Introduction:

This Capstone Project is to help people who are looking for better neighbourhood around them with better amenities. The project aims to strengthen the people's ability to help make them smarter decision in selecting the better neighbourhood in Scarborough, Toranto.

Business Problem:

Lately, most of the population across the globe are migrating to different states of Canada, in search of better lifestyle, which requires them to do lot of research for selecting a proper neighbourhood (such as housing prices, schools). This project would help people who are looking for better neighbourhoods, which is closer to school, super market, shopping mall, theatre, hospital, etc.

Target Audience:

This Capstone Project aim to create an analysis of features of different neighbourhoods in Scarborough, Toronto to find the most suitable neighbourhood. 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.

This project aims to create awareness about the area where the particular person prefers to move.

Data Section:

We would be using the same data set that was provided for week 3 assignment.

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

Methodology:

Foursquare API Data:

We would be needing data about different places across the neighbourhoods for that borough. We are going to use use "Foursquare" for information on location. Foursquare is a location data provider which provides information on different venues and events within an area of interest. The information includes venue names, locations, menus and even photos. With API, the foursquare location platform is used as the sole data source.

Post fetching the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues across different neighbourhoods in the list with the radius to be 100 meter.

The data retrieved from Foursquare API, contains information of venues within a specified distance of the longitude and latitude of the postcodes.

The information obtained are 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

We use clustering approach to compare different cities and explore different neighbourhoods and segment them. After segmentation, we group them into different clusters to find similar neighbourhoods in Toronto which are similar to those in NY. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

We then find the most common venue across neighbourhood using clustering

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.

Results Section:

We have used Scarborough location as it is one of the popular destination for new immigrants.It is one of the most diverse and multicultural areas in the Greater Toronto. Even with more restrictions on immigrants and refugees, people migrating to Canada has been on rise and the is the reason we have selected the location for out analysis

Map of Clusters in Scarborough

Avg. House Price in Scarborough

Rating of different schools across in Scarborough
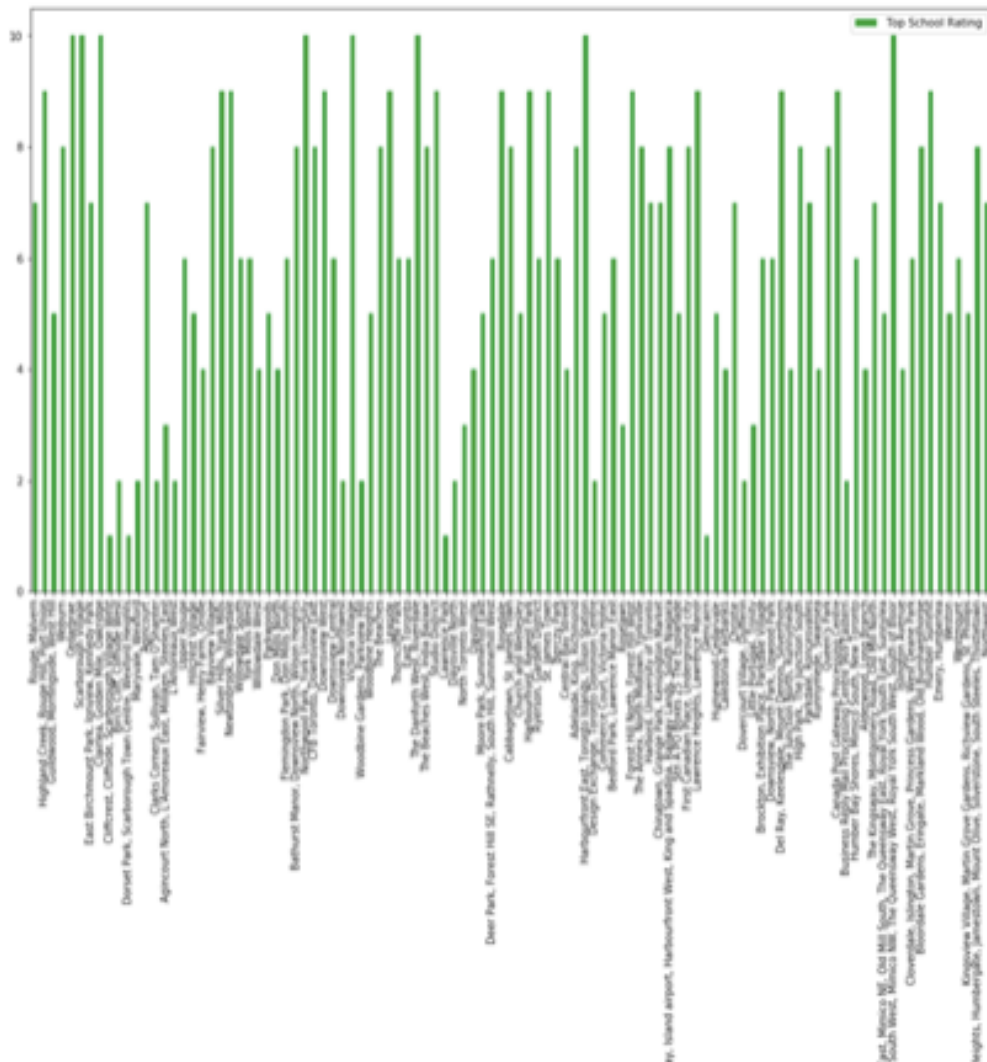
Discussion:

Main Purpose of the project was to suggest the best neighbourhood in the city for a person to migrate. We shortlisted the pricing of housing and schools based on fees, reviews and location.The particular person can select house with an average pricing and presence of better amenities across the neighbourhood.

Conclusion:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 10 clusters based on their similarities, and lastly

providing recommendations to the relevant stakeholders i.e. property developers and buyers regarding the best locations to migrate. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 2 are the most preferred location for migrating to Scarborough, Toronto. The findings of this project will help the individuals to look for the places with better amenities.

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.

Reference Libraries used:

1. Pandas: For creating and manipulating dataframes.
2. Folium: Python visualization library would be used to visualize the neighbourhoods 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.