ANALYZING THE AVERAGE HOUSE PRICES AND SCHOOL RATINGS FOR SCARBOROUGH, TORONTO, CA FOR IMMIGRANTS

Applied Data Science Capstone

IBM Data Science Professional Certificate

Location: Scarborough, Toronto, Canada

Part 1: Problem 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.
- The projects aim to create an analysis of features for a neighborhood as a comparative analysis between neighborhoods. The features include median house price and school ratings and 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.

Part 2: Data We Need

Scarborough / Coordinates

43.7764° N, 79.2318° W

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.
- A dataset comprising latitude and longitude, postal codes is 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...

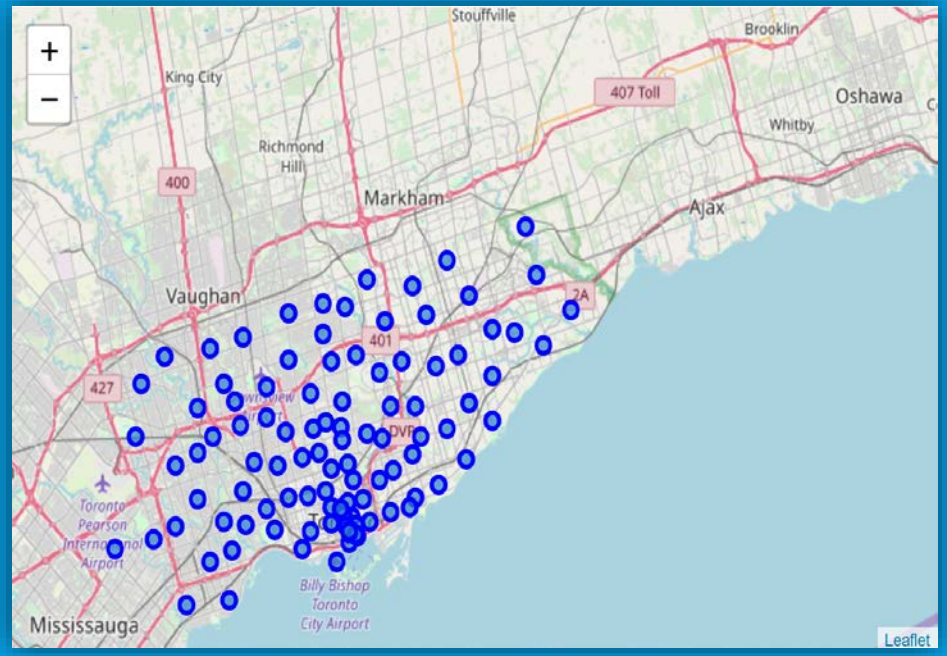
School Rates and Average House Prices Data

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

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

- A dataframe comprising the average house prices is provided within the project.

Map of Scarborough and its Neighborhood



	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale, Wexford	43.750071	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577
15	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389

Foursquare API

After connecting to Foursquare, we can retrieve locational data for each venue in the neighborhood.

- ✓After finding the list of neighborhoods, we connect to the Foursquare API to gather information about venues inside each and every neighborhood. For this project, we have chosen the radius to be 300 meter in each neighborhood.

Data Preprocessing

Processing the retrieved data and creating a dataframe for all the venues inside the Scarborough

- ✓When the data is 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 DataFrame

Processing the retrieved data and creating a dataframe for all the venues inside the Scarborough

	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

Finding The Most Common Venues near Each Neighborhood

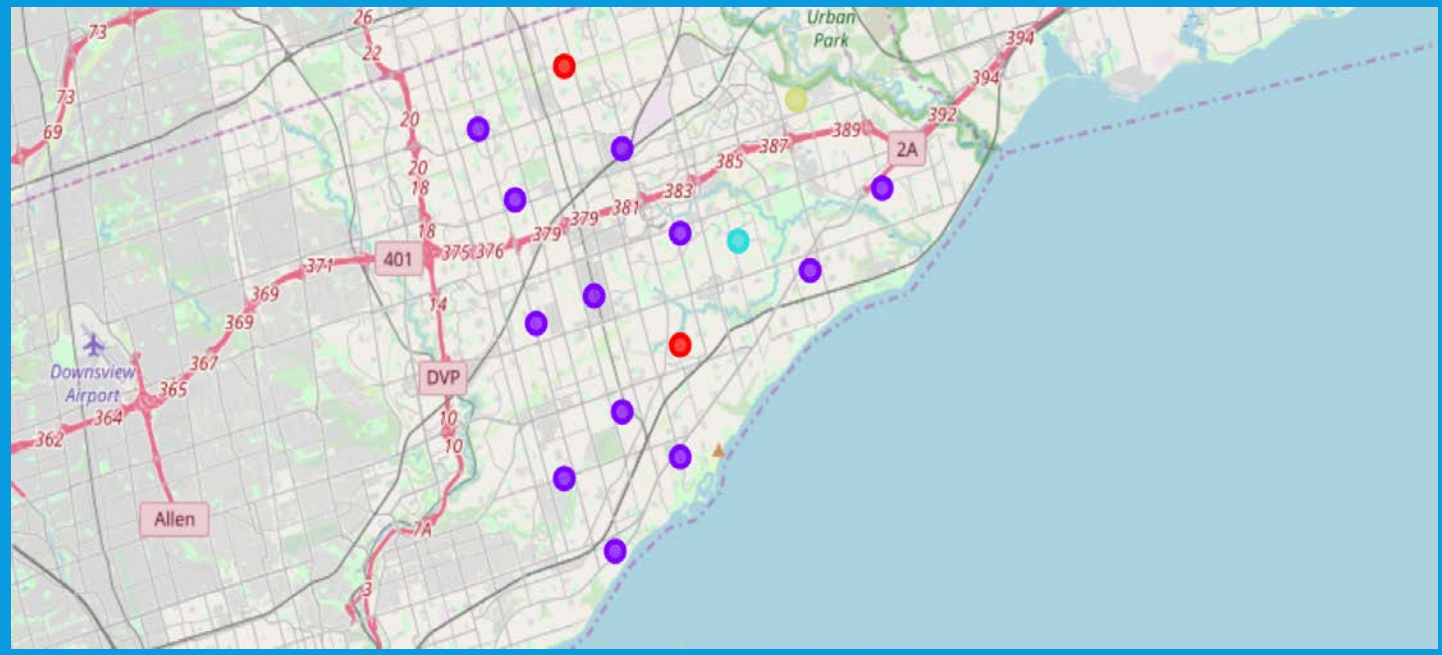
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Skating Rink	Sandwich Place	Breakfast 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

Main Article: Clustering

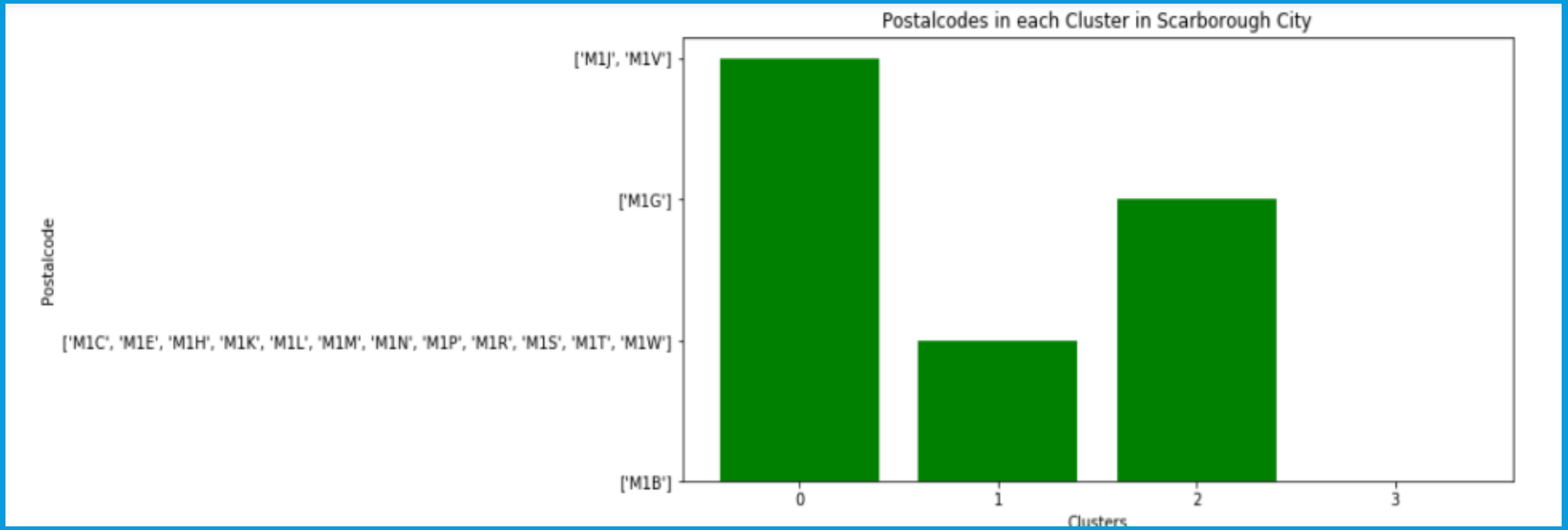
```
# set number of clusters
n_clusters = 4
# Using k-means to cluster the neighborhood into 4 clusters.

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_
```

	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

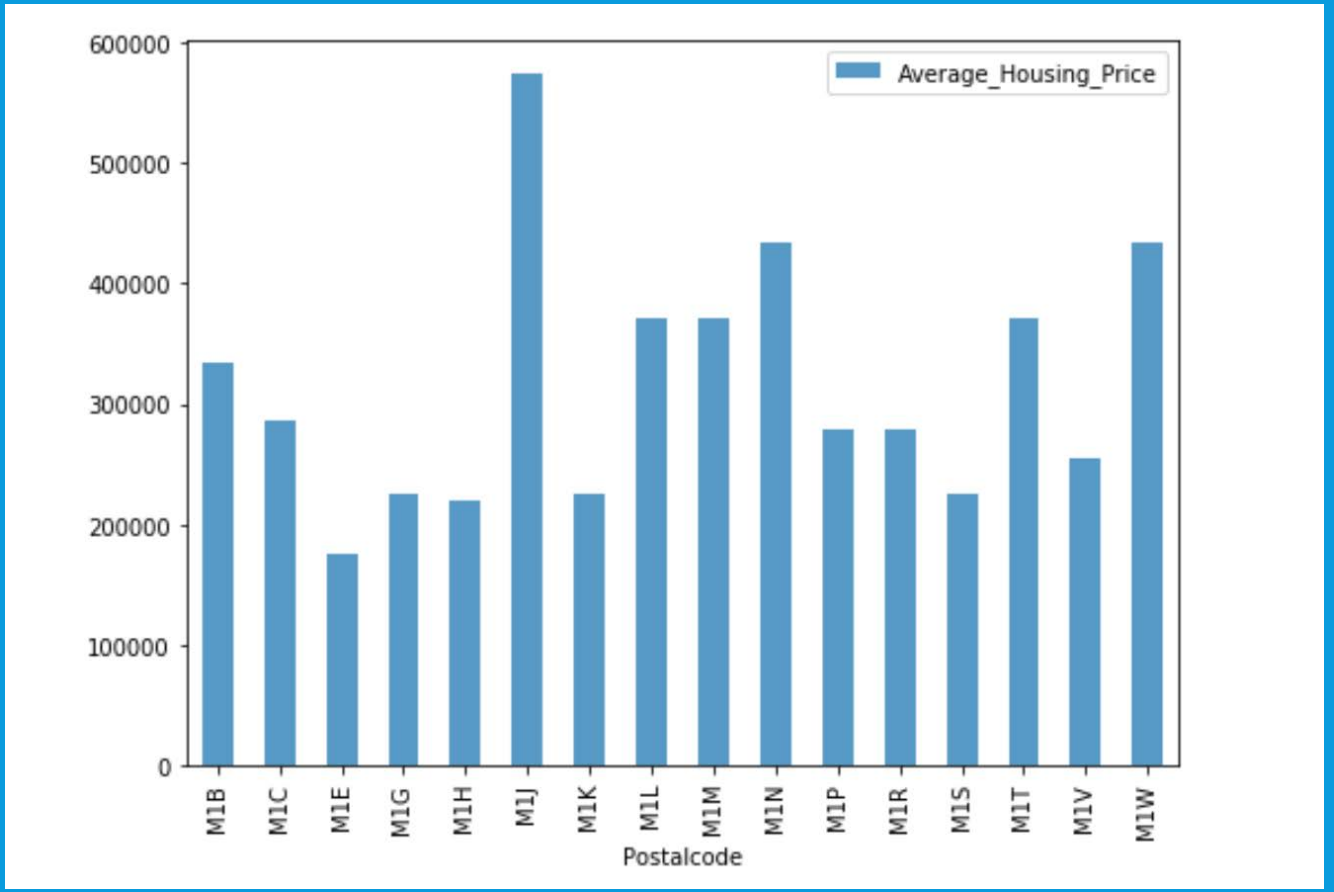
Visualize the Resulting Clusters



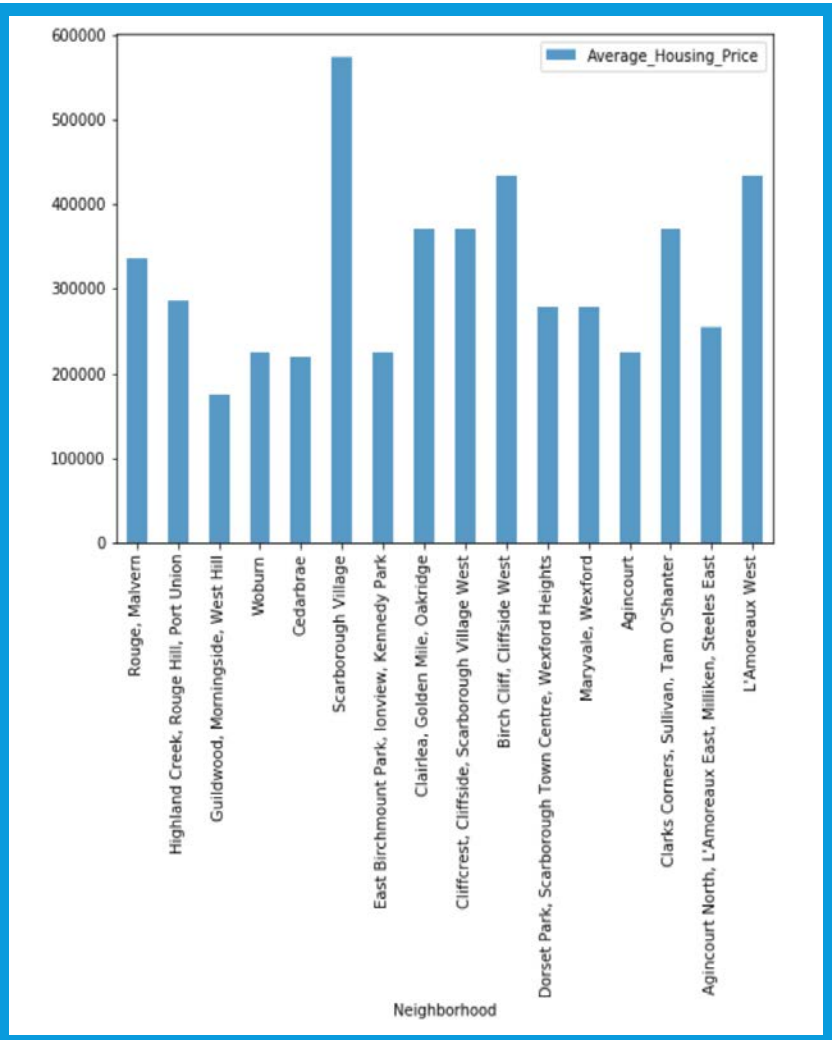
Postal Codes In Each Cluster In Scarborough



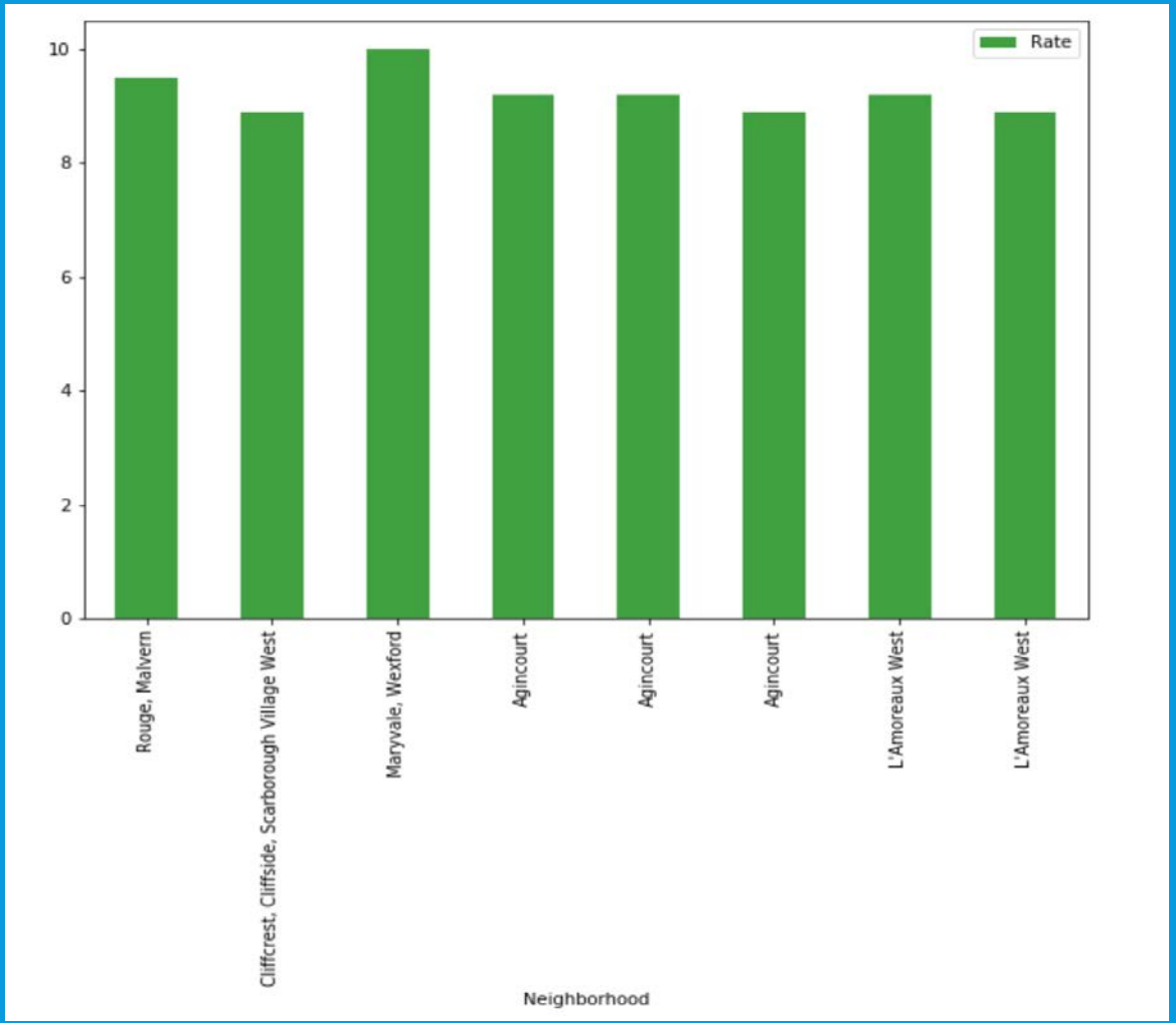
Average Housing Prices in Each Postal Code



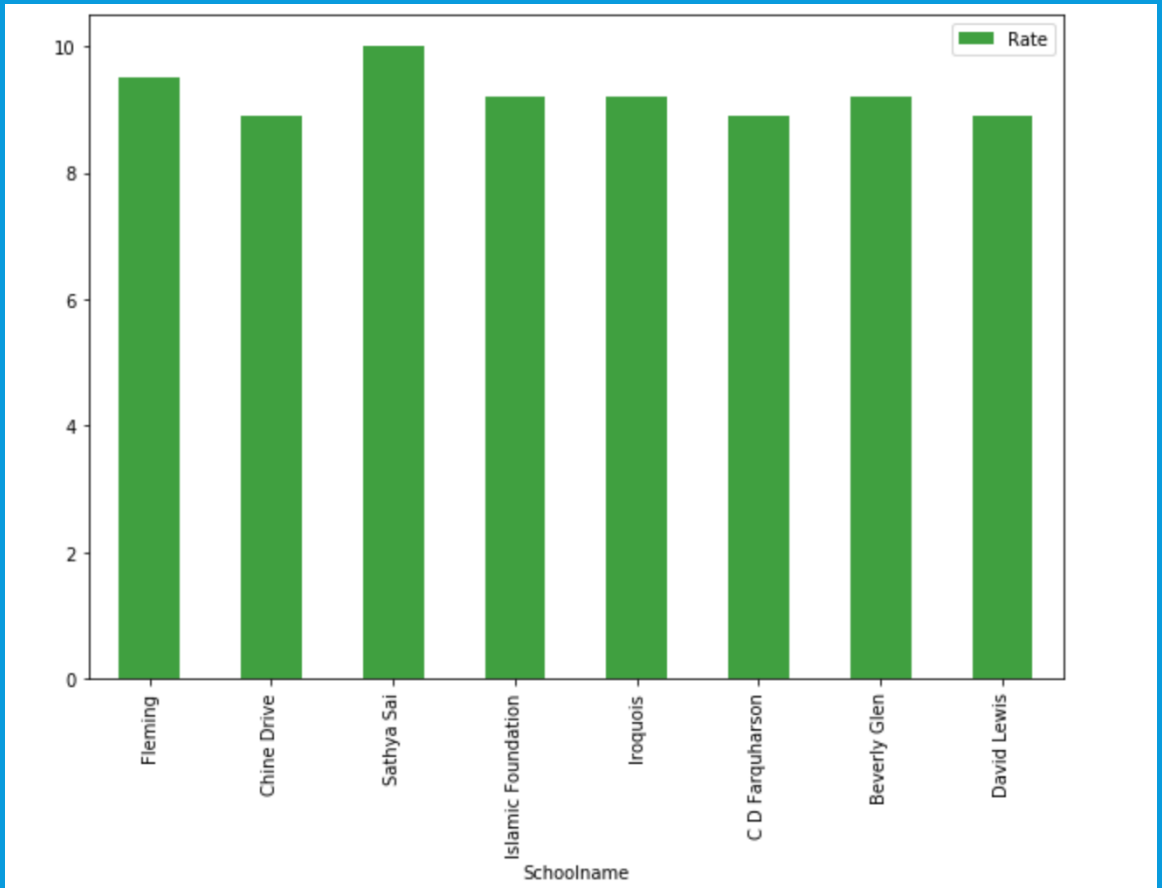
Average Housing Prices in Each Neighborhood



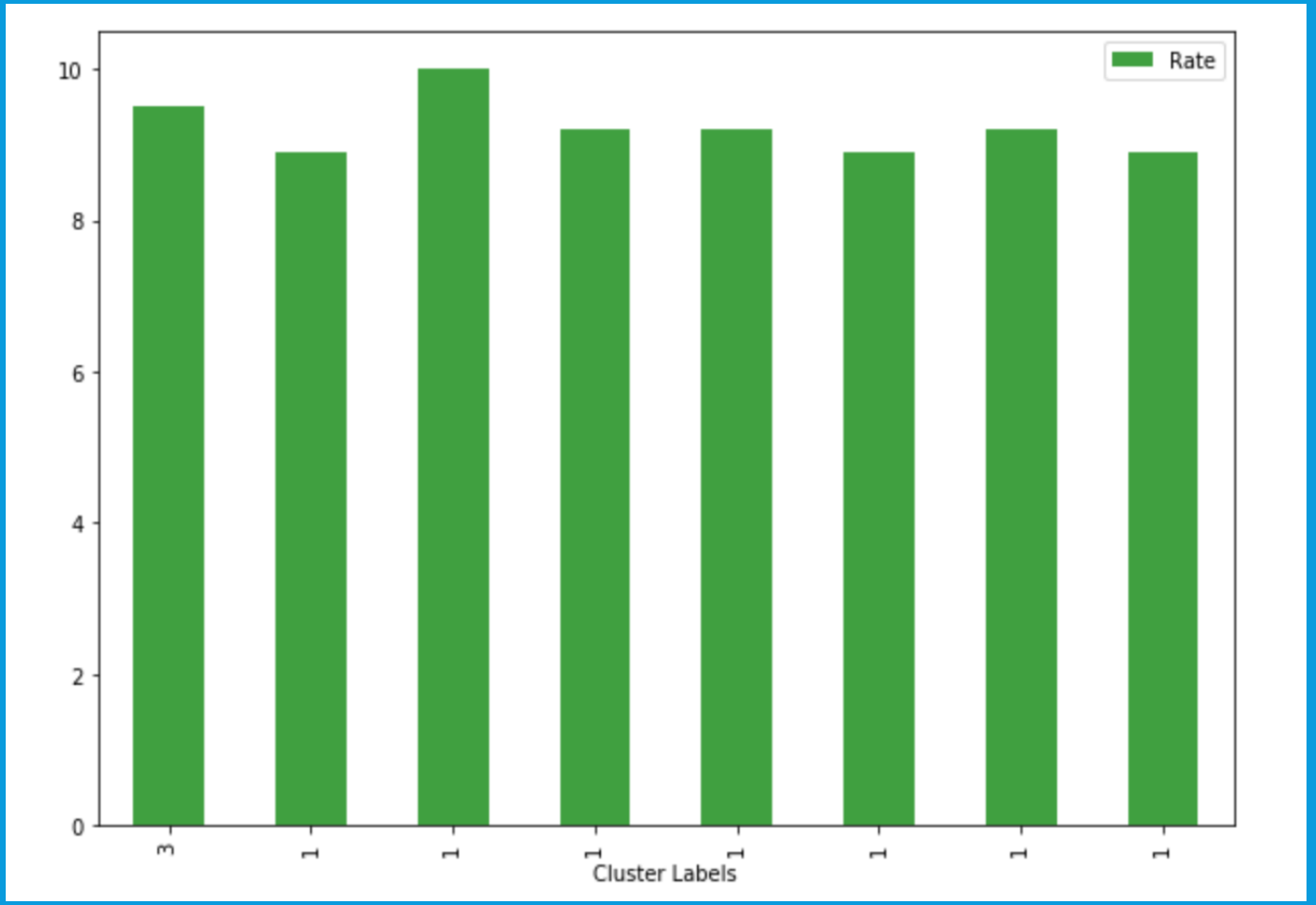
Top school ratings in the neighborhood



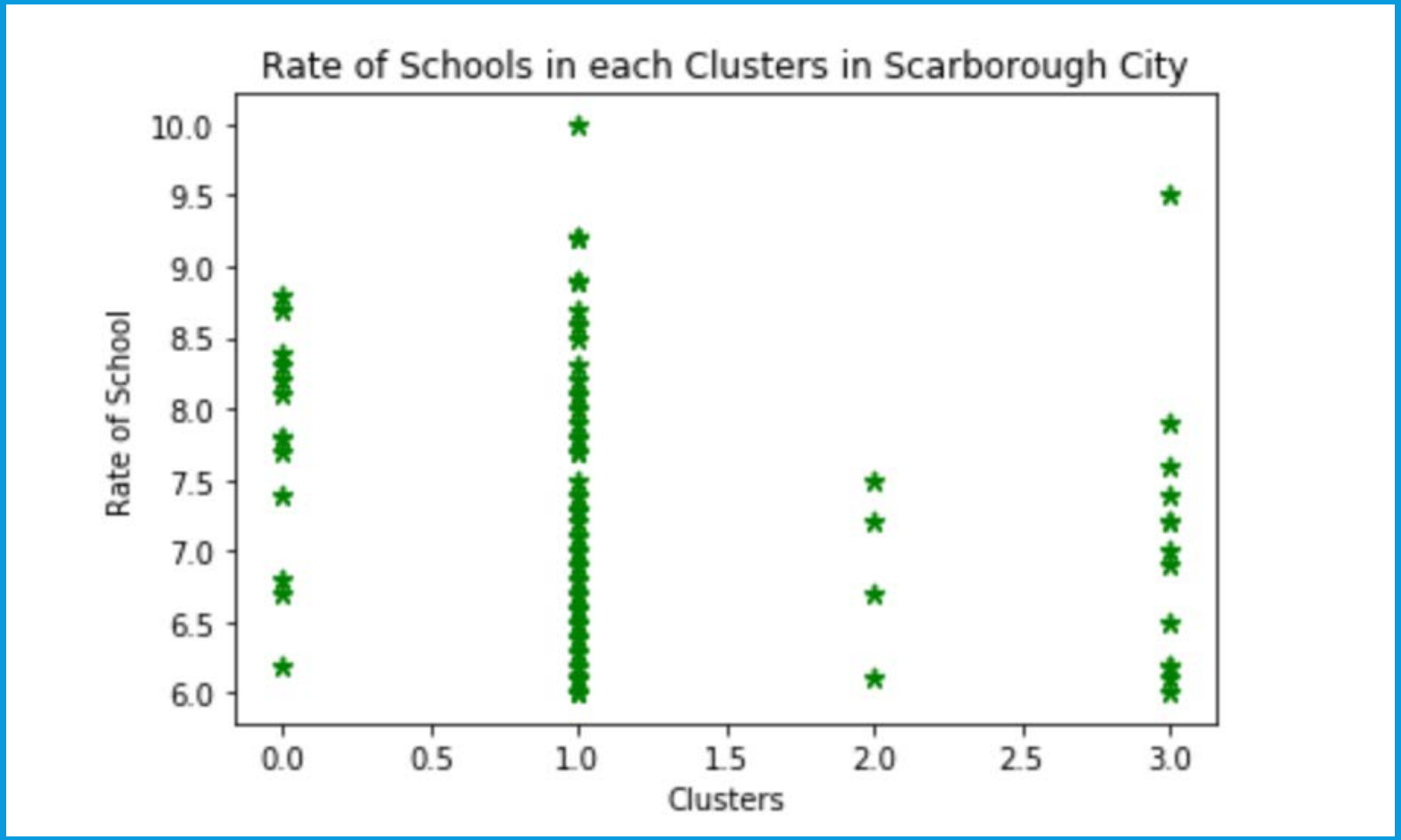
Top School Rating Schools



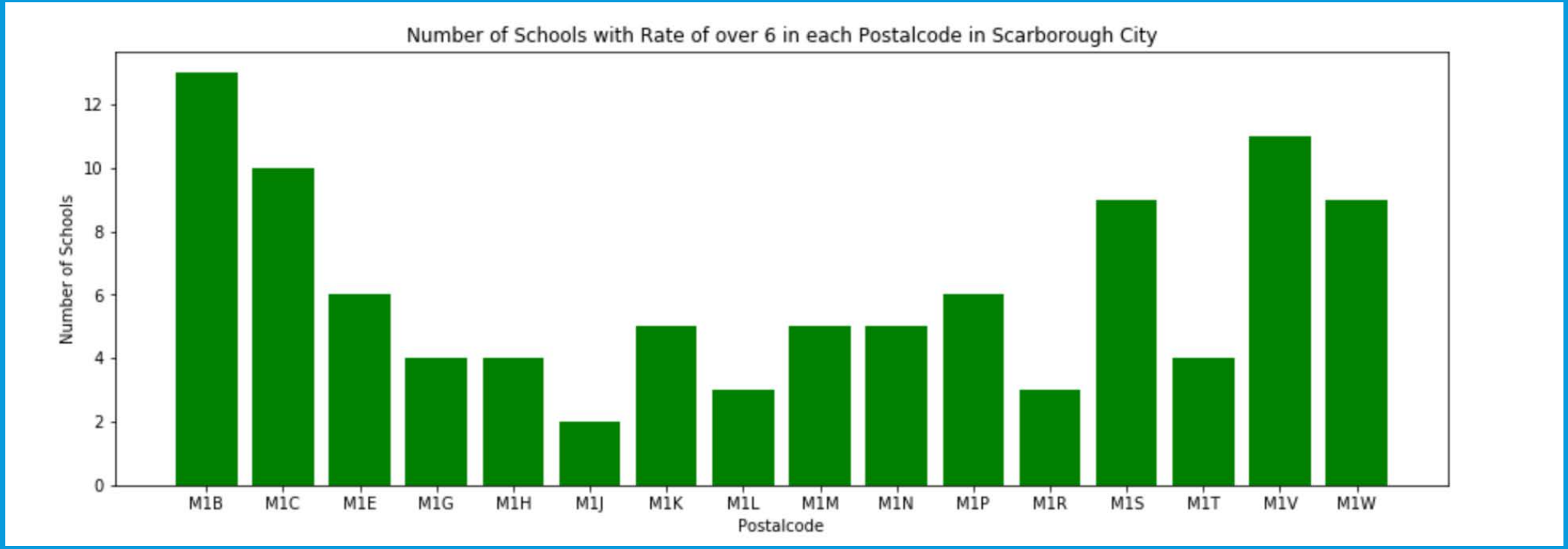
Top School Ratings Clusters



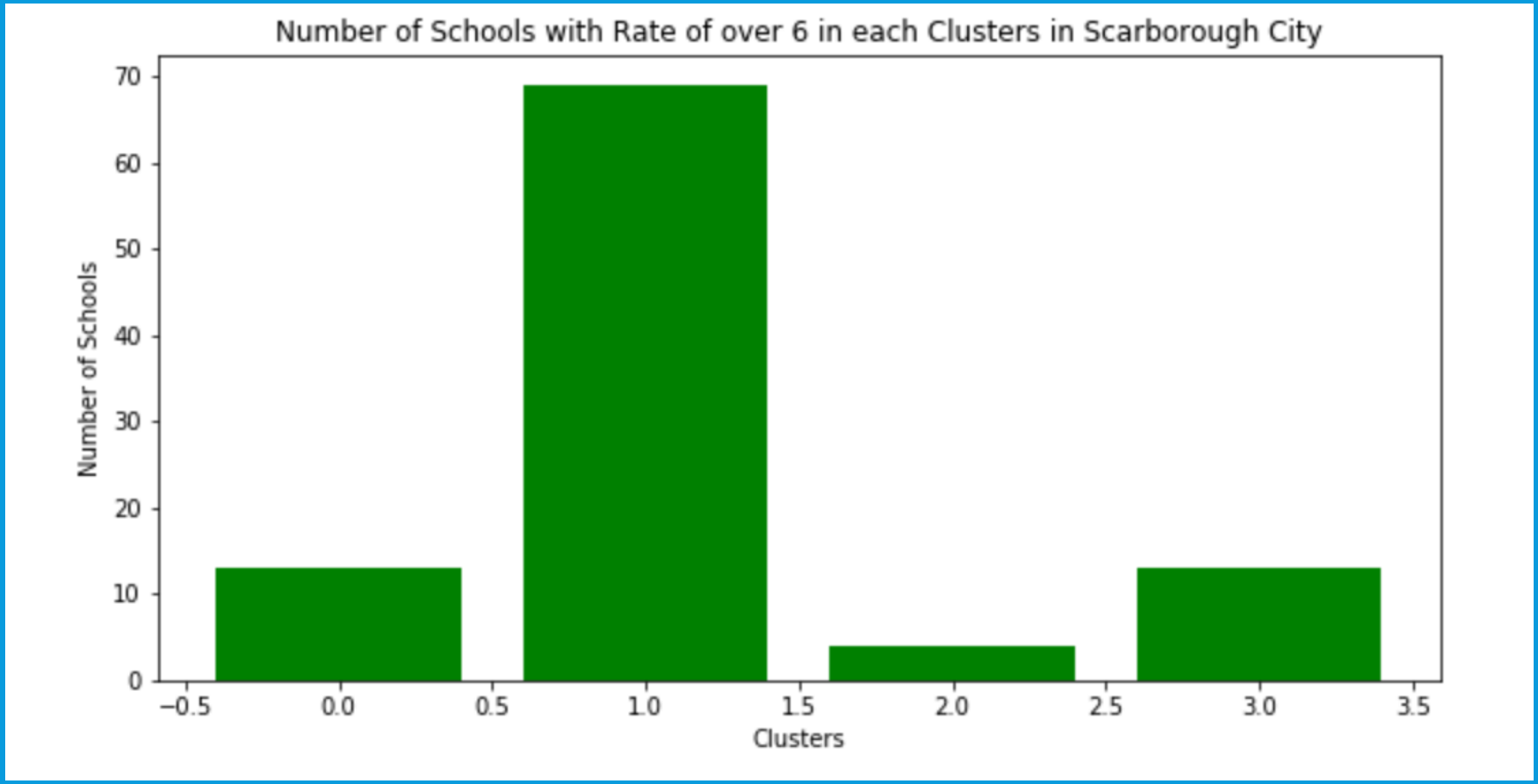
School Rating in Each Cluster



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



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.

Thank you