

1. Introduction:

The purpose of this Capstone Project is to help people in exploring better facilities around their neighbourhood. It will help people making smart and efficient decisions on selecting a great neighbourhood out of numbers of other neighbourhoods in Scarborough, Toronto.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighbourhoods. For ease of accessing to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This Capstone Project aims to create an analysis of features for people migrating to Scarborough to search for the best neighbourhood as a comparative analysis between neighbourhoods. 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 an emergency, water resources both fresh and wastewater and excrement conveyed in sewers and recreational facilities.

It will help people to get the awareness of the area and neighbourhood before moving to a new city, state, country or place for their work or to start a new fresh life.

2. Data Section

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

Will use the Scarborough dataset which we scraped from Wikipedia on Week 3.

Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

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

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.

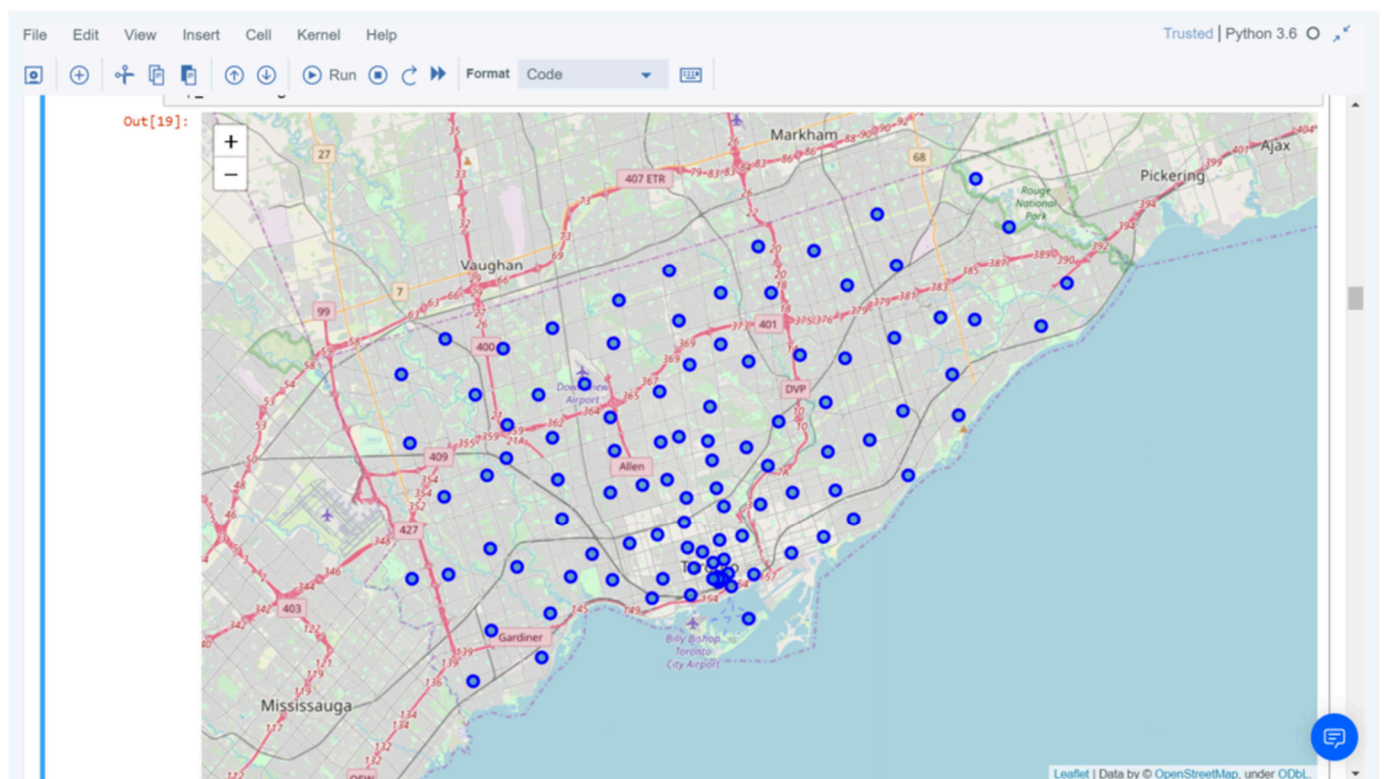
After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meters.

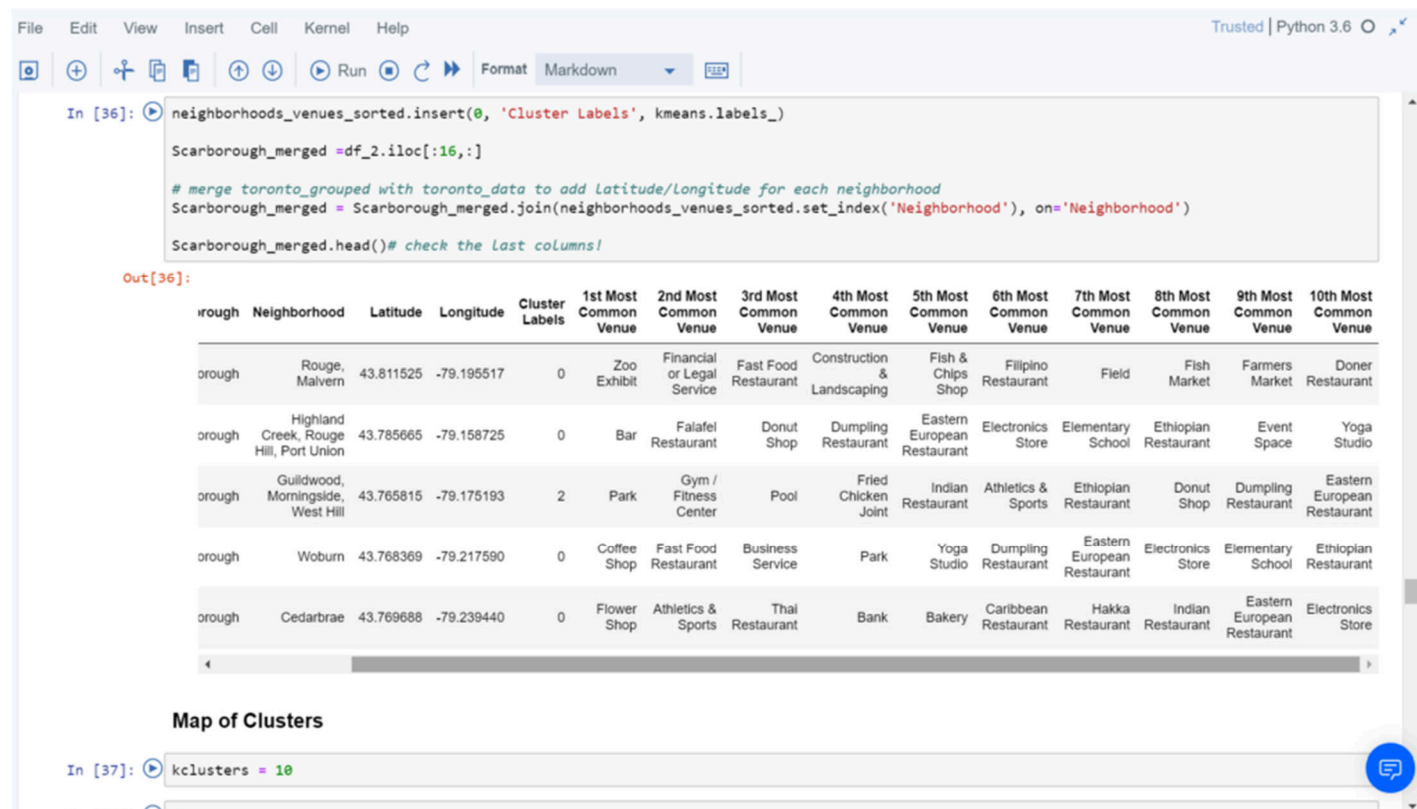
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

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





Most Common Venues near Neighborhood | Using Clustering

File Edit View Insert Cell Kernel Help Trusted | Python 3.6

```
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, Beaumont Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Aldenwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

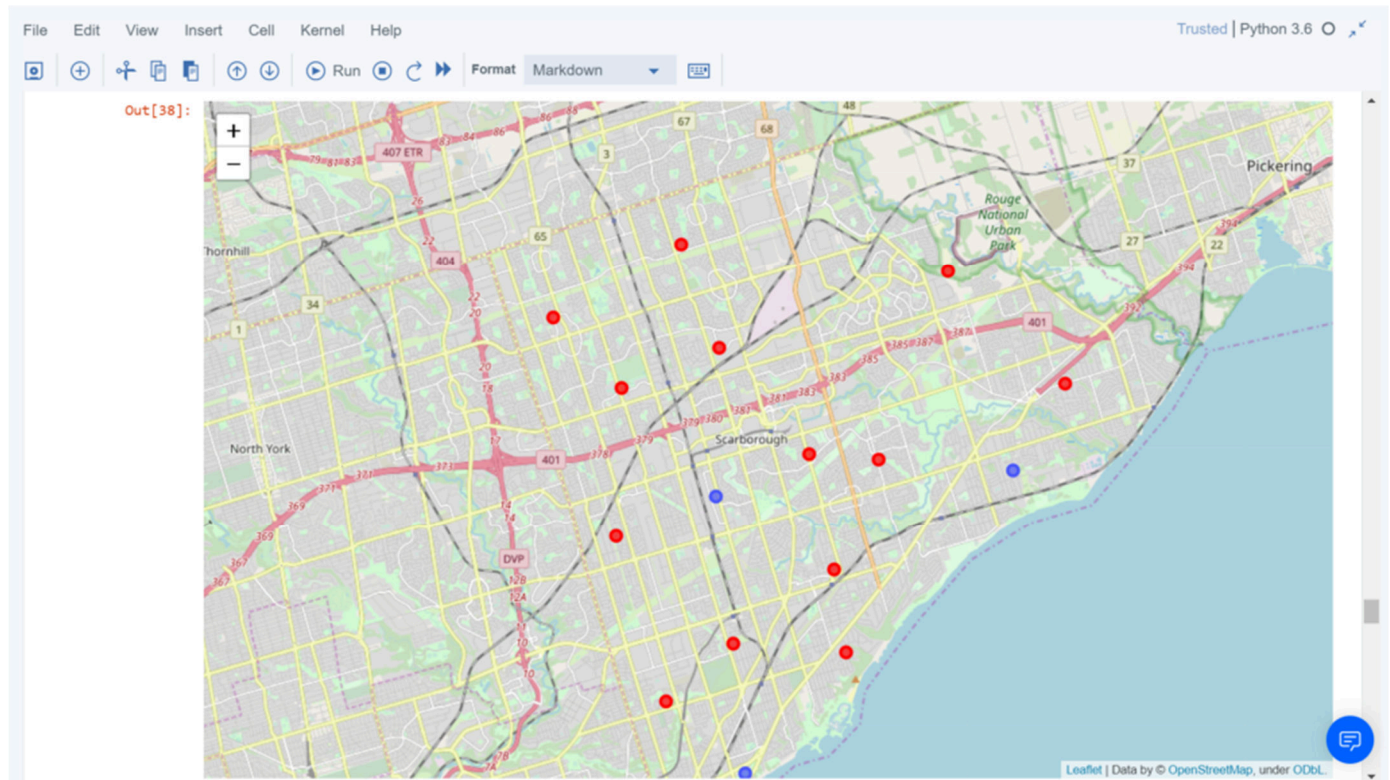
Work Flow:

Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to HTTP request limitations, the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

would be set to 500.

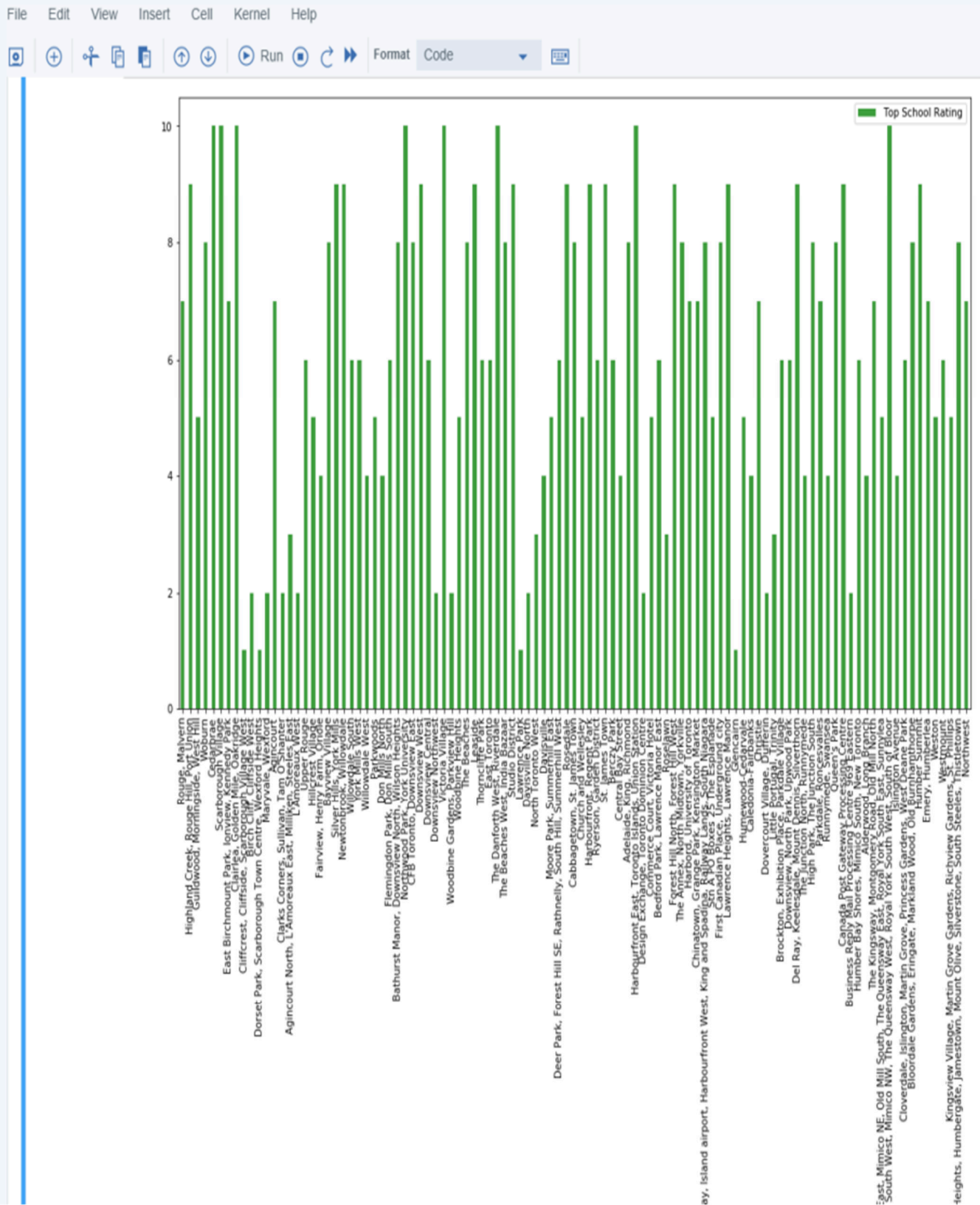
4. Results Section

Map of Clusters in Scarborough



A screenshot of the JupyterLab top toolbar. The menu bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. Below the menu bar, the toolbar contains icons for file operations (new, open, save, save as), cell navigation (previous, next), and execution (run, interrupt, restart). The 'Format' dropdown menu is open, showing 'Code' as the selected option. A 'Undo' button is also visible on the right side of the toolbar.

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 on the rise.

Foursquare API:

This Capstone project has 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 neighbourhood in a new city for the person who is 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 the house in terms of housing prices in an ascending or descending order
- Sorted list of schools in terms of location, fees, rating and reviews

6. Conclusion Section

In this Capstone project, using the k-means cluster algorithm I separated the neighbourhood into 10(Ten) different clusters and for 103 different latitudes and longitude from the dataset, which has very-similar neighbourhoods around them. Using the charts above results presented to a particular neighbourhood 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 the best house in Scarborough. Best means based on all required things(daily needs or things we need to live a better life) around and also in terms of cost-effectiveness.

Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the neighbourhood cluster distribution using an 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.