The Battle of Neighborhoods - Finding a Better Place in Scarborough, Toronto

Capstone Project

@mdr

Introduction:

- The purpose of this Project is to help people in exploring better facilities around their neighborhood
- 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 area, road connectivity, weather conditions, good management for emergency etc.
- 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

Data Selection

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

- Scarborough dataset which was scrapped from Wikipedia is used. Dataset consisting of latitude and longitude, zip codes.

- Foursquare is a location data provider with information about all manner of venues and events within an area of interest.

- 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 neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

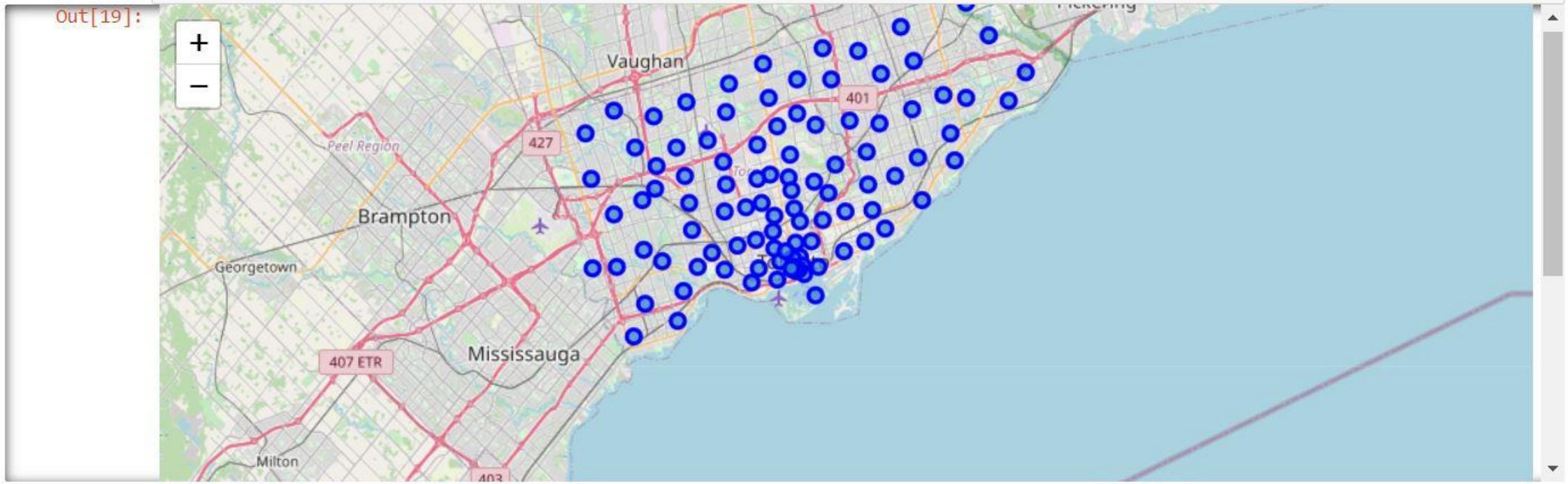
- The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

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

Map of Scarborough

```
map_Scarborough
```

Out[19]:



Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

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 Data.

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

Matplotlib: Python Plotting Module.

Methodology

- To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods.

- To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

- 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 500.

Using K-Means Clustering Approach

In [36]:
```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Scarborough_merged =df_2.iloc[:16,:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Scarborough_merged.head()# check the last columns!
```

Out[36]:

	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	7th Most Common Venue	8 C
0	M1B	Scarborough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	E Re
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	
3	M1G	Scarborough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Ele
4	M1H	Scarborough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Re

Most Common venues near Neighborhood

```
neighborhoods_venues_sorted.head()
```

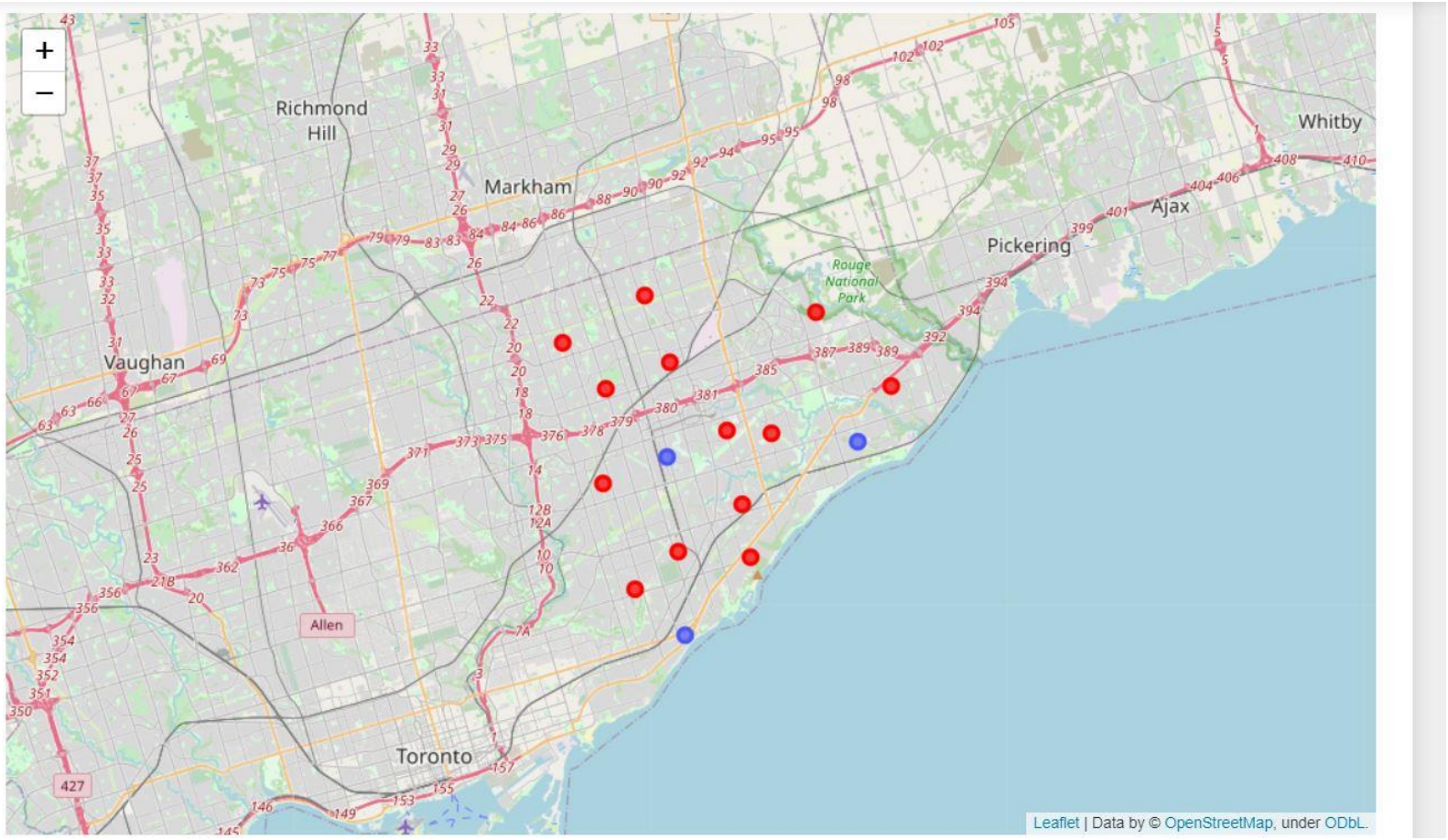
Out[34]:

	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	Café	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Steakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Print Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Sushi Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

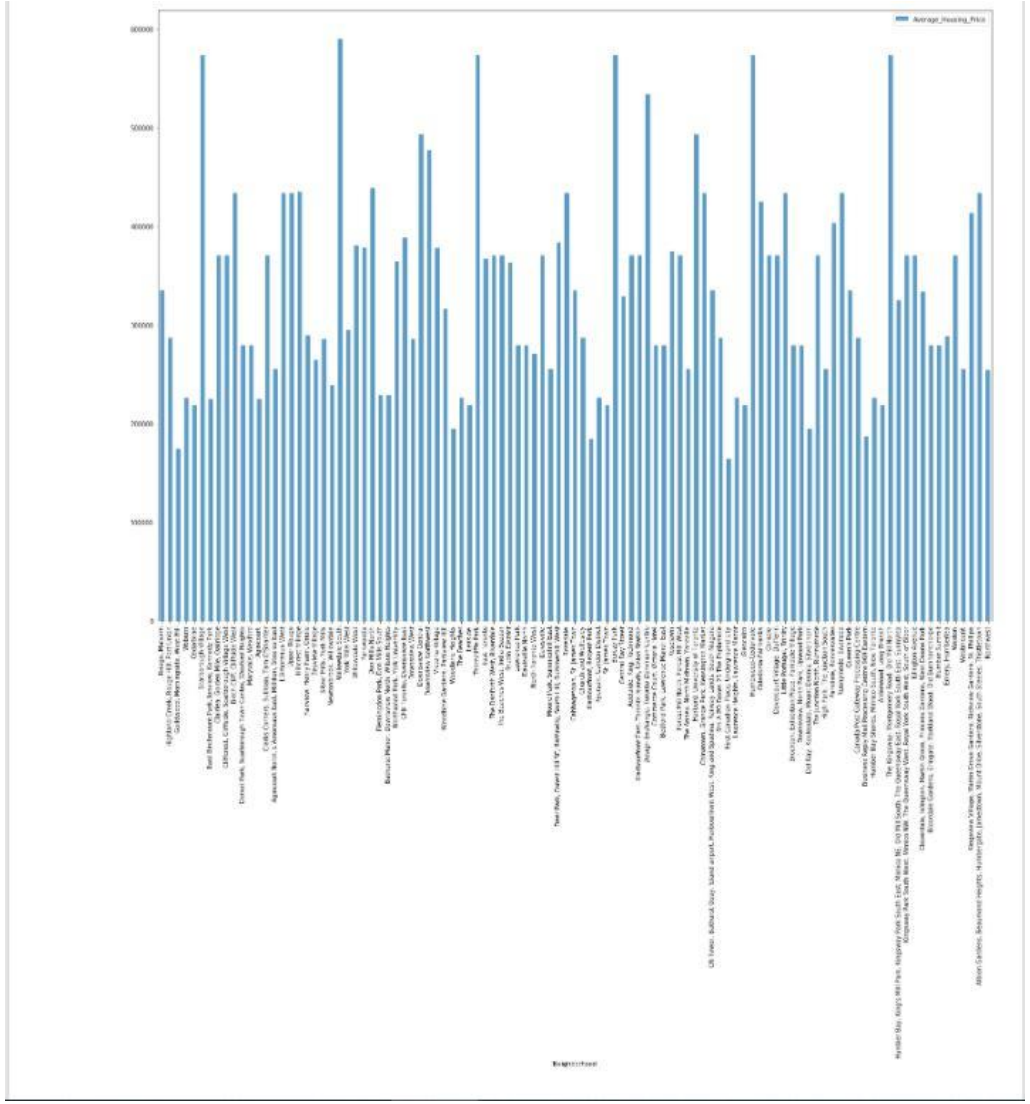
Results

Map of Clusters in Scarborough

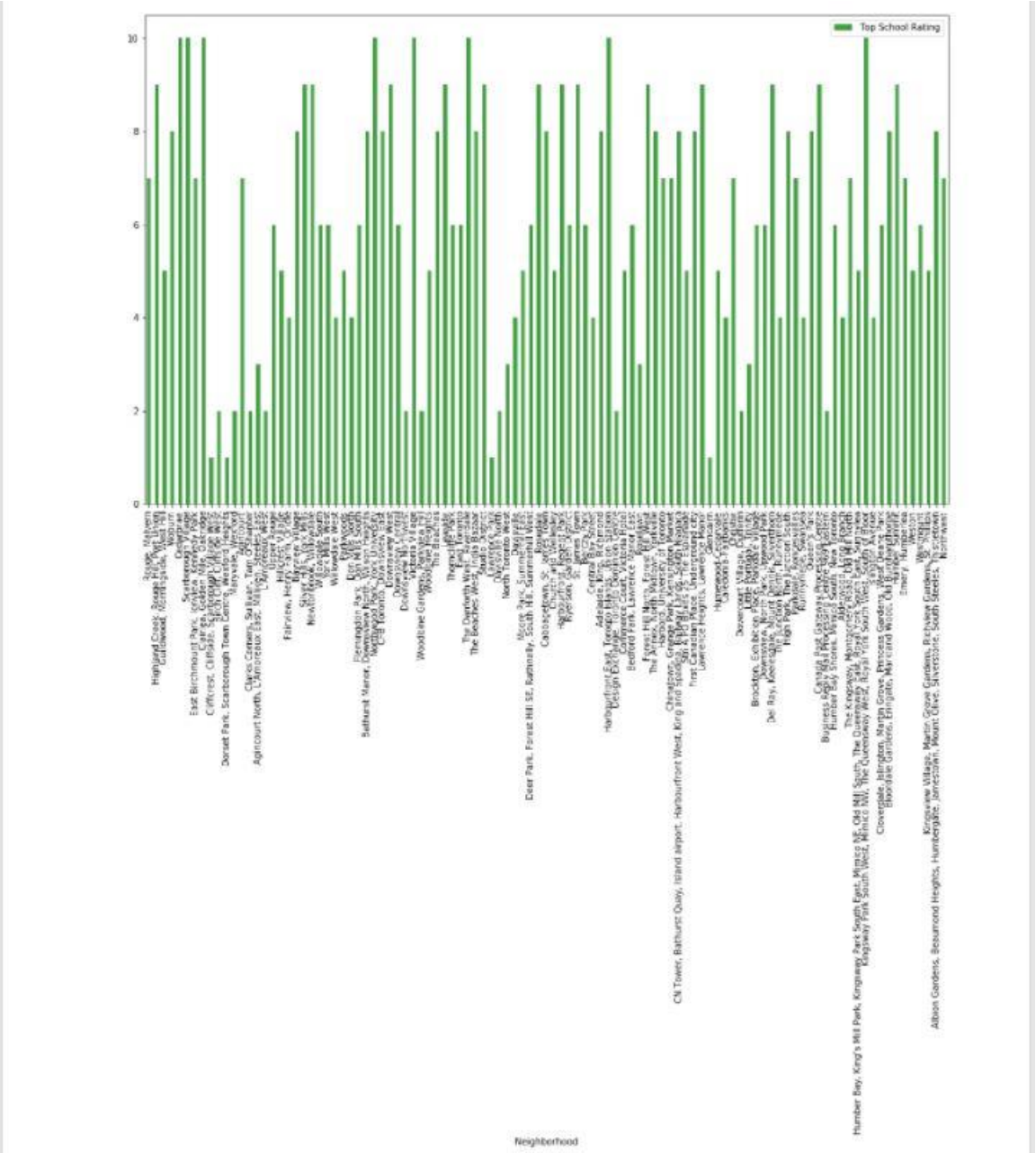
Out[38]:



Average Housing Price by Clusters in Scarborough



School Ratings by Clusters in Scarborough



Discussion:

- The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting to the neighborhoods considering various factors like 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, etc.

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

Conclusion

- Thus, in this project, using k-means cluster algorithm I separated the neighborhood into 10 different clusters and for 103 different latitude and longitude from dataset, which have very-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.

- The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

- This project can be continued for making it more precise in terms to find best house in Scarborough. Best means based on all required things(daily needs or things we need to live a better life) around and in terms of cost effective.

THANK YOU