IBM Applied Data Science Capstone Week 5

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods

Muhammad Bilal Tahir

Table of Contents:

1. Introduction:

The motivation behind this Capstone Venture is to assist individuals with bettering offices around their area. It will help individuals settling on brilliant and effective choice on choosing extraordinary neighborhood out of quantities of different neighborhoods in Scarborough, Toronto.

Bunches of individuals are moving to different conditions of Canada and required loads of exploration at great lodging costs and rumored schools for their youngsters. This venture is for those individuals who are searching for better areas. For simplicity of getting to Bistro, School, General store, clinical shops, and staple shops, shopping center, theater, emergency clinic, likeminded individuals, and so on

This Capstone Task intend to make an examination of highlights for a group moving to Scarborough to look through a best neighborhood as a near investigation between neighborhoods. The highlights incorporate middle lodging cost and better school as indicated by appraisals, crime percentages of that specific region, street availability, climate conditions, great administration for crisis, water assets both new and waste water and fecal matter passed on in sewers and sporting offices.

It will assist individuals with getting consciousness of the region and neighborhood prior to moving to another city, state, country or spot for their work or to begin another new life.

2. Data:

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

Will use Scarborough dataset which we scrapped from Wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:
We will require information about various settings in various neighborhoods of that particular precinct.

To acquire that data we will utilize "Foursquare" locational data. Foursquare is an area information supplier with data pretty much all way of scenes and occasions inside a region of interest. Such data incorporates setting names, areas, menus and even photographs. In that capacity, the foursquare area stage will be utilized as the sole information source since all the expressed required data can be acquired through the Programming interface.

Subsequent to finding the rundown of neighborhoods, we at that point interface with the Foursquare Programming interface to accumulate data about scenes inside every single area. For every area, we have picked the range to be 100 meter.

The information recovered from Foursquare contained data of scenes inside a predefined distance of the longitude and scope of the postcodes. The data acquired per scene 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:
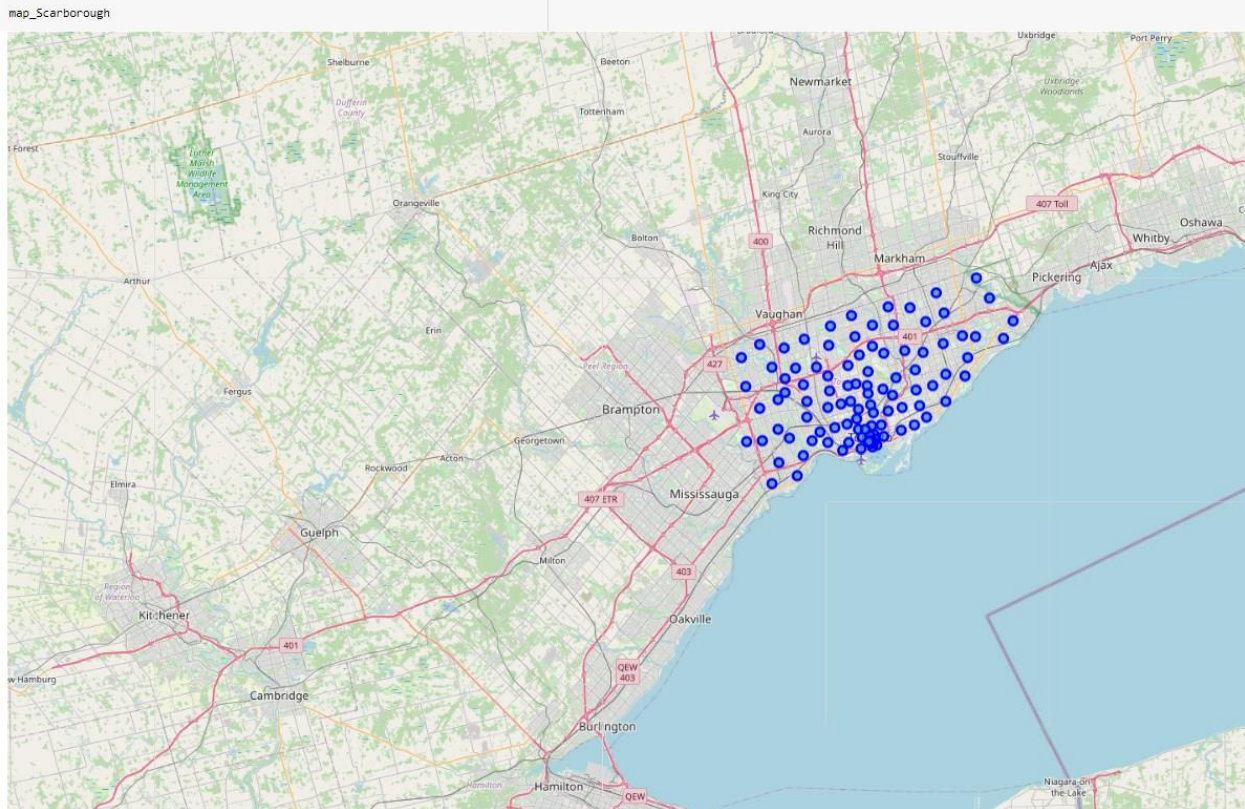
3. Methodology:

Clustering Approach:
To look at the likenesses of two urban communities, we chose to investigate neighborhoods, fragment them, and gathering them into bunches to discover comparative neighborhoods in a major city like New York and Toronto. To have the option to do that, we need to bunch information which is a type of solo AI: k-implies grouping calculation.

Using K-Means Clustering Approach | Most Common Venue:

K-Means Clustering Approach

```
[36] # Using K-Means to cluster neighborhood into 3 clusters
     Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
     kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
     kmeans.labels_

     array([1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
            1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1,
            1, 0, 1, 1, 1, 1, 1, 0, 1, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1,
            1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
            2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2], dtype=int32)
```

```
[37] 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!
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M1A\n	Not assigned\n	Not assigned\n	43.64869	-79.38544	1	Coffee Shop	Hotel	Café
1	M1B\n	Scarborough\n	Malvern, Rouge	43.81139	-79.19662	1	Zoo Exhibit	Fast Food Restaurant	Business Service
2	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	1	Fish & Chips Shop	Bar	Doner Restaurant
3	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.76575	-79.17470	2	Park	Gym / Fitness Center	Athletics & Sports
4	M1G\n	Scarborough\n	Woburn	43.76812	-79.21761	2	Park	Fast Food Restaurant	Chinese Restaurant

Map of Clusters

Most Common Venues near Neighborhood | Using Clustering:

Most Common venues near neighborhood

```
[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()
```
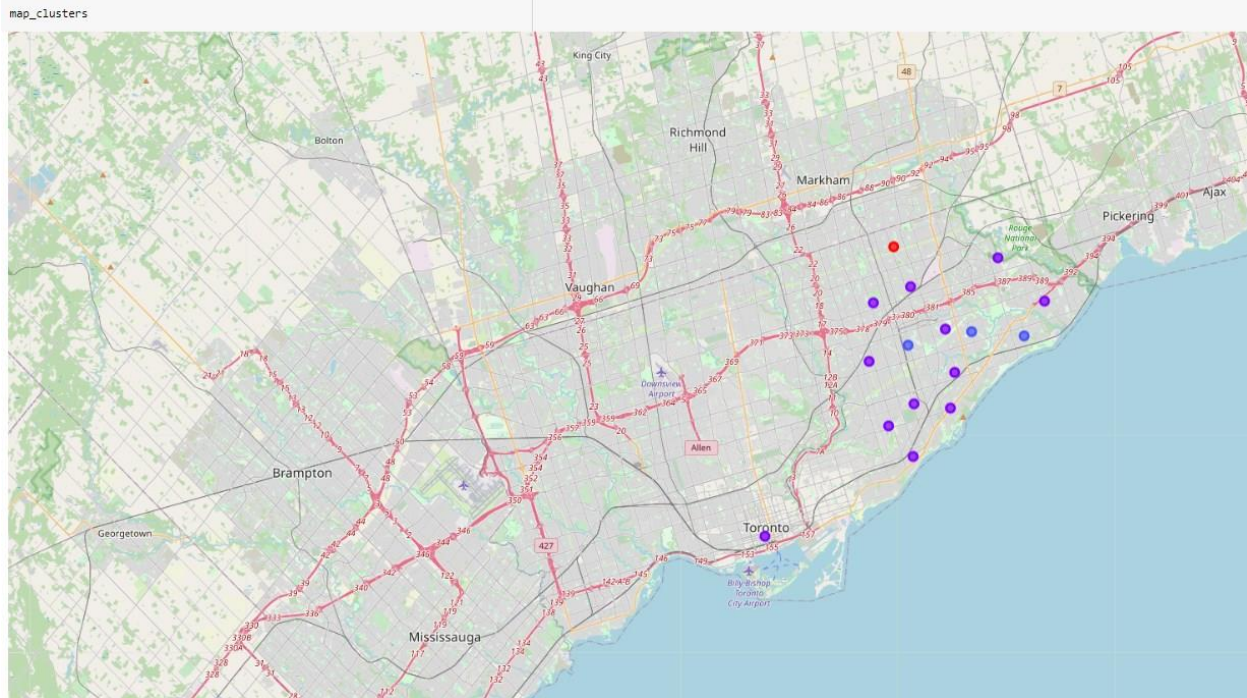
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Agincourt	Shopping Mall	Pizza Place	Clothing Store	Bubble Tea Shop
1	Alderwood, Long Branch	Pub	Gym	Convenience Store	Coffee Shop
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Pizza Place	Middle Eastern Restaurant	Restaurant
3	Bayview Village	Trail	Park	Gas Station	Asian Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Sandwich Place	Coffee Shop	Restaurant

Work Flow:

Utilizing qualifications of Foursquare Programming interface highlights of close by spots of the areas would be mined. Because of http demand limits the quantity of spots per neighborhood boundary would sensibly be set to 100 and the span boundary would be set to 500.
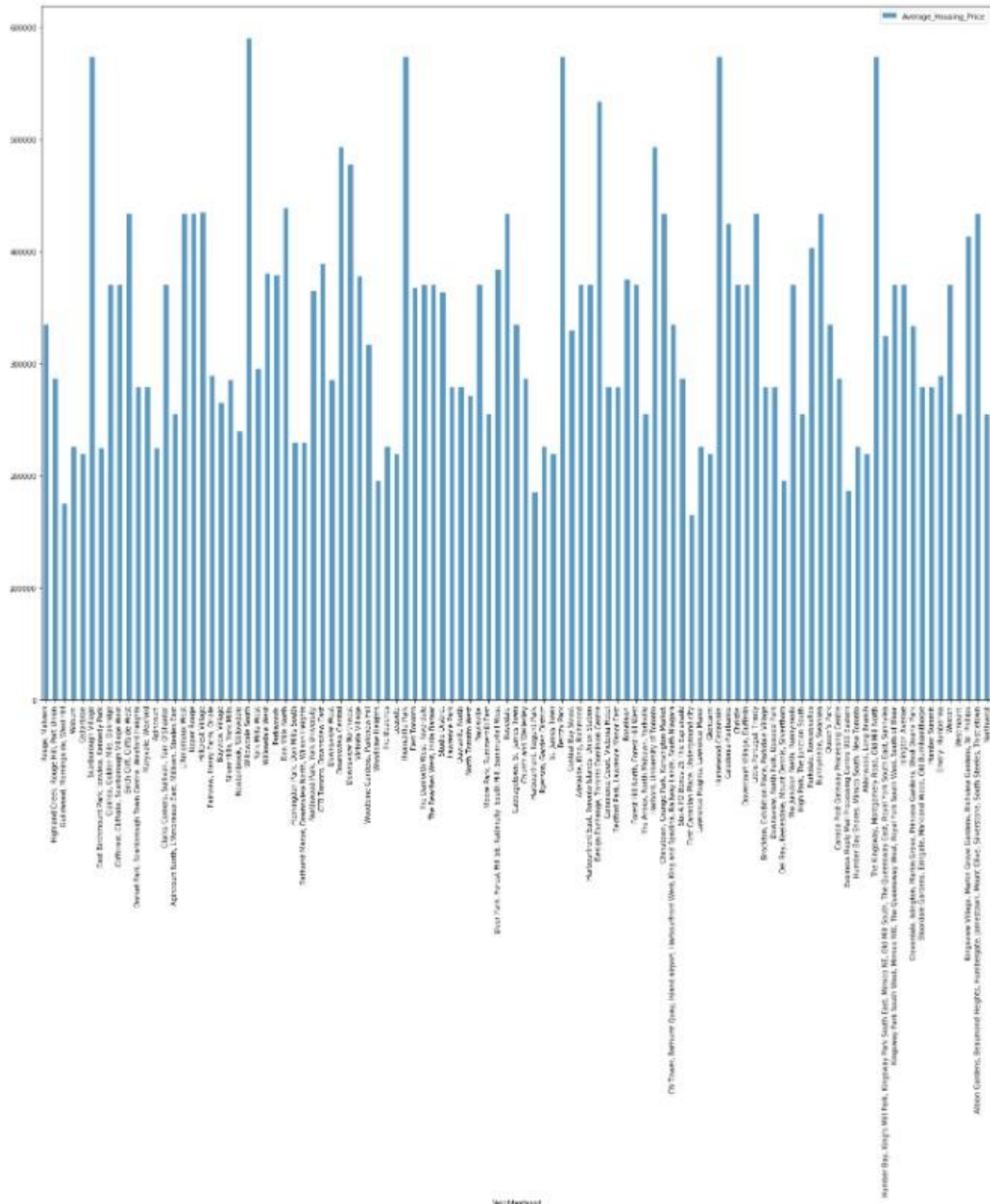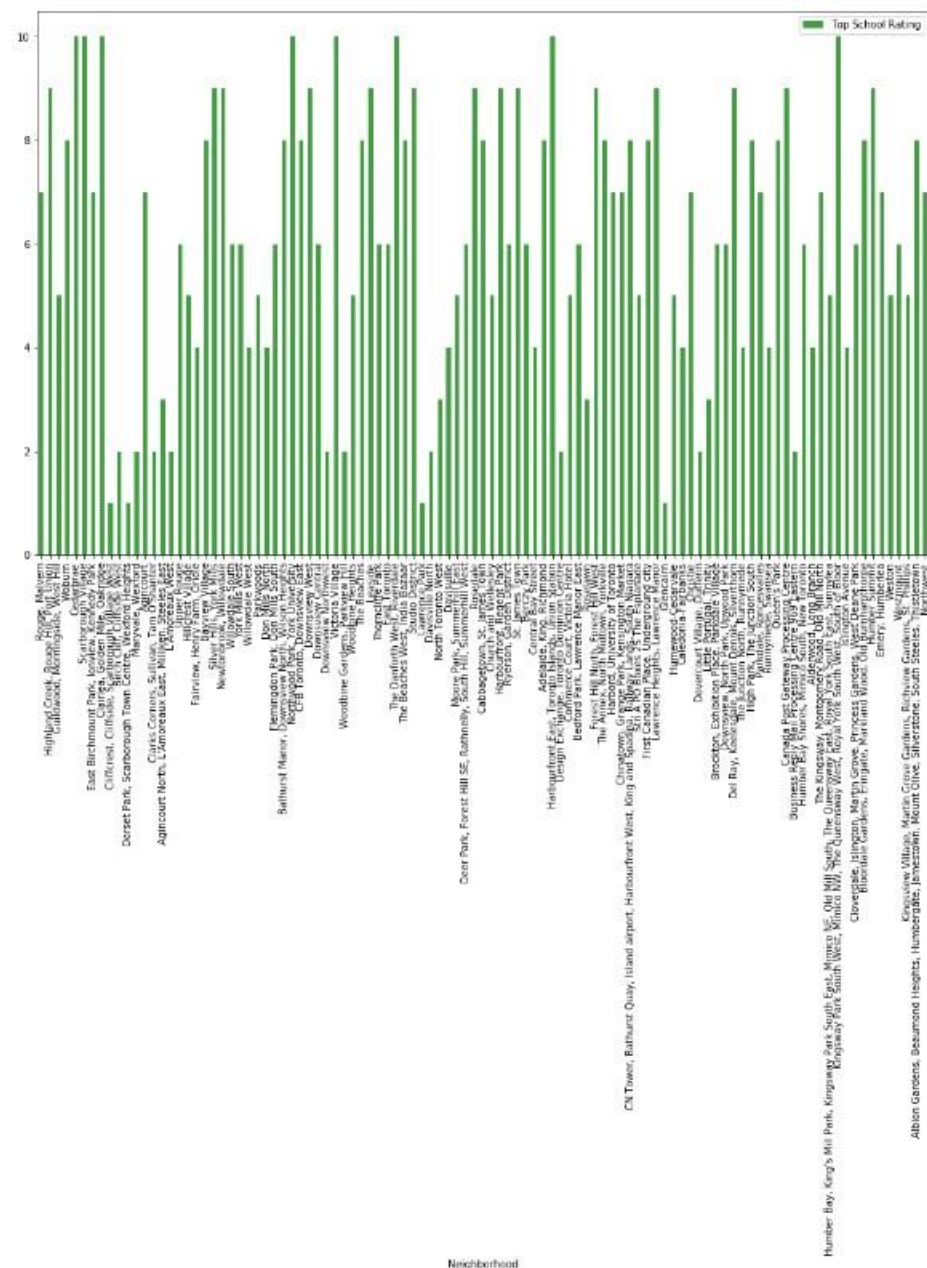
4. Results:

Map of Clusters in Scarborough:

Average Housing Price by Clusters in Scarborough:

School Ratings by Clusters in Scarborough:

The Location:
Scarborough is a famous objective for new workers in Canada to dwell. Thus, it is quite possibly the most different and multicultural territories in the More noteworthy Toronto Territory, being home to different strict gatherings and spots of love. In spite of the fact that movement has become an interesting issue in the course of recent years with more governments looking for additional limitations on migrants and exiles, the overall pattern of migration into Canada has been one of on the ascent.

Foursquare API:
This Capstone project have utilized Four-square Programming interface as its great information gathering source as it has a data set of millions of spots, particularly their places Programming interface which gives the capacity to perform area search, area sharing and insights concerning a business.

5. Discussion:

Problem to be solved:
The significant motivation behind this task, is to recommend a superior neighborhood in another city for the individual who are moving there. Social presence in the public arena regarding likeminded individuals. Network to the air terminal, transport stand, downtown area, markets and other every day needs things close by.

- Sorted rundown of house as far as lodging costs in a rising or sliding request
- Sorted rundown of schools as far as area, expenses, rating and surveys

6. Conclusion:

In this Capstone project, utilizing k-means cluster calculation I isolated the neighborhood into 10 various groups and for 180 distinctive scope and longitude from dataset, which have fundamentally the same as neighborhoods around them. Utilizing the outlines above outcomes introduced to a specific area dependent on normal house costs and school rating have been made.

I feel compensated with the endeavors and accept this course with all the subjects covered is well deserving of appreciation.

This task has shown me a pragmatic application to determine a genuine circumstance that has affecting individual and monetary effect utilizing Information Science devices.

The planning with Folium is an amazing procedure to solidify data and settle on the examination and choice better with certainty.

Libraries Which Are Used To Develop the Project:
Pandas: For creating and manipulating dataframes.

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.