

CAPSTONE PROJECT

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

INTRODUCTION

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Scarborough, Toronto.

DATA SECTION

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

Scarborough dataset which scrapped from Wikipedia

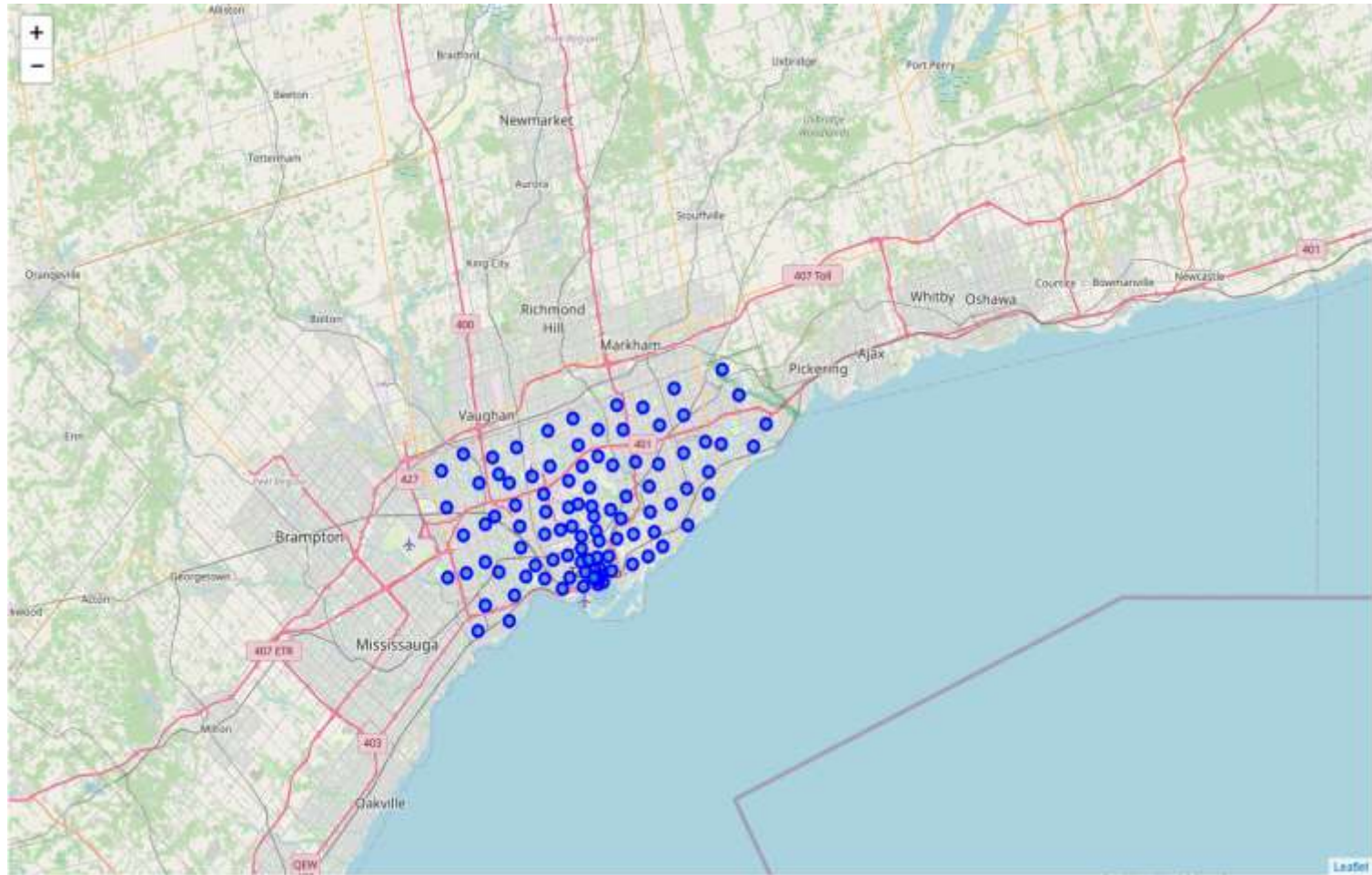
Dataset consisting of latitude and longitude, zip codes

FOURSQUARE API DATA

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



METHODOLOGY

K-Means Clustering Approach

```
[92]: 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!
```

[92]:	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
0	M1B\n	Scarborough\n	Malvern, Rouge	43.81153	-79.19552	1	Zoo Exhibit	Fast Food Restaurant	Paintball Field	Construction & Landscaping	History Museum	Ethiopian Restaurant
1	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	1	Fish & Chips Shop	Bar	Yoga Studio	Falafel Restaurant	Dumpling Restaurant	Eastern European Restaurant
2	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.76575	-79.17520	2	Park	Athletics & Sports	Gym / Fitness Center	Yoga Studio	Ethiopian Restaurant	Donut Shop
3	M1G\n	Scarborough\n	Woburn	43.76820	-79.21761	1	Chinese Restaurant	Fast Food Restaurant	Park	Coffee Shop	Yoga Studio	Event Space
4	M1H\n	Scarborough\n	Cedarbrae	43.76969	-79.23944	1	Flower Shop	Indian Restaurant	Caribbean Restaurant	Thai Restaurant	Gas Station	Athletics & Sports

MOST COMMON VENUES NEAR NEIGHBOURHOOD

```
[66]: 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()
```

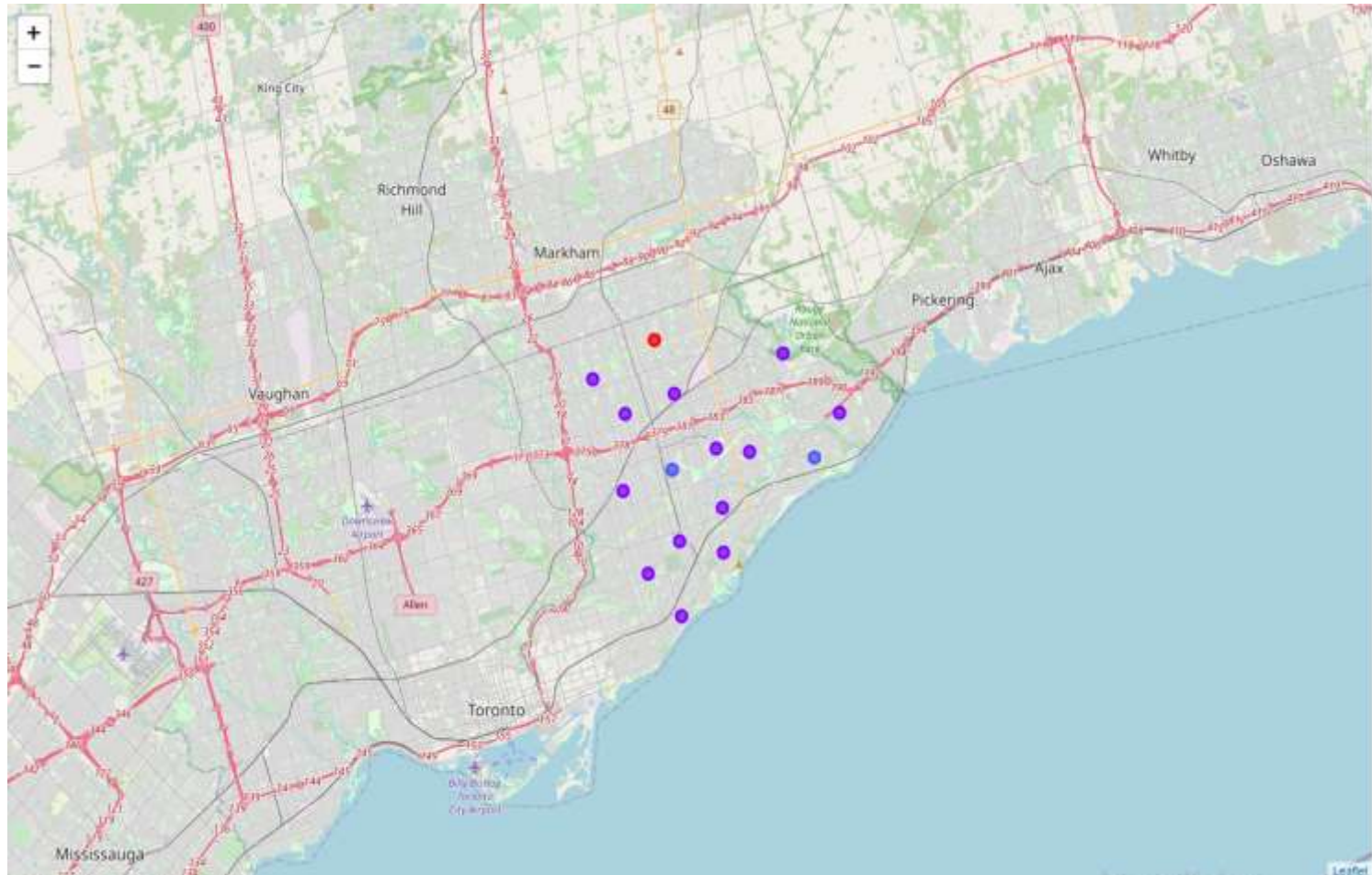
[66]:

	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
0	Agincourt	Chinese Restaurant	Pizza Place	Shopping Mall	Discount Store	Bakery	Bank	Sushi Restaurant
1	Alderwood, Long Branch	Pizza Place	Gas Station	Pool	Pharmacy	Pub	Sandwich Place	Coffee Shop
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Convenience Store	Other Great Outdoors	Yoga Studio	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant
3	Bayview Village	Asian Restaurant	Park	Trail	Yoga Studio	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Pizza Place	Restaurant	Sandwich Place	Pharmacy	Sports Club

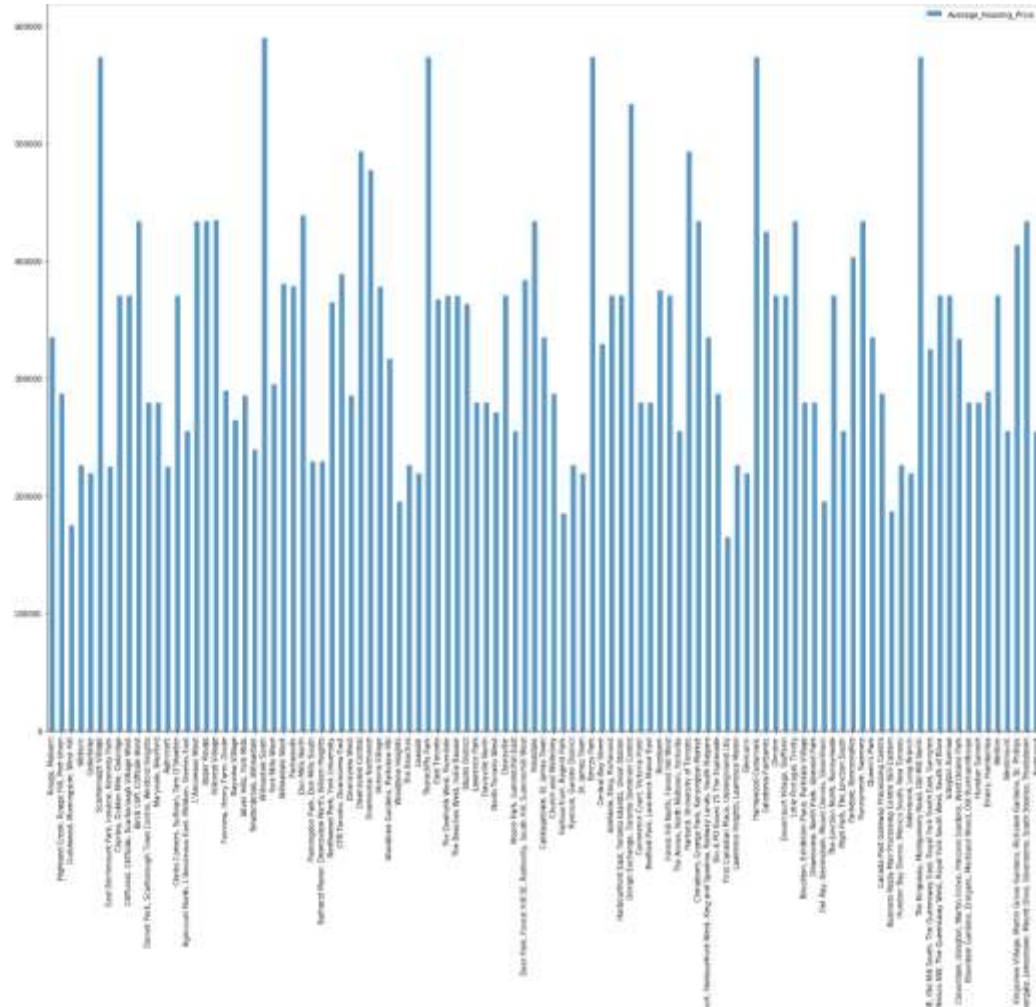
WORK FLOW

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.

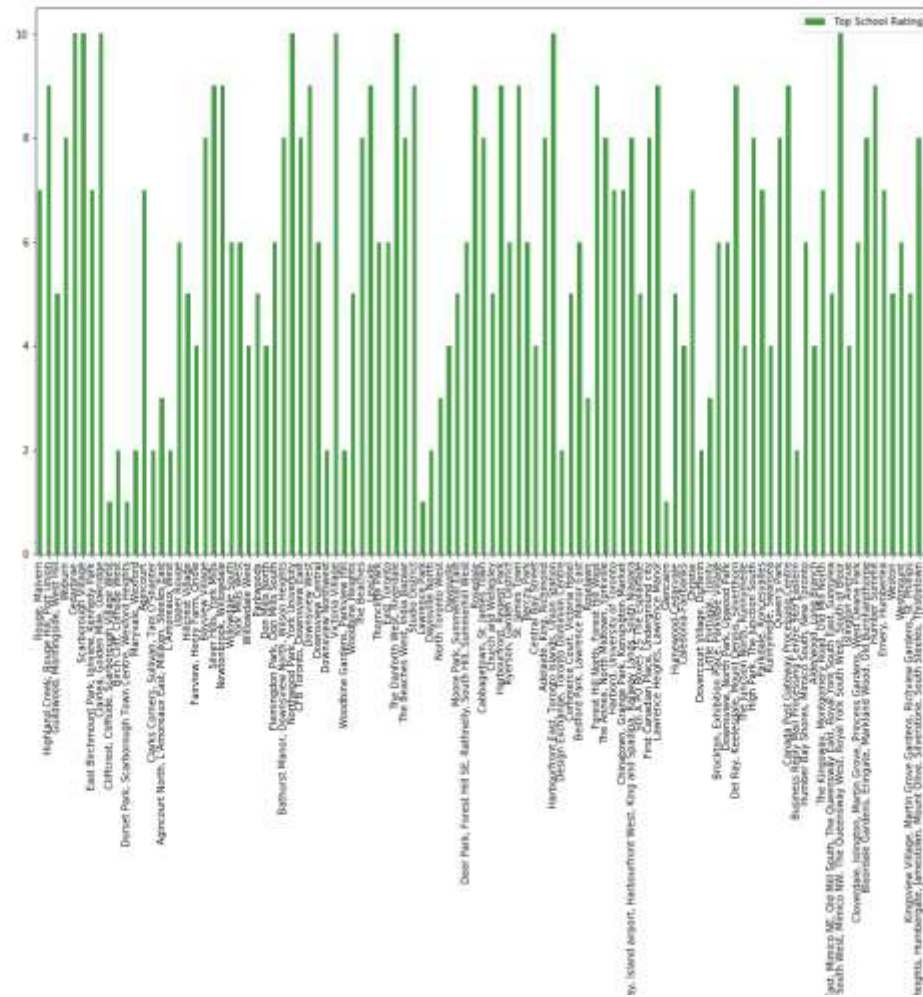
MAP OF CLUSTERS IN SCARBOROUGH



AVERAGE HOUSING PRICE BY CLUSTER IN SCARBOROUGH



SCHOOL RATINGS BY CLUSTERS IN SCARBOROUGH



LOCATION

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.

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.

PROBLEM WHICH TRIED TO SOLVE

The major purpose of this project is to suggest a better neighborhood in a new city for the person who is shifting there. Social presence in society in terms of likeminded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

Sorted list of house in terms of housing prices in a ascending or descending order

Sorted list of schools in terms of location, fees, rating and reviews

LIBRARIES USED

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.

CONCLUSION

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) 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 ratings have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worth of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.