Capstone Project – The Battle of Neighbourhoods Finding a Better Place in Scarborough, Toronto Report

1. Introduction:

The purpose of this Capstone Project is to help people in exploring better facilities around their neighbourhood. It will help people making smart and efficient decision on selecting great neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighbourhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This Capstone Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighbourhood as a comparative analysis between neighbourhoods. 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 fresh and waste water and excrement conveyed in sewers and recreational facilities.

It 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.

2. Data Section

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 Data:

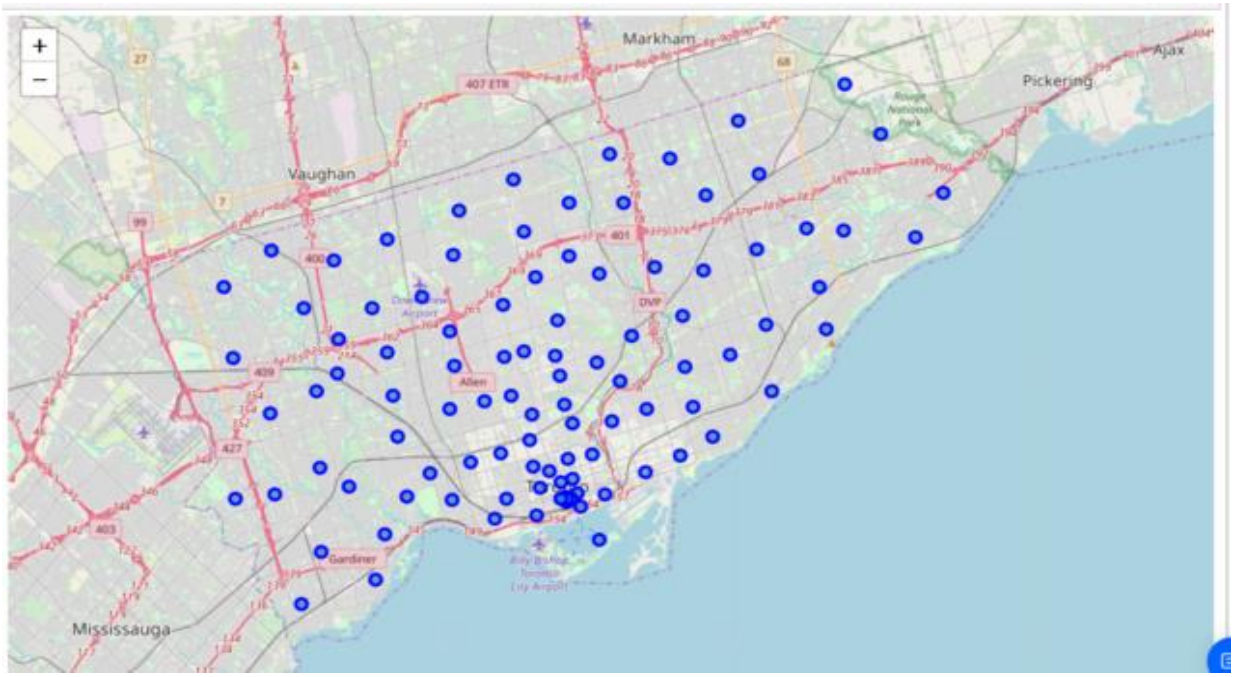
We will need data about different venues in different neighbourhoods of that specific borough.
In order to gain that information we will use “Foursquare” locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such 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 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

Map of Toronto



3. Methodology Section

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.

Using K-Means Clustering Approach | Most Common Venue

Out[47]:

index	Borough	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
21	York	0.0	Park	Women's Store	Pool	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
35	East York	0.0	Park	Convenience Store	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
45	North York	0.0	Park	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Drugstore	College Stadium
64	York	0.0	Park	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Drugstore	College Stadium
66	North York	0.0	Park	Convenience Store	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
85	Scarborough	0.0	Park	Playground	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop
91	Downtown Toronto	0.0	Park	Playground	Trail	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run

Most Common Venues near Neighbourhood | Using Clustering

Out[42]:

	Neighbourhood	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	Lounge	Latin American Restaurant	Skating Rink	Breakfast Spot	Clothing Store	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant
1	Alderwood, Long Branch	Pizza Place	Gym	Coffee Shop	Pharmacy	Sandwich Place	Pub	Distribution Center	Dessert Shop	Dim Sum Restaurant	Diner
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Sushi Restaurant	Pharmacy	Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Fried Chicken Joint	Diner	Deli / Bodega
3	Bayview Village	Café	Bank	Chinese Restaurant	Japanese Restaurant	Dim Sum Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop
4	Bedford Park, Lawrence Manor East	Coffee Shop	Restaurant	Sandwich Place	Italian Restaurant	Thai Restaurant	Pharmacy	Pizza Place	Pub	Café	Butcher

Work Flow:

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

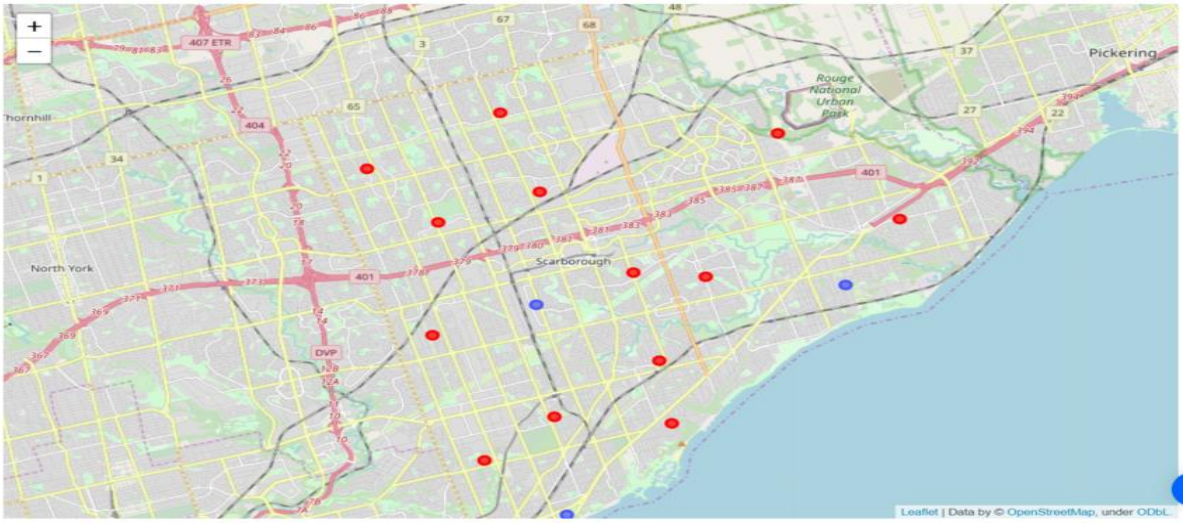
4. Results Section

Map of Clusters in Scarborough

```
In [49]: toronto_merged_nonan.loc[toronto_merged_nonan['Cluster Labels'] == 2, toronto_merged_nonan.columns[[1] + list(range(5, toronto_m
```

Out[49]:

index	Borough	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
6	Scarborough	2.0	Fast Food Restaurant	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Drugstore	Gym / Fitness Center



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 Capstone 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 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.

- Sorted list of houses in terms of housing prices in a ascending or descending order
- Sorted list of schools in terms of location, fees, rating and reviews

6. Conclusion Section

In this Capstone project, using k-means cluster algorithm I separated the neighbourhood into 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 Works:

This Capstone 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.