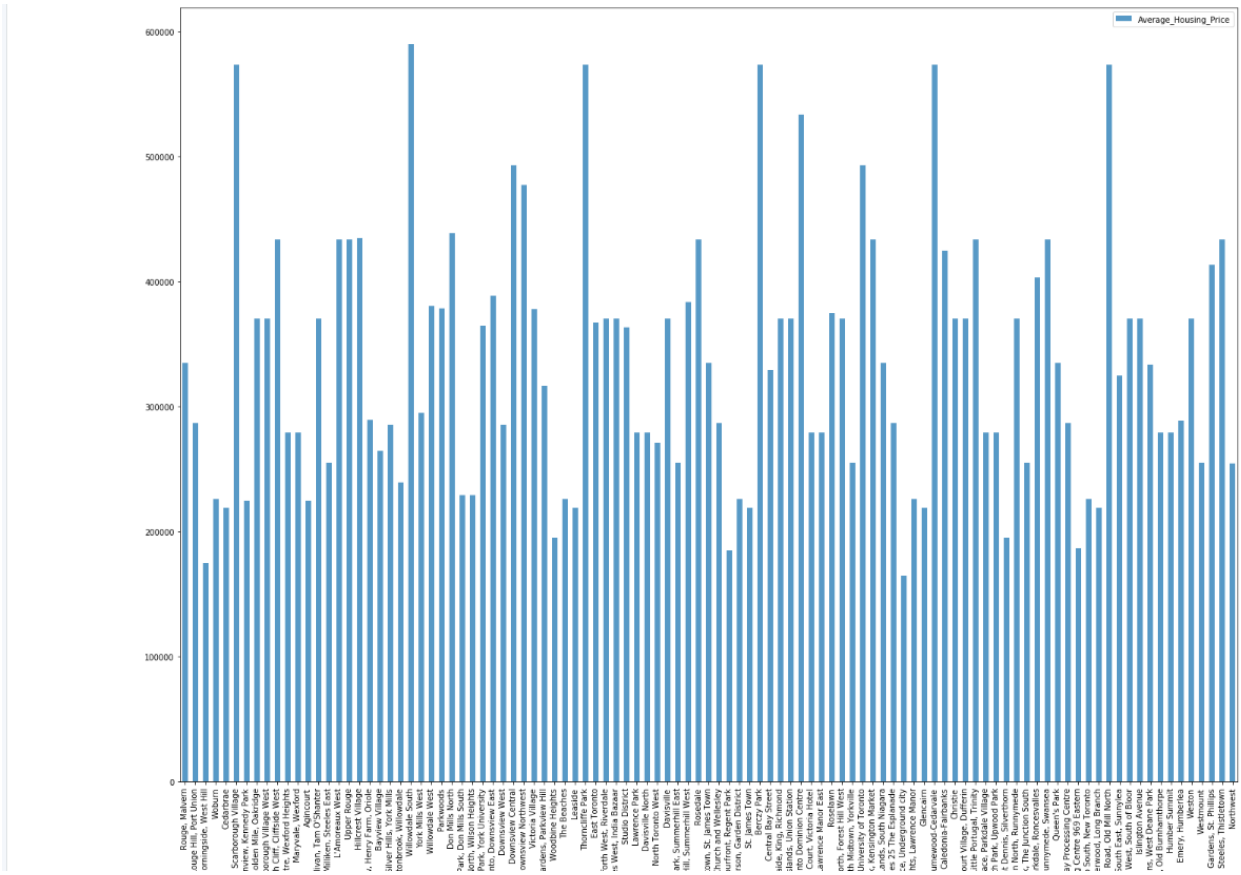


1. Introduction:

The reason of this Capstone Venture is to assist individuals in investigating way better facilities around their neighborhood. It'll offer assistance individuals making savvy and proficient choice on selecting awesome neighborhood out of numbers of other neighborhoods in Scarborough, Toranto.

Parcels of individuals are moving to different states of Canada and required parcels of inquire about for great lodging costs and reputated schools for their children. This venture is for those individuals who are searching for superior neighborhoods. For ease of getting to to Cafe, School, Super showcase, restorative shops, basic supply shops, shopping center, theater, healing center, like disapproved individuals, etc.

It'll help individuals to urge mindfulness of the range and neighborhood some time recently moving to a unused city, state, nation or put for their work or to begin a new life.

2. Data Section

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

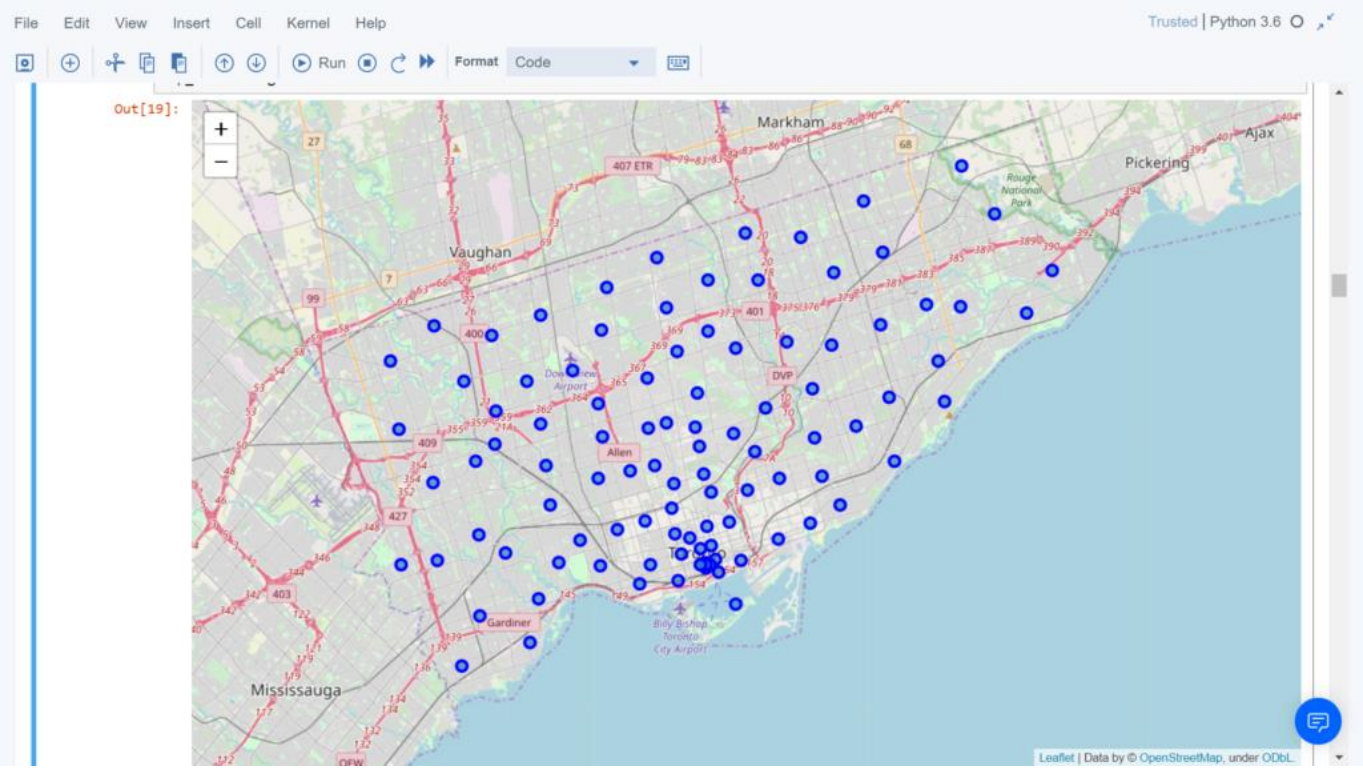
Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

We are going need information around diverse scenes totally different neighborhoods of that particular borough.

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

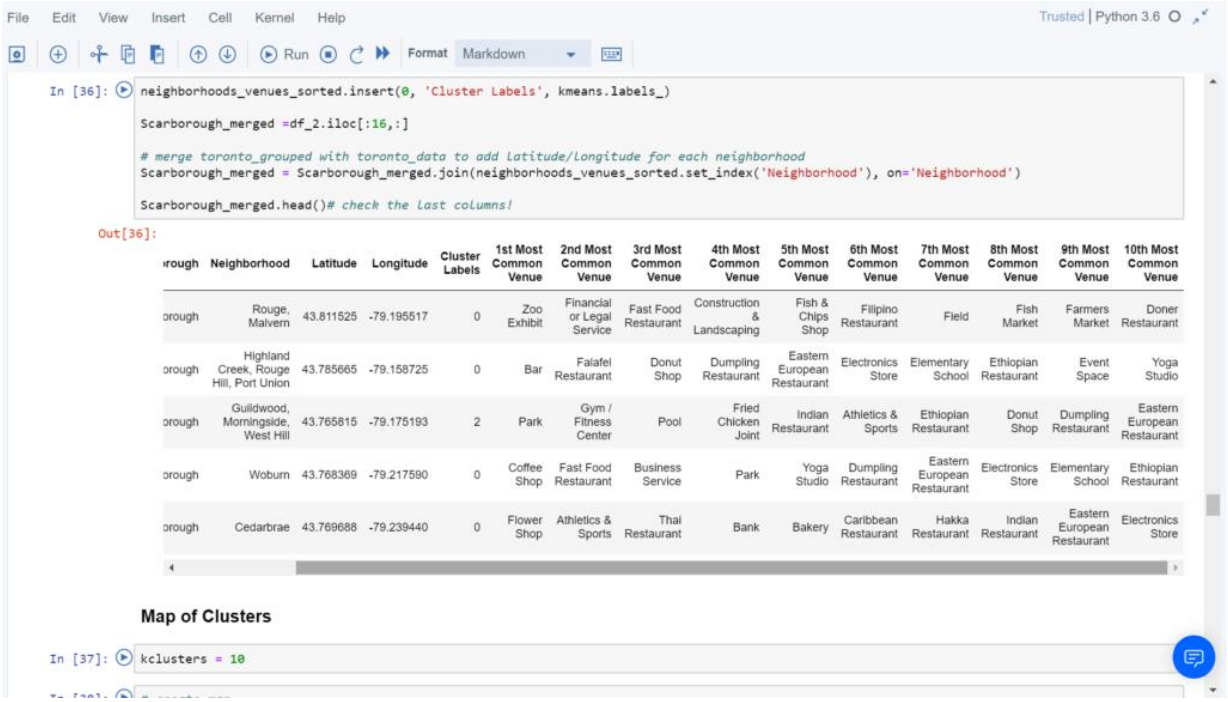


3. Methodology Section

Clustering Approach:

To compare the likenesses of two cities, we chosen to investigate neighborhoods, fragment them, and gather them into clusters to discover comparable neighborhoods in a enormous city like Modern York and Toronto. To be able to do that, we ought to cluster information which could be a frame of unsupervised machine learning: k-means clustering calculation.

Using K-Means Clustering Approach | Most Common Venue



```
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!
```

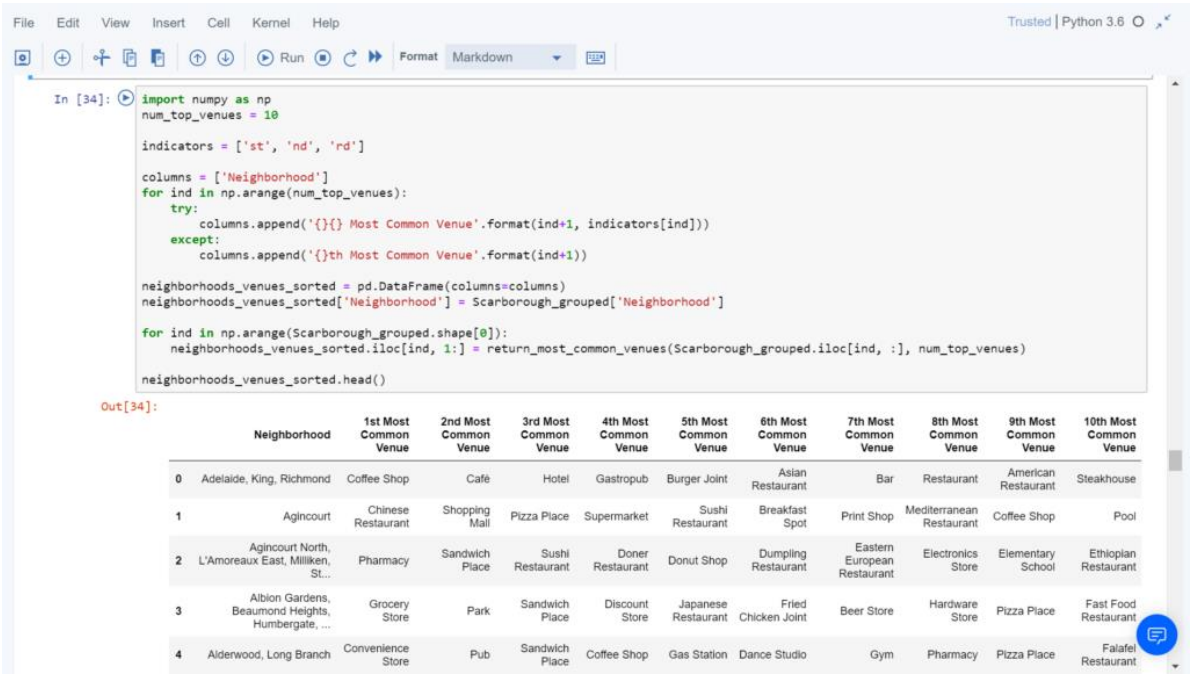
Out[36]:

rough	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
orough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Doner Restaurant
orough	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	Ethiopian Restaurant	Event Space	Yoga Studio
orough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
orough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
orough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

Map of Clusters

```
In [37]:
kclusters = 10
```

Most Common Venues near Neighborhood | Using Clustering



```
In [34]:
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[34]:

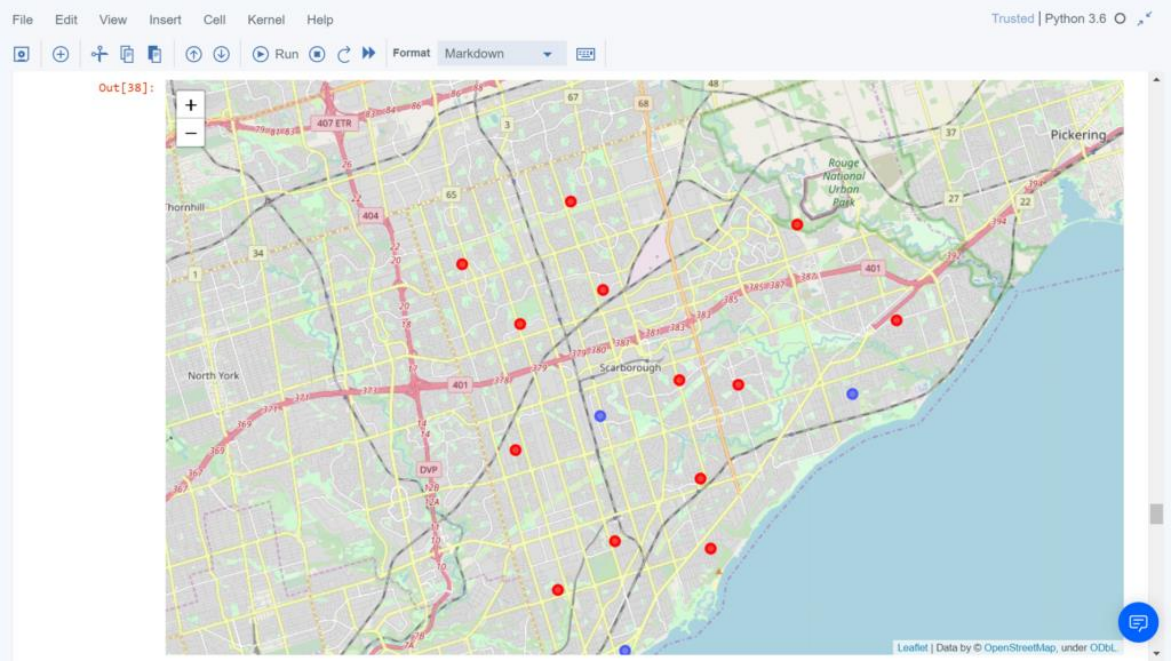
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger 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

Work Flow:

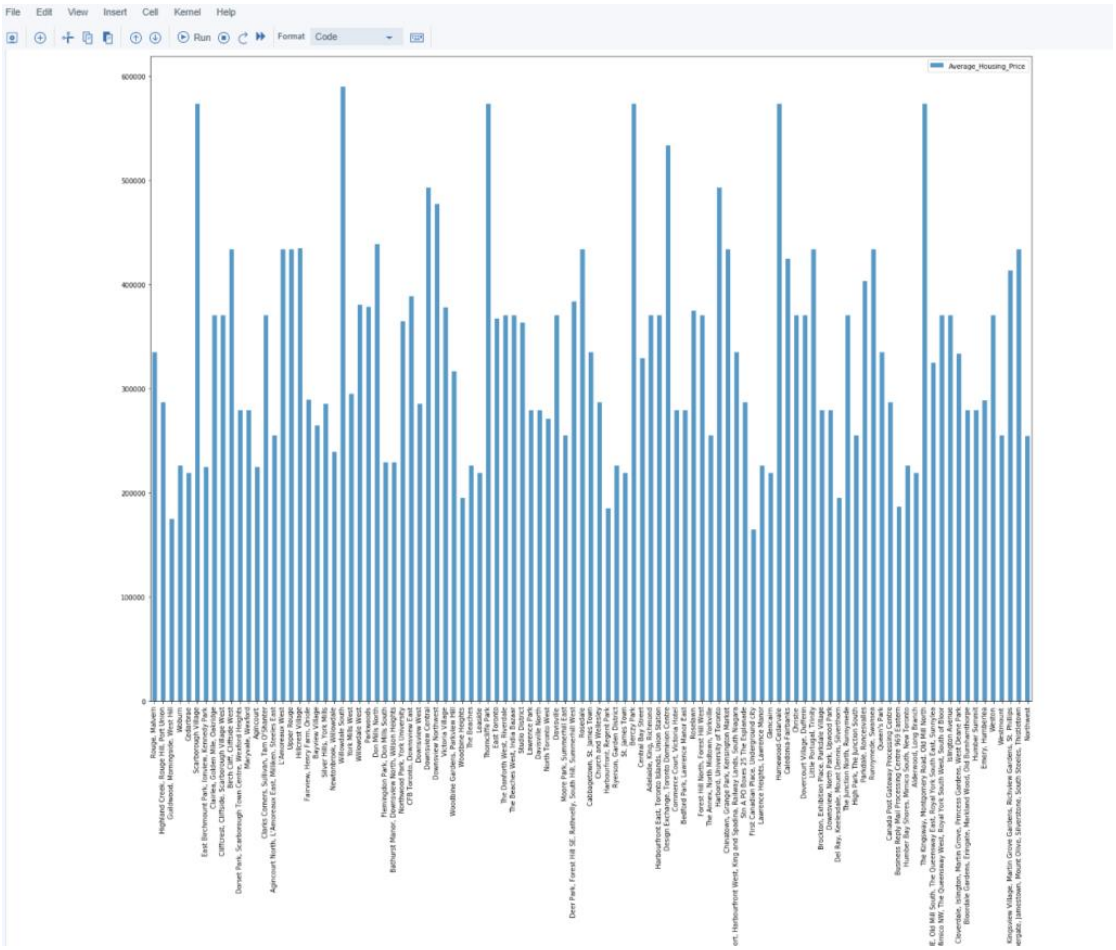
Utilizing qualifications of Foursquare API highlights of near-by places of the neighborhoods would be mined. Due to http ask impediments the number of places per neighborhood parameter would sensibly be set to 100 and the sweep parameter would be set to 500.

4. Results Section

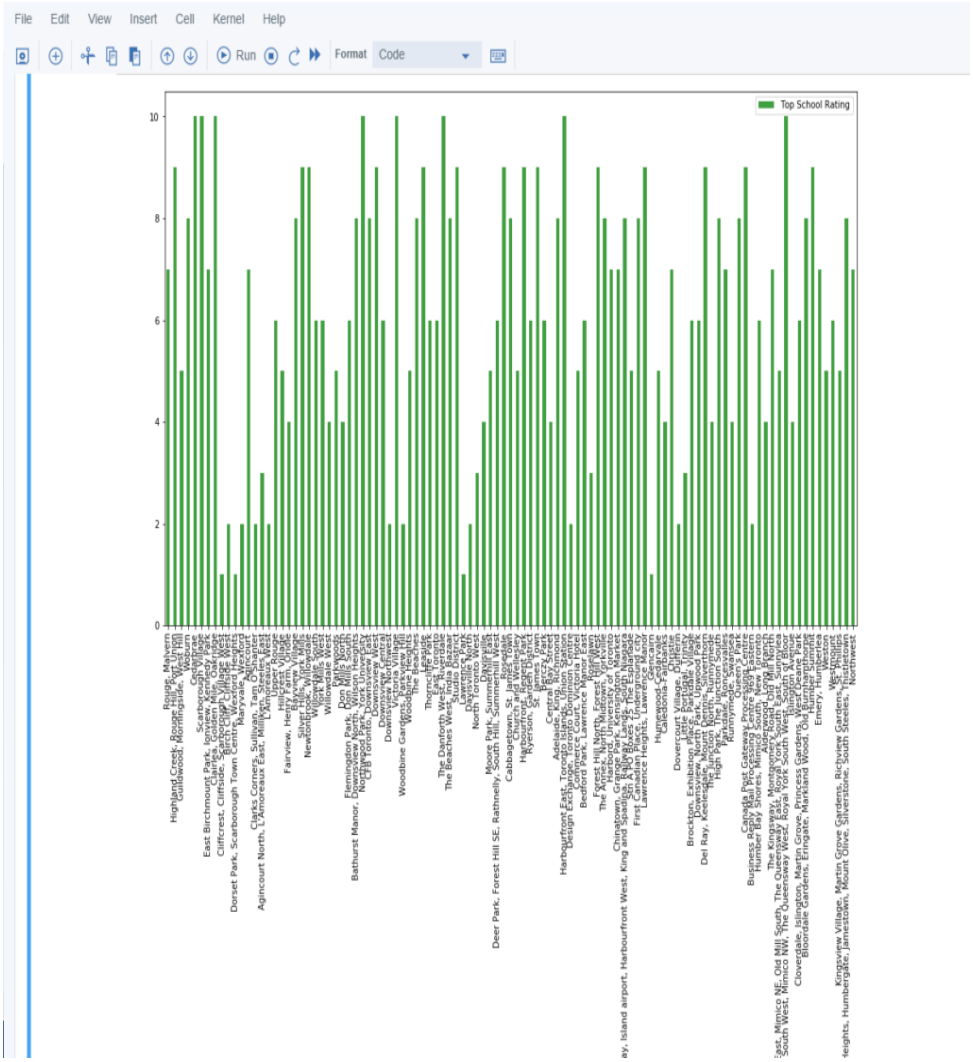
Map of Clusters in Scarborough



Average Housing Price by Clusters in Scarborough



School Ratings by Clusters in Scarborough



The Location:

Scarborough may be a prevalent goal for modern migrants in Canada to dwell. As a result, it is one of the foremost different and multicultural ranges within the More noteworthy Toronto Zone, being domestic to different devout bunches and places of revere. In spite of the fact that migration has ended up a hot point over the past few a long time with more governments looking for more limitations on migrants and outcasts, the common slant of migration into Canada has been one of on the rise.

Foursquare API:

This Capstone venture have utilized Four-square API as its prime information gathering source because it incorporates a database of millions of places, particularly their places API which gives the capacity to perform area look, area sharing and subtle elements approximately a commerce.

5. Discussion Section

Issue Which Attempted to Solve:

The major reason of this venture, is to propose a much better and an improved neighborhood in a unused city for the individual who are shifting there. Social nearness in society in terms of like disapproved individuals. Network to the airplane terminal, transport stand, city center, markets and other day by day needs things nearby.

6. Conclusion Section

In this Capstone venture, utilizing k-means cluster calculation I isolated the neighborhood into 10 distinctive clusters and for 103 diverse lattitude and logitude from dataset, which have very-similar neighborhoods around them.

Utilizing the charts over comes about displayed to a specific neighborhood based on normal house costs and school rating have been made. I feel compensated with the endeavors and accept this course with all the themes secured is well commendable of appreciation. This venture has appeared me a commonsense application to resolve a genuine circumstance that has affecting individual and monetary affect utilizing Information Science tools. The mapping with Folium could be a exceptionally capable procedure to solidify data and make the investigation and choice superior with confidence.

Future Works:

This Capstone extend can be proceeded for making it more exact in terms to discover best house in Scarborough. Best implies on the premise of all required things (daily needs or things we ought to live distant better around additionally in terms of fetched successful.)

Libraries

Which are Used to Develope the Project:

Pandas: For creating and manipulating dataframes.

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.

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

Matplotlib: Python Plotting Module.