IBM DATASCIENCE

CAPSTONE PROJECT PRESENTATION

BY SOURABH CHITRANSHI

INTRODUCTION

THE PURPOSE OF THE PROJECT

BUSINESS PROBLEM

- This Capstone Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.

DATA USED IN THE PROJECT

DEFINING THE DATA

- Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.
- Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

DEFINING THE DATA

- Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.
- Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The Map Scarborough

CLUSTERING APPROACH

METHODOLOGY

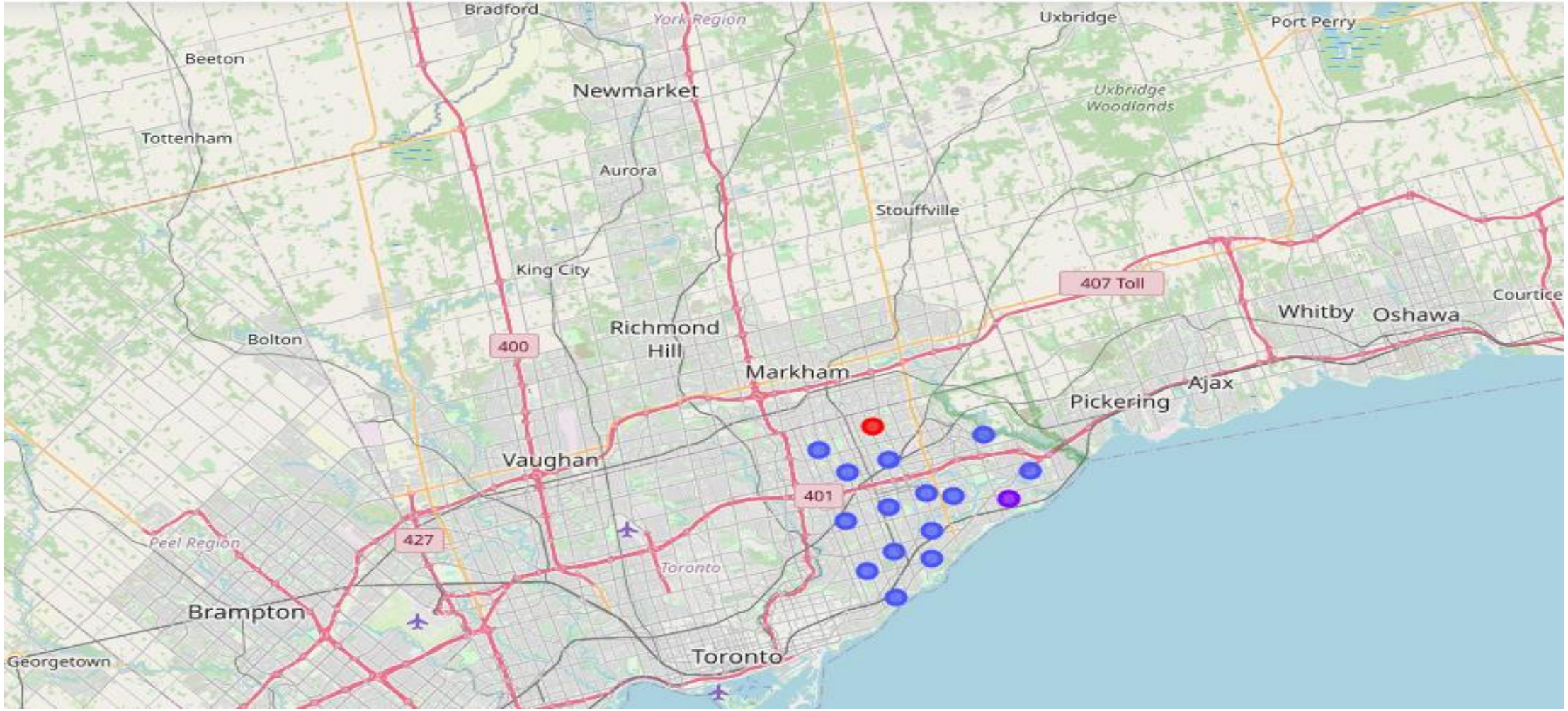
- To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.
- Using credentials of Foursquare API features of near-by 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 500.

Most Common Venues near Neighborhood

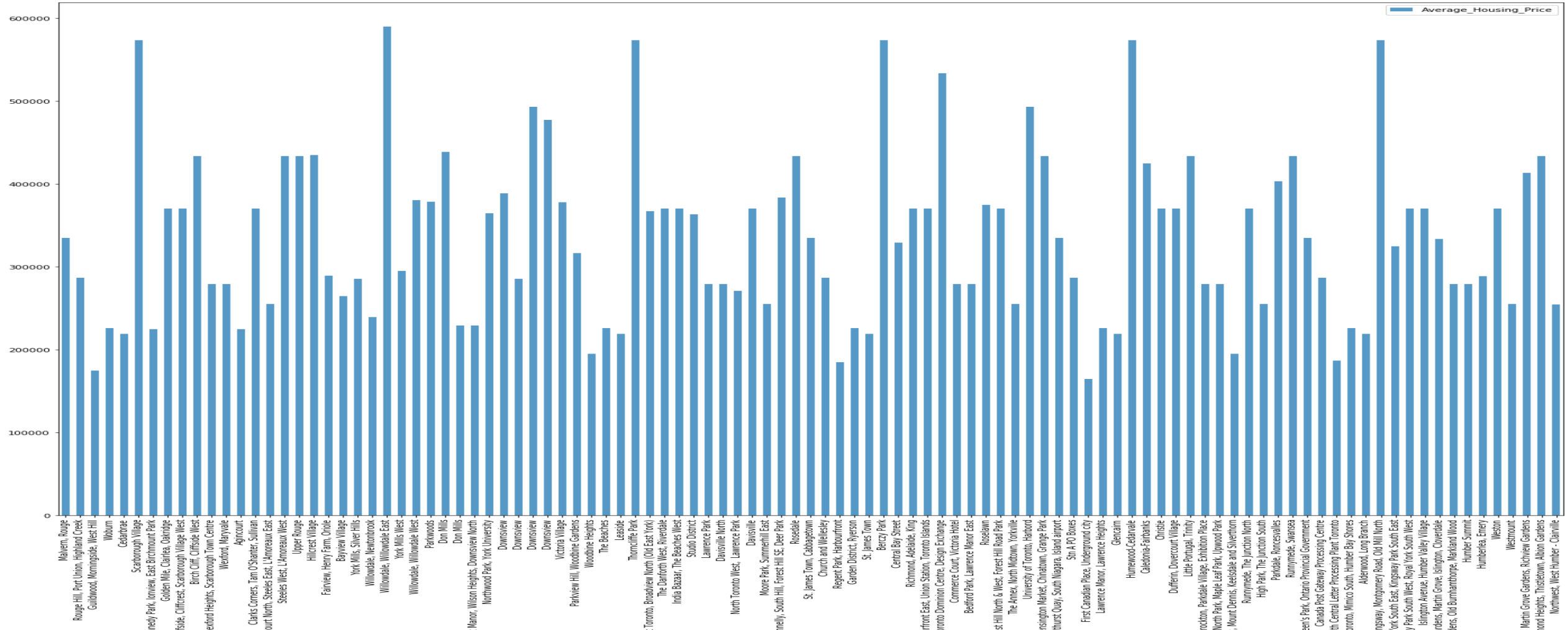
	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th M Comn Ve
0	M1B	Scarborough	Malvern, Rouge	43.81153	-79.19552	2	Zoo Exhibit	Fast Food Restaurant	Farmers Market	Construction & Landscaping	Flea Market	Fish Market	Fish & Chips Shop	Filip Restau
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	2	Bar	Home Service	Golf Course	Fish & Chips Shop	Yoga Studio	Event Space	Donut Shop	Dump Restau
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17520	1	Park	Athletics & Sports	Gym / Fitness Center	Yoga Studio	Dog Run	Donut Shop	Dumpling Restaurant	East Europ Restau
3	M1G	Scarborough	Woburn	43.76820	-79.21761	2	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Park	Yoga Studio	Event Space	Donut Shop	Dump Restau
4	M1H	Scarborough	Cedarbrae	43.76969	-79.23944	2	Thai Restaurant	Indian Restaurant	Flower Shop	Bank	Gas Station	Athletics & Sports	Caribbean Restaurant	Ha Restau

RESULTS

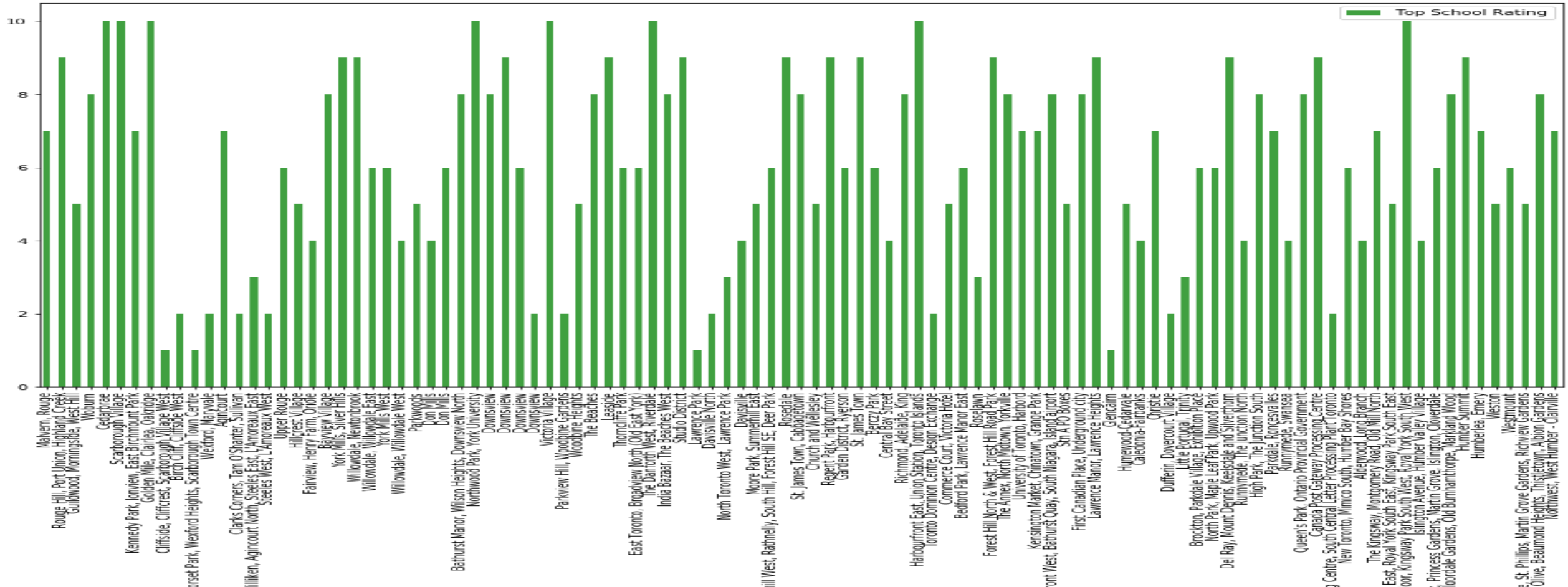
Clustering in Scarborough



Average Housing Price in Scarborough



Schools By Ratings in Scarborough



CONCLUSION

METHODOLOGY

- In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

THANK YOU