

Final Report | Capstone Project – The Battle of Neighborhoods

Hunting a convenient neighborhood in Scarborough, Toronto

1. Introduction:

The idea behind this project is to guide people in exploring convenient facilities around their neighborhood. It will help people choose most suitable neighborhoods amongst the crowded clusters in Scarborough, Toronto.

Canada migration is on a peak nowadays and needs a lot of research for good housing prices and reputed schools for their children etc. This project is for those people who are looking for better neighborhoods and for easy access to Cafes, Schools, Hospitals, Super markets, medical shops, grocery shops, mall, theatre, like minded people, etc.

This Project's aim is to create an analysis of features for a people migrating to Scarborough to search the 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 for both fresh and waste water.

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 Section

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

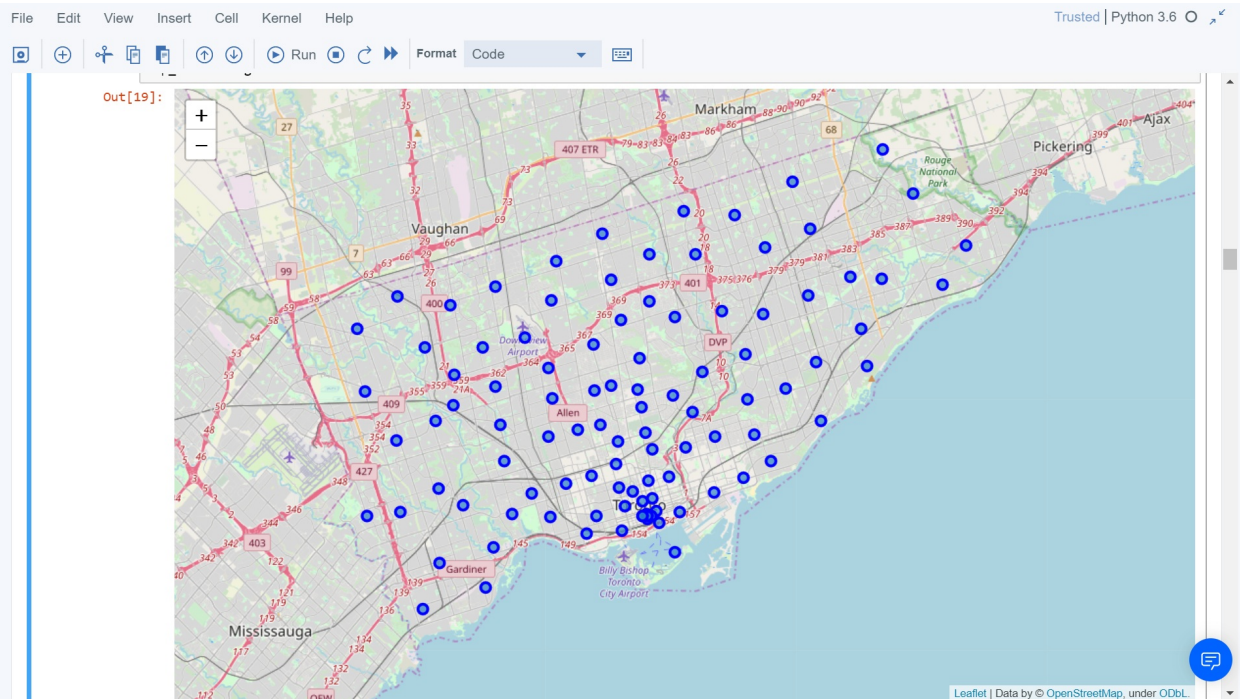
Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data: We will need data about different venues in different neighborhoods of that specific borough. This information will be extracted using "Foursquare" locational information. Foursquare is a location data provider with all the information about nearby venues and events within an area of interest using coordinate postitions. Such information includes venue names, geo locations, menus and even photos. As it is, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through this API.

After extracting the list of neighborhoods, we then retrieve information about different venues in each neighborhood using the Foursquare API. For each neighborhood, we have chosen the radius to be 100 meter. This can be a user defined approach to be it dynamic.

The data retrieved from Foursquare contains information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue is as following:

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

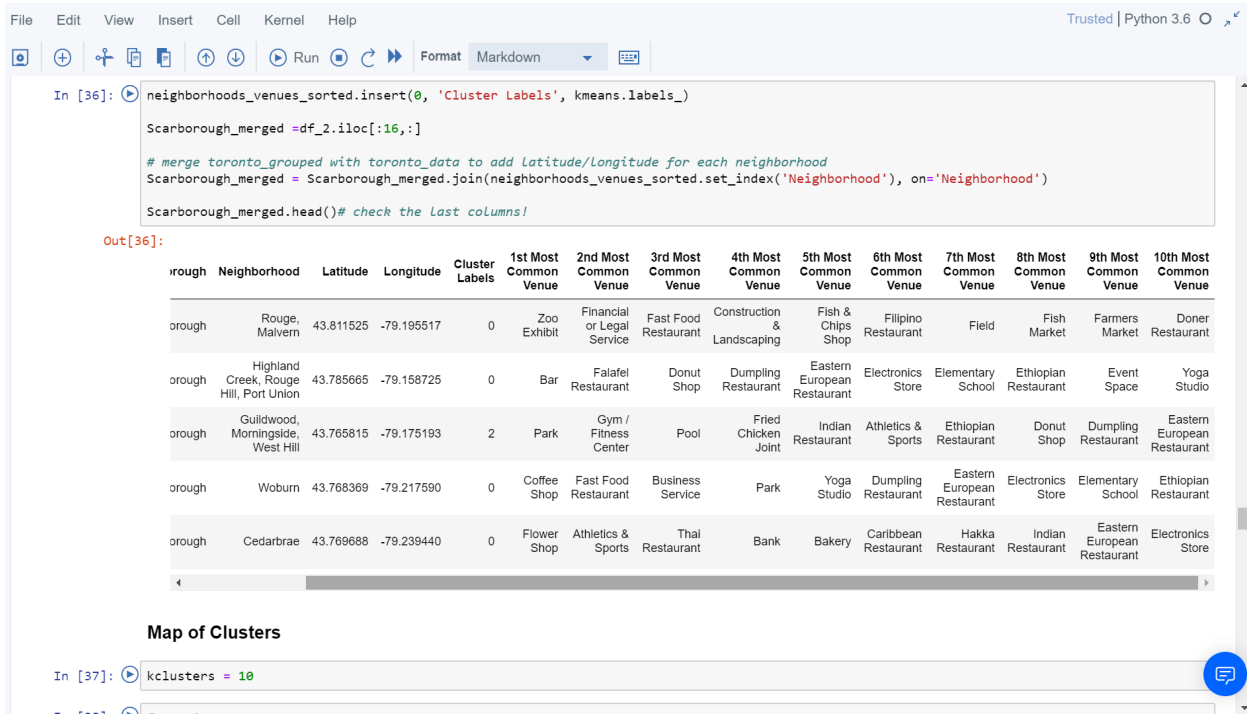


3. Methodology Section

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 achieve this, we have used an unsupervised machine learning algorithm - KMeans Clustering

Using K-Means Clustering Approach



Most Common venues near Neighborhood



```

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]:

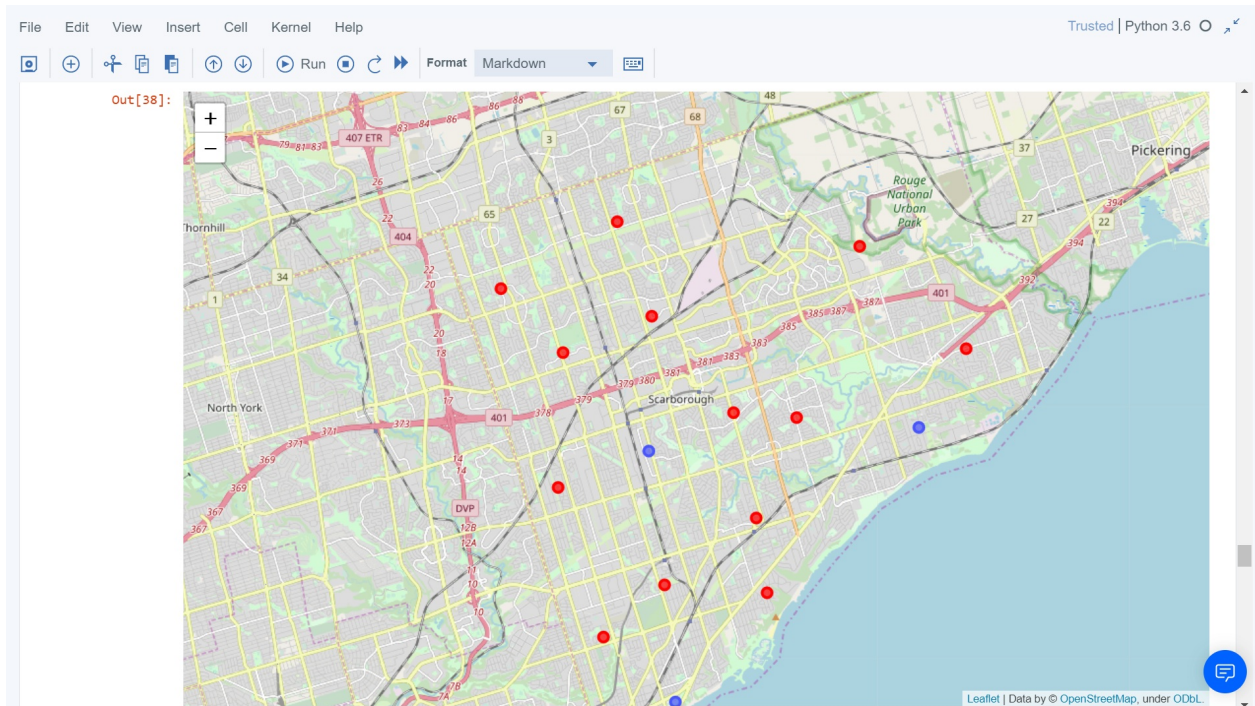
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger 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	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

Work Flow:

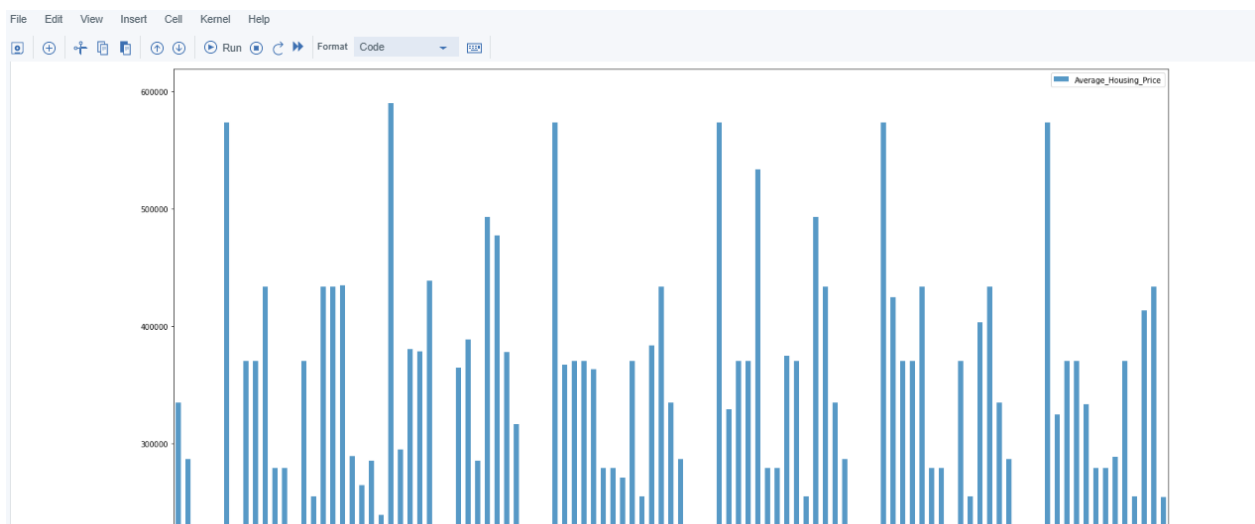
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

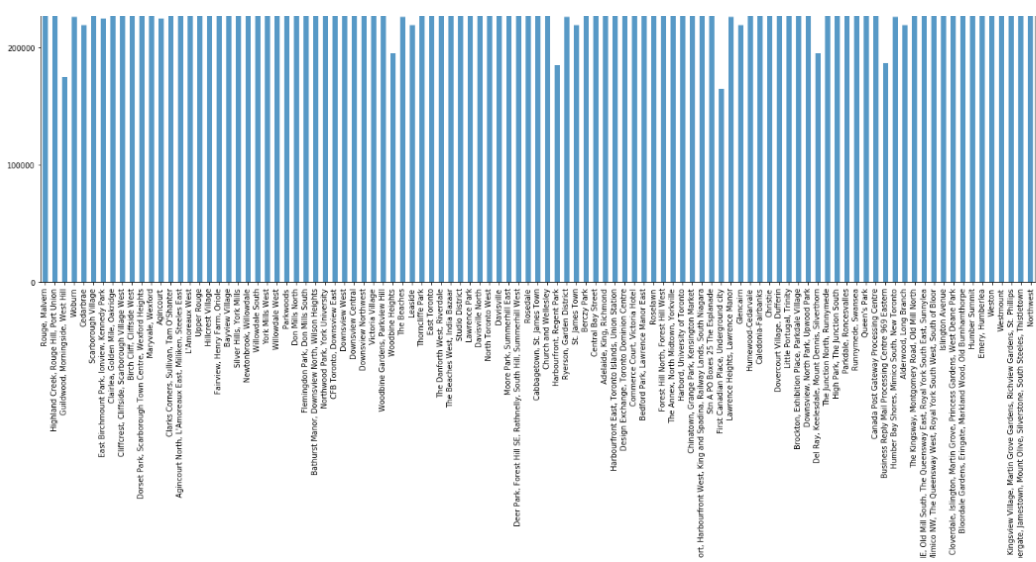
4. Results Section

Map of Clusters in Scarborough

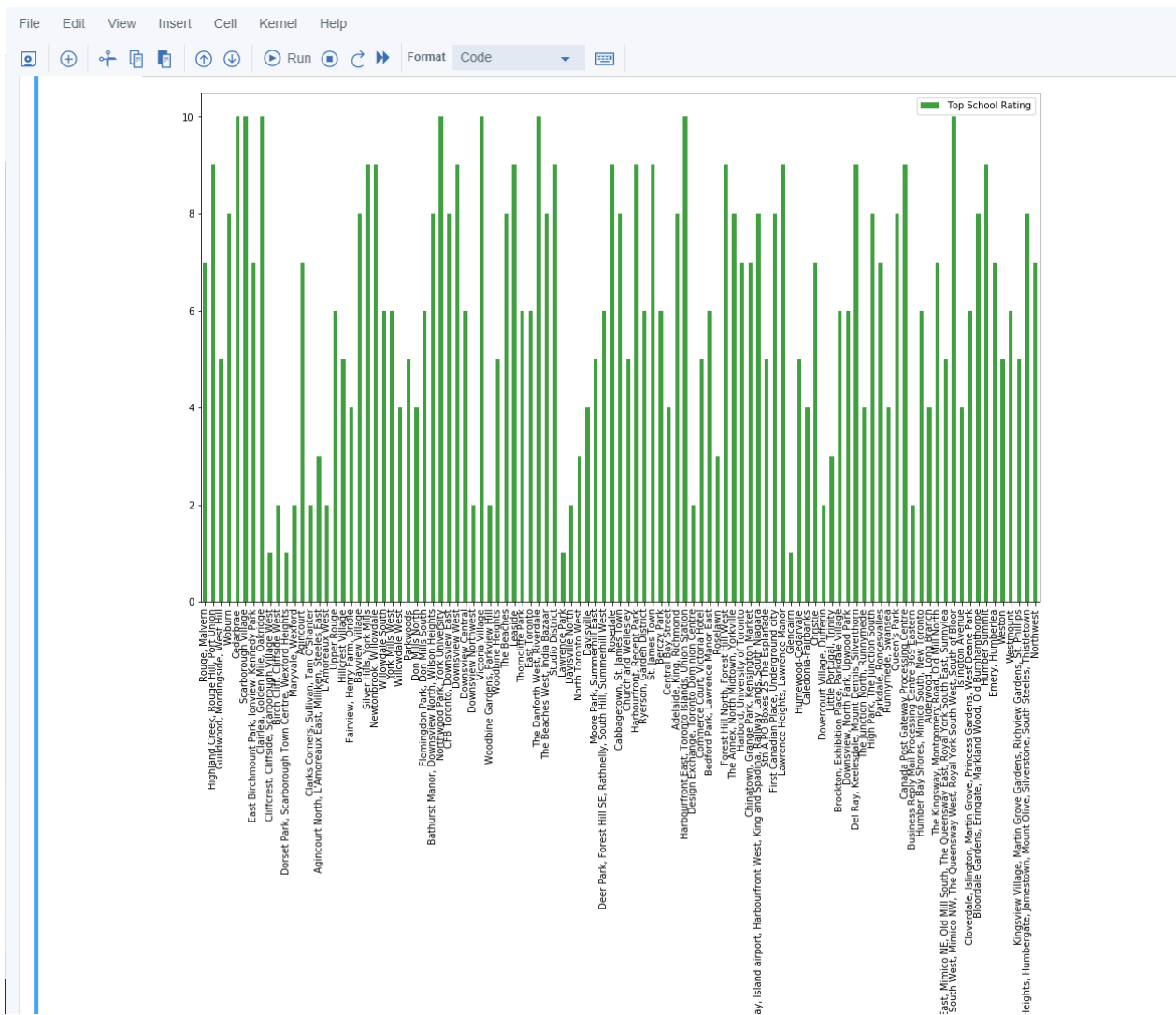


Average Housing Price by Clusters in Scarborough





School Ratings by Clusters in Scarborough



Why Scarborough ?:

Scarborough is a popular destination for new immigrants in Canada. It is well known to be one of the most diverse and multicultural areas in the Greater Toronto. It feels being at 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 on a rise in recent times.

Using Foursquare API:

This project has used Four-square API as its primary data extraction source as it has a database of millions of places using coordinates, 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 hunt for a convenient neighborhood for a person who is new to a city. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other essential needs.

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 this project, using k-means cluster algorithm we have segregated the neighborhood into 10(Ten) different clusters using the 180 different latitude and logitude from dataset, which have very-similar neighborhoods around them.

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 180 different latitude and logitude 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 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 Develope 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.

Presentation Link: