Final Report

Introduction

The purpose of this capstone project is to help people finding better facilities in and around their neighbourhood. It will help people in taking smart decision on selecting better neighbourhood in Scarborough, Toronto. The project aims to create an analysis of features for a people migrating to Scarborough to search a great neighbourhood as a comparative analysis between neighbourhoods. The features include house price, better school, low crime rates, good road connectivity and weather conditions, good services in case of emergencies, fresh water resources, etc. This project will help people to get awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Problem Which Tried to Solve:

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 like-minded 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

Data Section

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use Scarborough dataset from week 3. This dataset consists of latitude and longitude and zip codes.

Foursquare API: We will also need data about different venues in different neighbours of that specific borough. In order to that information, we will use Foursquare location information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. 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.

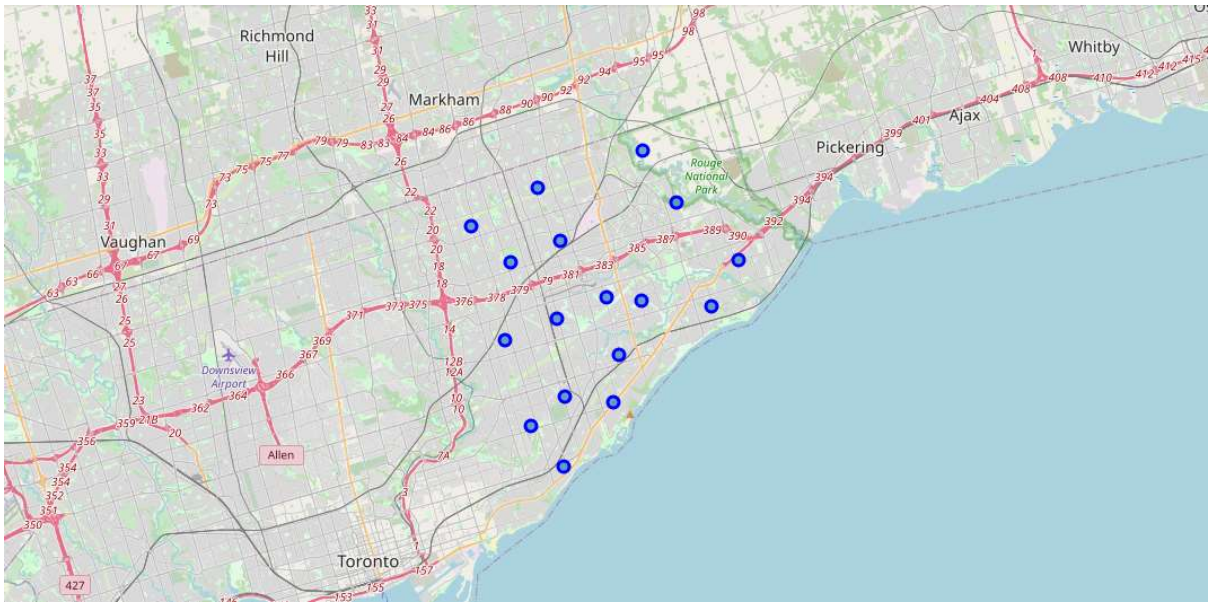
After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 500 meters and the number of venues to be explored as 100.

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.

Map of Scarborough:



Clustering Approach:

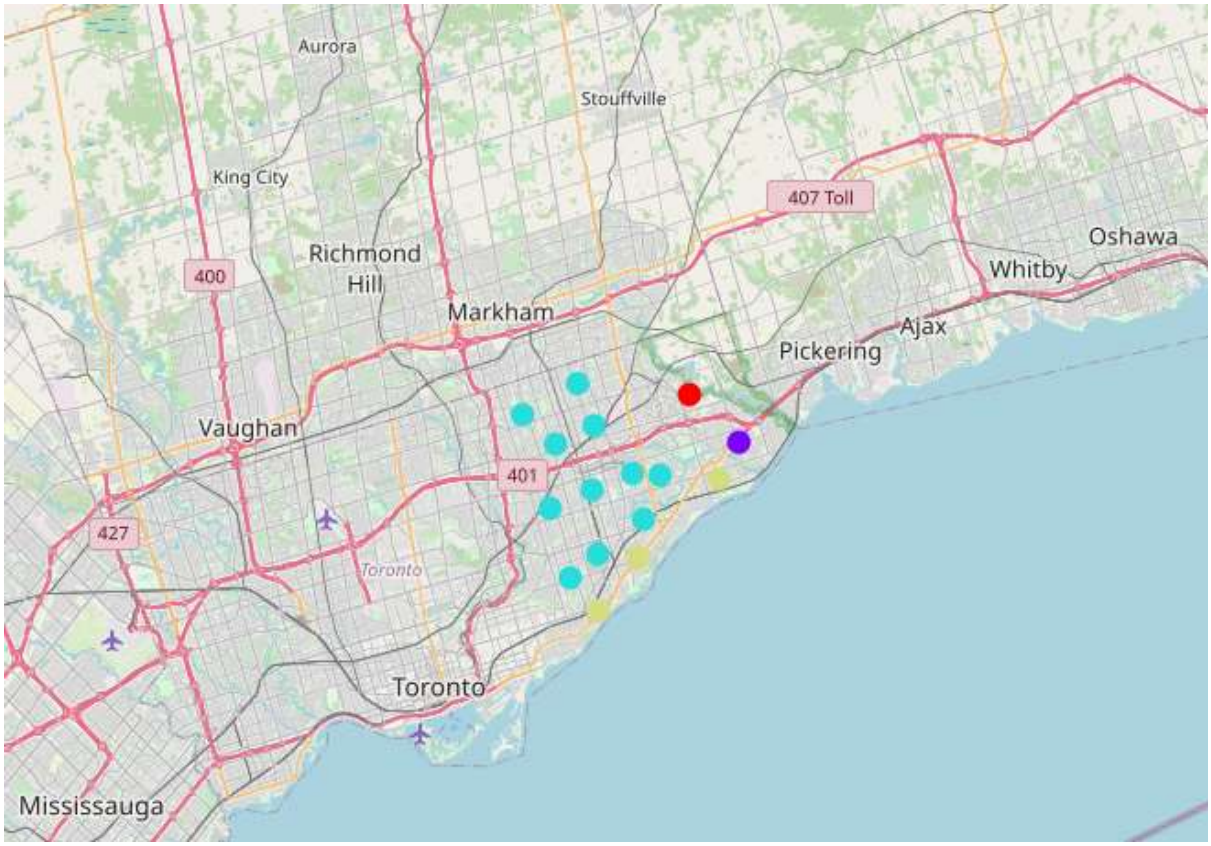
To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering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

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 1000.

Results Section:

Map of Clusters in Scarborough:



Different Neighborhood Clusters:

Cluster 1:

df1

	Neighbourhood	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	Score
0	Rouge Hill , Port Union , Highland Creek	0.0	Park	Italian Restaurant	Breakfast Spot	Burger Joint	Gaming Cafe	Dim Sum Restaurant	Diner	Discount Store	Electronics Store	Fast Food Restaurant	1.0

Cluster 2:

df2

	Neighbourhood	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	Score
2	Scarborough Village	1.0	Ice Cream Shop	Fast Food Restaurant	Sandwich Place	Convenience Store	Indian Restaurant	Coffee Shop	Big Box Store	Chinese Restaurant	Restaurant	Fried Chicken Joint	1.000000
0	Woburn	1.0	Indian Restaurant	Fast Food Restaurant	Park	Pizza Place	Supermarket	Department Store	Juice Bar	Discount Store	Chinese Restaurant	Sandwich Place	0.833333
3	Kennedy Park , Ionview , East Birchmount Park	1.0	Chinese Restaurant	Bus Line	Discount Store	Pizza Place	Grocery Store	Coffee Shop	Hobby Shop	Department Store	Pharmacy	Fast Food Restaurant	0.818182
9	Milliken , Agincourt North , Steeles East , L'Amoreaux East	1.0	Chinese Restaurant	Park	Pharmacy	BBQ Joint	Bakery	Dim Sum Restaurant	Shop & Service	Caribbean Restaurant	Shopping Plaza	Pizza Place	0.750000
1	Cedarbrae	1.0	Indian Restaurant	Coffee Shop	Hakka Restaurant	Burger Joint	Music Store	Flower Shop	Fried Chicken Joint	Chinese Restaurant	Caribbean Restaurant	Gas Station	0.700000
5	Dorset Park , Wexford Heights , Scarborough Town Centre	1.0	Electronics Store	Chinese Restaurant	Restaurant	Indian Restaurant	Furniture / Home Store	Bakery	Pizza Place	Coffee Shop	Park	Intersection	0.666667
10	Steeles West , L'Amoreaux West	1.0	Fast Food Restaurant	Intersection	Pharmacy	Pizza Place	Chinese Restaurant	Sandwich Place	Gym Pool	Nail Salon	Other Great Outdoors	Liquor Store	0.629630
6	Wexford , Maryvale	1.0	Middle Eastern Restaurant	Grocery Store	Pizza Place	Intersection	Burger Joint	Dessert Shop	Fish Market	Coffee Shop	Chinese Restaurant	Korean Restaurant	0.606061
4	Golden Mile , Clairlea , Oakridge	1.0	Intersection	Bakery	Park	Convenience Store	Coffee Shop	Pharmacy	Pizza Place	Fast Food Restaurant	Pub	Sandwich Place	0.571429
7	Agincourt	1.0	Chinese	Shopping	Coffee Shop	Sandwich	Newsagent	Pool Hall	Pool	Pizza Place	Pharmacy	Park	0.534884

Cluster 3:

df3

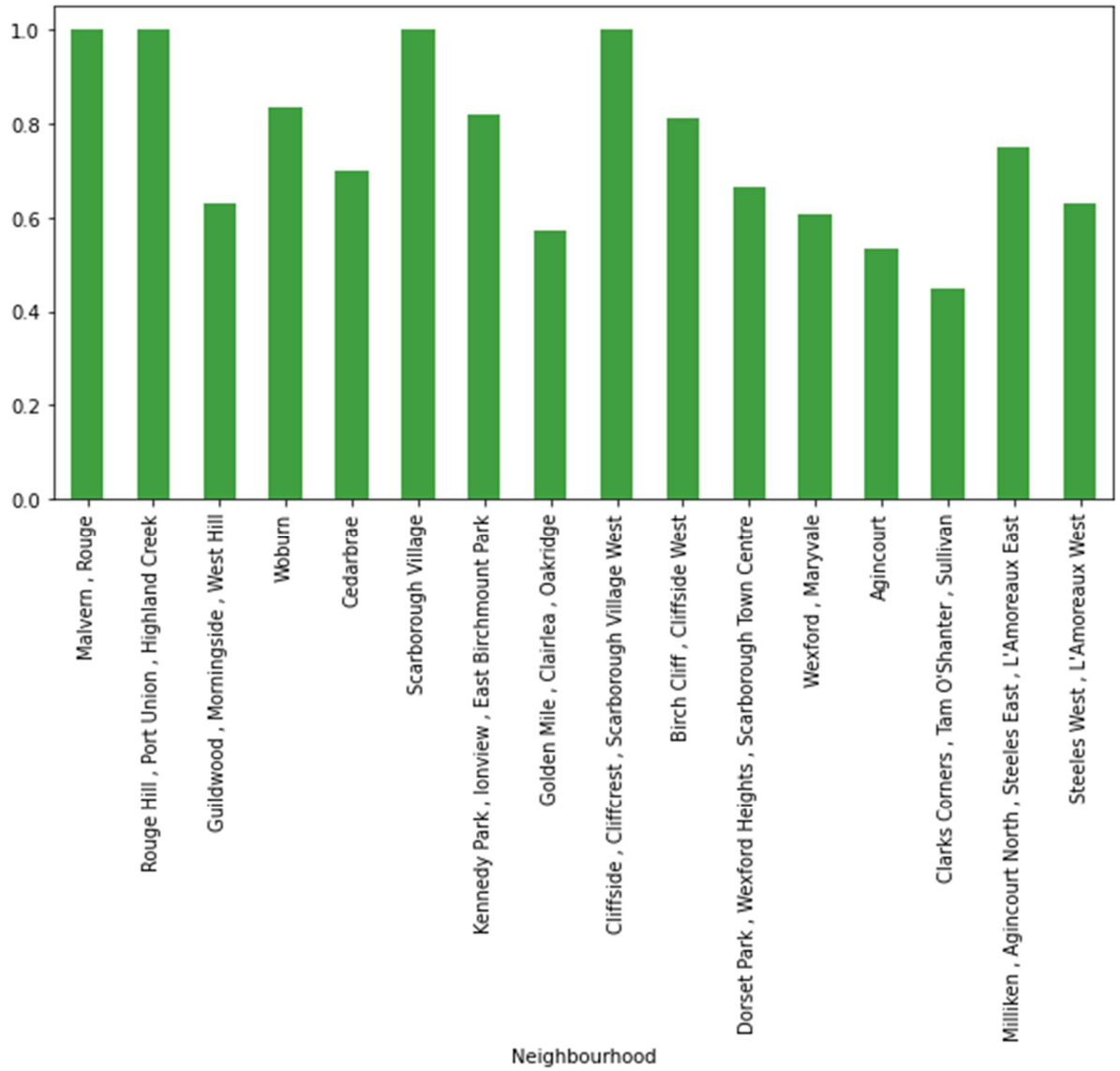
	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Score
1	Cliffside , Cliffcrest , Scarborough Village West	2.0	Ice Cream Shop	Pharmacy	Gift Shop	Coffee Shop	Sandwich Place	Park	Restaurant	Bank	Pizza Place	Bistro	1.000000
2	Birch Cliff , Cliffside West	2.0	Park	Skating Rink	General Entertainment	Convenience Store	College Stadium	Diner	Restaurant	Café	Dessert Shop	Gym Pool	0.812500
0	Guildwood , Morningside , West Hill	2.0	Park	Pharmacy	Coffee Shop	Smoothie Shop	Gymnastics Gym	Pizza Place	Discount Store	Fast Food Restaurant	Restaurant	Food & Drink Shop	0.631579

Cluster 4:

df4

	Neighbourhood	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	Score
0	Malvern , Rouge	3.0	Zoo Exhibit	Trail	Fast Food Restaurant	Construction & Landscaping	Hobby Shop	BBQ Joint	Badminton Court	Diner	Discount Store	Electronics Store	1.0

Comparing neighbourhoods on their scores:



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 has 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.

Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 3 different clusters. This dataset has very-similar neighborhoods around them.

I also get the scores of the neighbourhoods from the scores of the venues in the neighbourhoods and sorted them in their descending order.

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 Works:

This project can be continued for making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things (daily needs or things we need to live a better life) around and also in terms of cost effective.