REPORT ON CAPSTONE PROJECT –
THE BATTLE OF NEIGHBORHOODS-
RESOLVING ISSUE OF FINDING A BETTER PLACE IN SCARBOROUGH, TORONTO

1. INTRODUCTION:

The Objective of the project is to make it easy for people to explore better amenities, facilities and services in their neighborhood so as to make their life at a level of better comfort. The Project will pave way for the people to make a considerate and well planned decision in selecting their neighborhood among varied other neighborhoods in Scarborough, Toranto.

We observe that a large number of people individually and with family migrate to different states of Canada and need help in searching for good and affordable housing accommodations, availability of hospitals, public transport systems, recreational facilities and better school for their kids. The project is focused on all such individuals who are in search for a neighborhood to suit to their requirements and that fits their budget too.

The project will make it easy for the individuals to locate and access to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, places of their interest etcetera so that their time and resources are utilized optimally to the best of their satisfaction and enjoy life in their own way.

The Project emphasizes on analysis of features used in the study of data available on Wikipedia on a people migrating to Scarborough, Toranto in search of a better neighborhood as a comparative analysis among other available options of neighborhoods. Among other things features include availability of median housing price and better schools according to ratings, crime rates of that particular area, road connectivity, weather conditions, good emergency management, availability of fresh water resources and treatment of waste water and excrement conveyed in sewers alongwith recreational facilities and public transport.

The project will assist people get valuable information of their choice in the neighborhood and make them more aware about the area where they plan and intend to move for better life and prosperity.

2. DATA SECTION

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

Dataset consisting of latitude and longitude, zip codes about Scarborough, Toranto has been scrapped from the link of Wikipedia for the exclusive purpose of project and without any commercial considerations.

FOURSQUARE API DATA:

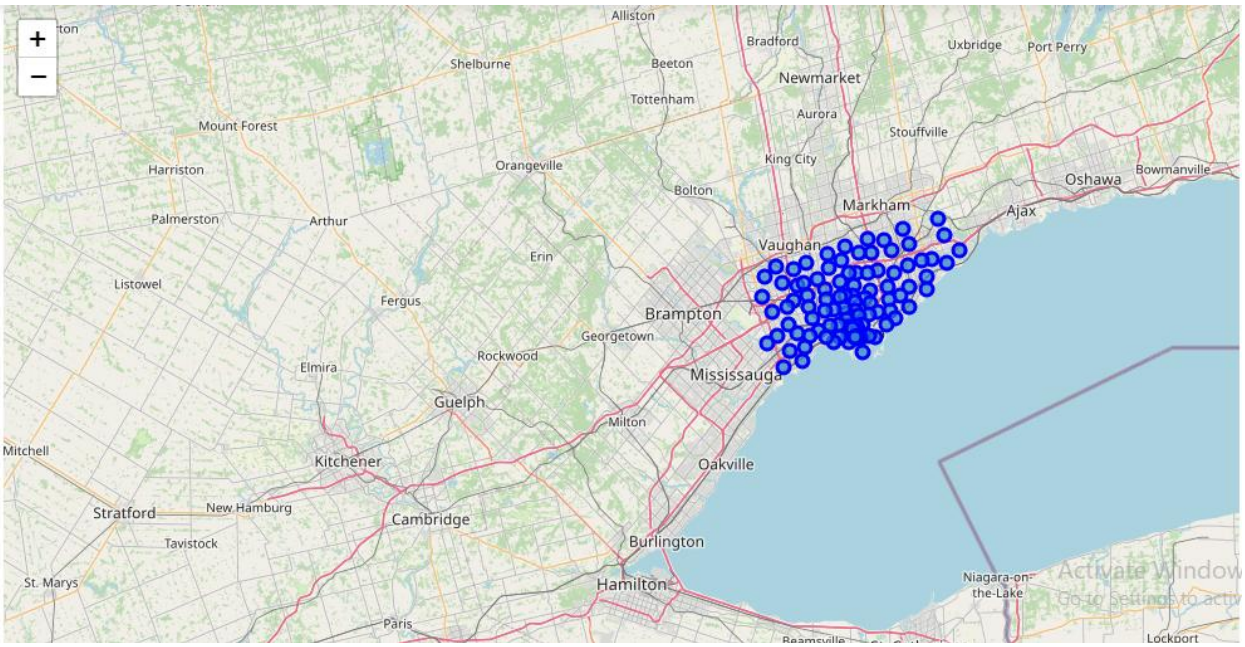
The Project uses Foursquare.com for the purpose of getting locational information as input. The project will need data about different venues in different neighborhoods of that specific borough. As informed in the instructional video of Coursera,I Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos etc. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be easily obtained through the API.

After finding the list of neighborhoods, the study then connect to the Foursquare API to procure information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter as default.

The data collected from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes (Zipcodes). The information obtained per venue is as on below mentioned points:

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 similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

USING K-MEANS CLUSTERING APPROACH

```
# 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([0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0])
```

MOST COMMON VENUES NEAR NEIGHBORHOOD

```
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!
```

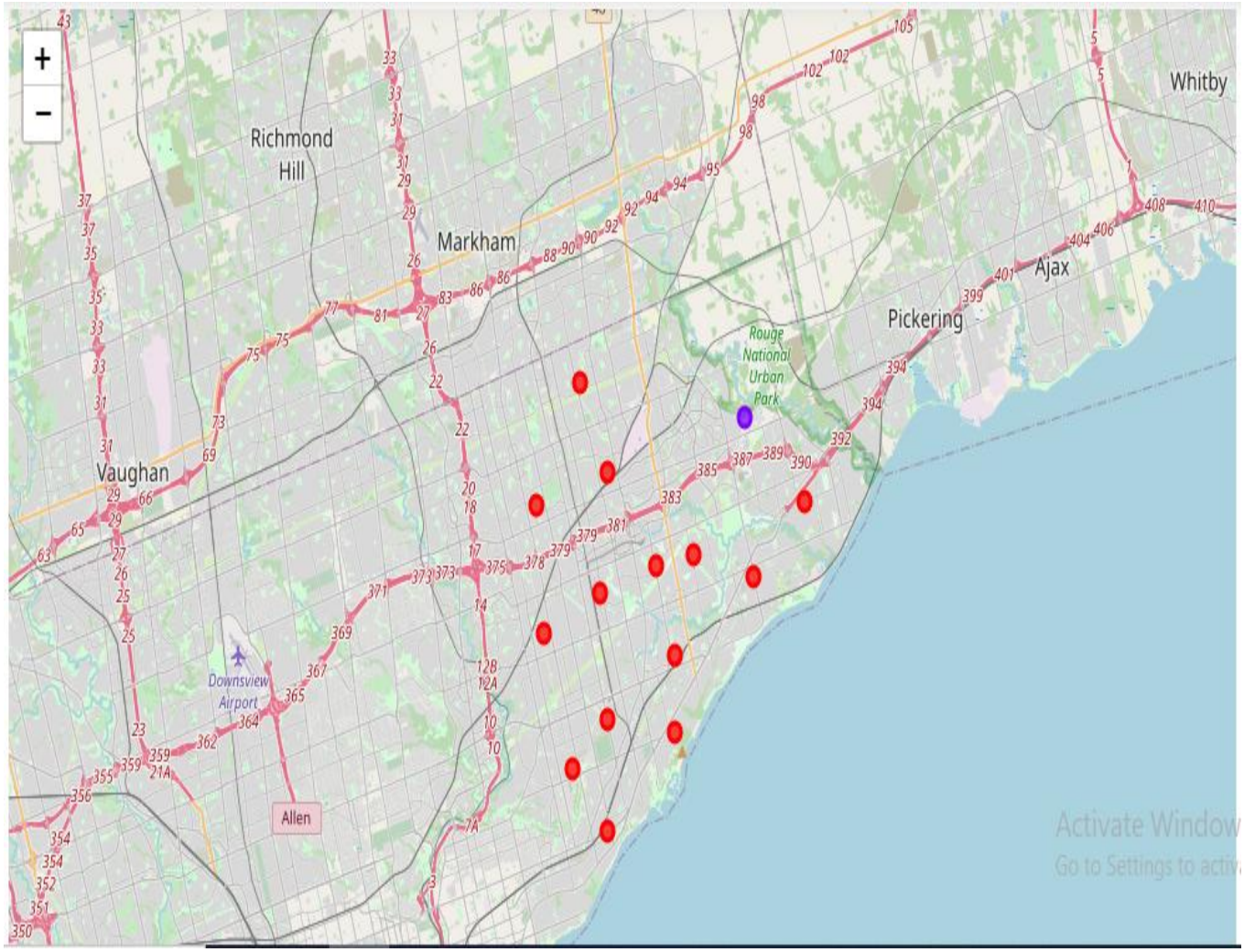
Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Not assigned\n	43.648690	-79.385440	0	Coffee Shop	Café	Hotel	Gym	Monument / Landmark	Tea Room	Movie Theater	Furniture / Home Store	Theater	Concert Hall
Malvern, Rouge	43.808626	-79.189913	1	Park	Trail	Women's Store	Electronics Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
Rouge Hill, Port Union, Highland Creek	43.785779	-79.157368	0	Moving Target	Park	Bar	Fish & Chips Shop	Filipino Restaurant	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
Guildwood, Morningside, West Hill	43.765806	-79.185284	0	Pizza Place	Coffee Shop	Park	Fast Food Restaurant	Greek Restaurant	Bank	Filipino Restaurant	Fried Chicken Joint	Discount Store	Supermarket
Woburn	43.771545	-79.218135	0	Coffee Shop	Park	Business Service	Dive Bar	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School

WORK FLOW:

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.

4. RESULTS SECTION

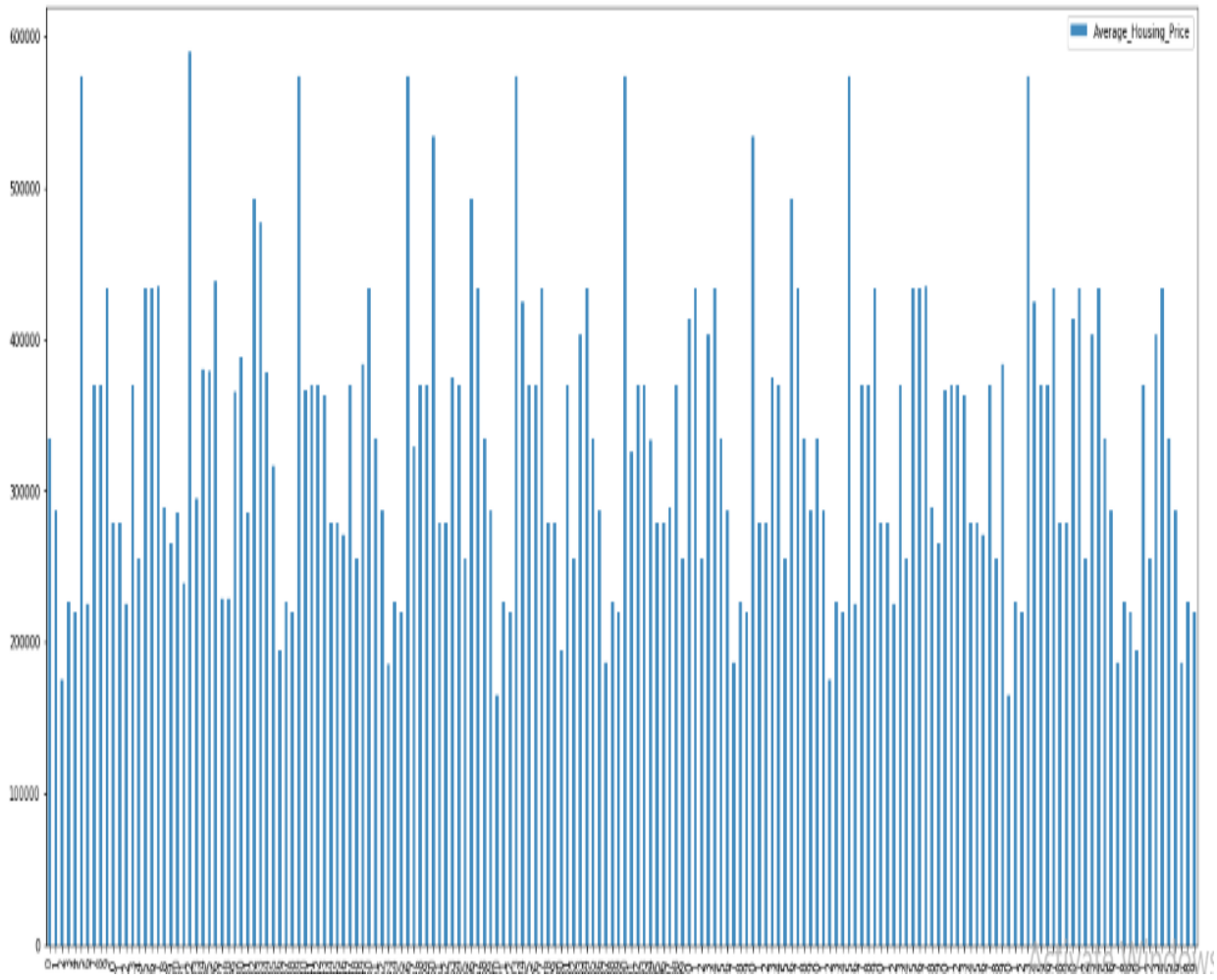
MAP OF CLUSTERS IN SCARBOROUGH



Average Housing Price by Clusters in Scarborough

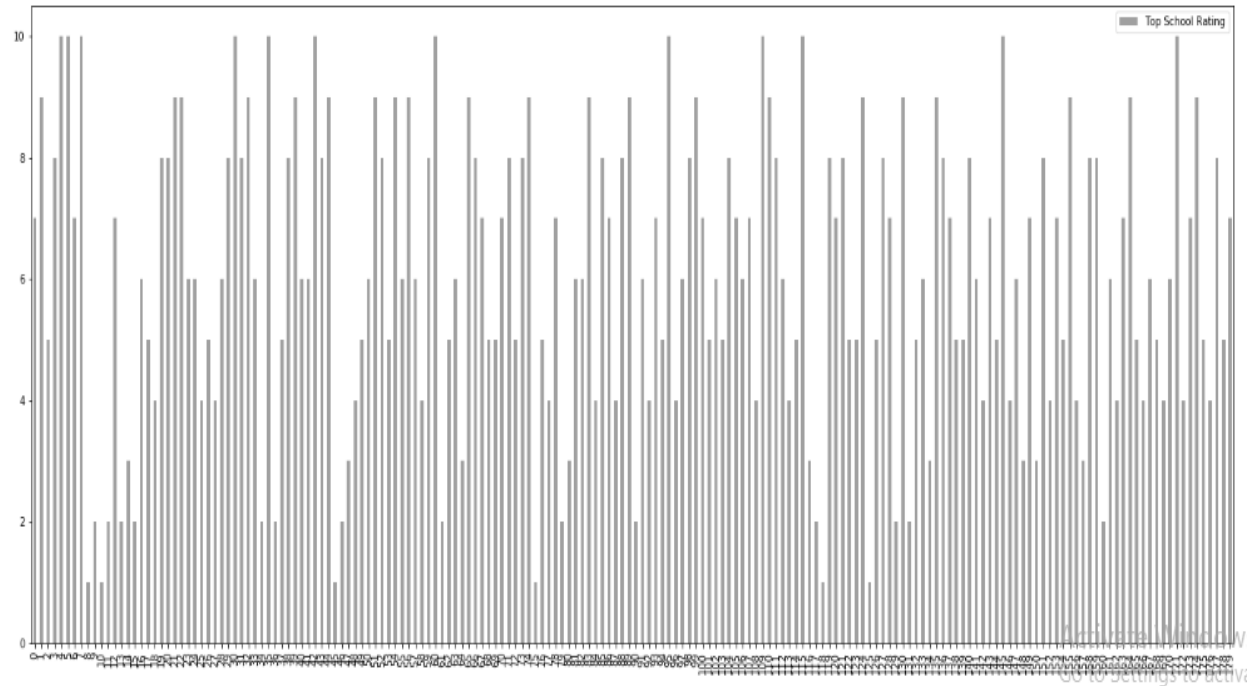
```
Scarborough_Avg_HousingPrice.plot(kind='bar',figsize=(25,10),alpha=0.85)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0xf331c90>
```



School Ratings by Clusters in Scarborough

```
Scarborough_school_ratings.plot(kind='bar',figsize=(24,10),color='grey',alpha=0.75);
```



The Location:

Scarborough is a paradise and most popular destination for new immigrants in Canada which has resulted into development of culturally diversified and economically prospering areas in the Greater Toranto Area as it is home for people belonging to various religious/cultural groups and following different ways of worship, due to which we see variety of places of worship.

Despite restrictions due to local issues and government concerning accommodating immigrants and refugees from different places, there is a upward trend and rise in immigration cases in search of better life and prospects of growth.

Foursquare API:

This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

Problem Addressed:

The Focus of this project, is to suggest a better neighborhood in a new city for the person who intends to move to Scarborough. Physical connectivity by Air, Rail and Road,Social presence in society in terms of like minded people, Market, Schools, medical facilities, grocery shops,recreational facilities etc have been taken into account.

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

6. Conclusion Section

In the project, using k-means cluster algorithm, learner has separated the neighborhood into 10(Ten) different clusters and for 103 different lattitude and logitude from dataset scrapped from wikipedia, which have very-similar neighborhoods around them. Using the charts, above results have been presented to a particular neighborhood based on average house prices and school rating.

Data Science being a very interesting field and requires passion to pursue it. Learner has made all out efforts to use to the best material provided during the course and available in public domain for analysis, interpretation and presentation. Data Science tools used in the project like matplotlib and Folium and website Foursquare.com have indeed made the task accomplishable.

Though I have attempted the project as per course fulfillment objective but trust that being Banker I would make good use of this skill for the betterment of data analysis for my academic purpose and become more useful for the organization I work in.

Future Works:

I would like to undertake a project using matplotlib, Folium and some statistical tools on financial inclusion and impact of thereof on women empowerment. I also intend to undertake a project which will focus on giving early warning signals to executives of Bank to arrest menace of increasing NPA.

Libraries Which are Used to Develope the Project:
Matplotlib: Python Plotting Module
Pandas: For creating and manipulating dataframes.
Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
Beautiful Soup and Requests: To scrap and library to handle http requests.
Geocoder: To retrieve Location Data.
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.
Website- Wikipedia and Foursqure
.

Thanks IBM, Thanks Coursera for wonderful learning experience.

----------END OF REPORT----------