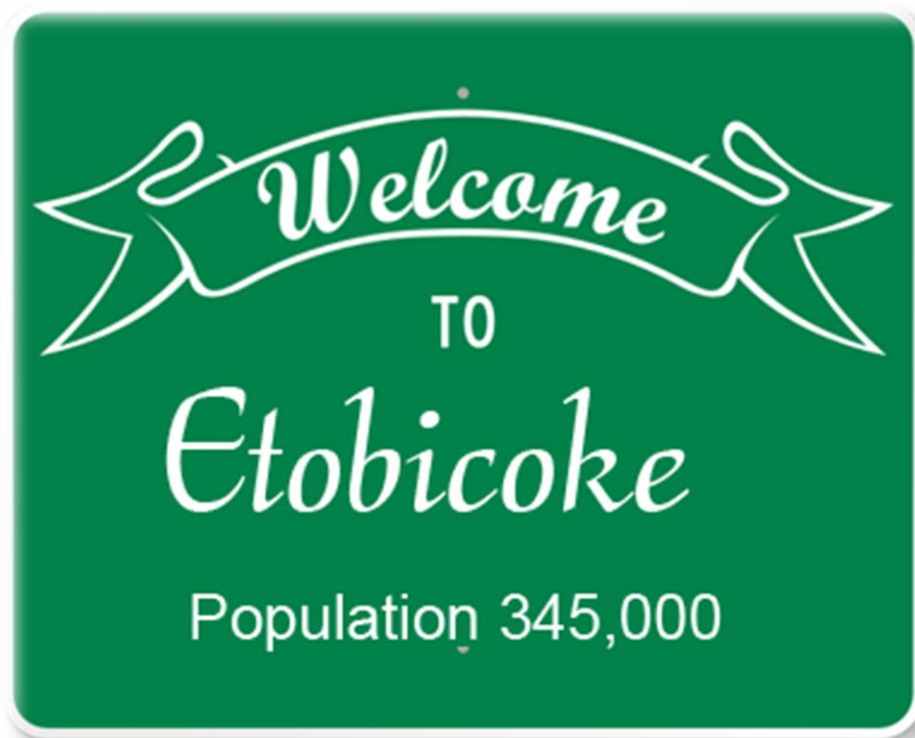


THE BATTLE OF NEIGHBORHOODS ETOBICOKE



IBM Coursera Applied Data Science Capstone PROJECT REPORT

Created by: Gajesh Solanki

ABSTRACT

Canada is one of the most diverse country in this world. It welcomes people from all over the world. Because of this welcoming nature, Canada is seeing a lot of migrants coming from different parts of the world. As the population grows, their demands grow too! Here, Jay who is the businessman and who belongs to an Indian community, he wants to open Plaza targeting to people who are coming from South Asia such as India, Pakistan, Bangladesh, Nepal, Sri Lanka, etc. But he is confused, and he needs help to decide. Using Machine Learning algorithm and other techniques we are going to do some analysis and help Jay to make his decision so he can open his plaza.

CONTENTS

1. Introduction

- 1.1. Background Information
- 1.2. Problem Statement: Where should Jay open South Asian Plaza?
- 1.3. Target Audience

2. Data

- 2.1. Data Description
- 2.2. Foursquare API Data

3. Methodology

4. Analysis

- 4.1. One Hot Encoding
- 4.2. K-Means Clustering
- 4.3. Data Wrangling
- 4.4. Visualize Map

5. Results

- 5.1. Analyze Clusters

6. Discussion

- 6.1. Cluster Comparison
- 6.2. Visualize Clusters in Map
- 6.3. Recommendation

7. Conclusion

8. References

1. INTRODUCTION

1.1. Background Information

Canada is a very diverse country which welcomes people from all over the world. There are several provinces in Canada in which Ontario is the biggest one by population and in this province GTA (Greater Toronto Area) is the most crowded and the most populous metropolitan area in Canada.

Etobicoke is one of the cities in this area, which has a highly diversified population, which totalled 365,143 in 2016. Out of that number more than 13% of people are from South Asia which means that the demand of their choice of foods and other items would be higher too!

In addition to these, Etobicoke is very near to the Pearson Airport (Biggest & busiest airport in Canada), Humber College (One of top colleges in Ontario for international students), The BAPS Shri Swaminarayan Temple (Religious place for Hindu community) and several other religious places (Gurudwara, Mosques, Temples)

1.2. Problem Statement: Where should Jay open South Asian Plaza?

In this report, we are trying to find a suitable neighborhood for Jay to make his decision that where should he open South Asian Plaza?

Jay is 40 years old who born in an Indian family. He came to Canada with his parents in 1990. He has been living in Etobicoke since then, but his problem is that since he came here to now, he could not find the one stop place where he can eat local Indian food, shop Indian groceries or other things such as buy sweets. He is a businessman who wants to help other people, not just Indian, but the people from the neighbour countries such as Pakistan, Nepal, Bangladesh, Sri Lanka who are homesick or who can find themselves at home and find things they want from one place.

Jay is looking for help from someone who can help him to decide where he can open this plaza. He wants this place to be located at very convenient location for everybody.

1.3. Target Audience

This report is an analysis of the neighborhoods of Etobicoke to help Jay to make his decision.

The information gathered from Foursquare in combination with Data Science methods are a good basis to derive data driven decisions regarding neighbourhoods that best fit the specific needs at hand. We are also going to use K-Means clustering algorithm to achieve the task.

Folium visualization library will be used to visualize the clusters superimposed on the map of Etobicoke. These clusters can be analyzed to help Jay to make his decision.

As we know already, Target Audience would be Jay who wants to open a South Asian Plaza in Etobicoke, but any stake holders can use this approach whether they want to open any shop, restaurants, or anything.

2. DATA

2.1. Data Description¶

We will use following Wikipedia page to scrape out postal codes of the Toronto area, their respective borough, and their neighborhoods. We will be focusing on Etobicoke borough only.

[Wikipedia page](#)

Here is data frame created from the above webpage after cleaning it:

	PostalCode	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

After getting postal codes, borough, and neighborhoods from the above Wikipedia page, we will be using Google maps Geocoding API to get the latitude and the longitude coordinates of each neighborhood. For simplicity of this project, I have gathered all the information in the CSV file. Here is a link to a csv file that has the geographical coordinates of each postal code:

[GeoSpatial Dataset](#)

Here is data frame after adding geospatial information to the above data frame:

	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

2.2. Foursquare API Data

We will need data about different venues in different neighborhoods of Etobicoke. To gain information, we will use Foursquare locational information. Foursquare is a location data provider with information about all manner of venues and events with an area of interest. Such information includes venue names, locations, menus and even photos. As such, the Foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to Foursquare API to gather information about venues inside each neighborhood. For each neighborhood, we have chosen the radius to be 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postal codes. The information obtained per venues as follows:

- i. Venue
- ii. Name of the venue
- iii. Venue Latitude
- iv. Venue Longitude
- v. Venue Category

Here is the data frame after adding above columns using Foursquare API:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	LCBO	43.602281	-79.499302	Liquor Store
1	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	New Toronto Fish & Chips	43.601849	-79.503281	Restaurant
2	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	Domino's Pizza	43.601583	-79.500905	Pizza Place
3	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	Delicia Bakery & Pastry	43.601403	-79.503012	Bakery
4	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	Lucky Dice Restaurant	43.601392	-79.503056	Café

Total 76 venues data have been obtained from Foursquare.

3. METHODOLOGY

Now, we have 12 neighborhoods data of Etobicoke. We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of 76 venues have been obtained in the whole city and 43 unique categories. Here we have only 12 neighborhoods so we will consider and analyze all.

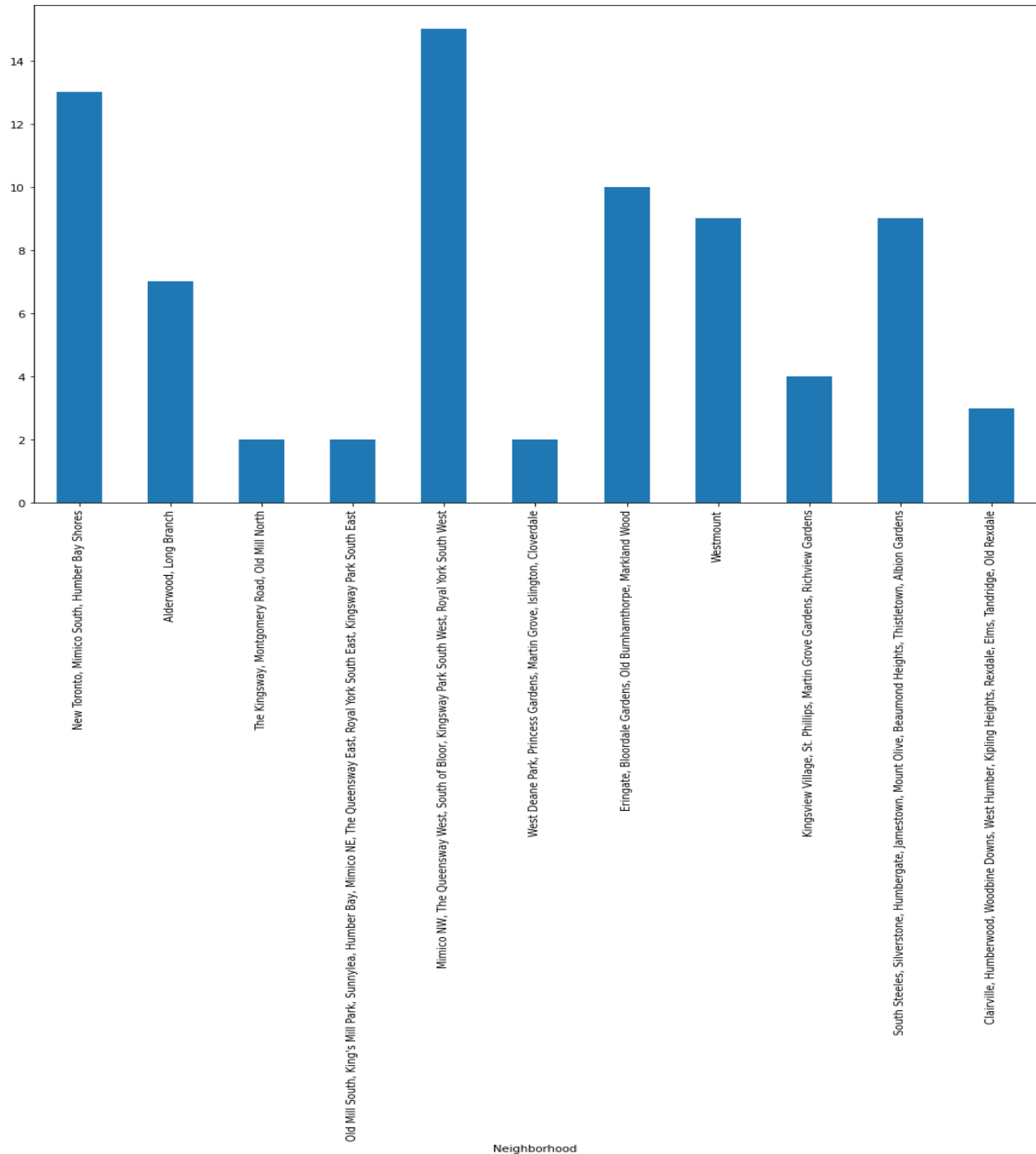
We can perform one hot encoding on the obtained data set and use it to find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. Here K-Means clustering technique have been used.

The clusters obtained can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest Jay some suitable locations based on the category.

4. ANALYSIS

We found that there were few neighborhoods with as low as 2 venues. But we did not remove those because we have only 12 neighborhoods, so we kept those.

See the bar chart given below:



4.1. One Hot Encoding

One hot encoding was performed on the data to obtain venue categories in each neighborhood. Then grouped data by neighborhood and take the mean value of the frequency of occurrences of each category.

See the output below:

	Neighborhood	American Restaurant	Bakery	Baseball Field	Beer Store	Burger Joint	Burrito Place	Bus Line	Café	Chinese Restaurant	Coffee Shop	Convenience Store	Discount Store	Drugstore
0	Alderwood, Long Branch	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.142857	0.000000	0.000000	0.000000
1	Clairville, Humberwood, Woodbine Downs, West H...	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333
2	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.000000	0.000000	0.0	0.100000	0.000000	0.000000	0.00	0.100000	0.000000	0.100000	0.100000	0.000000	0.000000
3	Kingsview Village, St. Phillips, Martin Grove ...	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Mimico NW, The Queensway West, South of Bloor,...	0.000000	0.066667	0.0	0.000000	0.066667	0.066667	0.00	0.000000	0.000000	0.000000	0.066667	0.066667	0.000000
5	New Toronto, Mimico South, Humber Bay Shores	0.076923	0.076923	0.0	0.000000	0.000000	0.000000	0.00	0.076923	0.000000	0.000000	0.000000	0.000000	0.000000
6	Old Mill South, King's Mill Park, Sunnylea, Hu...	0.000000	0.000000	0.5	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	South Steeles, Silverstone, Humbergate, Jamest...	0.000000	0.000000	0.0	0.111111	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	The Kingsway, Montgomery Road, Old Mill North	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	West Deane Park, Princess Gardens, Martin Grov...	0.000000	0.500000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Westmount	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.00	0.000000	0.111111	0.111111	0.000000	0.111111	0.000000

Then we created a new sorted data frame. In this data frame we put together top 10 venues for each neighborhood.

See the image below:

	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	Alderwood, Long Branch	Pizza Place	Pub	Sandwich Place	Pharmacy	Gym	Coffee Shop	Convenience Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant
1	Clairville, Humberwood, Woodbine Downs, West H...	Truck Stop	Drugstore	Rental Car Location	Convenience Store	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant
2	Eringate, Bloordale Gardens, Old Burnhamthorpe...	Liquor Store	Café	Park	Pet Store	Pharmacy	Pizza Place	Coffee Shop	Convenience Store	Shopping Plaza	Beer Store
3	Kingsview Village, St. Phillips, Martin Grove ...	Mobile Phone Shop	Park	Sandwich Place	Bus Line	Wings Joint	Discount Store	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant
4	Mimico NW, The Queensway West, South of Bloor,...	Wings Joint	Burger Joint	Grocery Store	Flower Shop	Fast Food Restaurant	Discount Store	Convenience Store	Gym	Burrito Place	Hardware Store
5	New Toronto, Mimico South, Humber Bay Shores	Liquor Store	Pizza Place	Bakery	Café	Fast Food Restaurant	Flower Shop	Gym	Mexican Restaurant	Pet Store	Pharmacy
6	Old Mill South, King's Mill Park, Sunnylea, Hu...	Pool	Baseball Field	Wings Joint	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore
7	South Steeles, Silverstone, Humbergate, Jamest...	Grocery Store	Pizza Place	Fried Chicken Joint	Beer Store	Sandwich Place	Fast Food Restaurant	Pharmacy	Wings Joint	Discount Store	Flower Shop
8	The Kingsway, Montgomery Road, Old Mill North	Park	River	Wings Joint	Convenience Store	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore
9	West Deane Park, Princess Gardens, Martin Grov...	Bakery	Filipino Restaurant	Wings Joint	Convenience Store	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Drugstore
10	Westmount	Pizza Place	Intersection	Middle Eastern Restaurant	Sandwich Place	Chinese Restaurant	Coffee Shop	Discount Store	Playground	Wings Joint	Fried Chicken Joint

4.2. K-Means Clustering

K-Means clustering technique was used to create different clusters from the given data. We chose 7 clusters for this data because it seems there are not much similarity between neighborhoods. However, we found some outliers while doing analysis which is described later in this report.

See the sample output of the new data frame below where new column was added to assign Cluster Labels:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Cluster Labels	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
New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	6	Liquor Store	Pizza Place	Bakery	Café	Fast Food Restaurant	Flower Shop	Gym	Mexican Restaurant	Pet Store	Pharmacy
Alderwood, Long Branch	43.602414	-79.543484	1	Pizza Place	Pub	Sandwich Place	Pharmacy	Gym	Coffee Shop	Convenience Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant
The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	4	Park	River	Wings Joint	Convenience Store	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore
Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	3	Pool	Baseball Field	Wings Joint	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore
Mimico NW, The Queensway West, South of Bloor,....	43.628841	-79.520999	6	Wings Joint	Burger Joint	Grocery Store	Flower Shop	Fast Food Restaurant	Discount Store	Convenience Store	Gym	Burrito Place	Hardware Store

4.3. Data Wrangling

We also did some data wrangling such as replacing all NaN values to 0 and add 1 to all cluster labels so 0 became 1, 1 became 2 and so on.

See the sample Python code below:

```
# In order to convert float to integer, we need to make sure that there is no NaN present
EB_merged['Cluster Labels'] = EB_merged['Cluster Labels'].replace(np.nan, 0)

#EB_merged['Cluster Labels'] = EB_merged['Cluster Labels'].apply(np.int64)

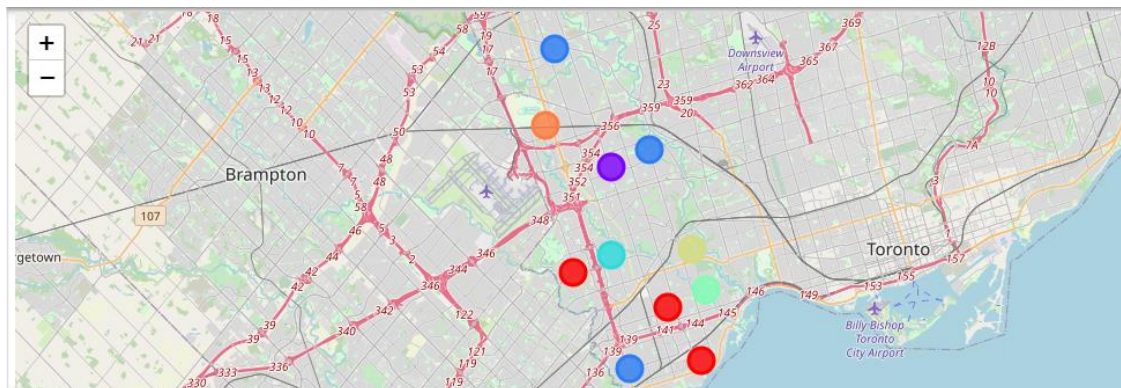
# +1 added to each cluster labels so 0 will become 1. 1 will become 2
EB_merged['Cluster Labels'] = EB_merged['Cluster Labels'] + 1

#EB_merged = EB_merged.drop(['PostalCode'], axis=1)

#EB_merged.head()
EB_merged
```

4.4. Visualize Map

See the below image for those 7 clusters which are flagged with circles with different colors:



5. RESULTS

As we seen above that we created 7 clusters. Let us analyze those below.

5.1. Analyze Clusters

Cluster 1: The top venues in cluster 1 are Mobile phone shop, Park, Sandwich plaza, Bus Line and Wings Joint.

	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
8	Kingsview Village, St. Phillips, Martin Grove ...	Mobile Phone Shop	Park	Sandwich Place	Bus Line	Wings Joint	Discount Store	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant

Cluster 2: The top venues in cluster 2 are Pizza places, Pub, Grocery Store, Intersection and Sandwich places.

	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
1	Alderwood, Long Branch	Pizza Place	Pub	Sandwich Place	Pharmacy	Gym	Coffee Shop	Convenience Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant
7	Westmount	Pizza Place	Intersection	Middle Eastern Restaurant	Sandwich Place	Chinese Restaurant	Coffee Shop	Discount Store	Playground	Wings Joint	Fried Chicken Joint
9	South Steeles, Silverstone, Humbergate, Jamest...	Grocery Store	Pizza Place	Fried Chicken Joint	Beer Store	Sandwich Place	Fast Food Restaurant	Pharmacy	Wings Joint	Discount Store	Flower Shop

However, there are many venues located multiple times in this cluster. See below:

Pizza Place	3
Fried Chicken Joint	3
Sandwich Place	3
Flower Shop	2
Discount Store	2
Coffee Shop	2
Pharmacy	2
Wings Joint	2

Cluster 3: The top venues in cluster 3 are Bakery, Filipino Restaurant, Wings Joint, Convenience Store and Gym.

	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
5	West Deane Park, Princess Gardens, Martin Grov...	Bakery	Filipino Restaurant	Wings Joint	Convenience Store	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Drugstore

Cluster 4: The top venues in cluster 4 are Pool, Baseball Field, Wings Joint, Grocery Store and Gym.

	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
3	Old Mill South, King's Mill Park, Sunnylea, Hu...	Pool	Baseball Field	Wings Joint	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore

Cluster 5: The top venues in cluster 5 are Park, River, Wings Joint, Convenience Store and Grocery Store.

	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
2	The Kingsway, Montgomery Road, Old Mill North	Park	River	Wings Joint	Convenience Store	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Drugstore

Cluster 6: The top venues in cluster 6 are Truck Stop, Drugstore, Rental Car Location, Convenience Store and Gym.

	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
10	Clairville, Humberwood, Woodbine Downs, West H...	Truck Stop	Drugstore	Rental Car Location	Convenience Store	Gym	Grocery Store	Fried Chicken Joint	Flower Shop	Filipino Restaurant	Fast Food Restaurant

Cluster 7: The top venues in cluster 7 are Pizza places, Liquor Store, Gym, Pharmacy and Café.

	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	New Toronto, Mimico South, Humber Bay Shores	Liquor Store	Pizza Place	Bakery	Café	Fast Food Restaurant	Flower Shop	Gym	Mexican Restaurant	Pet Store	Pharmacy
4	Mimico NW, The Queensway West, South of Bloor,...	Wings Joint	Burger Joint	Grocery Store	Flower Shop	Fast Food Restaurant	Discount Store	Convenience Store	Gym	Burrito Place	Hardware Store
6	Eringate, Bloordale Gardens, Old Burnhamthorpe...	Liquor Store	Café	Park	Pet Store	Pharmacy	Pizza Place	Coffee Shop	Convenience Store	Shopping Plaza	Beer Store

However, there are many venues located 2 times in this cluster. See below:

```

Pharmacy          2
Liquor Store      2
Convenience Store 2
Gym               2
Fast Food Restaurant 2
Pizza Place       2
Pet Store         2
Café              2
Flower Shop       2

```

Note: I created cluster 8 as the mimic of cluster 1 to create a better visualization. Please disregard it.

6. DISCUSSION

6.1. Cluster Comparison

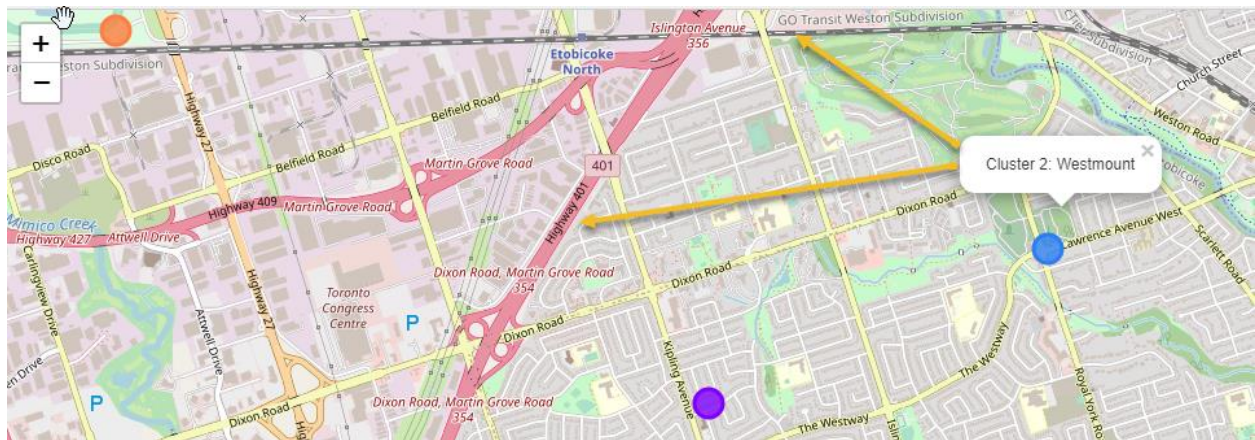
Now that we have the clusters and the top venue categories let us visualize the top 5 most common venue categories in each cluster for comparison.



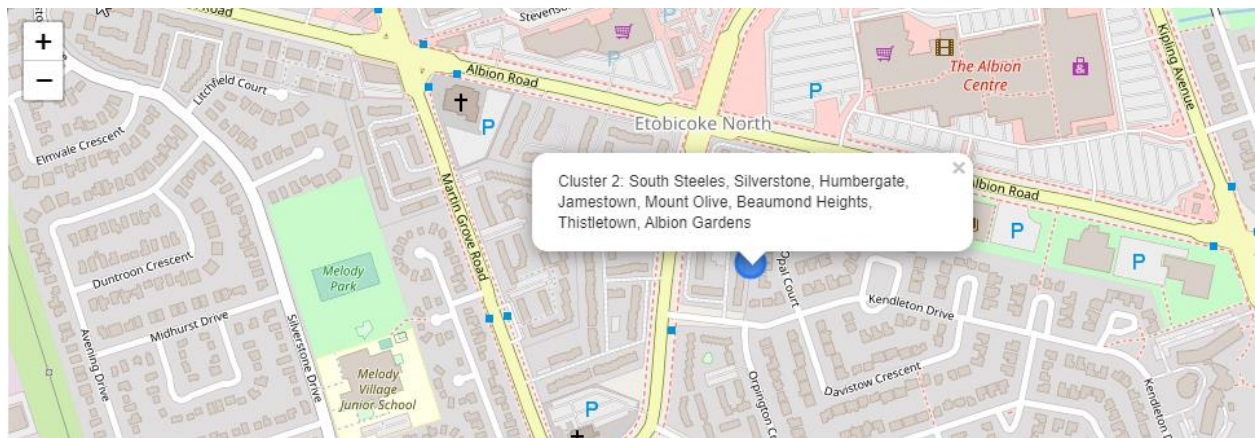
6.2. Visualize Clusters in Map

Let us visualize clusters in the map and then make a recommendation.

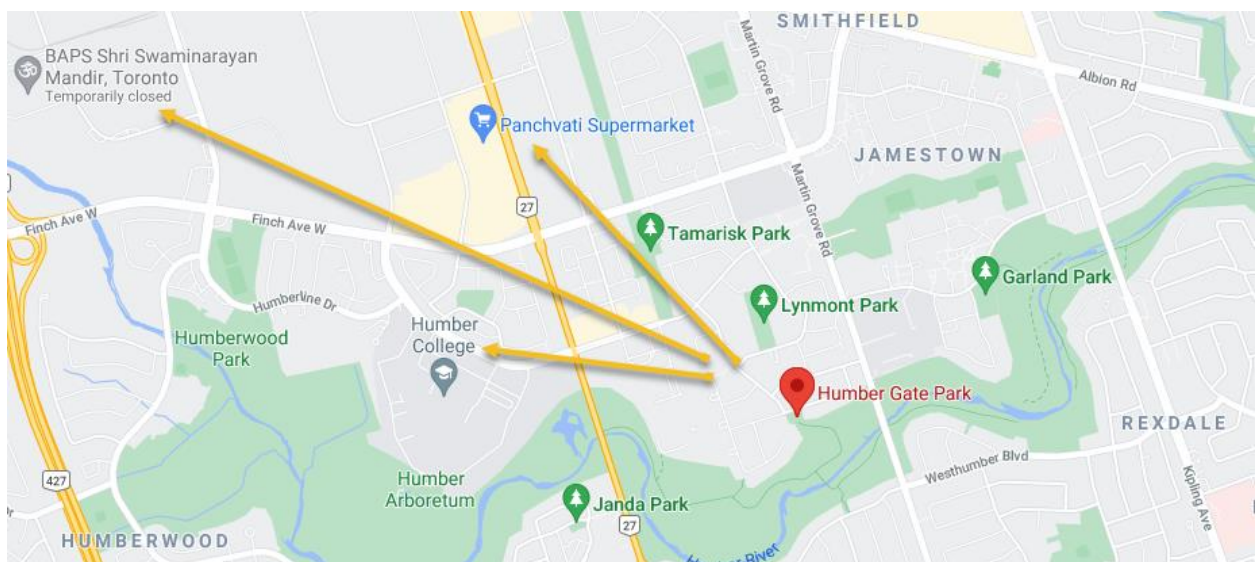
Cluster 2: Westmount neighborhood



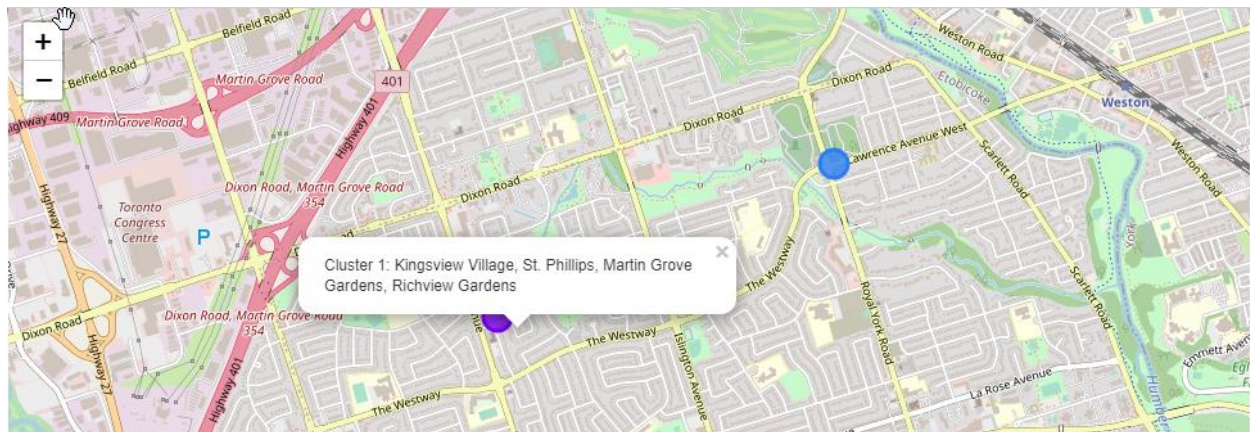
Cluster 2: Humber gate



Let us visualize this area in Google Maps as well.



Cluster 1: Kingsview Village



6.3. Recommendation

The above slides of bar charts and maps can be used to suggest valuable information to Jay. Let us discuss them below.

Cluster 2:

- It is very clear that there are many venues belong to this cluster such as few pizza places, some Asian restaurants, Fast food restaurants, Playground, Gym, etc. so it clearly gives an advantage to open any business in these neighborhoods. Jay can open his plaza in either Westmount neighborhood or near Humber gate. Let us discuss few advantages of these two neighborhoods.
- **Westmount:** As it shows in the map, it is very near to highway 401 which can attract daily commuters to this plaza. Moreover, this neighborhood is in residence area so it is possible that this plaza would attract many inhabitants and not just South Asian community.
- **Humber gate:** As it is depicted in the above map that this area is very near to Humber College and to very popular Hindu temple (BAPS table) which can attract many students and devotees.

Cluster 1:

- Kingsview village in this cluster is very similar to the Westmount area of cluster 2. In addition to that, it has Bus line too so it may attract some people who has no vehicles such as students.

7. CONCLUSION

- In this project, the purpose was to analyze all the neighborhoods located in Etobicoke city and cluster them using k-Means clustering techniques. We had to provide suggestion to Jay so he could open his plaza in suggested neighborhoods.
- We used Foursquare API to get necessary information of venues located near to each neighborhood and their respective geospatial data. As we saw in previous visualization in discussion section there are not many popular neighborhoods in Etobicoke except some.
- So, we took few neighborhoods into consideration and we provided suggestion that Jay can open his plaza either in Westmount or near to Humber gate.
- Kingsview village is very similar to Westmount neighborhood so we can remove Kingsview village from the consideration.
- We strongly recommend that Jay opens his plaza near to Humber gate area because it would boost his business. This place is very near to Humber college and some popular religious places nearby.

8. REFERENCES

1. Foursquare: <https://foursquare.com/>
2. Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_N
3. Geocoder: <https://geocoder.ca/>
4. Folium Map: <https://python-visualization.github.io/folium/>
5. Google Maps: <https://www.google.com/maps/place/Humber+Gate+Park>
6. Stack Overflow: <https://stackoverflow.com/>
7. Github: <https://github.com/Gajesh229/Segmenting-and-Clustering-Neighborhoods-in-Toronto>