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’.

	PostalCode	Borough	Neighborhood
1	M3A	North York	Parkwoods
2	M4A	North York	Victoria Village
3	M5A	Downtown Toronto	Regent Park, Harbourfront
4	M6A	North York	Lawrence Manor, Lawrence Heights
5	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
99	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
100	M4Y	Downtown Toronto	Church and Wellesley
101	M7Y	East Toronto	Business reply mail Processing Centre
102	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
103	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

103 rows × 3 columns

- 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.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
1	M3A	North York	Parkwoods	43.728102	-79.311890
2	M4A	North York	Victoria Village	43.650964	-79.353041
3	M5A	Downtown Toronto	Regent Park, Harbourfront	43.723265	-79.451211
4	M6A	North York	Lawrence Manor, Lawrence Heights	43.661790	-79.389390
5	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.667481	-79.528953
...	...	...	...	...	...
99	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.666659	-79.381472
100	M4Y	Downtown Toronto	Church and Wellesley	43.648700	-79.385450
101	M7Y	East Toronto	Business reply mail Processing Centre	43.632798	-79.493017
102	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.625490	-79.526000
103	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

103 rows × 5 columns

- 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.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.728102	-79.311890	Tim Hortons	43.725517	-79.313103	Coffee Shop
1	Parkwoods	43.728102	-79.311890	Portugril	43.725819	-79.312785	Portuguese Restaurant
2	Parkwoods	43.728102	-79.311890	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection
3	Parkwoods	43.728102	-79.311890	Pizza Nova	43.725824	-79.312860	Pizza Place
4	Parkwoods	43.728102	-79.311890	Wigmore Park	43.731023	-79.310771	Park
...	...	...	...	...	...	...	...
2266	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	RONA	43.629393	-79.518320	Hardware Store
2267	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	7-Eleven	43.629107	-79.517431	Convenience Store
2268	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Jim & Maria's No Frills	43.631152	-79.518617	Grocery Store
2269	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Value Village	43.631269	-79.518238	Thrift / Vintage Store
2270	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	Kingsway Boxing Club	43.627254	-79.526684	Gym

2271 rows × 7 columns

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

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	25	25	25	25	25	25
Alderwood, Long Branch	2	2	2	2	2	2
Bathurst Manor, Wilson Heights, Downsview North	11	11	11	11	11	11
Bayview Village	5	5	5	5	5	5
Bedford Park, Lawrence Manor East	4	4	4	4	4	4

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

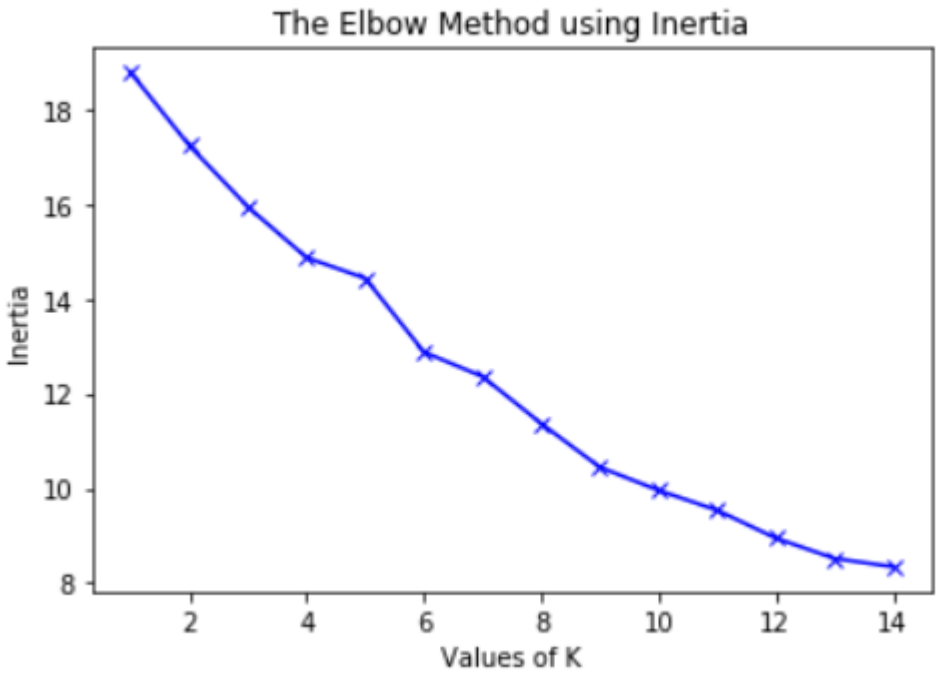
```
---- Agincourt ----
                venue  freq
0        Dessert Shop  0.12
1         Pizza Place  0.08
2  Italian Restaurant  0.08
3         Coffee Shop  0.08
4      Sandwich Place  0.08


---- Alderwood, Long Branch ----
                            venue  freq
0             Rental Car Location   1.0
1                     Yoga Studio   0.0
2          North Indian Restaurant   0.0
3       Modern European Restaurant   0.0
4  Molecular Gastronomy Restaurant   0.0
```

	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	Dessert Shop	Pizza Place	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Fast Food Restaurant	Farmers Market	Seafood Restaurant	Sushi Restaurant
1	Alderwood, Long Branch	Rental Car Location	Women's Store	Donut Shop	Food	Flower Shop	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market
2	Bathurst Manor, Wilson Heights, Downsview North	Indian Restaurant	Yoga Studio	Gas Station	Bridge	Sandwich Place	Supermarket	Coffee Shop	Intersection	Grocery Store	Park
3	Bayview Village	Electronics Store	Food Court	Airport	Shoe Store	Coffee Shop	Farm	Elementary School	Ethiopian Restaurant	Falafel Restaurant	Women's Store
4	Bedford Park, Lawrence Manor East	Construction & Landscaping	Print Shop	Coffee Shop	Fast Food Restaurant	Elementary School	Ethiopian Restaurant	Falafel Restaurant	Farm	Farmers Market	Field
...	...	...	...	...	...	...	...	...	...	...	...
90	Willowdale, Newtonbrook	Business Service	Women's Store	Food Truck	Food & Drink Shop	Food	Flower Shop	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant
91	Woburn	Sporting Goods Shop	Coffee Shop	Burger Joint	Bank	Sports Bar	Shopping Mall	Furniture / Home Store	Supermarket	Juice Bar	Bike Shop
92	Woodbine Heights	Coffee Shop	Café	Cocktail Bar	American Restaurant	Cosmetics Shop	Seafood Restaurant	Gastropub	Restaurant	Clothing Store	Italian Restaurant
93	York Mills West	Breakfast Spot	Park	Department Store	Gym / Fitness Center	Food & Drink Shop	Flower Shop	Fish Market	Fish & Chips Shop	Field	Food
94	York Mills, Silver Hills	Home Service	Caribbean Restaurant	Pool	Moving Target	Farmers Market	Elementary School	Ethiopian Restaurant	Falafel Restaurant	Farm	Fast Food Restaurant

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.

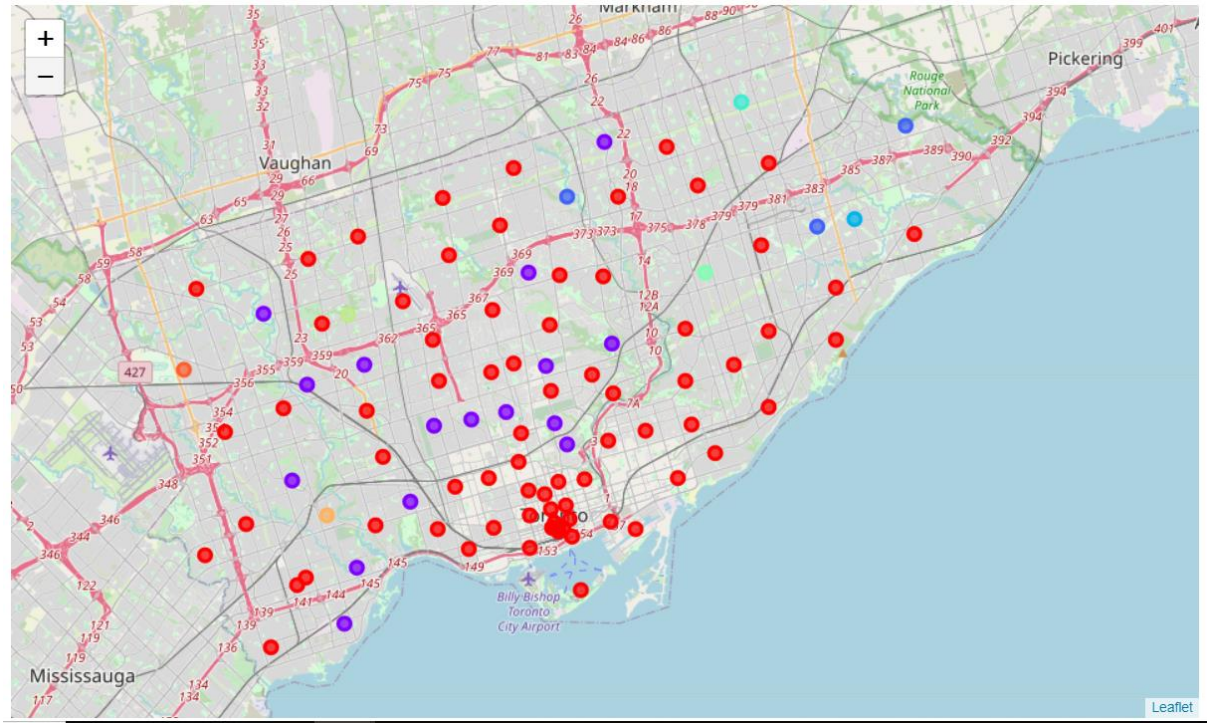


Using the above plot, we can conclude that the right k value for this data is 9.

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.