Final Report

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

1. Introduction

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toranto. 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 neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

This 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 fresh and waste water and excrement conveyed in sewers and recreational facilities. It will help people to get awareness of the area and neighborhood 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 neighborhoods 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 neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, 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.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
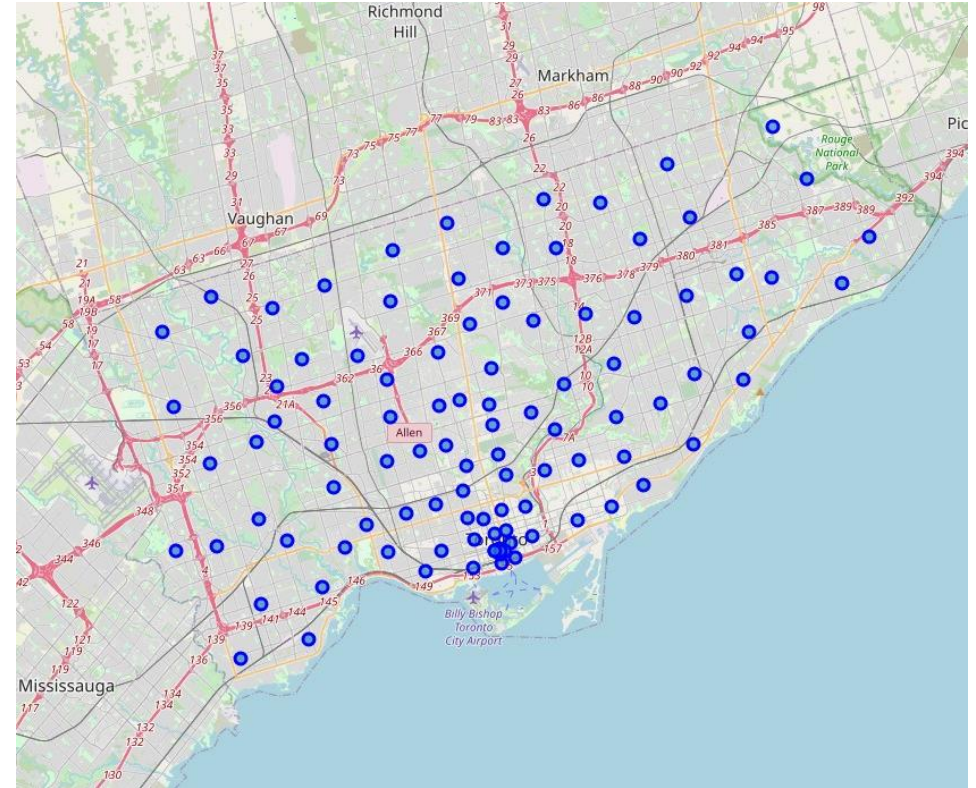
3. Methodology Section

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.

|	PostalCode	Borough	Neighborhood	Longitude	Latitude	Cluster Labels
0	M3A	North York	Parkwoods	-79.32991	43.75245	0
1	M4A	North York	Victoria Village	-79.31306	43.73057	1
2	M5A	Downtown Toronto	Regent Park, Harbourfront	-79.36264	43.65512	1
3	M6A	North York	Lawrence Manor, Lawrence Heights	-79.45042	43.72327	1
4	M7A	Queen's Park	Ontario Provincial Government	-79.39188	43.66253	1

Using K-Means Clustering Approach

|	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	Supermarket	Pool	Bakery	Pizza Place	Coffee Shop	Clothing Store	Print Shop	Sandwich Place
1	Alderwood, Long Branch	Gas Station	Coffee Shop	Pub	Sandwich Place	Pizza Place	Dog Run	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Pizza Place	Sushi Restaurant	Intersection	Sandwich Place	Convenience Store	Park	Yoga Studio	Doner Restaurant	Discount Store
3	Bayview Village	Asian Restaurant	Flower Shop	Trail	Dog Run	Park	Gas Station	Yoga Studio	Doner Restaurant	Discount Store	Distribution Center
4	Bedford Park, Lawrence Manor East	Coffee Shop	Restaurant	Sandwich Place	Liquor Store	Intersection	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Juice Bar	Sports Club

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 700.
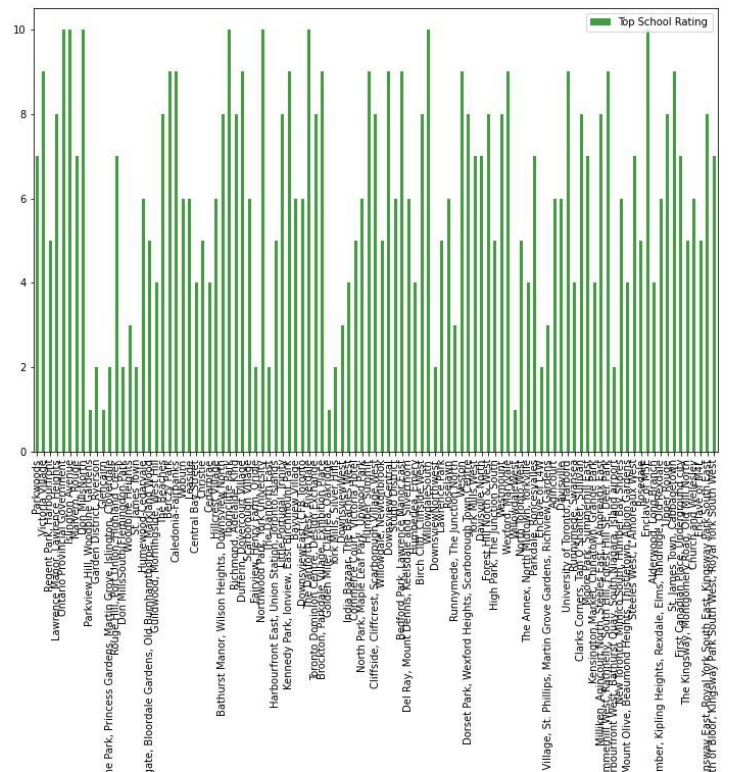
4. Results Section –

Map of Clusters in Scarborough
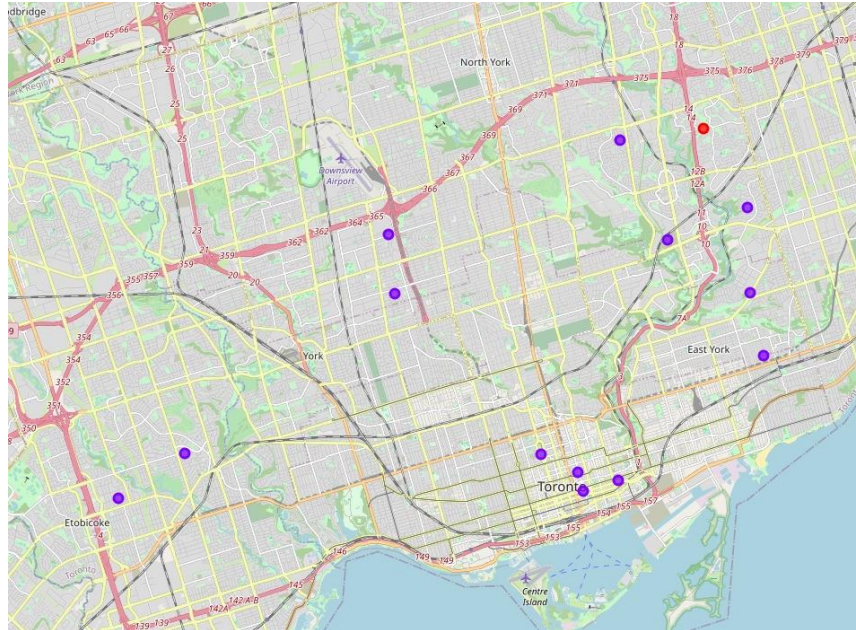
Average Housing Price by Clusters in Scarborough
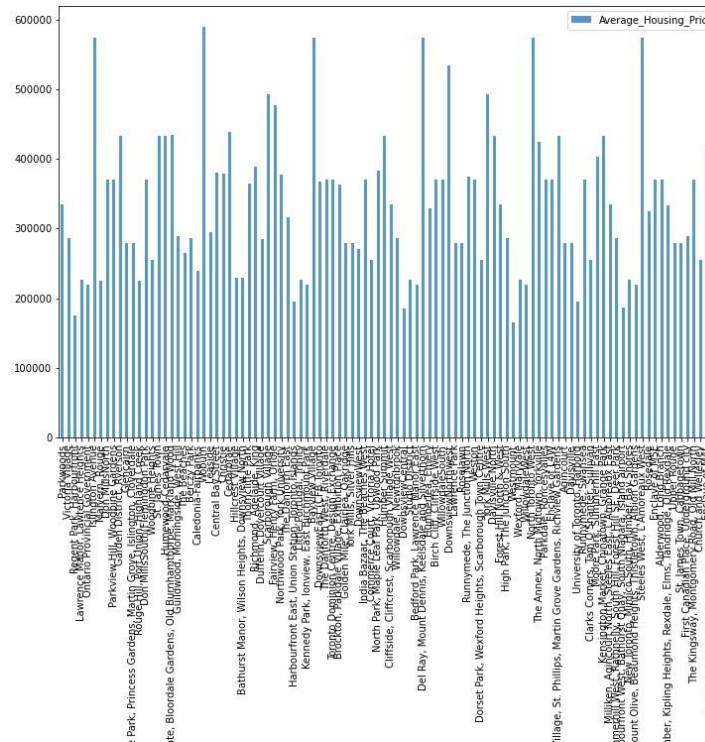
School Ratings by Clusters in Scarborough

The Location: Scarborough is a popular destination for new immigrants in Canada. 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.

Foursquare API: I 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 Statement: The purpose of this project is to suggest a better neighborhood in a new city for anyone who is shifting there. Social presence in society in terms of like minded 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 a ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

6. Conclusion Section

In this project, I separated the neighborhood into 10 different clusters using k-means cluster algorithm and for 103 different lattitude+logitude from dataset, which have 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.

Future Work: This project can be continued for making it more precise in terms to find best house in Scarborough. A dashboard can be created using input terms from a user on cost and preference indices and most probably neighborhoods can be suggested using the algorithm.

Libraries which were used to create the project -
1.Pandas: For creating and manipulating dataframes.
2.Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
3.Scikit Learn: For importing k-means clustering.
4.JSON: Library to handle JSON files.
5.XML: To separate data from presentation and XML stores data in plain text format.
6.Geocoder: To retrieve Location Data.
7.Beautiful Soup and Requests: To scrap and library to handle http requests.
8.Matplotlib: Python Plotting Module.