Battle of the Neighbourhoods Assignment Presentation

1. Introduction

This project was first established in order to provide individuals with a method of analysing their neighbourhood thus allowing them to make smart decision in regards to the area's venues and points of interest.

Every year many people miragrate and move homes, this project aimed to help people migrating from various states of Canada to the neighbourhoods of Scarborough, Toronto; and lets them analyse said neighbourhoods in order to asses the different points of venues which might be of importance to them. Venues include: house prices, school ratings, cafes, hospitals, supermarkets, cineamas, shoppinng mall, etc.

Therefore; people migrating can choose between the best neighbourhoods of Scarborough via this comparative analysys thus, this project will aid individuals gain awareness of the area and the neighbourhood before making concrete decisions when moving to a new city, state or country for whichever reason they may have.

2. Data

Data Link:

'https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&direction=prev&oldid=946126446'.

This project will be utilising the Scarborough stataset which was previously scrapped from the week 3 assignment. The sataaset will includes coordinates and zip codes.

Foursquare

The Foursquare API was usedd for this project since specific data from different venues in the Scarborough area was required. 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.

First a list of the neighbourhoods in Scarborough is made, then after connecting to the Forusqaure API information of each venue in the given area can be gathered. The data was chosen from a 100 metre radius within each neighbourhood.

The following data was obtained from Foursquare about each venue:

- Venue Name

- Venue Category

- Venue Coordinates

- Neighbourhood Coordinates

Scarborough Map
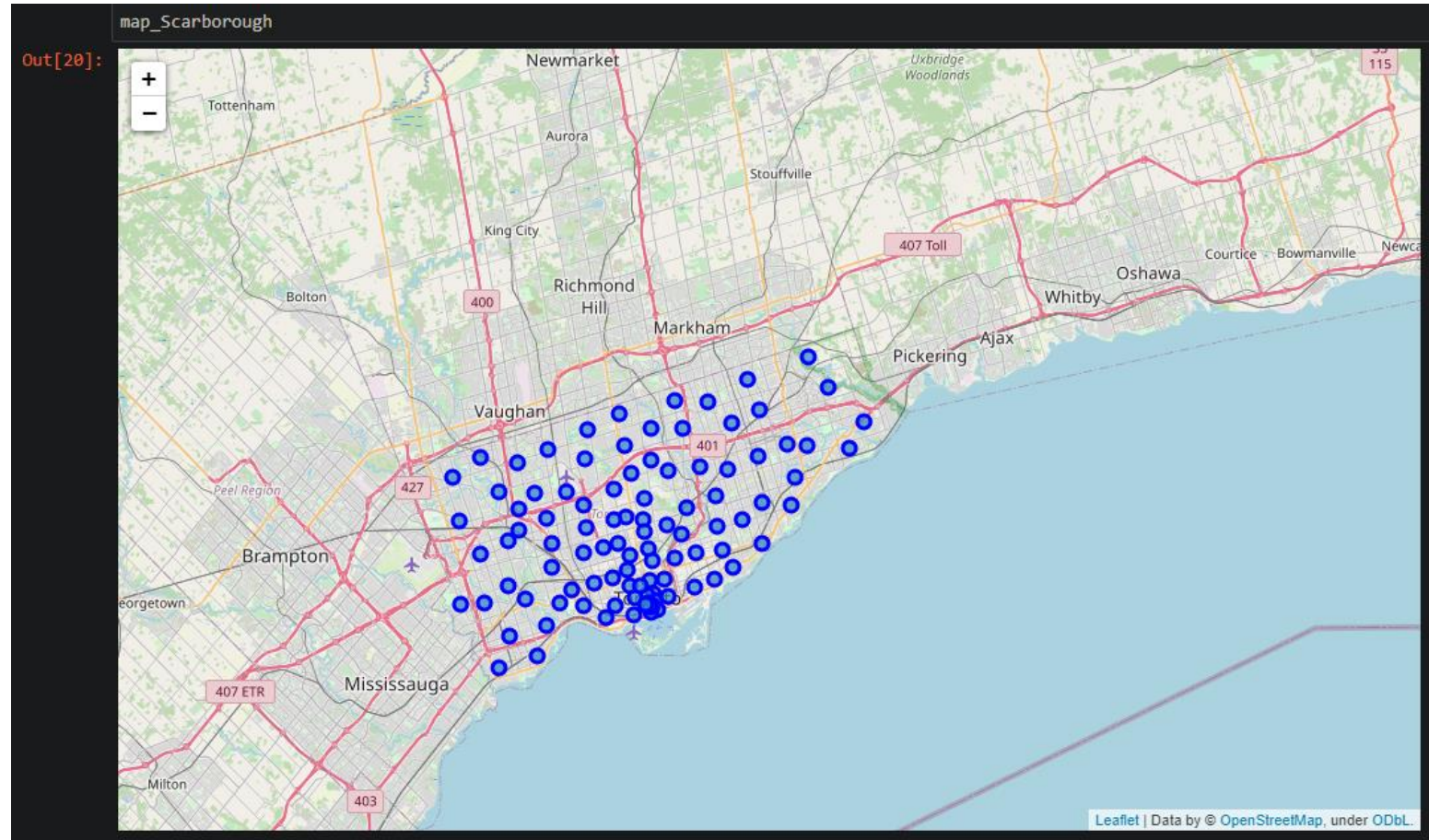
3. Methodology

In order to compare cities the k-means clustering method was utilized. Therefore; allowing two or more cities to be compared by segmenting their neighbourhoods and then grouping them into clusters to find their similarities.

K-Means Clustering Method

```
In [37]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Scarborough_merged =df_2.iloc[:16,:]

Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Scarborough_merged.head()
```

Out[37]:

	Postalcode	Borough	Neighborhood	Latitude	Longitude	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	7tl Co
0	M1B	Scarborough	Rouge, Malvern	43.81139	-79.19662	0	Zoo Exhibit	Fast Food Restaurant	Electronics Store	Dive Bar	Dog Run	Doner Restaurant	Do Sh
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.78574	-79.15875	2	Fish & Chips Shop	Home Service	Bar	Electronics Store	Dog Run	Doner Restaurant	Do Sh
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	0	Park	Gym / Fitness Center	Athletics & Sports	Gymnastics Gym	Eastern European Restaurant	Dive Bar	Do
3	M1G	Scarborough	Woburn	43.76812	-79.21761	0	Fast Food Restaurant	Coffee Shop	Park	Chinese Restaurant	Dumpling Restaurant	Dive Bar	Do
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892	2	Thai Restaurant	Playground	Gas Station	Bank	Caribbean Restaurant	Hakka Restaurant	Ath & !

Most Common Venues

```
In [35]: import numpy as np
         num_top_venues = 10

         indicators = ['st', 'nd', 'rd']

         columns = ['Neighborhood']
         for ind in np.arange(num_top_venues):
             try:
                 columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
             except:
                 columns.append('{}th Most Common Venue'.format(ind+1))

         neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
         neighborhoods_venues_sorted['Neighborhood'] = Scarborough_grouped['Neighborhood']

         for ind in np.arange(Scarborough_grouped.shape[0]):
             neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)

         neighborhoods_venues_sorted.head()
```

Out[35]:

|	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	Adelaide, King, Richmond	Coffee Shop	Hotel	Café	Japanese Restaurant	Restaurant	Seafood Restaurant	Gym	Gastropub	Theater	Thai Restaurant
1	Agincourt	Shopping Mall	Coffee Shop	Pool	Bubble Tea Shop	Skating Rink	Latin American Restaurant	Sandwich Place	Supermarket	Sushi Restaurant	Bank
2	Agincourt North, L'Amoreaux East, Milliken, St...	Intersection	Pharmacy	Arts & Crafts Store	Sandwich Place	Yoga Studio	Dry Cleaner	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Pizza Place	Fried Chicken Joint	Caribbean Restaurant	Auto Garage	Beer Store	Sandwich Place	Park	Fast Food Restaurant	Video Store
4	Alderwood, Long Branch	Gas Station	Pub	Sandwich Place	Dance Studio	Gym	Pizza Place	Pharmacy	Coffee Shop	Comic Shop	Dumpling Restaurant

4. Results

Scarborough Map of Clusters:
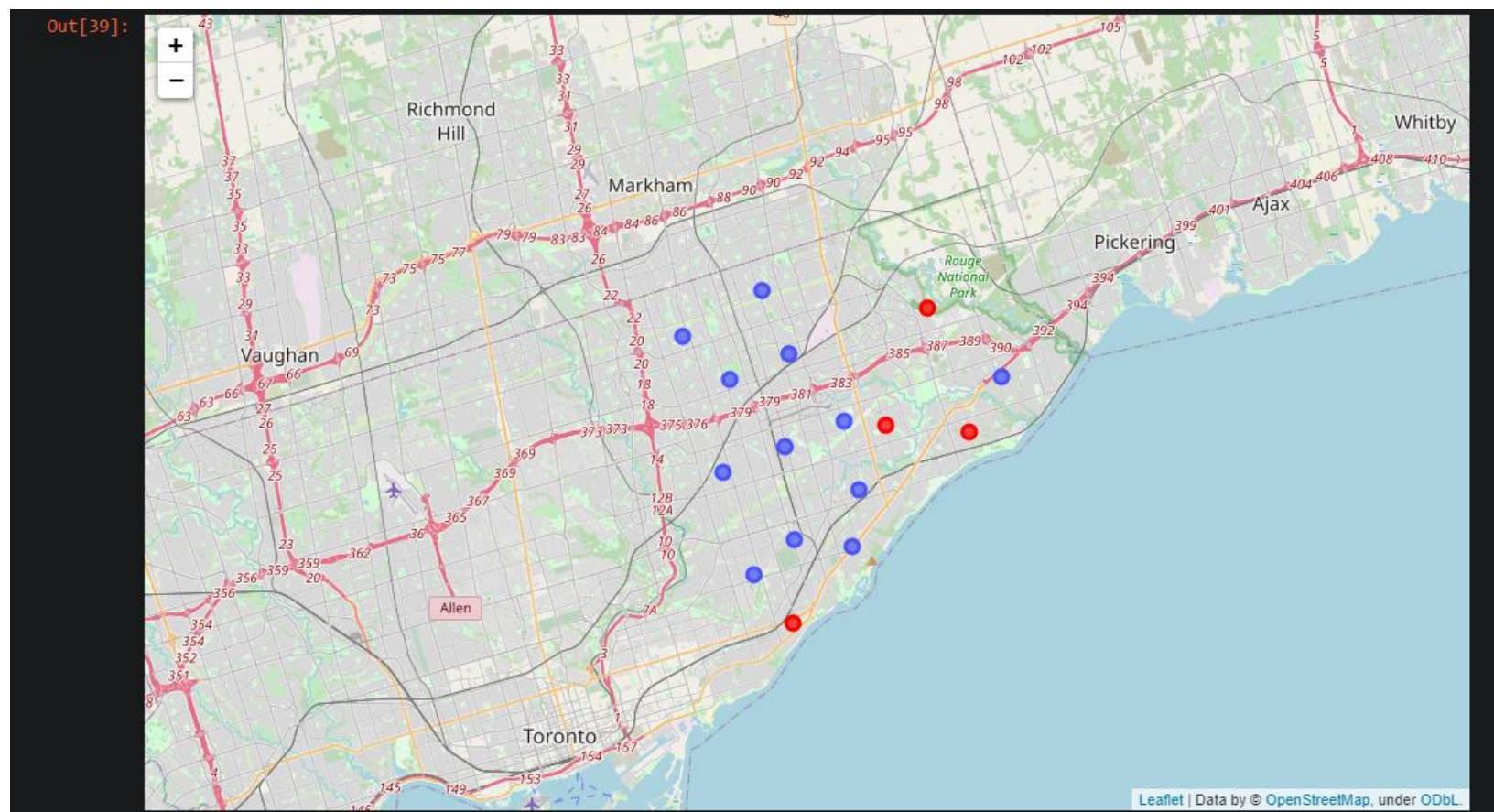
Average House Pricing in Scarborough:
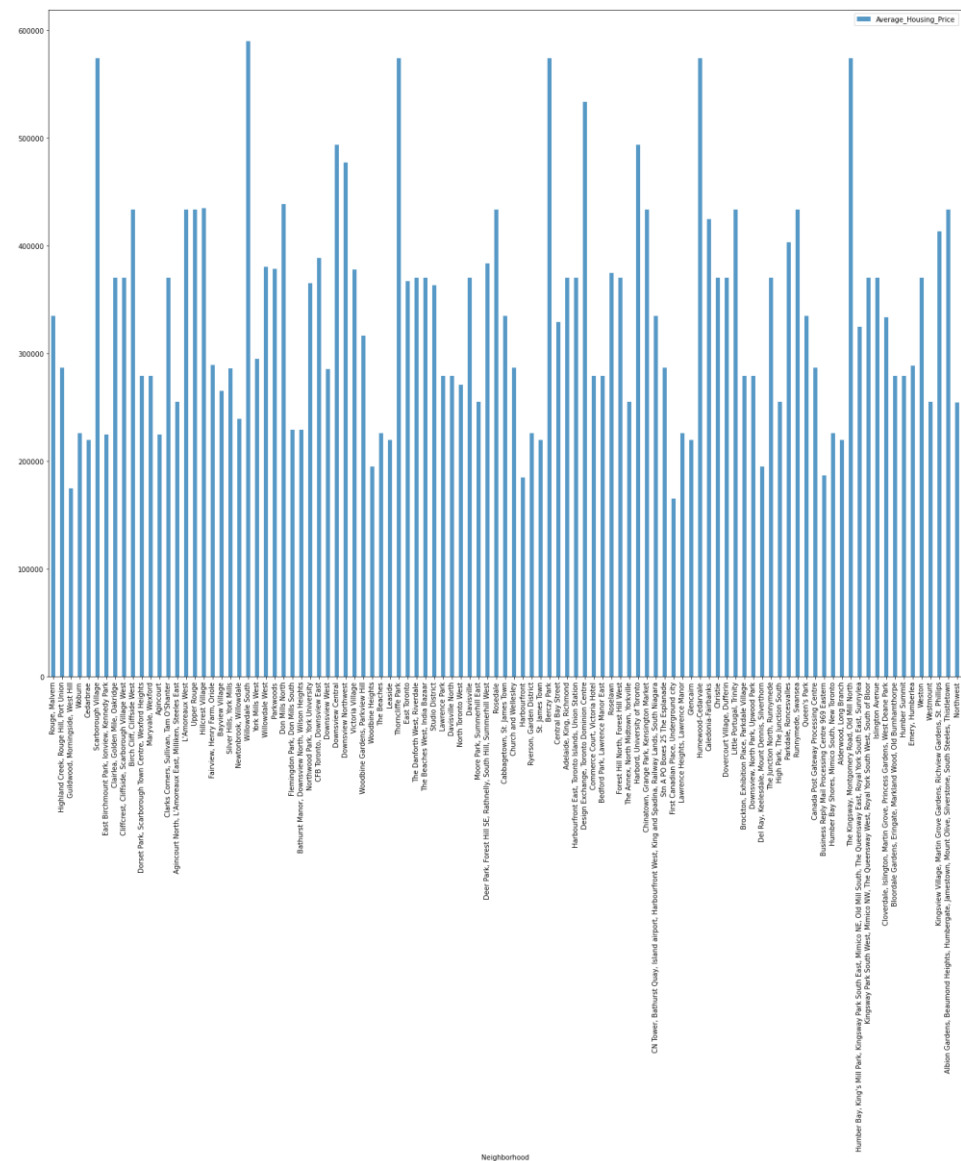
Average School Rating in Scarborough:
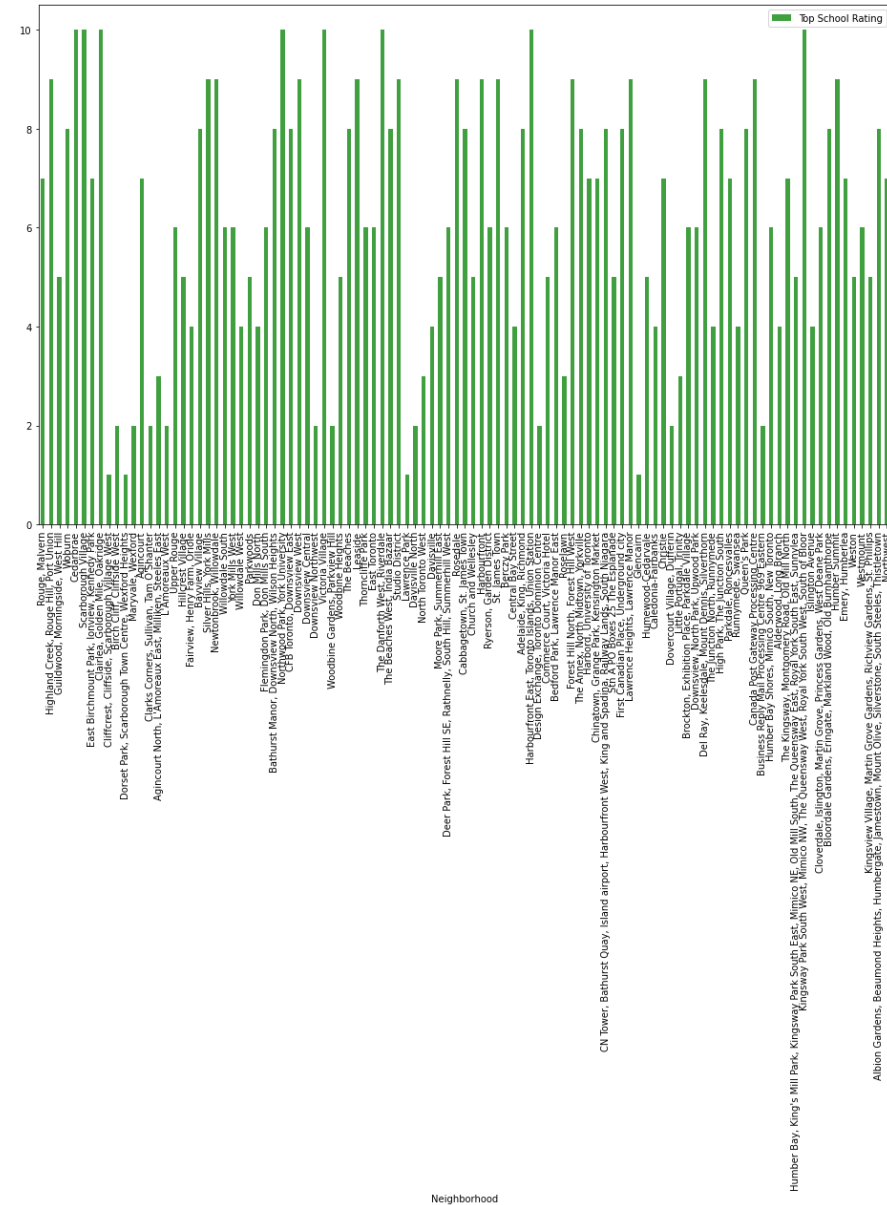
Foursquare API and Scarborough

Foursquare was chosen for this project as a data gathering source because it has a database on various points of interest, as well details on all of these venues. Whereas, the Scarborough location was chosen because of its ever increasing rate of immigration and thus; rising number of individuals which wish to learn more about the locale before choosing which neighbourhood to move to.

5. Discussion

The aim of this project was to provide a means in which an individual can compare neighbourhoods and actively choose the better neighbourhood for their circumstances, based on the points of interest and venues in the area such as: schools; coffee shops; hospitals; airports; housing and etc.

In turn; this project was able to provide a comparison of the average housing prices, as well as, the average school ratings across the neighbourhoods of Scarborough as a sample of how the better neighbourhood could then be determined.

6. Conclusion

During this project the K-means clustering algorithm was utilized to separate the neighbourhoods into 10 different clusters and for 103 different coordinates. Then, the average house pricing and school rating reviews were used for the charts above thus; allowing for the presentation of a specific neighbourhood based on preferences, while the Folium mapping tool allowed for a more summarised view of the tested neighbourhoods in Scarborough.

Overall, this project was quite enjoyable as it truly allowed me to understand the importance of data science in today's world and provided a means of using what was learnt from the Applied Data Science course.

In addition, in the future the project could be further improved by analysing specific houses and potentially other variables for better living arrangements hence; the best houses for an individual in Scarborough can be found, as well as other necessities for better and more cost effective living.

Libraries Used:

Pandas: For creating dataframes.

Folium: Python visualization library to visualize the neighbourhoods cluster distribution in map format.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Database.

Beautiful Soup and Requests: To scrap data to handle http requests.

Matplotlib: Python Plotting Module.