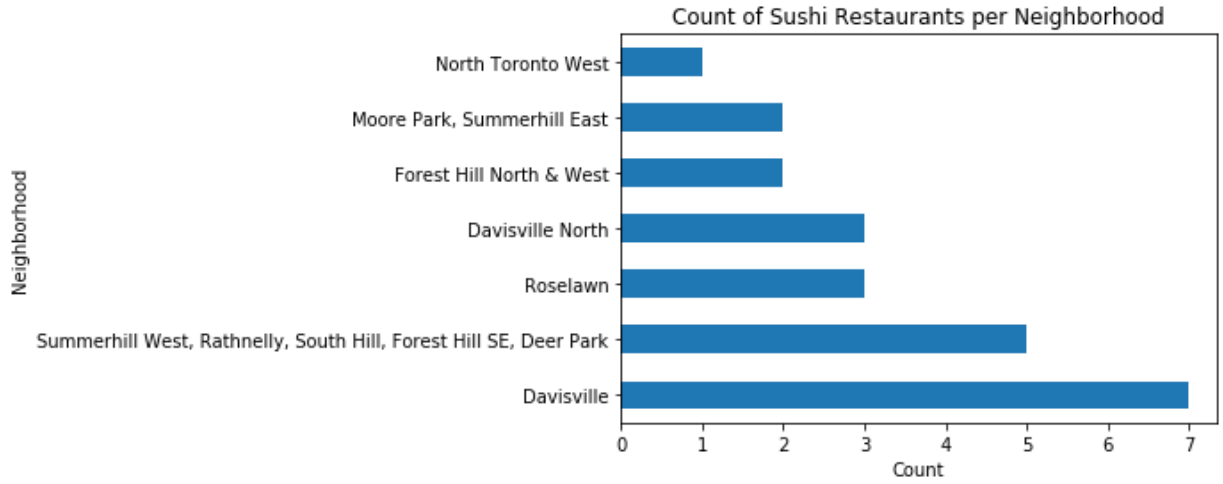
The Cluster with the most count is Cluster 1 which is Italian Restaurant.



Below bar graph shows which among the Neighborhood has the most number of Italian Restaurants.



The second highest restaurant type is Sushi Restaurant (Cluster 2). It can be seen from below that Davisville has the most number of Sushi Restaurants.



V. Discussion

There are some recommendations that may be used to make the study more accurate. As the data from the study relies on Foursqaure!, there are some Restaurants where in their specific category or cuisine was not listed. It was just Restaurant. After observation, this came in third. Thus, we can make it more accurate by conducting some survey on these restaurants. Also, it is essential to note here that Sushi Restaurant is different from Japanese Restaurant. Lastly, people there are a lot of factors that affect people's preference in cuisine. This study just based on the available restaurants.

VI. Conclusion

There are several conclusions from this study. One, the central Toronto Area is populated with Italian Restaurants. This can say a lot about a community. It may mean people prefer Italian Food especially people from the neighborhood of Davisville. Thus, if a chef would put up an Italian restaurant, he/she can assure that people from Central Toronto would be familiar with it. Also, it may be advisable to put up in Roselawn or Forest Hill North & west as there are fewer restaurants in this area.

The second highest was Sushi Restaurants. It can be concluded that there are a lot of Sushi Restaurants in Davisville. If this will be compared with the number of Italian Restaurants in Davisville, we can see that that Italian Restaurants are just ahead by 2. This may mean that there is a good demand in Davisville when it come to Italian and Sushi.

If Roselawn will be used to compare, it is seen that it was the least when it comes to Italian Restaurants. However, it came is third when it comes to Sushi Restaurants. It is ahead by 1. It may mean that people from Roselawn would prefer Sushi over Italian.