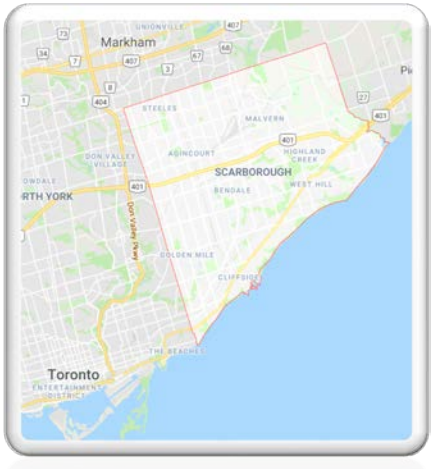
Capstone Project - The Battle of Neighborhoods

Analyzing Average House Prices and School Ratings for Scarborough, Canada for Immigrants



Project Description

Every year, many people migrating to various states of Canada require to search for a great location with the good housing price; those families with kids are under even greater pressure to find an ideal neighborhood to suit their family's needs. Whether a neighborhood has great schools or is located in a coveted school district is the main consideration for 39 per cent of Canadians with aspirations to purchase a home.

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship.

The project aims to create an analysis of features for the neighborhood as a comparative analysis between neighborhoods. The features include median house price and school ratings, crime rates, weather conditions, recreational facilities. This would help people to get awareness of the places before moving to a new country, state, city or place for their work or to start a new life. The aim of this project is to help the people explore different possibilities and make a better decision on choosing their

best neighborhood out of many neighborhoods in Scarborough city based on the distribution of various facilities in and around the neighborhood.

Location

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship.

Data We Need

This project will require knowledge of the different neighborhoods in Toronto, school ratings and median house prices. As such the neighborhood data required will be:

- Neighborhood location in terms of latitude and longitude
- School Ratings
- Median House Prices

Longitude and Latitude Data

We will need Geo-locational information about that specific borough and the neighborhoods in that borough. It is "Scarborough" in Toronto, CA.

A dataset comprising latitude and longitude, postal codes were already available through the previous notebook. The location of Scarborough would be filtered using the same:

https://github.com/Nazaninr/Coursera_Capstone/blob/master/Battle%20of%20Neighborhoods%20-%20Final%20...

School Rates Data

A data comprising the school rates of over 6 in the Scarborough were already available through the link bellow:

http://ontario.compareschoolrankings.org/elementary/SchoolsByRankLocationName.aspx

Average House Prices Data

A data frame comprising the average house prices based on postal codes and neighborhoods in the Scarborough is provided within the project.

Foursquare API Data

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within 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 for all the stated required information can be obtained through the API. After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues in each and every neighborhood. For each neighborhood, we have chosen the radius to be 300 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 venue as follows:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Name of the venue e.g. the name of a store or restaurant
- Venue Latitude
- Venue Longitude
- Venue Category

Work Flow

Using credentials of Foursquare API features of nearby places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood

parameter would reasonably be set to 100 and the radius parameter would be set to 300. The steps taken were:

1. Data acquisition and cleansing
2. Data preparation
3. Feature selection
4. Clustering

Data acquisition and cleansing

Data acquisition of this project was a 2-step process:

1. Obtaining the postal codes for neighborhoods in Toronto
2. Obtaining venues within these neighborhoods

Data Preparation

To do so, we need to process the retrieved data and creating a data frame for all the venues inside the Scarborough. When the data are completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" will be One-hot encoded and different venues will have different feature-columns.

Creating a Data Frame

Processing the retrieved data and creating a data frame for all the venues inside the Scarborough. It can be seen in Figure 1.

	Neighborhood	American Restaurant	Athletics & Sports	Bakery	Bank	Bar	Breakfast Spot	Bus Line	Bus Station	Café	Caribbean Restaurant	Chinese Restaurant	Coffee Shop	College Stadium	Convenience Store	Department Store	Discount Store	Electronics Store	Fast Food Restaurant	Fried Chicken Joint	Gaming Cafe
0	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1	Highland Creek, Rouge Hill, Port Union	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

Figure 1 creating a data frame for all the venues inside the Scarborough

Finding The Most Common Venues near Each Neighborhood

As the aim of this project is to create an analysis of features in each neighborhood and doing a comparative analysis between them. In order to the comparative analysis, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in the Scarborough city. The features included are average house price and school ratings.

This would help people to get awareness of the places before moving to a new place for their work or to start a new life. For this means, we find the most common venues near each neighborhood as it is shown in Figure 2

	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	Agincourt	Chinese Restaurant	Skating Rink	Sandwich Place	Breakfast Spot	Lounge	College Stadium	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint
1	Agincourt North, L'Amoreaux East, Milliken, St...	Playground	Park	Vietnamese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
2	Birch Cliff, Cliffside West	College Stadium	General Entertainment	Skating Rink	Café	Vietnamese Restaurant	Hakka Restaurant	Grocery Store	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
3	Cedarbrae	Thai Restaurant	Athletics & Sports	Bakery	Bank	Hakka Restaurant	Caribbean Restaurant	Fried Chicken Joint	Department Store	Hobby Shop	Grocery Store
4	Clairlea, Golden Mile, Oakridge	Bakery	Bus Line	Intersection	Soccer Field	Bus Station	Fast Food Restaurant	Park	Metro Station	Convenience Store	Grocery Store

Figure 2 The most common venues near each neighborhood

Feature Selection

For the purposes of this project, the definition of a good neighborhood is the one that has an appreciable commercial presence within a given community as well as:

- Compare average housing prices in each postal code and neighborhood
- Compare school ratings in each postal code and neighborhood

Clustering Approach

The project is intended to help the people to explore different possibilities and make the best decision about choosing the best neighborhood out of many neighborhoods in Scarborough city based on the distribution of various facilities in and around the neighborhood. To be able to do that, we need to use a clustering approach which is a form of unsupervised machine learning: k-means clustering algorithm as it shows in Figure 3.

```
# set number of clusters
n_clusters = 4
# Using k-means to cluster the neighborhood into 4 clusters.

Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters, random_state=0).fit(Scarborough_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

Figure 3 K-Means Clustering

We create a data frame and display the top 10 venues for each neighborhood with its cluster belonging to as shown in Figure 4.

	Postalcode	Borough	Neighborhood	Latitude	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
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	3	Fast Food Restaurant	Vietnamese Restaurant	College Stadium	Hakka Restaurant	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Electronics Store	Discount Store
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	1	Moving Target	Bar	Vietnamese Restaurant	Hakka Restaurant	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1	Intersection	Spa	Rental Car Location	Breakfast Spot	Pizza Place	Electronics Store	Medical Center	Mexican Restaurant	College Stadium	General Entertainment
3	M1G	Scarborough	Woburn	43.770992	-79.216917	2	Coffee Shop	Korean Restaurant	Pharmacy	College Stadium	Grocery Store	General Entertainment	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	1	Thai Restaurant	Athletics & Sports	Bakery	Bank	Hakka Restaurant	Caribbean Restaurant	Fried Chicken Joint	Department Store	Hobby Shop	Grocery Store

Figure 4 Ten Most Common Venues in each Neighborhood

Visualize the Resulting Clusters

In this project, the neighborhoods are clustered into 4 groups by the use of k-means clustering algorithm as it can be seen in Figure 5.

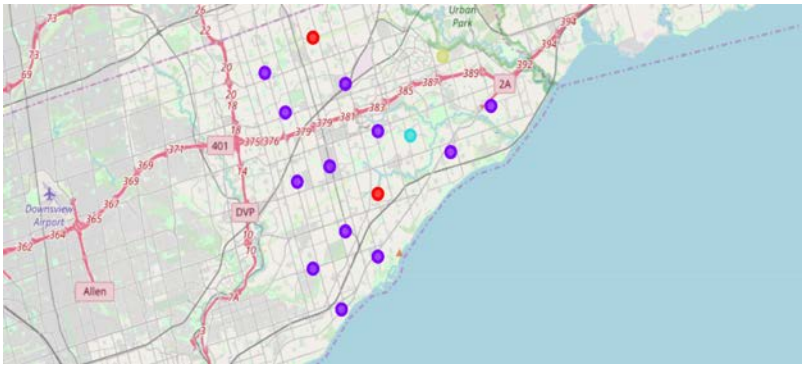


Figure 5 Visualizing the Clusters in Scarborough

Postal Codes in Each Cluster in Scarborough

In the Figure 6, the neighborhoods of the Scarborough in each cluster are represented.

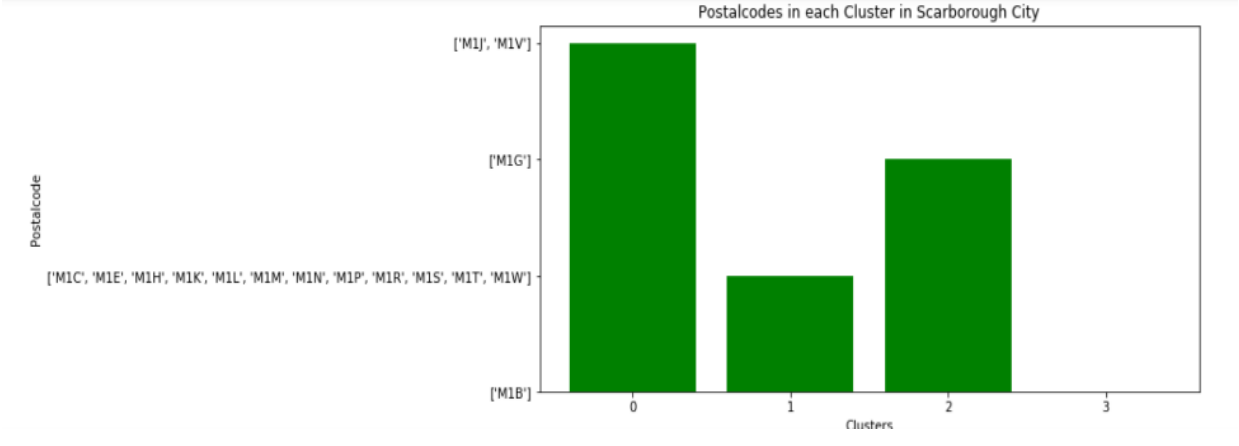


Figure 6 Postal Codes in Each Cluster in Scarborough

Average Housing Prices in Each Postal Code

As it can be seen from the Figure 7, the average house prices in each Postal code, it is obvious that the houses which are located in M1J postal code have the highest average price among the other postal code in Scarborough city. On the other hand,

the houses which are located in M1E postal code have the lowest average price among the other postal code in Scarborough city.

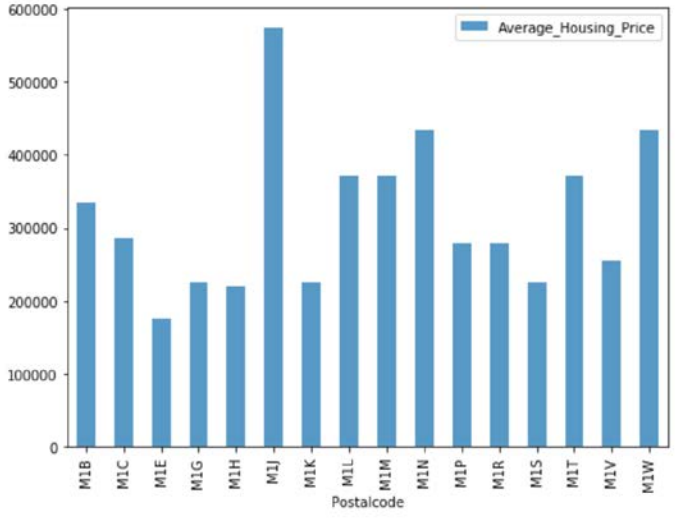


Figure 7 Average Housing Prices in Each Postal Code

Average Housing Prices in Each Neighborhood

As it can be seen from the Figure 8, it is obvious that the houses which are located in "Scarborough Village" have the highest average price among the other postal code in Scarborough city.

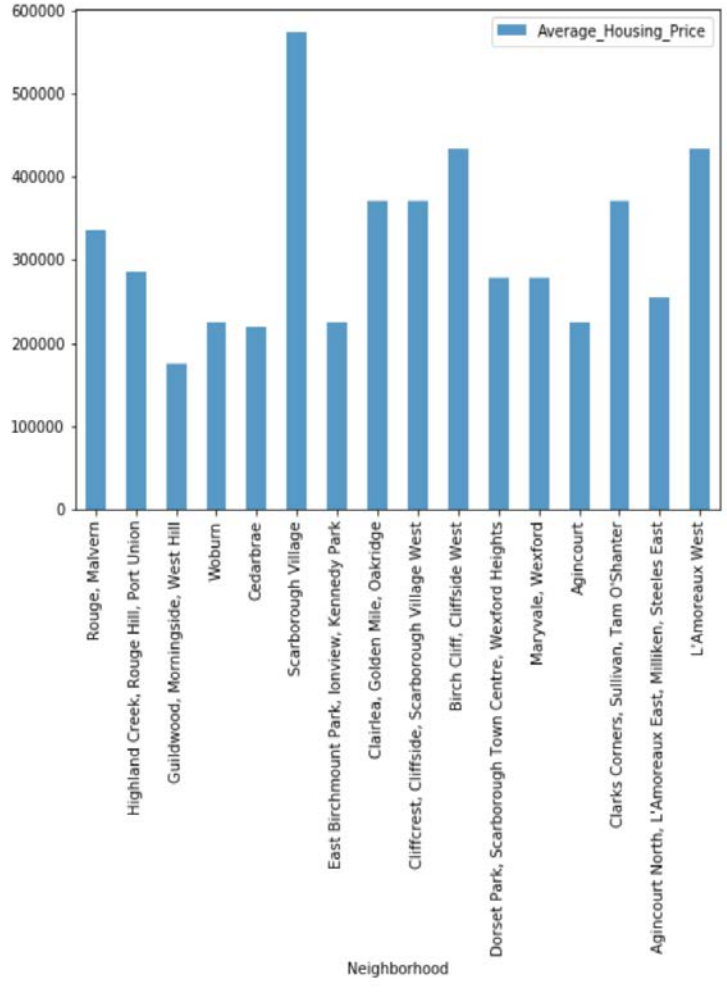


Figure 8 Average Housing Prices in Each Neighborhood

Top school ratings in the neighborhood

The Data regards school ratings in Scarborough is given from this link: http://ontario.compareschoolrankings.org/elementary/SchoolsByRankLocationName.aspx. Based on the Figure 9, top rated schools within each neighborhood can be seen.

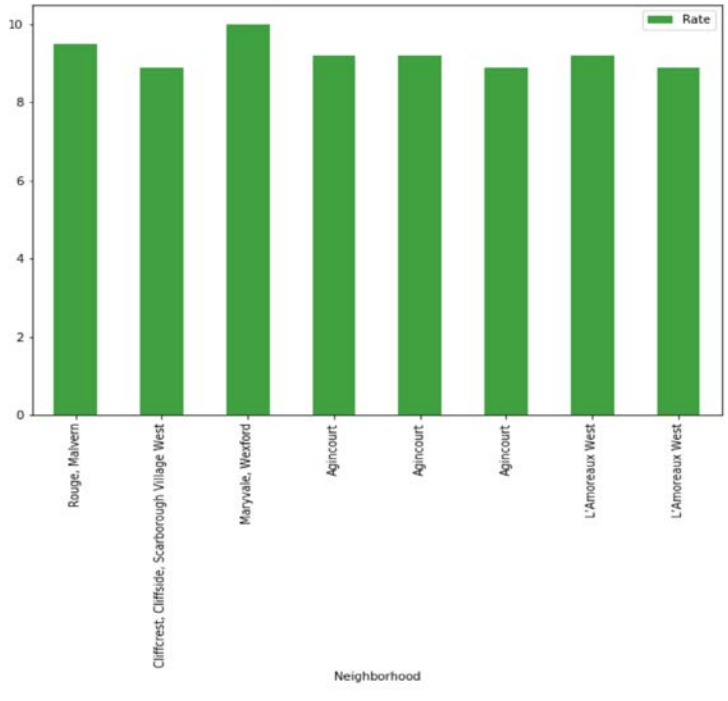


Figure 9 Top school ratings in the neighborhood

Top School Rating Schools

There are very good schools in Scarborough. The highest rank ones are represented in the Figure 10.

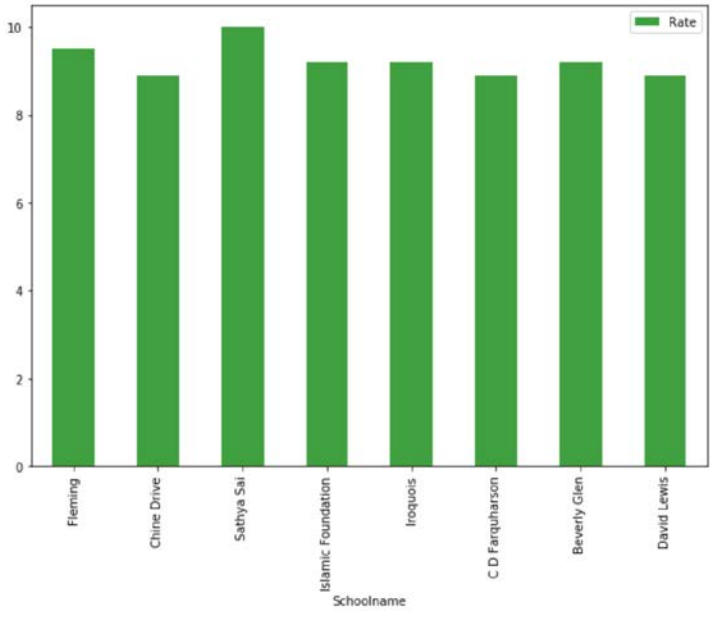


Figure 10 Top School Rating Schools

Top School Ratings Clusters

The Figure 11 shows the top school ranking in each cluster.

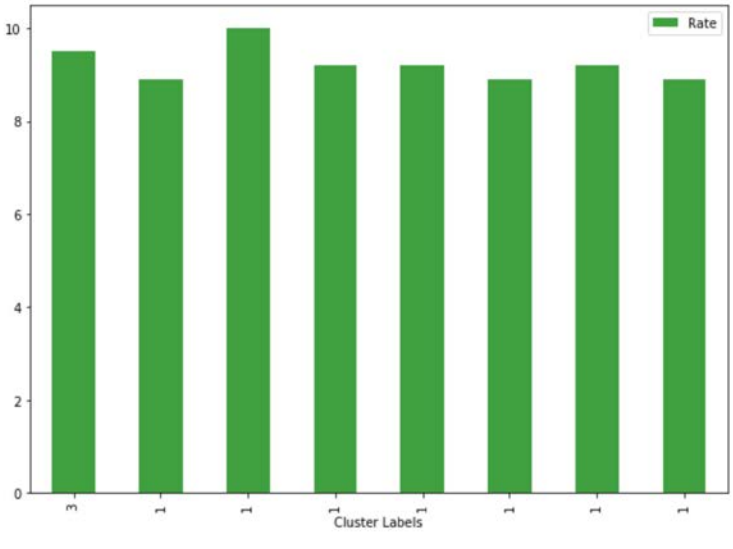


Figure 11 Top School Ratings Clusters

School Rating in Each Cluster

The average high school rates data are represented for the means of this project, which can be seen based on clusters in the scatter plot in the Figure 12.

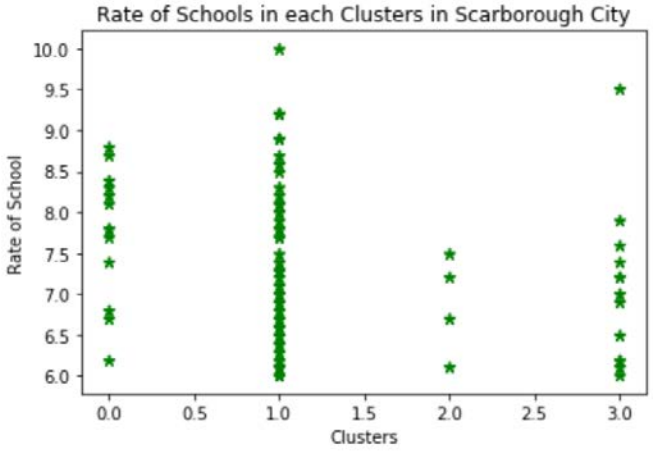


Figure 12 School Rating in Each Cluster

Number of Schools with Rate of Over 6 in Each Postal Code

As it can be seen in the Figure 13, there are a lot of schools with average highs in the Scarborough. Although, some postal codes have more good schools in their vicinities likes “M1B” postal code with the most number of quality schools within.

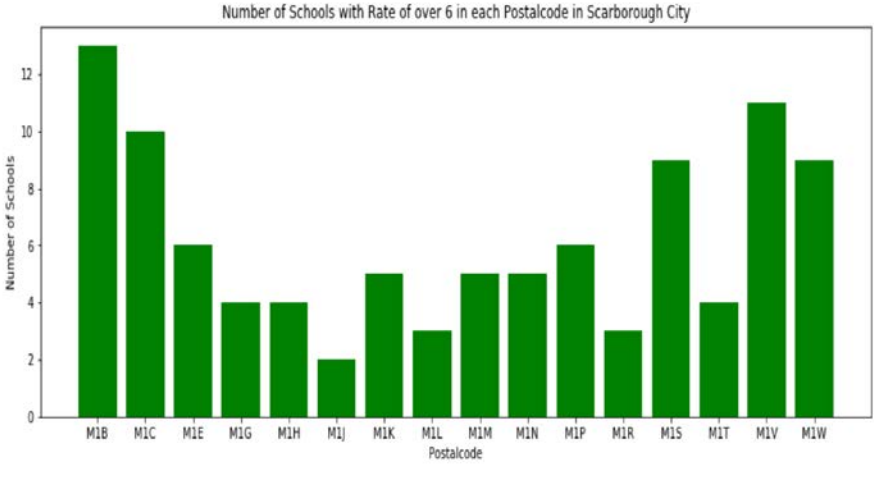


Figure 13 Number of Schools with Rate of Over 6 in Each Postal Code

Number of Schools with Rate of over 6 in Each Clusters

Among the 4 clusters existed, cluster 1 has contained the most number of quality schools, as it it obvious in Figure 14.

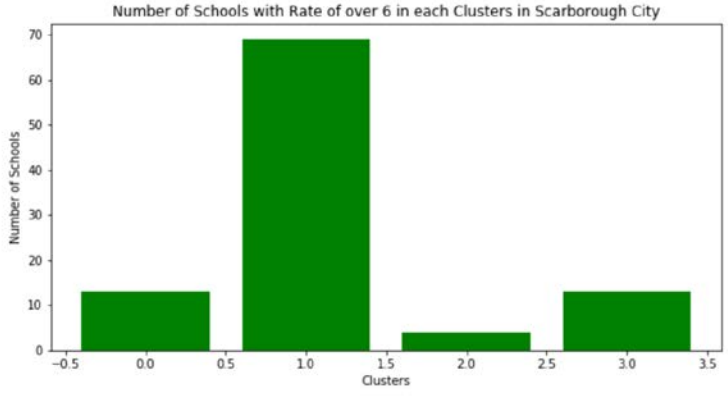


Figure 14 Number of Schools with Rate of Over 6 In Each Clusters

Conclusion

In this project, through a k-means clustering algorithm, we separate the neighborhood into 4 clusters, which the similar venues around them.

By using the visualization techniques like the bar charts as well as the scatter plot we can find the neighborhood with affordable price of houses and high quality school rate.

The average house prices in cluster 1 have the lowest amount among other clusters which makes it a good choice for the families looking for the place to live.

Furthermore, the existence of the majority of high quality schools in this cluster will definitely make sure the families with kids to select this neighborhood as their home in the Scarborough City.