

A dark blue vertical bar is on the left. A blue arrow points right from it, containing the date.

2/16/2021

Capstone Project

The Battle of Neighborhoods

Selecting a Better Place

in Scarborough, Toronto

Several thin, curved lines in dark blue and light grey originate from the bottom left and sweep upwards and to the right.

Ravi Prakash Gupta

rabiprakas@gmail.com

1. Introduction:

This Capstone Project is mainly concerned to help people in exploring better amenities around their neighborhood. It will assist people in smart and efficient decision making on selecting neighborhood in Scarborough, Toronto.

People who are looking for better places to live in and want to have easy access to transportation, schools, colleges, supermarket, medical shops, hospitals, grocery, fitness centers, theater, amusement park and society of people belonging to same group.

Every year lots of people are migrating to various states of Canada. So, they need lot of research for good housing prices and good school for their children.

This Capstone 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 particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

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.

2. Data Extraction Section

Data was extracted from the Wikipedia. Scarborough dataset was used which we scrapped from Wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes. The link is given below:

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

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough.

In order to gain that information we will use “Foursquare” locational information. 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. 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 and every 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



3. Methodology Section

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods by segmenting and grouping 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 | Most Common Venue

```
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	4th Most Common Venue	5th Most Common Venue
0	M1B	Scarborough	Malvern, Rouge	43.81139	-79.19662	1	Zoo Exhibit	Fast Food Restaurant	Construction & Landscaping	Financial or Legal Service	Health & Beauty Service
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	2	Fish & Chips Shop	Bar	Electronics Store	Dog Run	Doner Restaurant
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	0	Park	Gymnastics Gym	Athletics & Sports	Gym / Fitness Center	Yoga Studio
3	M1G	Scarborough	Woburn	43.76812	-79.21761	0	Fast Food Restaurant	Park	Coffee Shop	Chinese Restaurant	Yoga Studio
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892	1	Gas Station	Lounge	Bakery	Bank	Caribbean Restaurant

Most Common Venues near Neighborhood | Using Clustering

```
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	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	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Breakfast Spot	Skating Rink	Latin American Restaurant	Supermarket	Sushi Restaurant	Sandwich Place	Discount Store
1	Alderwood, Long Branch	Gas Station	Coffee Shop	Sandwich Place	Pizza Place	Gym	Convenience Store	Pub	Print Shop	Ethiopian Restaurant	Doner Restaurant
2	Bathurst Manor, Wilson Heights, Downsview North	Pizza Place	Coffee Shop	Men's Store	Deli / Bodega	Fried Chicken Joint	Middle Eastern Restaurant	Mobile Phone Shop	Mediterranean Restaurant	Sandwich Place	Restaurant
3	Bayview Village	Dog Run	Park	Flower Shop	Trail	Asian Restaurant	Gas Station	Yoga Studio	Dry Cleaner	Dive Bar	Doner Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Thai Restaurant	Coffee Shop	Sandwich Place	Boutique	Juice Bar	Sushi Restaurant	Liquor Store	Sports Club	Pizza Place

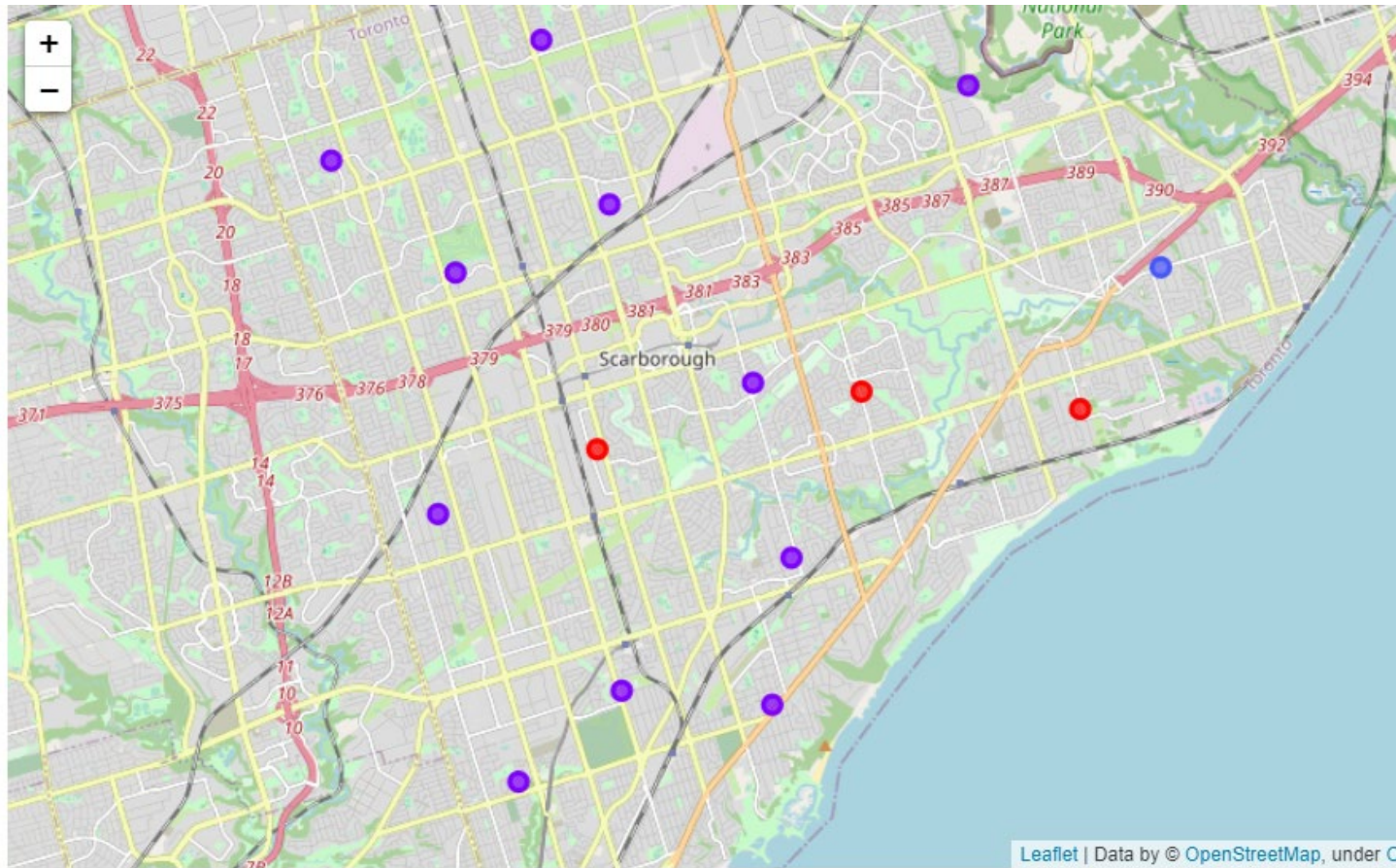
Work Flow:

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.

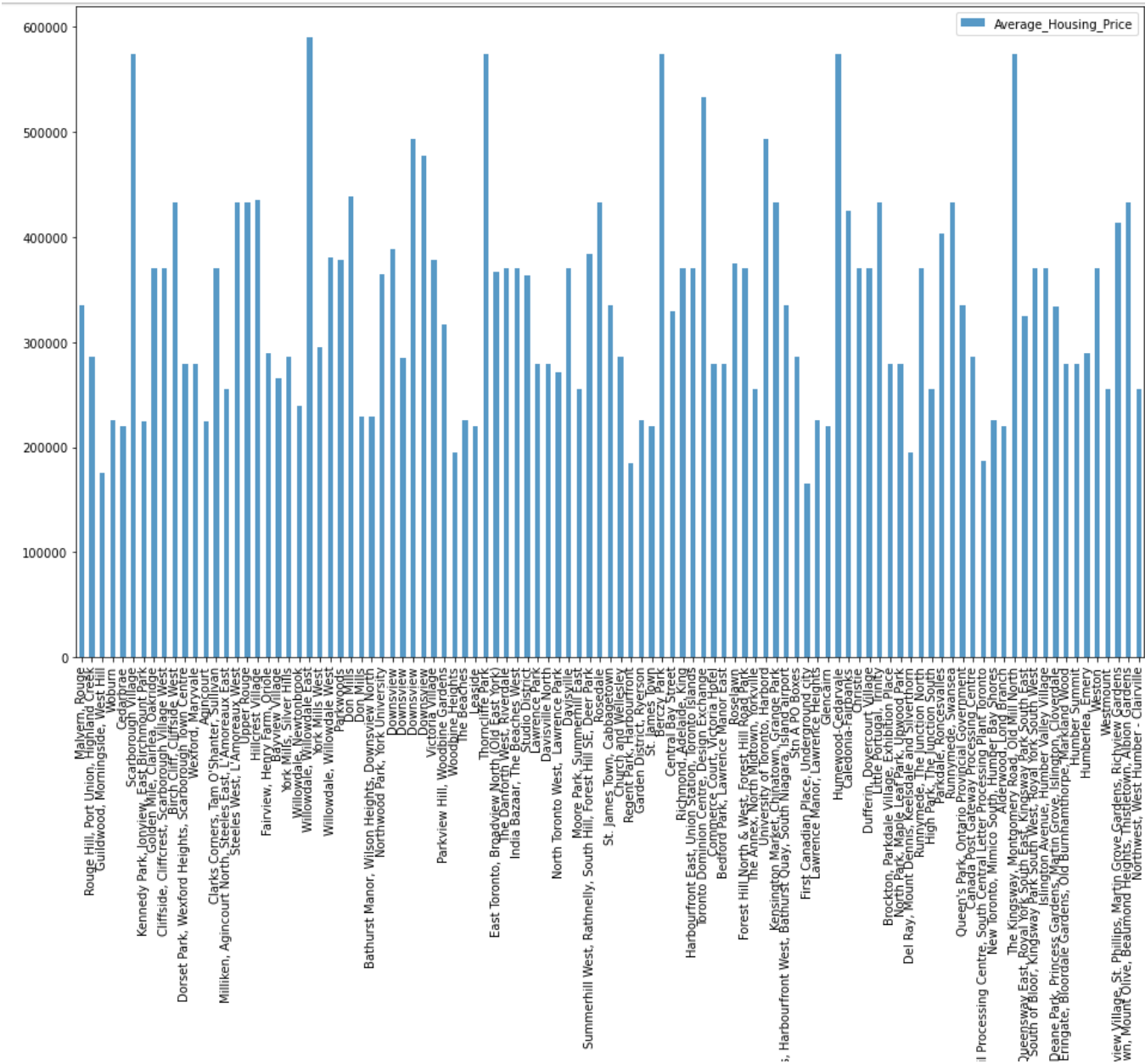
would be set to 500.

4. Results Section

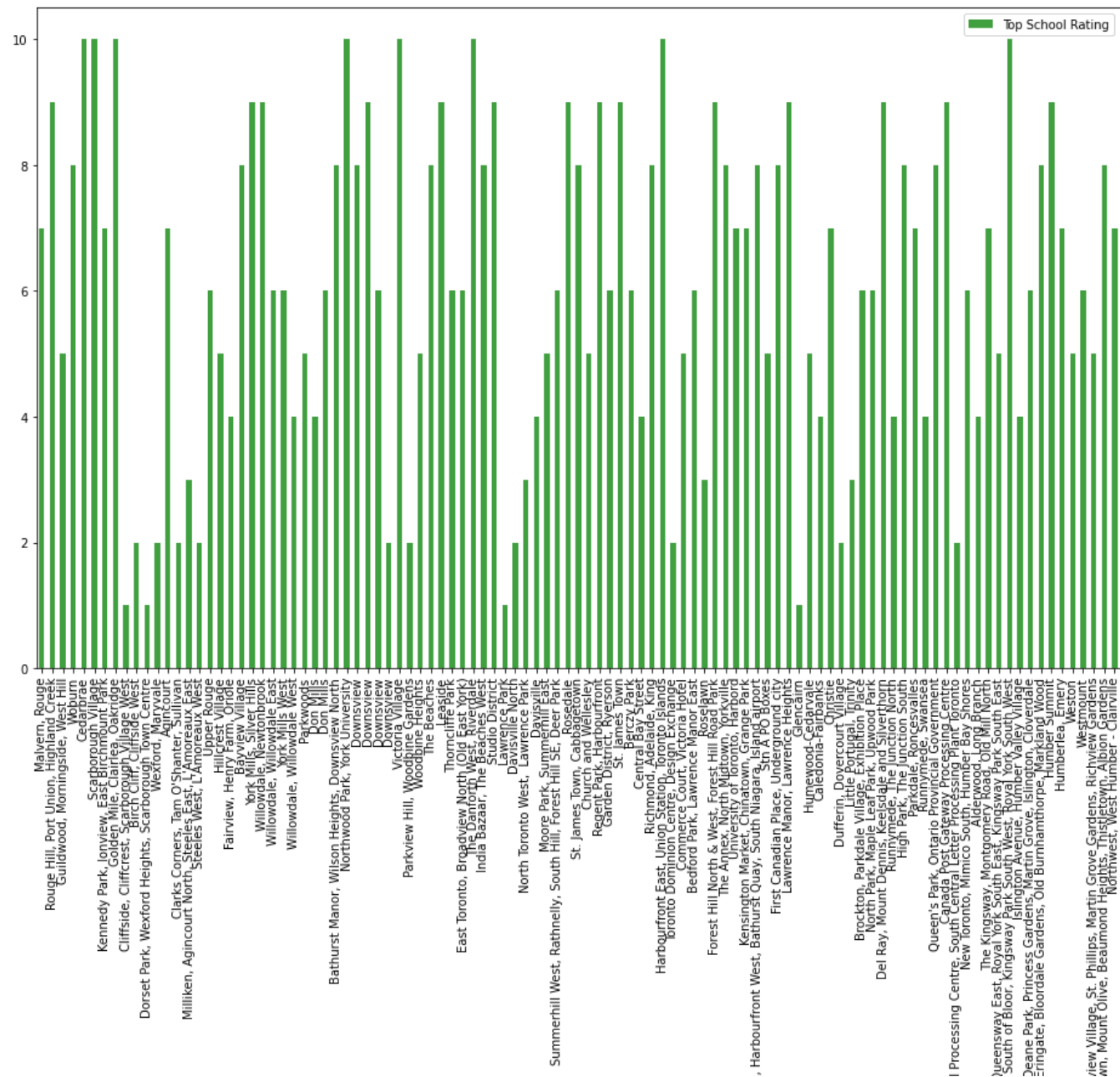
Map of Clusters in Scarborough



Average Housing Price by Clusters in Scarborough



School Ratings by Clusters in Scarborough



The Location:

Scarborough 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 Capstone 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 Which Tried to Solve:

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shifting there. 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.

- 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

6. Conclusion Section

In this Capstone 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 rating have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy 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.

Future Works:

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

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.