Opening a Filipino restaurant in Toronto

Mohammed Salem

24th of June 24, 2020

1.0 Introduction

1.1 Background & business problem

With over 250 ethnicities and 170 languages presented, Toronto has always been considered one of the most diverse multicultural cities. Out of those backgrounds, The Filipino community comes in third place behind China and India among the largest minorities in Canada according to 2016 census with about 27% growth since 2011 compared to the rest. Its estimated that this immigration trend will only move upward with the majority of the Filipino community residing in Toronto (62 %), consequently such increasing flow of settling requires an equivalent increase in services and amenities (ex. Restaurants).

Thus, this paper will utilize the foursquare API along with machine learning techniques (K-mean Clustering) to explore & analyze the different neighborhoods of Toronto. It will determine the most optimum location to open a Filipino restaurant that will serve a bigger audience and hence stand as a profitable choice.

1.2 Target Audience

This project will be mainly aimed at the following:

1) Existing Filipino Restaurants owners who are willing to expand their restaurants to new neighborhoods and enhance their reach.
2) Entrepreneurs of the Filipino community hoping to start a new venture in Toronto.
3) Foodies/ Tourists interested in finding neighborhoods with Filipino cuisines.
4) Filipino Newcomers/Residents who are considering to relocate to Toronto.

2.0 Data Processing

In this section, we will list & describe the various data resources that were used in this analysis, in order to give the reader a brief overview to the upcoming sections.

- List of Toronto`s Neighborhoods

First we used the Wikipedia page (en.Wikipidea.org/wiki/List_of_postal_codes_of_canada:M) which was used to list all the neighborhoods, Postal codes and Boroughs of Toronto . the detailed coding can be viewed in the attached notebook. In brief, beautiful soup package was imported to extract Toronto Neighborhood data from the website and converted to a dataframe. the neighborhoods were listed and cleaned into the following table:

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Table 1 List of Toronto's Boroughs & neighborhoods by Postal code

- The coordinates of the Neighborhoods

The csv file, “https://cocl.us/Geospatial_data” , was used to get all the geographical coordinates of the neighborhoods and then merged with the above table and the results is below:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Table 2: Adding the locations of the Neighborhoods

- The Filipino Demographic in Toronto

Another source used was the wiki page (en.Wikipidea.org/wiki/Demographic_of_Toronto) .It contains the ethnicity percentages in different ridings / Boroughs in Toronto. We scrapped this page using the Wikipedia package in order to find out the neighborhoods with the highest percentages of Filipino which will be a good place for a Filipino restaurant. Again the codes are shown in the notebook file , however I’ll explain briefly . The Wikipedia page shows the ethnicity percentages as per riding. we confined our ridings to include the ones with Filipino population in them. The final table looks like this:

	Riding	Population	Ethnic Origin #1	Ethnic Origin 1 in %	Ethnic Origin #2	Ethnic Origin 2 in %	Ethnic Origin #3	Ethnic Origin 3 in %	Ethnic Origin #4	Ethnic Origin 4 in %	Ethnic Origin #5	Ethnic Origin 5 in %
0	Eglinton-Lawrence	112925	Canadian	14.7	English	12.6	Polish	12.0	Filipino	11.0	Scottish	9.7
1	York Centre	103760	Filipino	17.0	Italian	13.4	Russian	9.5	Canadian	8.6	NaN	NaN
2	Don Valley East	93170	East Indian	10.6	Canadian	10.4	English	10.1	Chinese	8.9	Irish	8.1
3	Scarborough Centre	110450	Filipino	13.1	East Indian	12.2	Canadian	11.2	Chinese	10.7	English	7.8
4	Scarborough Southwest	108295	Canadian	16.2	English	14.3	Irish	11.5	Scottish	10.9	Filipino	9.5
5	Scarborough-Rouge Park	101445	East Indian	16.7	Canadian	11.8	Sri Lankan	11.1	English	9.8	Filipino	9.3
6	Scarborough-Guildwood	101115	East Indian	18.0	Canadian	11.6	English	9.7	Filipino	8.5	Sri Lankan	7.8

Table 3: The Filipino population in Ridings

- Locating venues in Toronto's neighborhoods

Foursquare API is among the most popular location data platform, it helps discover and locate the popular venues & attractions around the world. In this case, it was used to extract and explore the Filipino restaurants in Toronto and get their coordinates.

In order to retrieve a good number of venues, The search radius was expanded to 1700m and the venues limit to 900. Inputting the longitude and latitude of Toronto, and filtering the venues to only include "Restaurant" Category, the results were as followed:

	index	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	2	Malvern, Rouge	43.806686	-79.194353	Harvey's	43.800020	-79.198307	Restaurant
1	5	Malvern, Rouge	43.806686	-79.194353	Caribbean Wave	43.798558	-79.195777	Caribbean Restaurant
2	6	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.802008	-79.198080	Fast Food Restaurant
3	8	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
4	13	Malvern, Rouge	43.806686	-79.194353	Swiss Chalet	43.800236	-79.198366	Restaurant

Table 4: List of venues

3.0 Methodology

In this section, Using the above data, We will go thru the exploratory analysis & machine learning techniques (K –mean clustering) that were utilized in this project .

3.1 Filipino restaurants distribution

Location is one of the most vital factors for a successful business, Using Foursquare, we can determine the neighborhoods that are highly densed with Filipino restaurants in order to avoid the big & well established competition. First we used the table above and converted the categorical variables (Restaurants) into Numerical for better prediction later on, this process is called one hot encoding.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cantonese Restaurant	Caribbean Restaurant
0	Agincourt	0.0	0.000000	0.000000	0.000000	0.060606	0.000000	0.000000	0.000000	0.090909	0.060606
1	Alderwood, Long Branch	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.000000	0.142857	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Bayview Village	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Bedford Park, Lawrence Manor East	0.0	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Berczy Park	0.0	0.000000	0.111111	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table 5: One hot encoding

After that, we filtered the Filipino restaurants from the above table and merged it with the list of neighborhoods of Toronto and their coordinates. Then display the results on a bar graph as below

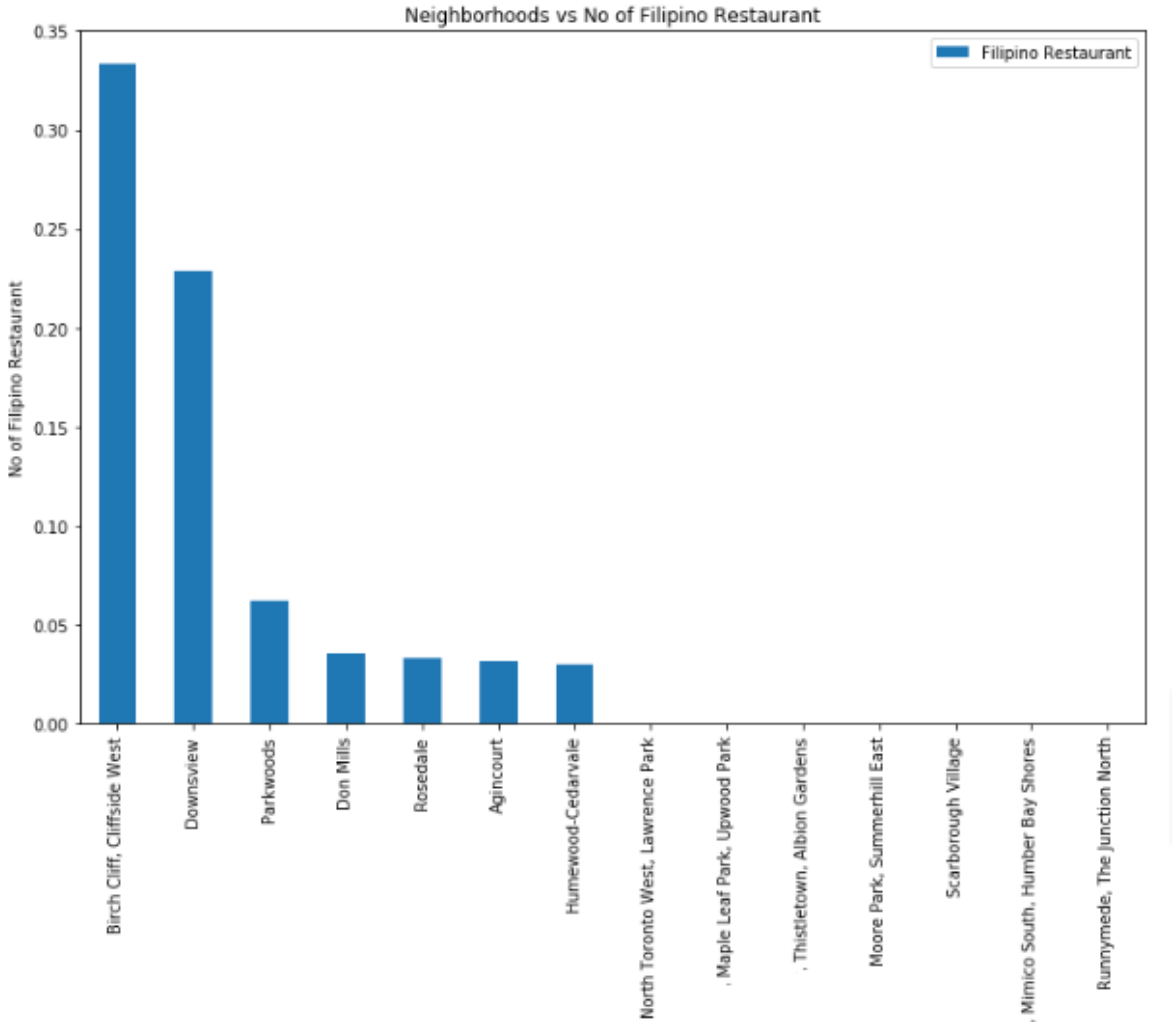

Figure 1: Numbers of Filipino Restaurants

3.2 Filipino Population distribution

Another instinctive factor in choosing a location for an ethnic restaurant is how popular that group of people are in that area. Thus, in this section, we extracted the ethnicity data of Toronto found in the table (table-3) and tried to sum all the percentages in each riding into total population in their respective neighborhoods. as the graph below shows. Placing a Filipino restaurant in a neighborhood with a very high Filipino population will definitely draw a higher traffic, which results in more profitability.

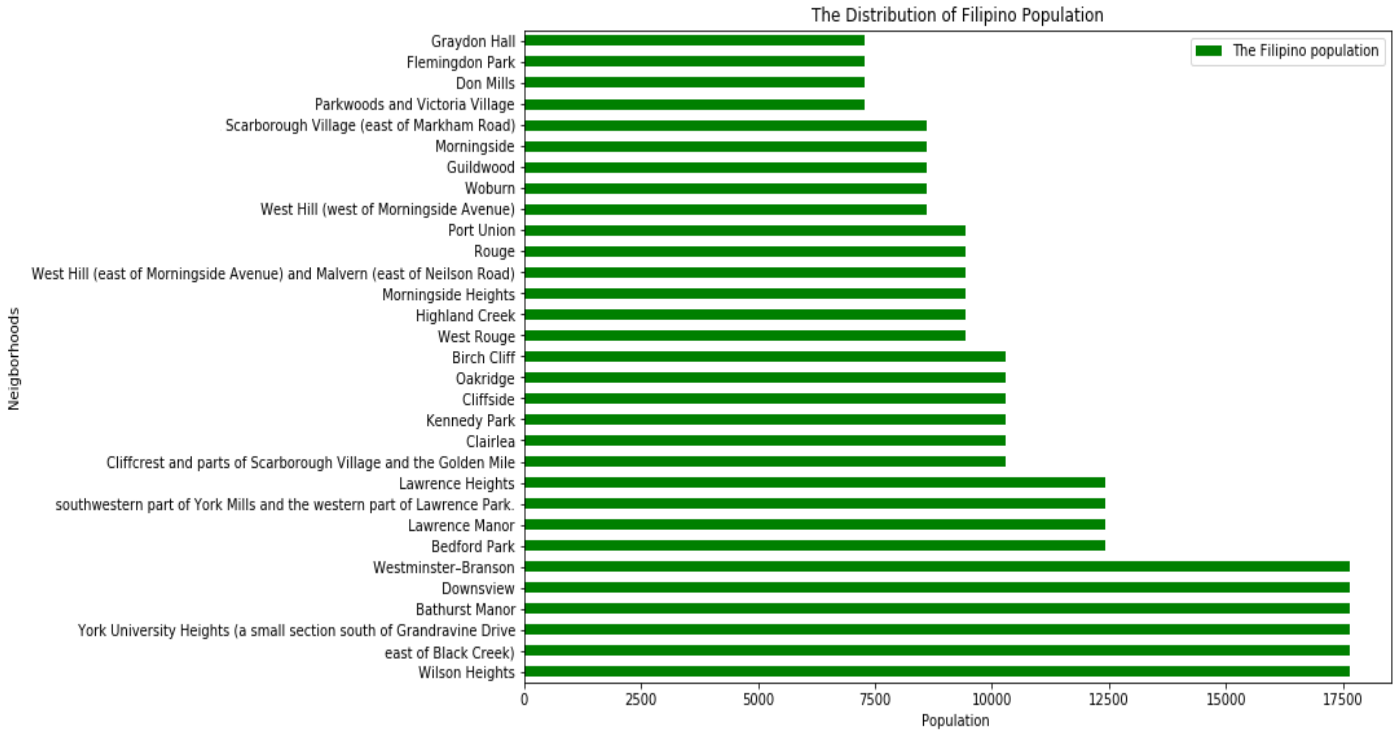

Figure 2: Neighborhoods with the highest Filipino population

3.3 Population v.s Restaurants

Here, after some data cleaning and merging, we linked the Filipino restaurants found earlier in section 3.1 with the highly populated neighborhoods. The results can be seen in the table below

	The Filipino population	Neighborhood	Filipino Restaurant
0	10288.025	Oakridge	0.0
1	9434.385	Rouge	0.0
2	12421.750	Lawrence Heights	0.0
3	17639.200	Wilson Heights	0.0
4	9434.385	Port Union	0.0
5	8594.775	Morningside	0.0
6	10288.025	Clairlea	0.0
7	9434.385	Highland Creek	0.0

Table 6: Filipino Population vs Restaurants

Unfortunately, the above results seem inconclusive where there is no clear relationship between the population and restaurants that can be observed. Hence it won't be considered in our analysis.

3.4 Neighborhoods clusters modeling

K-mean clustering is an unsupervised machine learning algorithm which groups objects in such a way that similar samples go into a cluster and dissimilar samples fall into different clusters. we used this algorithm, in this case, to cluster Toronto's neighborhoods based on the Filipino restaurants in each.

First, we used the elbow method to determine the optimum k that would minimize our modeling error. According to the graph below, we used K = 6 .

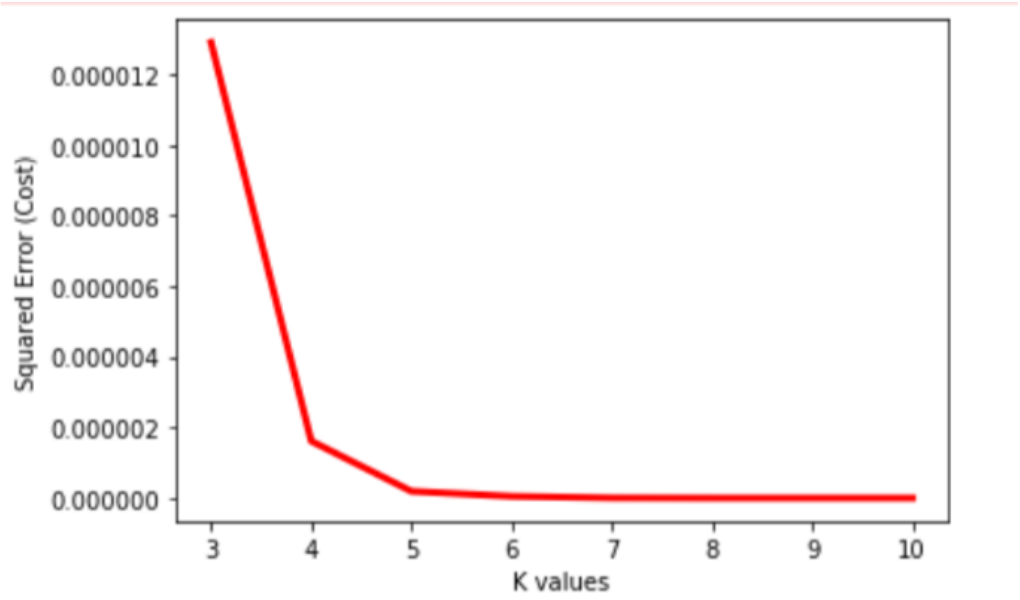

Figure 3: Elbow method

After running the modelling , we ended up with a total of 6 clusters. Next ,we merged the cluster labels with the previous list that contains Toronto`s neighborhoods with the corrsponding numbers of Filipino restaurants and their locations so we can examine them in the next section .

4.0 Results

In this section, we will highlight the most important findings from all the modelling, analysis & visualization carried out in the previous section.

Modelling findings

Firstly we will examine the 6 clusters resulting from the k mean cluster modelling above , where it can be observed that Cluster 1 doesn't contain any data points nor neighborhoods, hence we won't consider it in any decision making.

Next we have cluster 0 , It has the neighborhoods with no amount of Filipino Restaurants, Its shown in red color on the map

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Filipino Restaurant
0	M1G	Scarborough	Woburn	43.770992	-79.216917	0.0	0.0
1	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0.0	0.0
2	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0.0	0.0
4	M2H	North York	Hillcrest Village	43.803762	-79.363452	0.0	0.0
5	M2K	North York	Bayview Village	43.786947	-79.385975	0.0	0.0
6	M2P	North York	York Mills West	43.752758	-79.400049	0.0	0.0
7	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	0.0
14	M4A	North York	Victoria Village	43.725882	-79.315572	0.0	0.0
15	M4C	East York	Woodbine Heights	43.695344	-79.318389	0.0	0.0
16	M4E	East Toronto	The Beaches	43.676357	-79.293031	0.0	0.0
17	M4G	East York	Leaside	43.709060	-79.363452	0.0	0.0
18	M4H	East York	Thorncliffe Park	43.705369	-79.349372	0.0	0.0
19	M4M	East Toronto	Studio District	43.659526	-79.340923	0.0	0.0
20	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0.0	0.0
21	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0.0	0.0

Table 7: Cluster 0 results

Cluster 2 contains the neighborhoods with a wide range of distributed Filipino restaurants, Shown in blue.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Filipino Restaurant
3	M1S	Scarborough	Agincourt	43.794200	-79.262029	2.0	0.031250
23	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	2.0	0.033333
32	M6C	York	Humewood-Cedarvale	43.693781	-79.428191	2.0	0.033333

Table 8: : Cluster 2 results

Cluster 3 contains the neighborhoods that are highly occupied with Filipino restaurants, Shown in turquoise.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Filipino Restaurant
7	M3A	North York	Parkwoods	43.753259	-79.329656	3.0	0.0625

Table 9: : Cluster 3 results

Cluster 4 contains the neighborhoods that are relatively the least occupied with Filipino restaurants, Shown in olive green .

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Filipino Restaurant
8	M3B	North York	Don Mills	43.745906	-79.352188	4.0	0.017857

Table 10:: Cluster 4 results

Cluster 5 contains the neighborhoods that are relatively medially occupied Filipino restaurants, Shown in orange.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	Filipino Restaurant
10	M3K	North York	Downsview	43.737473	-79.464763	5.0	0.057143

Table 11:: Cluster 5 results

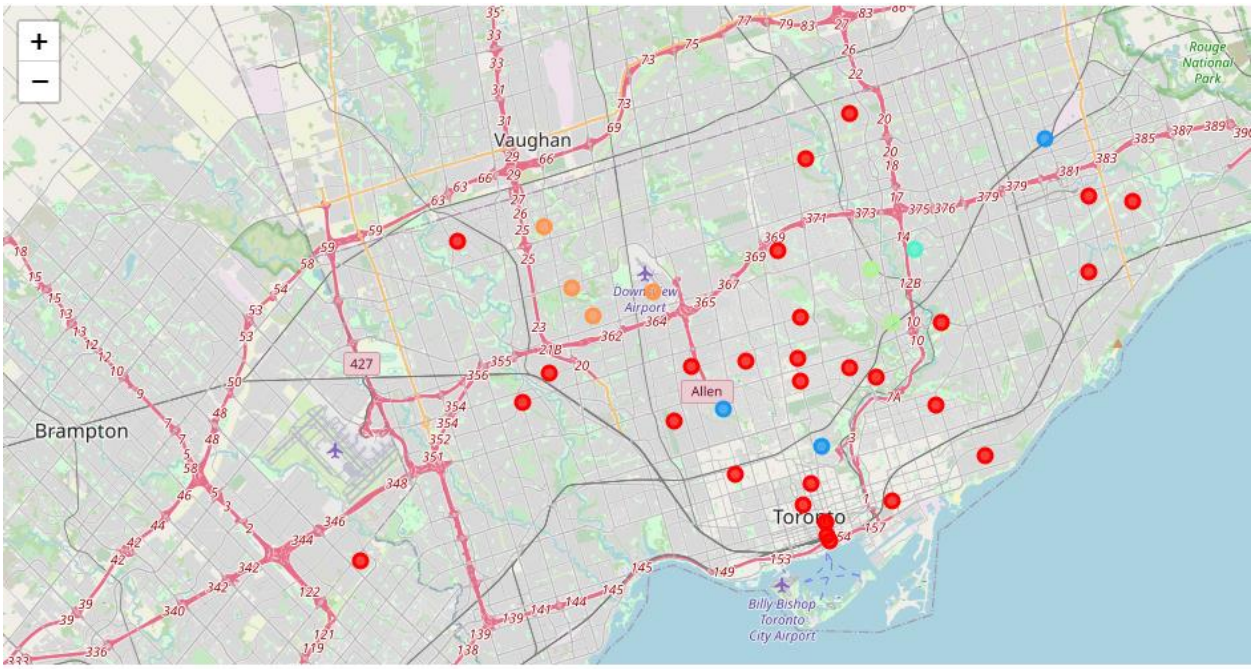

Figure 4: CLusters in a map

Analysis findings

After looking into the Filipino population distribution in Toronto, we can see that the majority of the Filipino residents are situated in the ridings of York Centre, Scarbourgh Centre and Eglinton Lawrence.

In terms of neighborhoods, some of the neighborhoods with the highest population of the Filipino community in descending order are Wilson Heights, East of Black Creek, York University Heights, Bathurst Manor, Westminster-Branson, Bedford Park, Lawrence Manor, Birch Cliff, Rough, Morningside Heights, Guildwood, Flemingdon Park and Dons Mills.

Using the foursquare API, refer to (Figure-1), It can be perceived that the highest number of Filipino Restaurants is located in the neighborhoods of Birch Cliff, Cliffside west, Parkwood, Downsview, Humewoods-Cedarvale, Don Mills, Rosedale and Agincourt.

5.0 Discussion

After taking a close look into the analysis & modelling findings, we will limit our location selection primarily to be in a cluster with low density of Filipino Restaurants, namely Cluster 0 & 4 while discarding the rest to avoid the fierce competition. Then we will choose the option with the neighborhood that contains the highest number of Filipino community.

Accordingly, four neighborhoods can be identified which are Scarborough village, Woburn, Victoria village & Don Mills. These neighborhoods should be a profitable option since they provide very low competition with high customer traffic.

Some of the points to take into consideration for a better future study are as follow:

1) Even with expanding the search parameters to 1700 m radius and venues limit to 900, the number of Filipino restaurants was still considered low comparing to some research done n Google maps. This could be due to mislabeling in Foursquare, where some Filipino restaurants may be simply referred to as “Asian Restaurant” or “Fast food Restaurant” or just “Restaurant. Thus its advisable to use another location data platform such as Google or Kaggle.

2) The Filipino population taken from Wikipedia only represents the neighborhoods of the highest percentage and not all Filipino communities, moreover it’s based on the 2016 census data which is over 5 years old.

3) Renting prices, safety and Public transportation availability should be considered to refine the options more. investors might be interested to opt for a cheap, safe neighborhood with a good transit connection.

6.0 Conclusion

In conclusion, this paper presented some exploratory analysis of foursquare & Wikipedia data along with machine learning techniques (K-mean Clustering). It was utilized to explore & examine the different neighborhoods of Toronto in order to determine the most optimum location to open a Filipino restaurant that will serve a bigger audience. Based on the clustering modelling, Filipino population & number of Filipino restaurant established, four neighborhoods were suggested for a successful restaurant location namely, Scarborough village, Woburn, Victoria village & Don Mills. For further refinement criteria, it was recommended for investors to look into the renting prices in these neighborhoods, in addition to the safety & transportation efficiency.