1. Introduction:

Nowadays myriads of people are migrating to Canada and they are struggling with problems like finding a good neighborhood, house pricing, good school or universities for their children. Moreover, this was exactly the problems I faced last year when I started the process of applying for master's degree in Canada. It takes a lot of time to find the good school or house with reasonable pricing especially in such large cities like Toronto. As me myself has come up against this problem I know how helpful would be to have program which analyses all these data.

The objective of this project is to sort list of houses based on house pricing and schools based on ratings and reviews to help people migration to Scarborough, Toronto.

- Scarborough is a large, multicultural area that contains the Scarborough Bluffs, huge cliffs overlooking Lake Ontario, lined with parks, beaches, and hiking trails. Inland, the sprawling Toronto Zoo features global animal pavilions, close-up encounters, and a wildlife health center. The area is also known for its diverse spread of restaurants, including regional Southeast Asian, Chinese, and Indian cuisine
- This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.
- 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
- Libraries, which are used: Pandas, Folium, Scikit Learn, JSON, XML, Beautiful Soup and Matplotlib.

2. Data:

I will use Scarborough dataset from Wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Foursquare API Data:

A data about different venues in different neighborhoods of that specific borough is needed. In order to gain that information we will use "Foursquare" locational information. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since 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 inside every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

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. Neighborhood
2. Neighborhood Latitude
3. Neighborhood 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

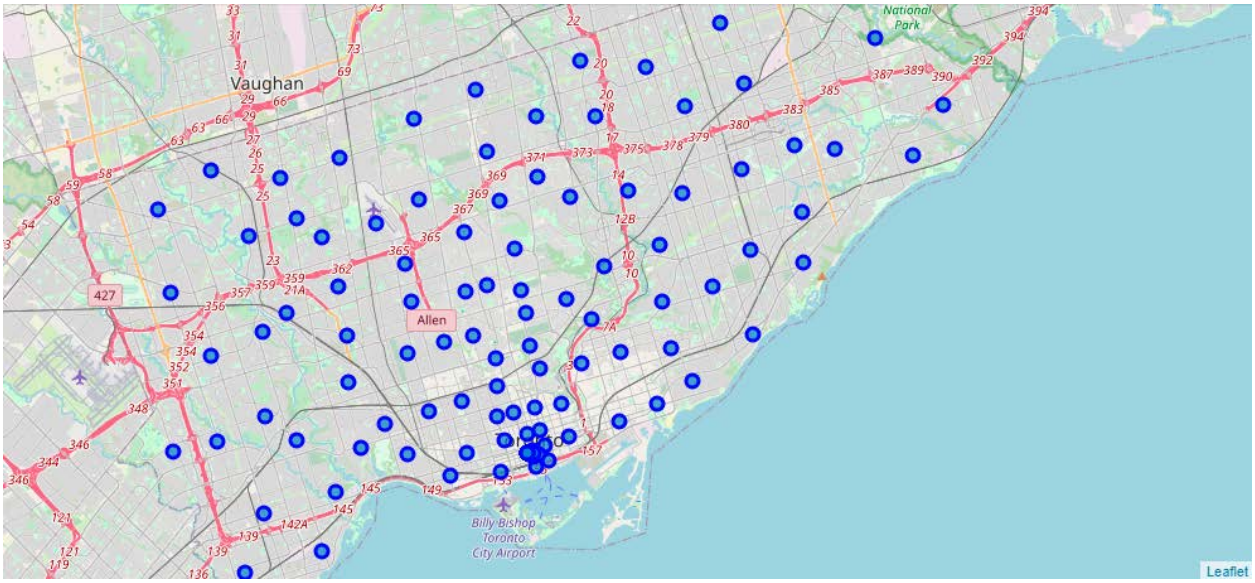


Figure 1: Scarborough Map

3. Methodology:

Clustering Approach:

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 K-Means Clustering Approach

ɔugh	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
Not ıed\n	Not assigned\n	43.64869	-79.38544	2	Coffee Shop	Hotel	Café	Restaurant	Beer Bar	Japanese Restaurant	Tea Room	Arts & Crafts Store	Steakhouse	Pizza Place
gh\n	Malvern, Rouge	43.81153	-79.19552	2	Zoo Exhibit	Fast Food Restaurant	Farmers Market	Paintball Field	Ethiopian Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
gh\n	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	1	Bar	Home Service	Fish & Chips Shop	Yoga Studio	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
gh\n	Guildwood, Morningside, West Hill	43.76575	-79.17520	0	Park	Athletics & Sports	Gym / Fitness Center	Yoga Studio	Dive Bar	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
gh\n	Woburn	43.76820	-79.21761	0	Park	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Yoga Studio	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store

Most Common venues near Neighborhood

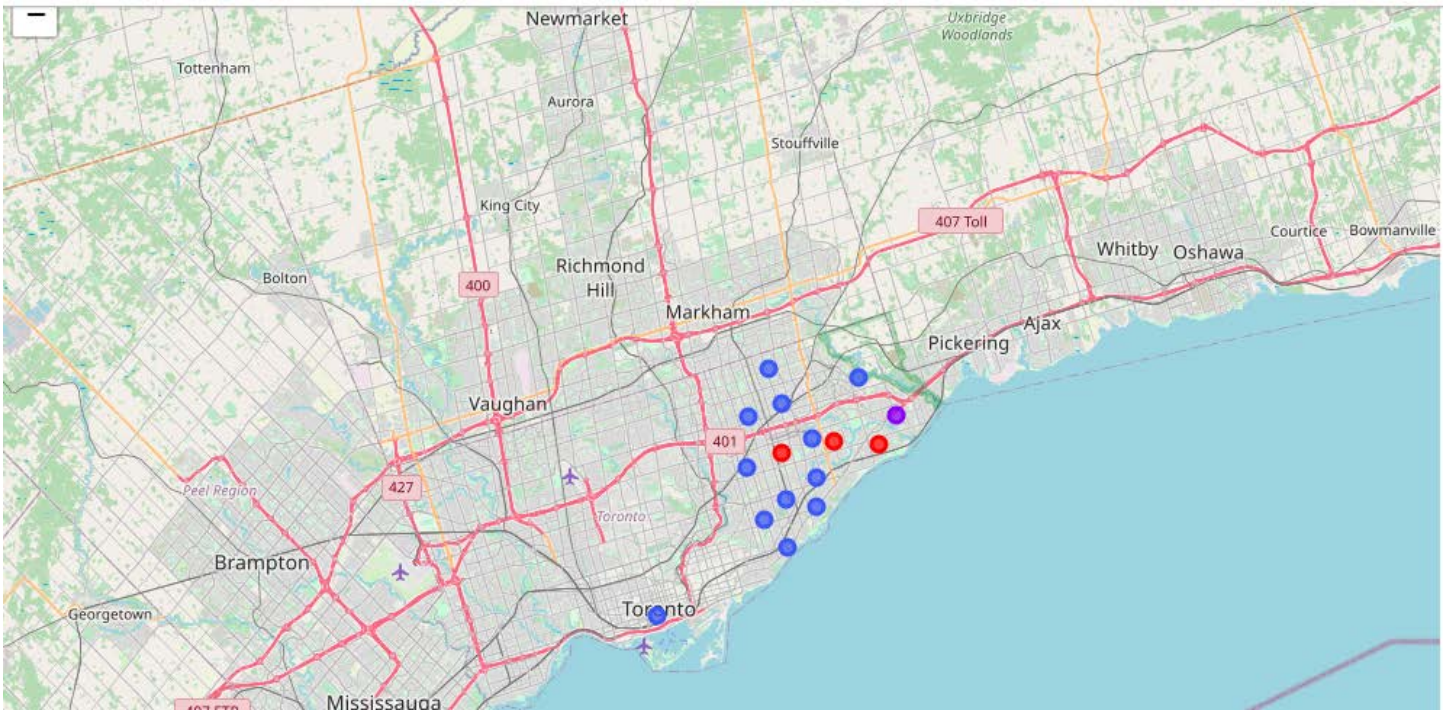
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Seafood Restaurant	Sushi Restaurant	Supermarket	Latin American Restaurant	Malay Restaurant	Mediterranean Restaurant	Skating Rink
1	Alderwood, Long Branch	Pizza Place	Gym	Sandwich Place	Gas Station	Convenience Store	Pub	Coffee Shop	Athletics & Sports	Dog Run	Doner Restaurant
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Convenience Store	Other Great Outdoors	Yoga Studio	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
3	Bayview Village	Park	Asian Restaurant	Locksmith	Trail	Event Space	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
4	Bedford Park, Lawrence Manor East	Pizza Place	Restaurant	Italian Restaurant	Sandwich Place	Thai Restaurant	Coffee Shop	Pub	Pharmacy	Pet Store	Butcher

Work Flow:

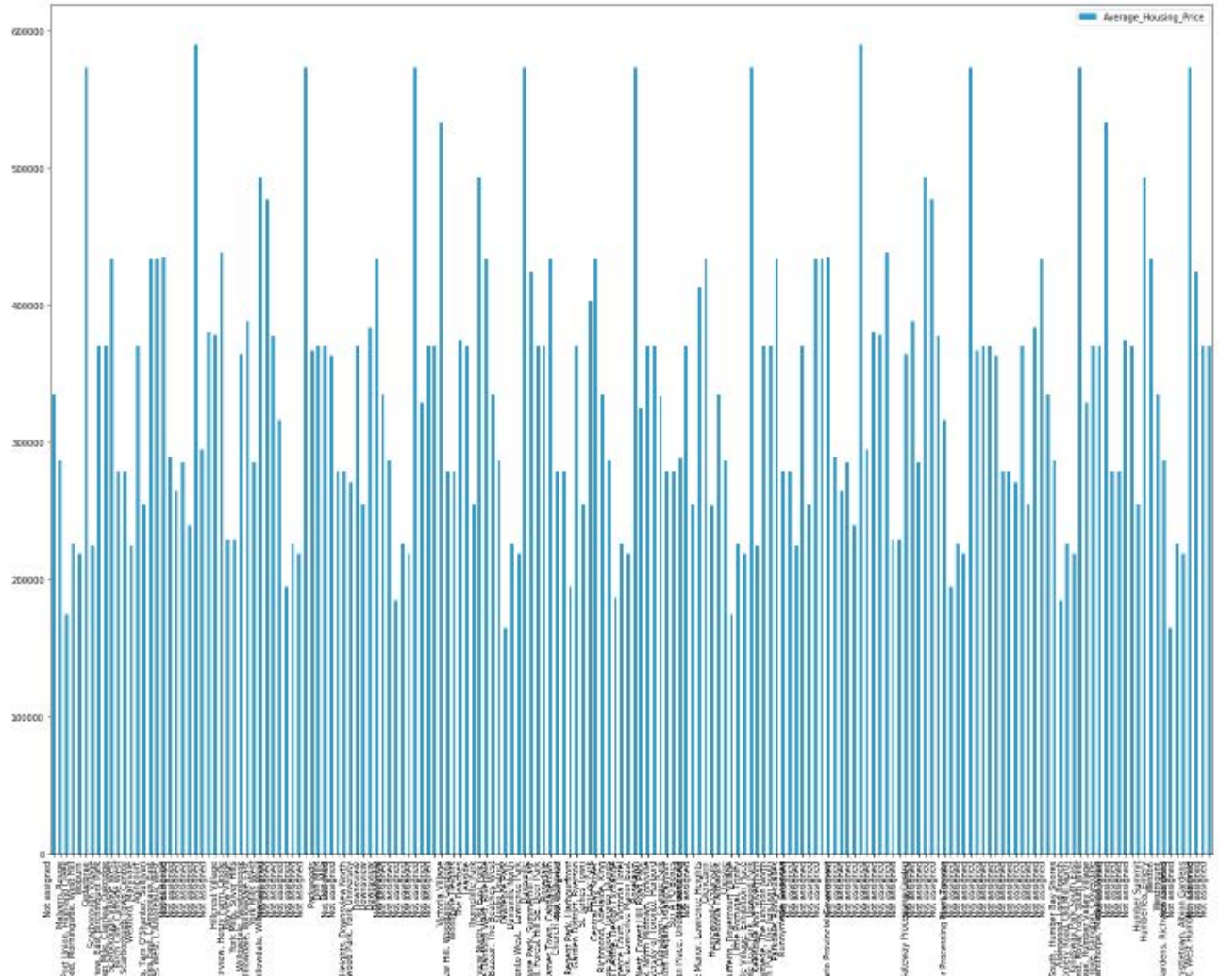
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

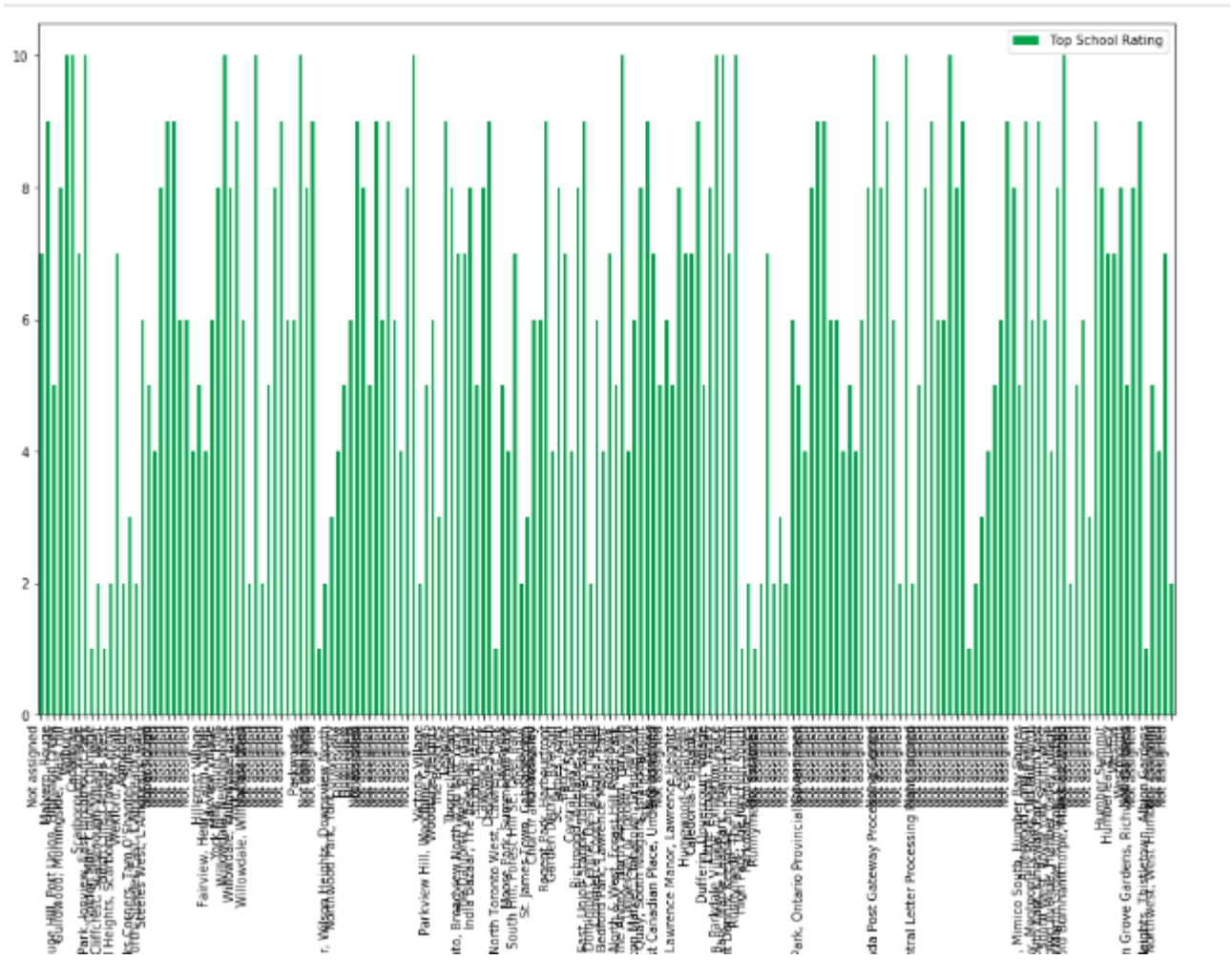
Map of Clusters in Scarborough



Average Housing Price by Clusters in Scarborough



School Ratings by Clusters in Scarborough



The 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. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.

Foursquare API:

This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

Problem, Which Tried to Solve:

The major purpose of this project is to suggest a better neighborhood 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 center, markets and other daily needs things nearby.

1) Sorted list of house in terms of housing prices in an ascending or descending order

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

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 180 different latitude and longitude 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.