

# **Final Report | Capstone Project – The Battle of Neighbourhoods Finding a Better Place in North York, Toronto**

## **1. Introduction: Business Problem**

The purpose of this project is to help people getting the best neighbourhood who wants to migrate to Canada. Not only this but they can also refer through different locations which they have near that place. So this will help people to get the best environment according to their needs.

Nowadays lots of people are migrating to Canada either for study or business purposes, so this is my main motto to choose this topic as my capstone project. Especially for a student who wants to do an immense amount of work in a very short period of time. They will need some essential things at their fingertips and it will help them to get work done immediately. For a better experience, it is more important to have a maximum of essentialities near their residence.

This project is for those people who are migrating to Canada for study purposes or think to settle there. With the help of this project, they can easily find things like Schools, colleges, malls, Metro stations, Bus stations, theatre, hospital and so on and so forth.

By creating an analytical feature of the North York area to help people choose a better neighbourhood in Canada. It will give a brief introduction to things which they will get in their neighbourhoods at the chosen place and will help them to start a fresh life.

## **2. Data Section**

Data Link: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Will use the North York 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. 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.

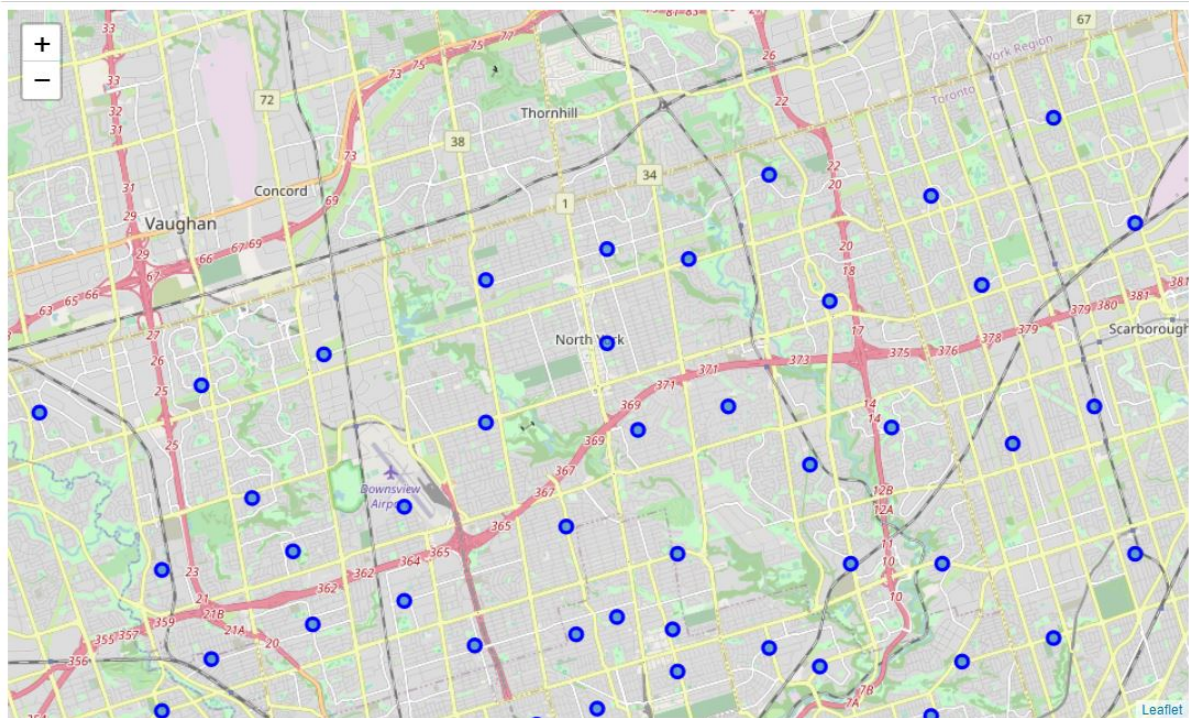
After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meters.

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. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store
6. Venue Latitude
7. Venue Longitude
8. Venue Category

## Maps of North York.

Add alt text



## 3. Methodology Section

### Clustering approach

To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods 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

## k-means clustering approach | most common venue

```
#neighbourhoods_venues_sorted.insert(0, 'ClusterLabels', kmeans.labels_)

NY_merged = df.iloc[:16,:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
NY_merged = NY_merged.join(neighbourhoods_venues_sorted.set_index('Neighbourhood'), on='Neighbourhood')

NY_merged.head()# check the last columns!
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	ClusterLabels	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	M1B	North York	Malvern, Rouge	43.806686	-79.194353	0	Fast Food Restaurant	Coffee Shop	Construction & Landscaping	Business Service	Bus Station	Spa	Electronics Store
1	M1C	North York	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	3	Breakfast Spot	Burger Joint	Bar	Ethiopian Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
2	M1E	North York	Guildwood, Morningside, West Hill	43.763573	-79.188711	3	Restaurant	Park	Fast Food Restaurant	Rental Car Location	Bank	Fried Chicken Joint	Moving Target
3	M1G	North York	Woburn	43.770992	-79.216917	1	Coffee Shop	Park	Construction & Landscaping	Business Service	Yoga Studio	Eastern European Restaurant	Dog Run
4	M1H	North York	Cedarbrae	43.773136	-79.239476	3	Indian Restaurant	Coffee Shop	Flower Shop	Bakery	Gas Station	Thai Restaurant	Chinese Restaurant

## k-means clustering approach | most common venue in the neighbourhood

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighbourhood']
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))

# create a new dataframe
neighbourhoods_venues_sorted = pd.DataFrame(columns=columns)
neighbourhoods_venues_sorted['Neighbourhood'] = NY_grouped['Neighbourhood']

for ind in np.arange(NY_grouped.shape[0]):
    neighbourhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(NY_grouped.iloc[ind, :], num_top_venues)

neighbourhoods_venues_sorted.head()
```

	Neighbourhood	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	Aldenwood, Long Branch	Pharmacy	Pizza Place	Coffee Shop	Yoga Studio	Cupcake Shop	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Department Store
1	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Pizza Place	Middle Eastern Restaurant	Fried Chicken Joint	Deli / Bodega	Restaurant	Sushi Restaurant	Yoga Studio	Cupcake Shop	Dim Sum Restaurant
2	Bedford Park, Lawrence Manor East	Italian Restaurant	Comfort Food Restaurant	Sushi Restaurant	Breakfast Spot	Sandwich Place	Coffee Shop	Juice Bar	Cosmetics Shop	Fast Food Restaurant	Discount Store
3	Birch Cliff, Cliffside West	Café	Yoga Studio	Department Store	Dumpling Restaurant	Dog Run	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop	Deli / Bodega
4	Business reply mail Processing Centre, South C...	Brewery	Auto Workshop	Department Store	Electronics Store	Dumpling Restaurant	Dog Run	Discount Store	Diner	Dim Sum Restaurant	Dessert Shop

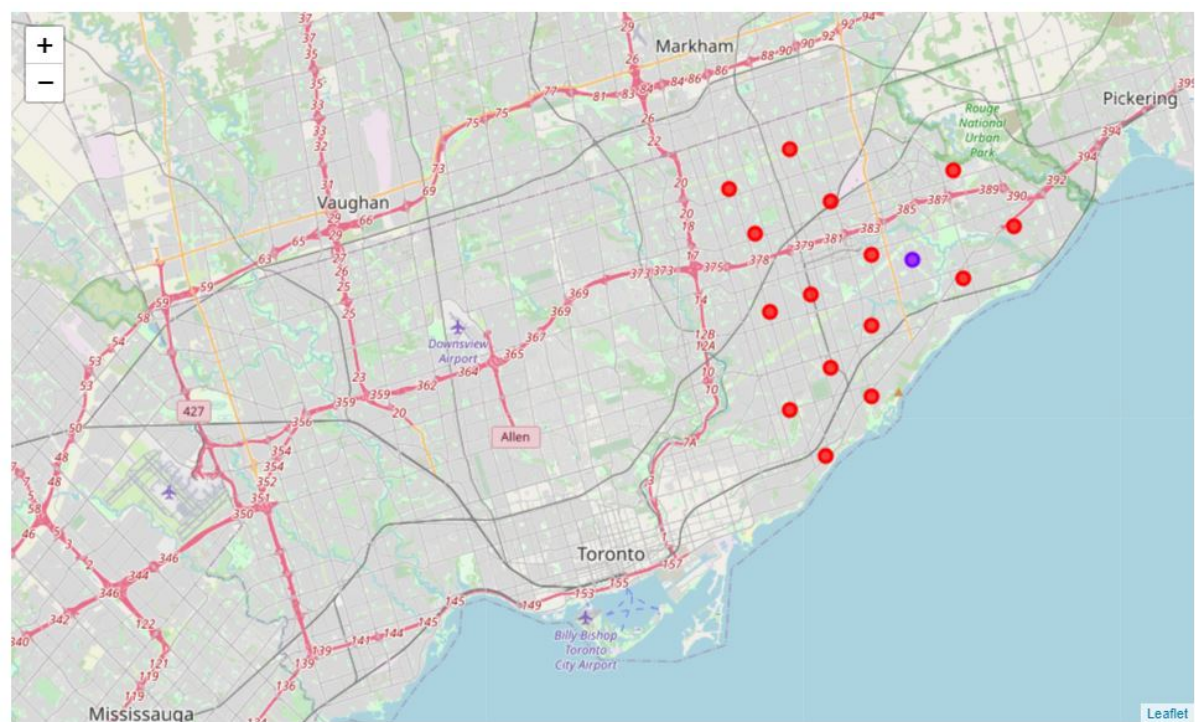
## Work Flow:

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.

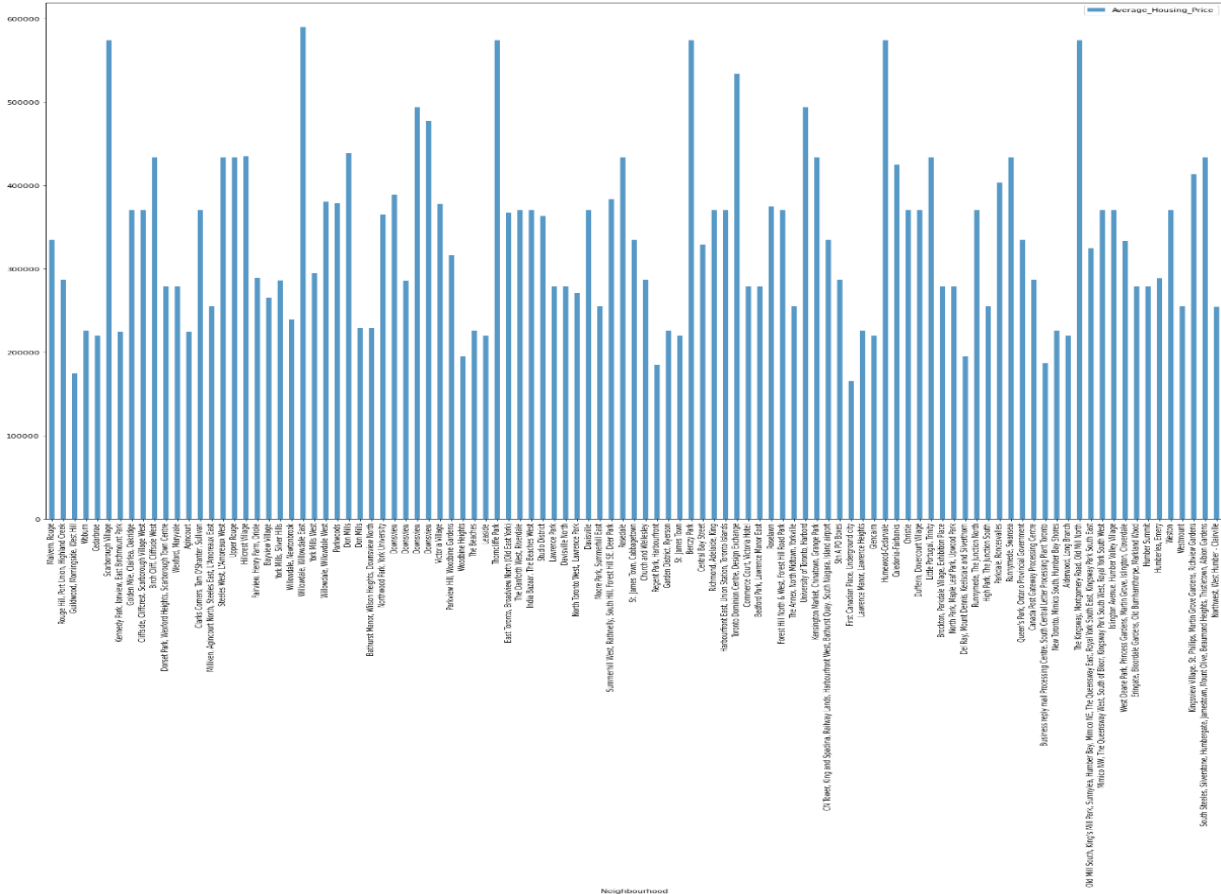


# 4. Result Section

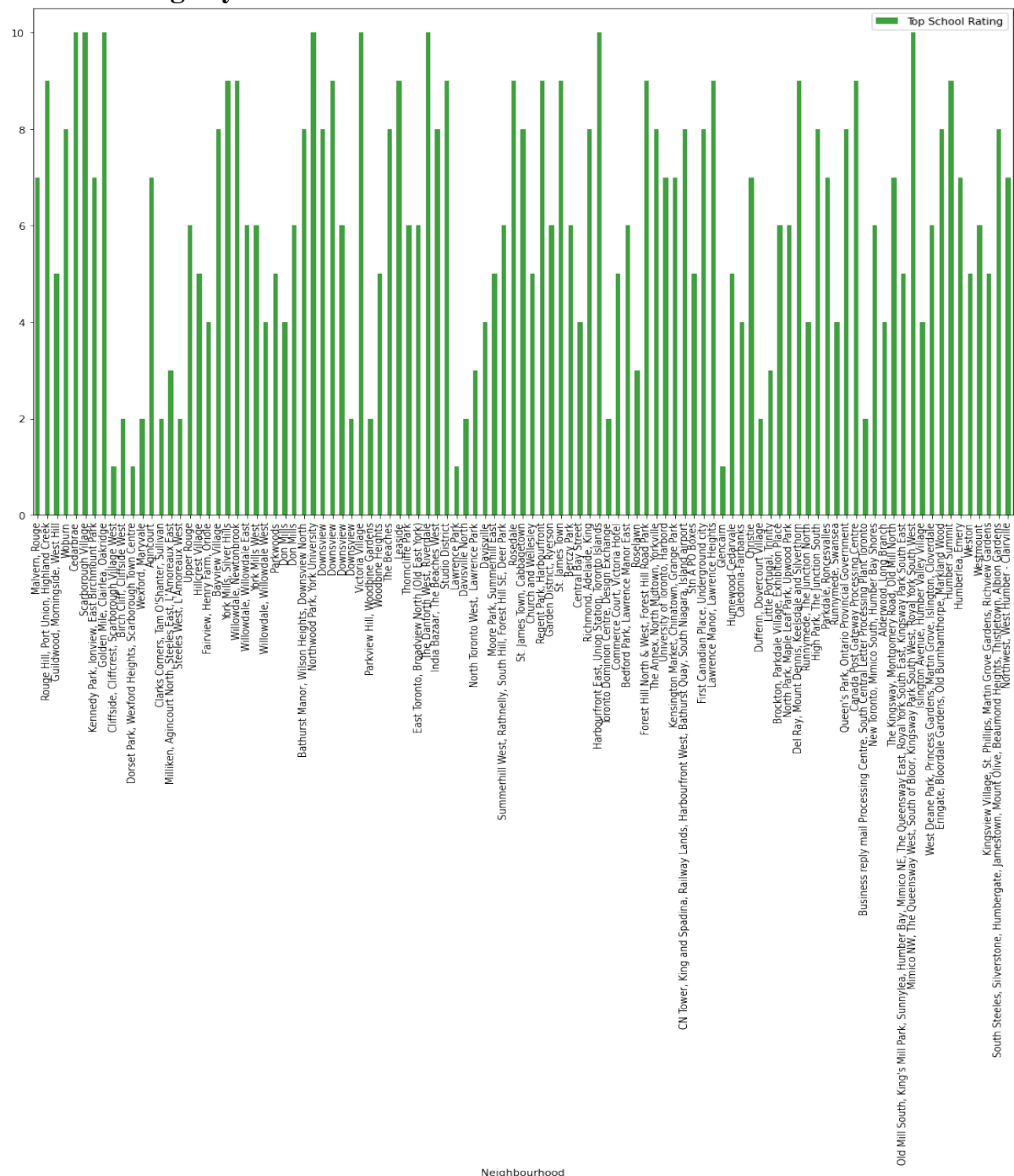
Map of cluster near North York



Average Housing Price by Clusters in North York



## School Ratings by Clusters in North York



## The Location:

North York, is a very popular location for students as a top-ranked university (YORK UNIVERSITY) is settled there. And not only for this but it also very suitable for a business class person as it very near to the capital of Toronto. So it is useful for those newcomers who want to settle their business.

## **Foursquare API**

With the help of this API, we can gather information about hospitals, schools, bus stand, metro station, airport, colleges, theatre, mall and many more.

## **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 centre, markets and other daily needs things nearby.

1. Sorted list of the house in terms of housing prices in an ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

## **6. Conclusion Section**

In this capstone project, using the k-means cluster approach I separated the neighbourhood into 3 different clusters which help people to investigate their taste into the various fields.

### **Future Works:**

This Capstone project can be continued for making it more precise in terms to find the best house in North York. 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-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 of 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.