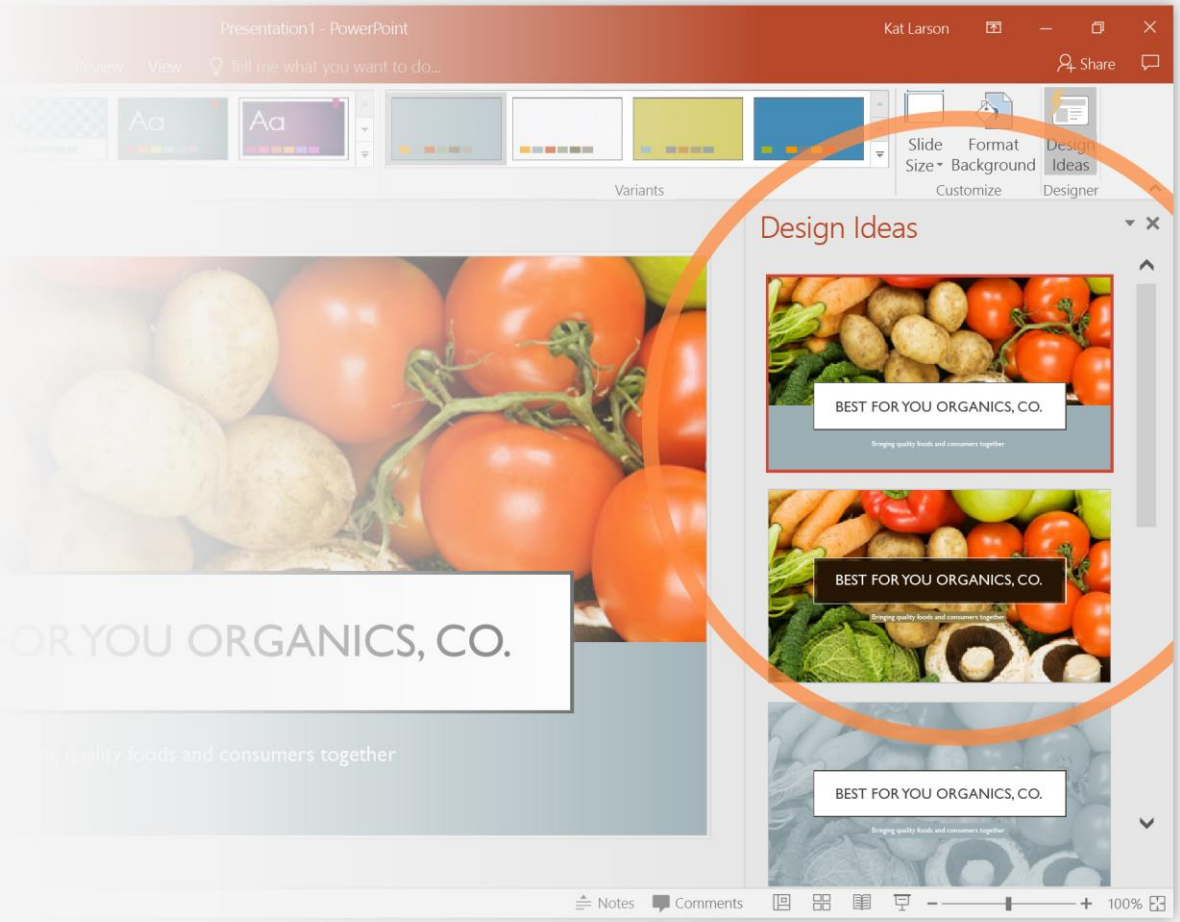
The Battle of Neighbourhoods

Swapnil Gaikwad

How to determine a better neighbourhood?

Ease of accessing-

- Cafe,
- School,
- Supermarket,
- medical shops,
- grocery shops,
- mall,
- theatre,
- hospital,
- likeminded people, etc.



Data acquisition and cleaning

Data source: For the Toronto neighbourhood data, a Wikipedia page exists following information we need to explore and cluster the neighbourhoods in Toronto.

Columns Names	Meaning	Absent Values
Postal Code	It is compulsory and should be present.	NA
Borough	It is present or absent.	Not assigned
Neighbourhood	It is present or absent.	Not assigned

Postalcode	Borough	Neighborhood
M1B	Scarborough	Malvern, Rouge
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
M1E	Scarborough	Guildwood, Morningside, West Hill
M1G	Scarborough	Woburn
M1H	Scarborough	Cedarbrae

Foursquare contained information of venues

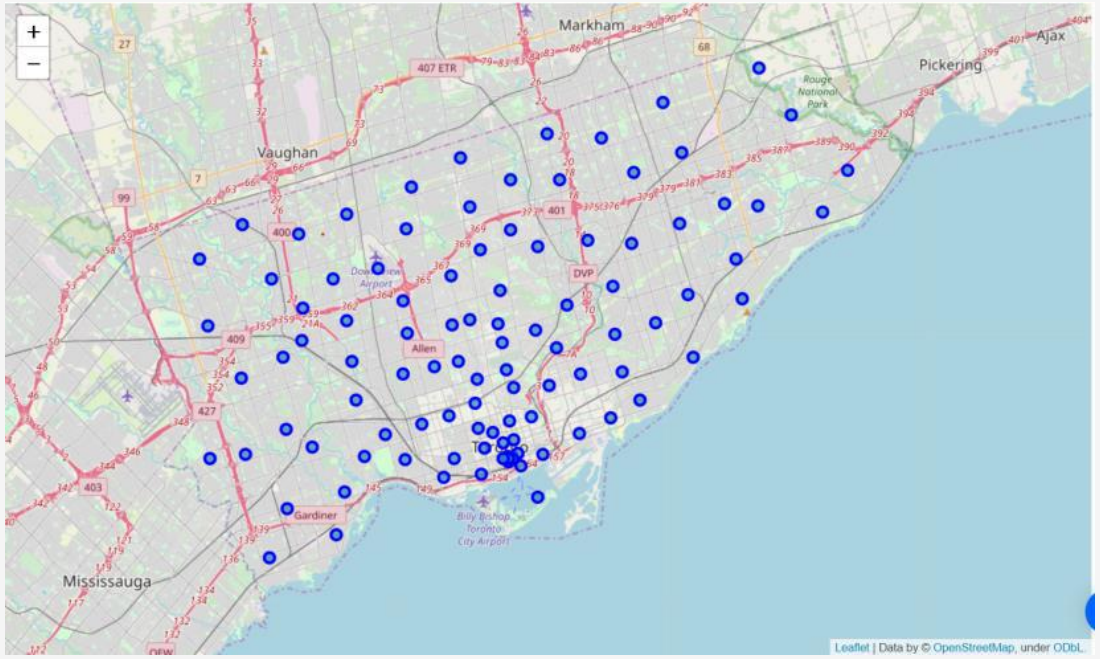
Data extraction using Foursquare API Data: After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

The information obtained per venue as follows:
1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Foursquare contained information of venues

Postalcode	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875
M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470
M1G	Scarborough	Woburn	43.76812	-79.21761
M1H	Scarborough	Cedarbrae	43.76944	-79.23892
M1J	Scarborough	Scarborough Village	43.74446	-79.23117
M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.72582	-79.26461
M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.71289	-79.28506
M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.72360	-79.23496
M1N	Scarborough	Birch Cliff, Cliffside West	43.69510	-79.26466

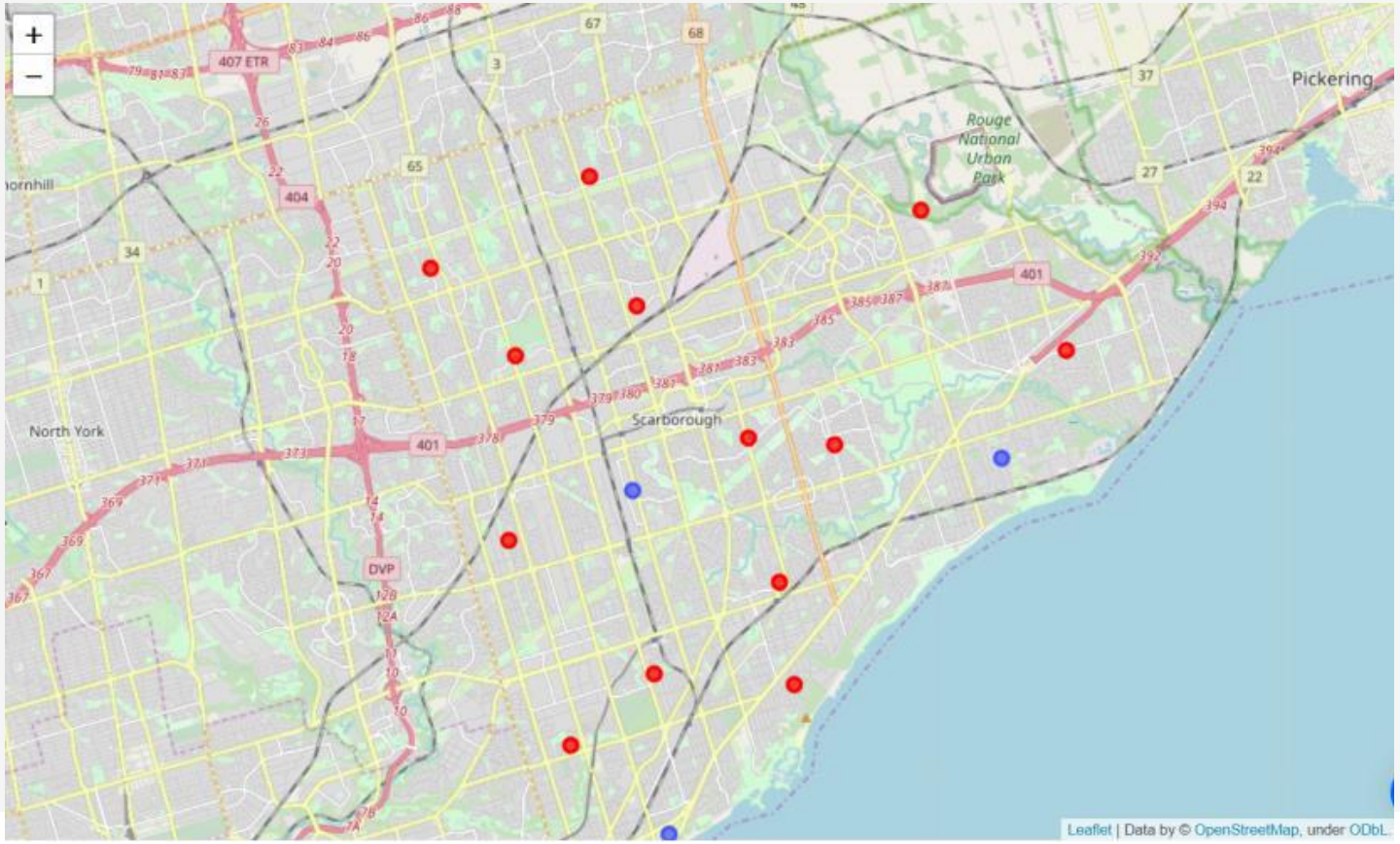


Clustering Approach

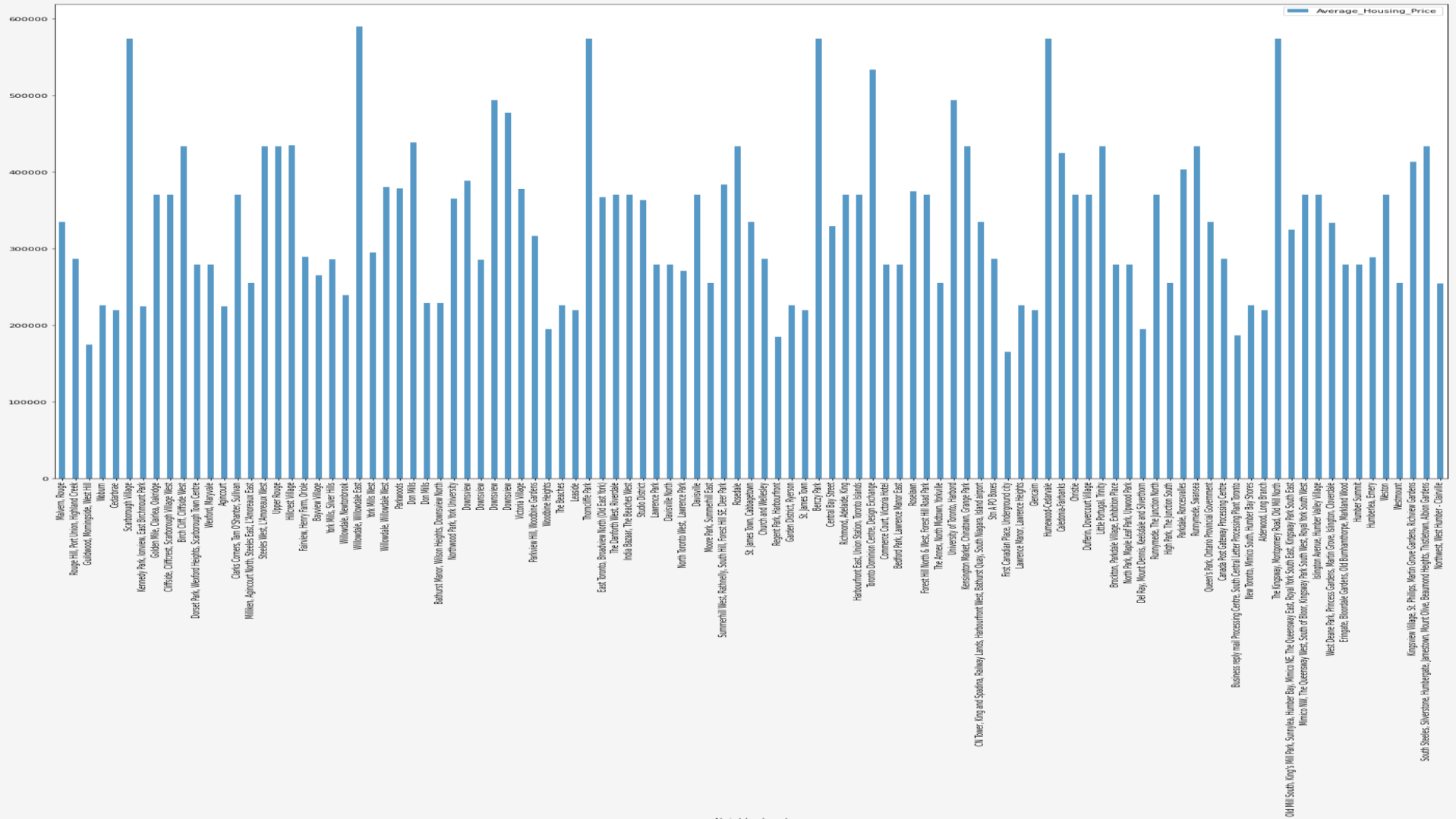
To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods in a big city like New York and Toronto.

To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

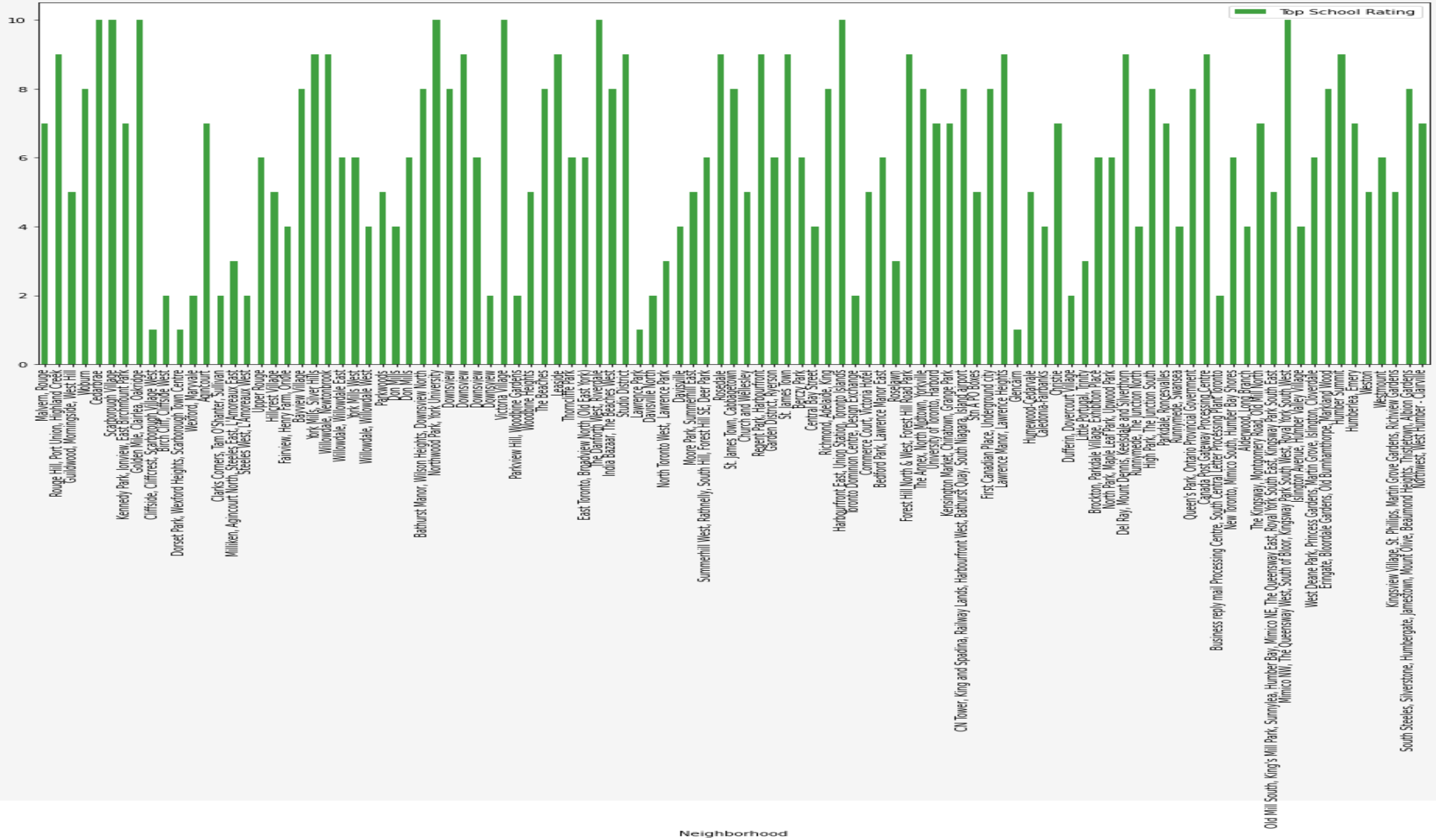
name	categories	lat	lng
SEPHORA	Cosmetics Shop	43.775017	-79.258109
Disney Store	Toy / Game Store	43.775537	-79.256833
American Eagle Store	Clothing Store	43.776012	-79.258334
St. Andrews Fish & Chips	Fish & Chips Shop	43.771865	-79.252645
Tommy Hilfiger	Clothing Store	43.776015	-79.257369



Average Housing Price by Clusters



School Ratings by Clusters



Predictive Modelling

The major purpose of this project is to suggest a better neighbourhood in a new city for the person who are shifting there.

Social presence in society in terms of likeminded people.

Connectivity to the airport, bus stand, city centre, markets, and other daily needs things nearby.

1. **Sorted list of houses in terms of housing prices in an ascending or descending order**

2. **Sorted list of schools in terms of location, fees, rating, and reviews**

Conclusion

In this project, using k-means cluster algorithm I separated the neighbourhood into 10 (Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighbourhoods around them.

Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation.

This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Future directions

This project can be continued for making it more precise in terms to find best house in Scarborough.

Best means based on all required things (daily needs or things we need to live a better life) around and in terms of cost effective.