Recommendation of Neighbourhoods in Scarborough

**Introduction :**

For many people, planning for migrating to a new city is still a hard process, as it is a hectic and time-consuming process to find a suitable place and neighbourhood, to accommodate their families. With varying budgets and needs, the person searching for a place, would want to know all the available neighbourhoods and pricing to make a decision. The person would also like to consider about the accessibility of schools, supermarkets, restaurant, weather conditions, etc., This report will focus on exploring the neighbourhoods in the city of Toronto.

The main objective of this project is to recommend a better neighbourhood in a new city of Toronto and will help in exploring the city and get the awareness of the neighbourhoods in the new city. Using data science methodology and machine learning techniques, this project aims to recommend neighbourhood for people moving to the city Toronto.

**Target Audience:**

Potential clients look to buy property in Toronto, who have lack of knowledge about the neighbourhoods in the city.

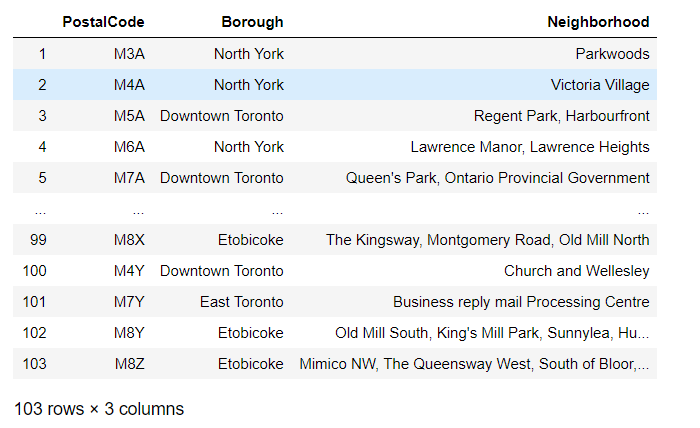
People who are interested in investing on Real estate

The business that wants to use the information to enhance the user experience and advertisements

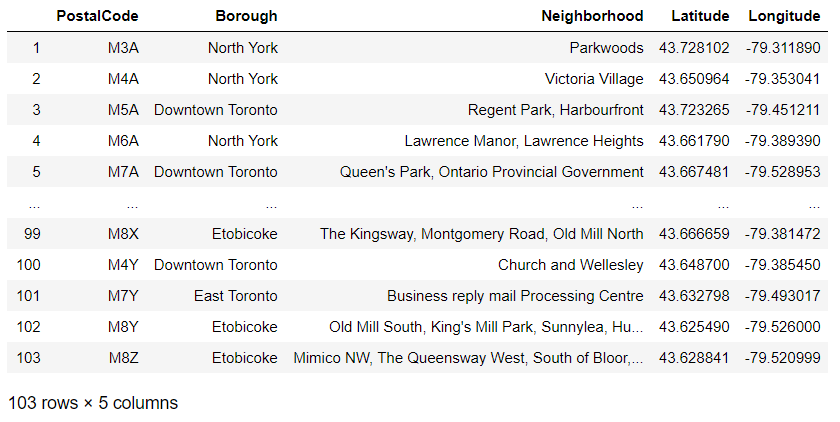
**Data Description:**

* Scarborough, a neighbourhood in Toronto, is chosen for as the observation target, As It is a favourite destination for new immigrants in Canada to reside in and is one of the most distinct and multicultural areas in the Greater Toronto Area.
* ‘https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M’

The above mentioned url is used for getting all the neighbourhood info for Toronto. Below is the python data frame that has the Neighbourhood information, which is scrapped from the above-mentioned web page using ‘Beautiful Soup’.



* Google geocoder API will be used to get location coordinates, which can be further used to visualize onto a map. The following data has the coordinates of each neighbourhood, which has collected form the geocoder API.
* There is a csv file (Geospatial\_Coordinates) from the previous ‘Segmenting and Clustering’ section, which contains the coordinates of all the neighbourhoods in Toronto, used if the geocoder package does not return coordinates for certain neighbourhoods.



* Foursquare API is used as a data source to get the location and other information about various venues in Scarborough. Using this, nearby venues and its features can be fetched and will be used for analysis.

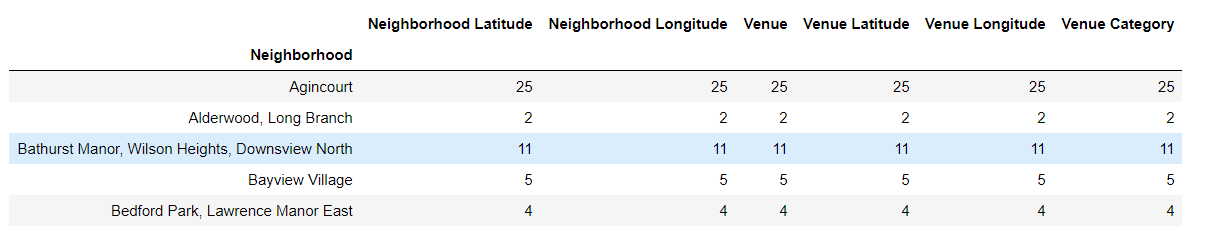
**Methodology:**

With the available data frame that contains the Neighbourhood data with coordinates, the analysis can be started by plotting a map with the help of the package folium.

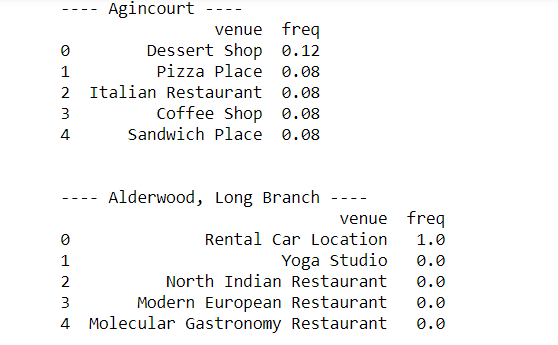
Then, we can use the Foursquare API, to get the nearby venues. Below is the data frame that is obtained for exploring the venues.

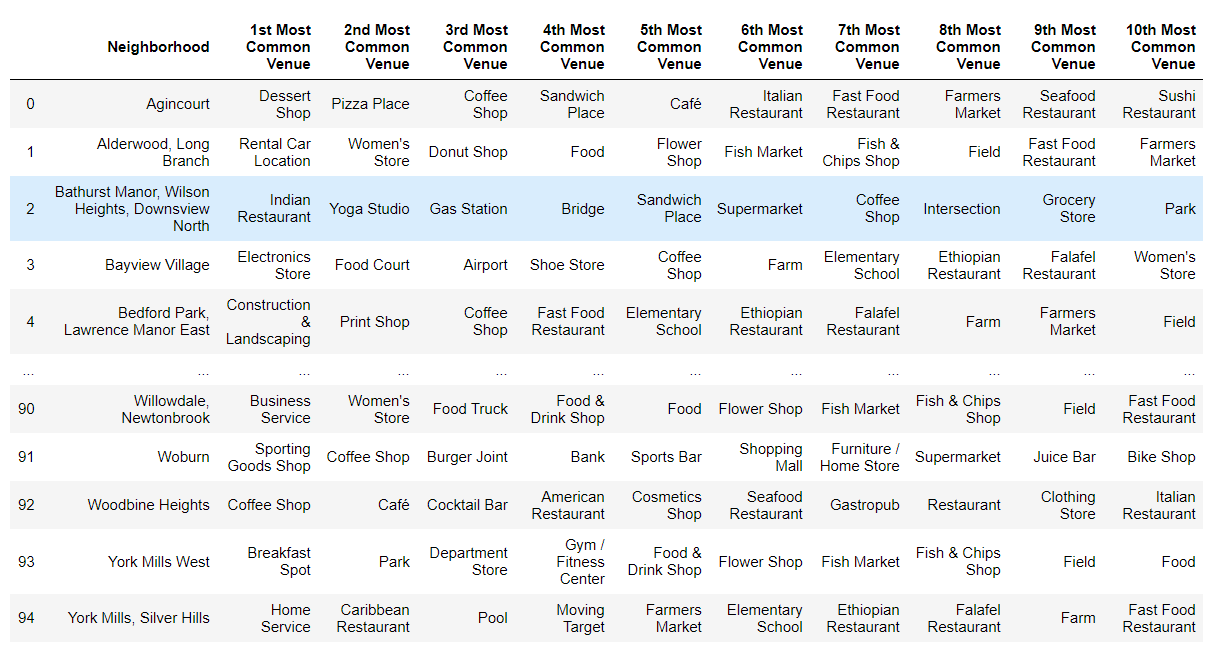


Further we can derive the venue count based on a neighbourhood, and that is shown below.



After that, we can use ‘One hot encoding’, to find the top 5 venues in each neighbourhood, as below.

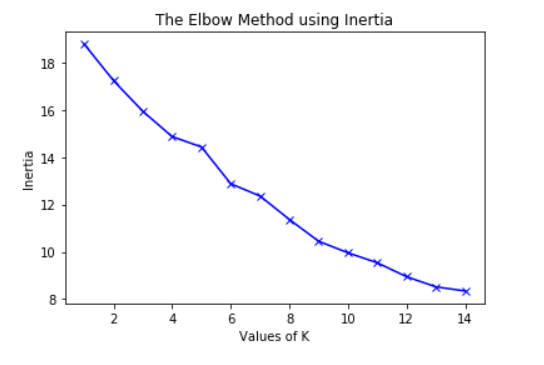




Now, we use Clustering algorithm to find similar places in the City. In this project, ‘K Means Clustering’ is used for clustering the similar neighbourhoods.

Since, the goal is to group similar neighbourhoods, k means will be more appropriate technique.

For finding the k value, ‘Elbow method using inertia’ is used. The plot between inertia and k value can be used to find the right k value.

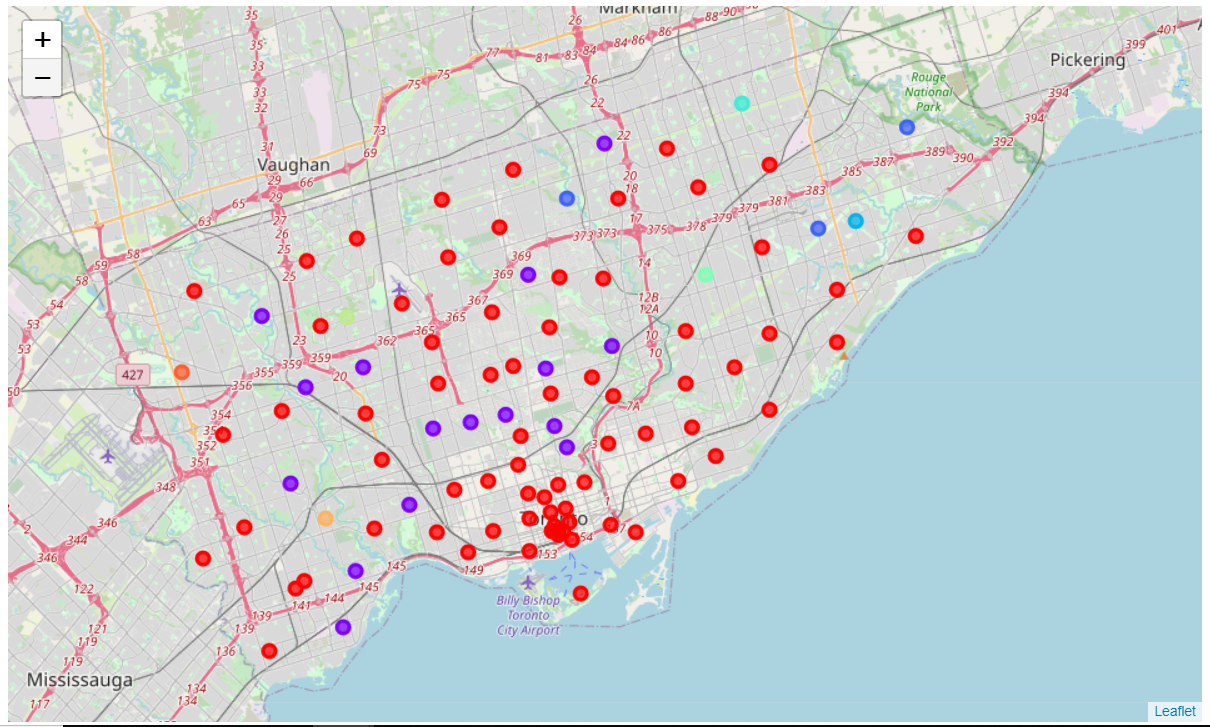


Using the above plot, we can conclude that the right k value for this data is 9.

After clustering the neighbourhoods, the data frame can be updated with the house price and Top school rating in the neighbourhood and the analysis can be done on each cluster with the available data.

**Results:**

Below is the plot of the clustered neighbourhoods of Scarborough.



**Discussion:**

Based on the results calculated we can make the following observations and recommendations:

* The clusters 1 and 2 has a greater number of venues with the Top school rating of more than 5 and Average house price of more than 330000
* Cluster 6 has only one venue ( Westmount ), where the Average house price is 165000 (which is lesser than any other venue in other clusters) and Top school rating is 8
* For a person, who loves to have food outside most of the time (especially bachelors) and spend time in pub, cluster 1 is the obvious choice. As it has lot of such venues with Restaurant, Pub, Café, Hotel, Bakery and all.
* For a person, who often go to recreational sites such as Park, Playground, etc., Cluster 2 is the obvious choice.
* For a person who has his family along, will go for the clusters based on their budget. If the budget is low, then ‘Westmount’ will be recommended. Others, it may vary with their needs.

**Result:**

Purpose of this project is to help a person who is migrating to a new city (Scarborough), who is lack of knowledge about the city. Foursquare and Toronto neighbourhood data are the primary data source used. Using ‘K means clustering’ algorithm, the neighbourhoods are clustered.

With the help of the clustered data, this project will be helpful in recommending neighbourhoods to consider and explore. The final decision on selecting a neighbourhood will be made by the client based on their budget and primary needs.