

# Analyzing, Segmenting & Clustering of Toronto Boroughs:

Feasible Recommendations for  
Peoples & City Managers

March 2020

Business Proposal by,  
Fabio Melle

---

# Scenario and Place : Scarborough Toronto, Ontario (Canada)

Scarborough (Ontario), is a municipality in Canada and a borough of Toronto. Located above the Scarborough Bluffs, occupying the eastern part of the city. Scarborough is bordered by Victoria Park Avenue to the west, Steeles Avenue to the north, Rouge River and Pickering City to the east and Lake Ontario to the south. The city is named after Scarborough (United Kingdom). The city is one of the most multicultural of the Greater Toronto Area, hosting various religions and cultures. The city has also been declared the greenest of the Greater Toronto Area [1].



Figura 1 - Toronto and its Boroughs [2]

# Introduction / Business Problem

Some Italian people have realized that new job opportunities have opened up in Scarborough, having relatives living there. So they decided to move to this city. Before leaving, they would like to know more about this city and its neighborhoods so that they can evaluate where to live. Before leaving, they would like to have more information about this borough and its neighborhoods in order to evaluate where to go to live.





# Goal

The aim of this work is to give these families the opportunity to settle in a good neighbourhood by taking into account some evaluation parameters such as the services that are inside, the quality of education guaranteed by schools and the average cost of housing.



# Target Audience

In addition to the actors mentioned in this project, this work may also refer to the following figures:

- Professionals or Individuals relocating to Scarborough for work
- Mayors of cities
- City Managers
- International Students & Overseas Nationals
- Small Businesses
- Policy Makers
- Law Enforcement
- Families & Communities
- Housing & Real Estate Sector
- Direct Investors

# Data & Data Sources

This project is focused on the city of Scarborough, a borough of Toronto. In order to implement this work, the geo-location of this district and its neighborhoods will be necessary. This project will also require more information about the different districts of Scarborough, average house prices and school evaluations. Below are the required data for each neighborhood:

1. Location of the neighborhood in terms of latitude and longitude
2. Average prices of the apartments
3. School Ratings

The data set containing position data and postcodes is present in the previous project (Segmenting and Clustering Neighborhoods2.ipynb). The location of Scarborough and its neighborhoods will be obtained by filtering the available information:

<https://github.com/FMelle-DataScientist/Applied-Data-Science-Capstone/blob/master/Segmenting%20and%20Clustering%20Neighborhoods2.ipynb>

# Data & Data Sources

I used GitHub repository as a database. My master data which has the main components *Borough*, *Average House Price*, *School Ratings* and *Latitude* and *Longitude* informations of city. I used python folium library to visualize geographic details of Scarborough and its neighborhood and I created a map of Scarborough with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:

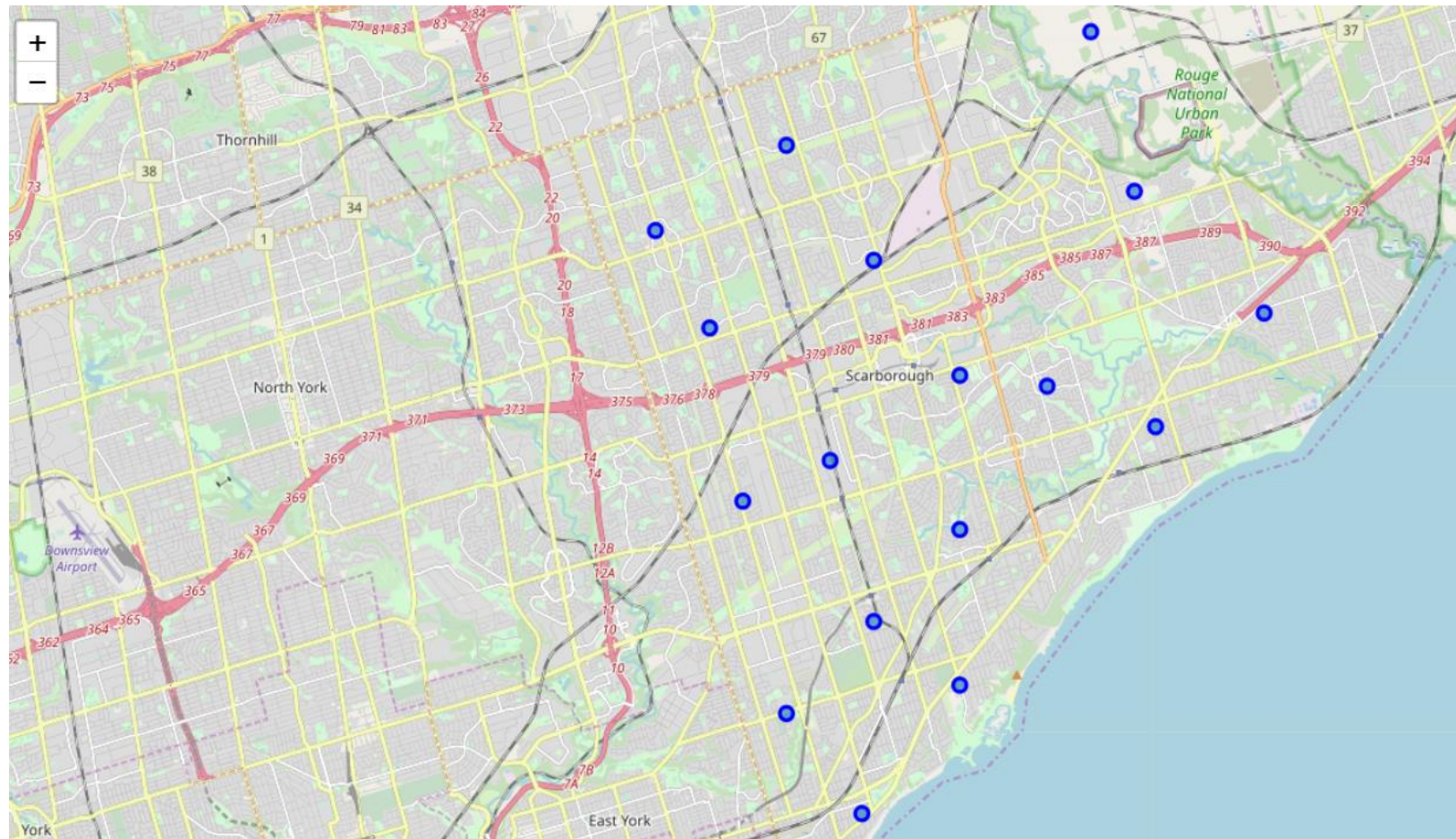
	Postal Code	Latitude	Longitude	Borough	Neighborhood
0	M1B	43.806686	-79.194353	Scarborough	Rouge, Malvern
1	M1C	43.784535	-79.160497	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	43.763573	-79.188711	Scarborough	Guildwood, Morningside, West Hill
3	M1G	43.770992	-79.216917	Scarborough	Woburn
4	M1H	43.773136	-79.239476	Scarborough	Cedarbrae
5	M1J	43.744734	-79.239476	Scarborough	Scarborough Village
6	M1K	43.727929	-79.262029	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	43.711112	-79.284577	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	43.716316	-79.239476	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	43.692657	-79.264848	Scarborough	Birch Cliff, Cliffside West
10	M1P	43.757410	-79.273304	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...
11	M1R	43.750071	-79.295849	Scarborough	Maryvale, Wexford
12	M1S	43.794200	-79.262029	Scarborough	Agincourt
13	M1T	43.781638	-79.304302	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter
14	M1V	43.815252	-79.284577	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...
15	M1W	43.799525	-79.318389	Scarborough	L'Amoreaux West
16	M1X	43.836125	-79.205636	Scarborough	Upper Rouge

Figura 2 - Info geo-location of neighborhoods of Scarborough



# Methodology

I used GitHub repository as a database. My master data which has the main components *Borough*, *Average House Price*, *School Ratings* and *Latitude* and *Longitude* informations of city. I used Python Folium library to visualize geographic details of Scarborough and its 17 neighborhoods and I created a map of Scarborough with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below:





## Foursquare API:

As the main source of data collection I used the Foursquare, I will be able to retrieve Location data for each venues in every neighborhood.

After finding the list of neighbourhoods, I will use the Foursquare API to collect information on the venues inside each and every neighborhood. For each neighbourhood, I will set a radius to 500 meter.

# Preprocessing

Processing the Retrieved Data and Development a Dataframe for Venues inside Scarborough.

After the complete data collection, I will carry out the processing on these raw data to extract the desired features for each venue. The main feature is category of that venues.

After this step, the "Venue's Category" column will be One-hot encoded and there will be several columns of features for the different venues.

# CREATING DATAFRAME

Below is the dataframe created for Venues inside Scarborough.

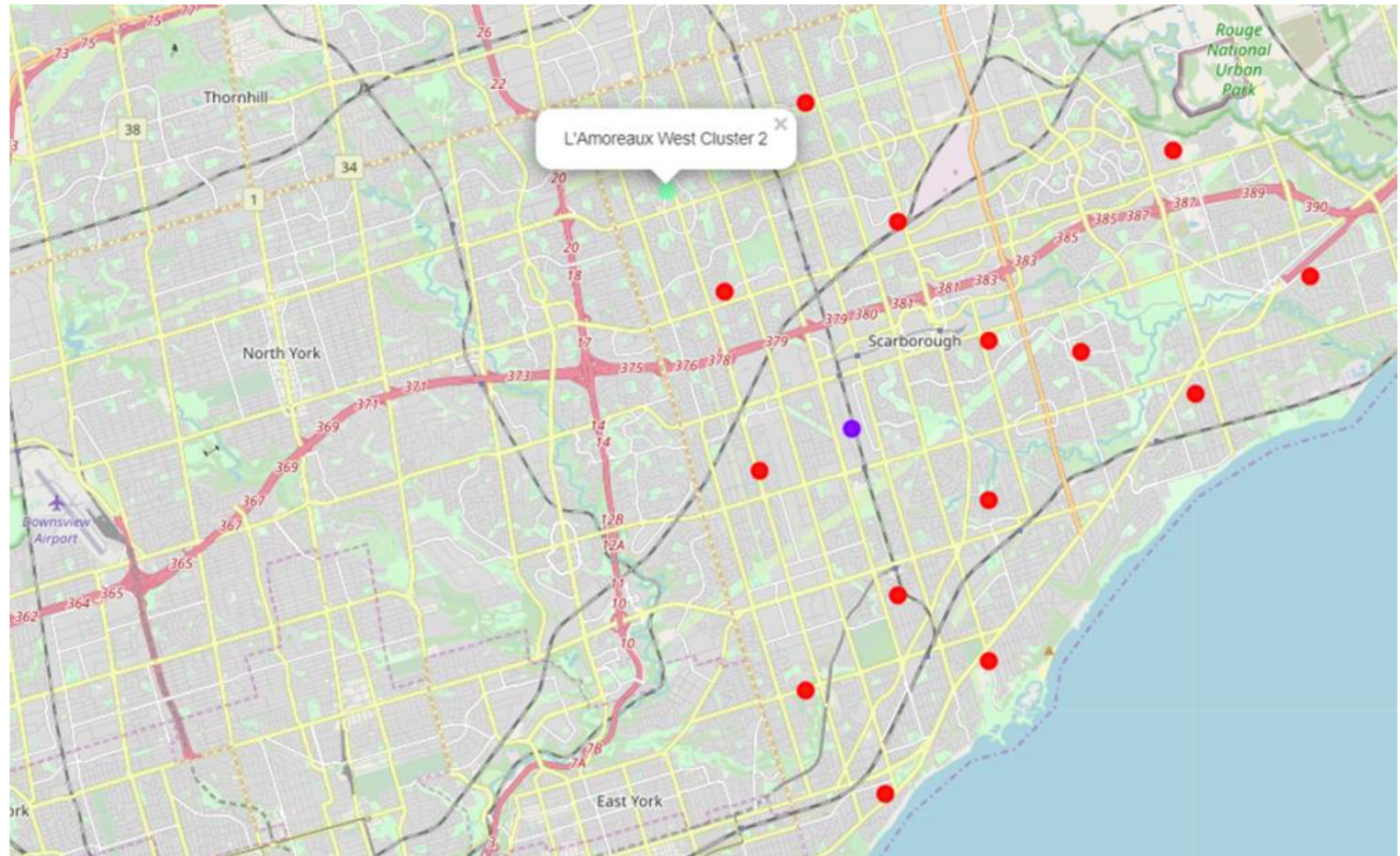
	Neighborhood	Athletics & Sports	Badminton Court	Bakery	Bank	Bar	Beer Store	Breakfast Spot	Burger Joint	Bus Line	Bus Station	Business Service	Café	Caribb Resta
0	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0	0



# Machine Learning Algorithm

In this study I will apply a Clustering algorithm - k-means

Based on the common venue categories & characteristics, neighborhoods are grouped into 3 major Clusters using the Folium Map as shown. For each neighborhood is released:  
Name of the neighborhood,  
Cluster label,



# Segmenting & Classification of Clusters:

From the different observations, each cluster is named according to its characteristics, respectively.

Cluster 0 -> Characteristics: Breakfast Spot, Bar, etc.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0	Breakfast Spot	Bar	Burger Joint	Wings Joint

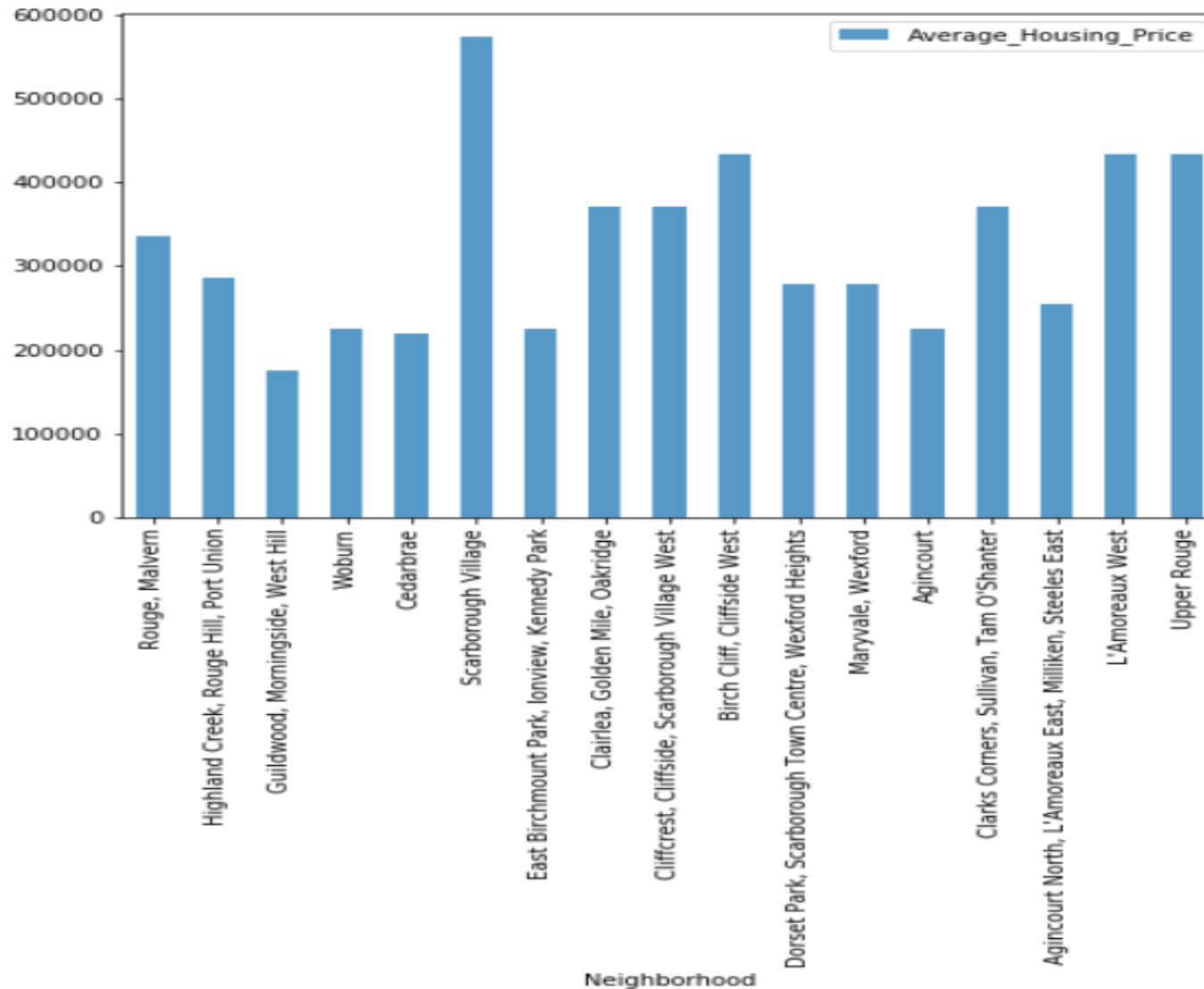
Cluster 1 -> Characteristics: Indian Restaurant , Electronics Store, etc.

10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	1	Indian Restaurant	Electronics Store	Wings Joint	Chinese Restaurant
----	-----	-------------	---	-----------	------------	---	-------------------	-------------------	-------------	--------------------

Cluster 2 -> Characteristics: Grocery Store, Chinese Restaurant, etc.

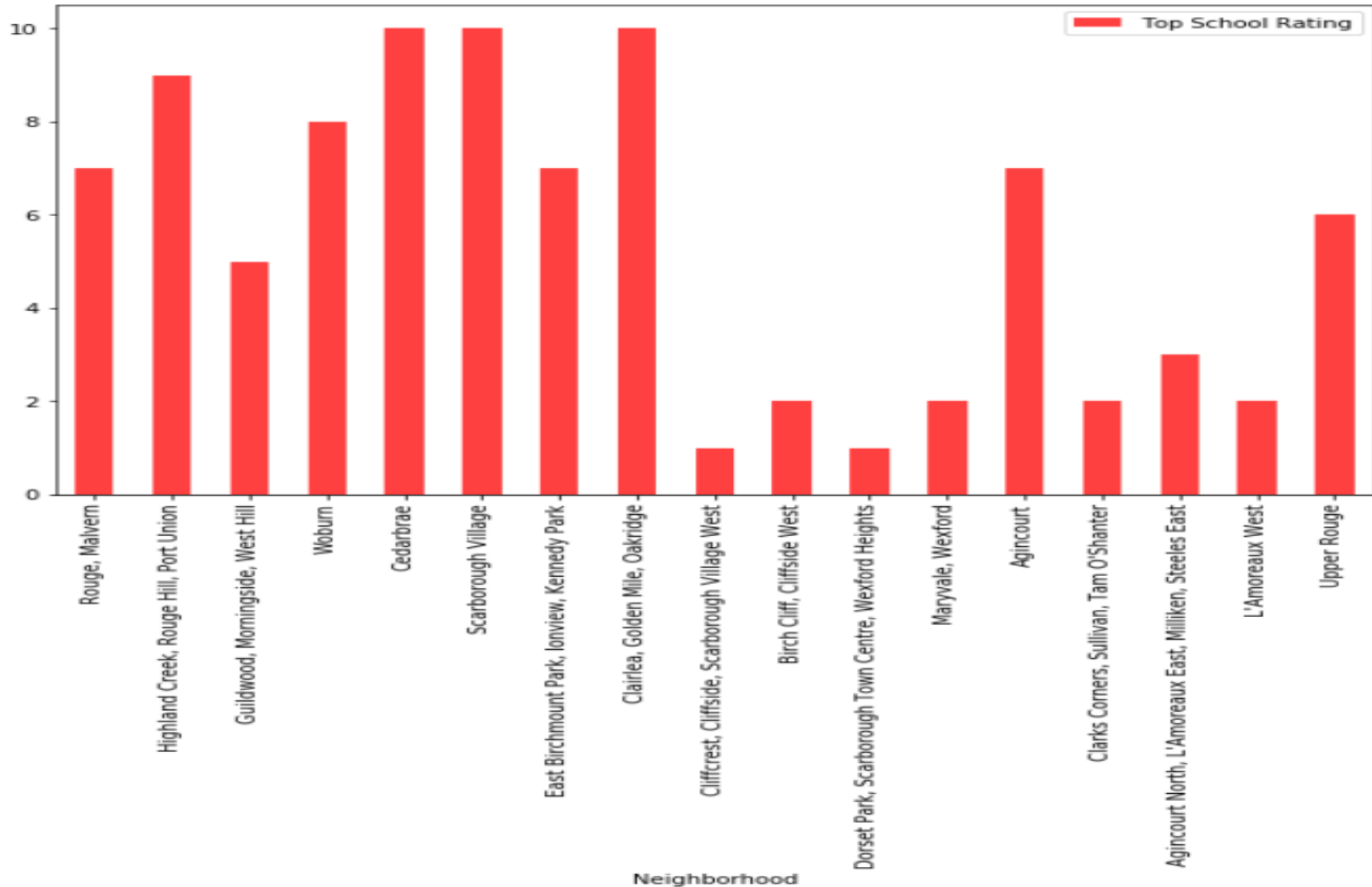
15	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389	2	Grocery Store	Chinese Restaurant	Fast Food Restaurant	Nail Salon
----	-----	-------------	-----------------	-----------	------------	---	---------------	--------------------	----------------------	------------

# Neighborhood Average Sales Housing





# School rating by clusters



## Discussion:

As I said before, Scarborough is a big city that has been attracting a lot of people in recent years as a new residential destination, with its 17 districts offering different services and venues. Because there is such complexity, you can try very different approaches in clustering studies.

I used the k-means algorithm as part of this clustering study. I divide the borough into 03 clusters, which have similar neighbourhoods around them.

I also performed the data analysis through this information by adding the district coordinates and average house prices and school ratings as statistic data on GitHub.

I concluded the study by displaying the data and clustering information on the map of Scarborough. In the future, web or telephone applications can be developed for direct investors



## Conclusion:

So, a lot of people are heading to big cities to change their lives, find a job or start a business. For this reason, people can get better results through access to platforms where this kind of information is provided.

Not only investors, but also mayors of cities can manage the city more effectively and efficiently by exploiting this kind of analysis, visualization and knowledge extraction.





THANK YOU

# Sitography and Bibliography

- [1] Scarborough - Wikipadia: [https://en.wikipedia.org/wiki/Scarborough,\\_Toronto](https://en.wikipedia.org/wiki/Scarborough,_Toronto)  
[2] Toronto Map: <http://www.johomaps.com/na/canada/ontario/toronto/to1.html>