

# Final Report | Capstone Project – The Battle of Neighborhoods, Toronto

## 1. Introduction:

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

Lots of people are migrating to various states of Canada and need lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighborhoods.

This Project aims to create an analysis of features for people migrating to Toronto to search for a best neighborhood to settle down. 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 wastewater and excrement conveyed in sewers and recreational facilities and especially restaurants! Because good food is equally important for survival.

## 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 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. 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

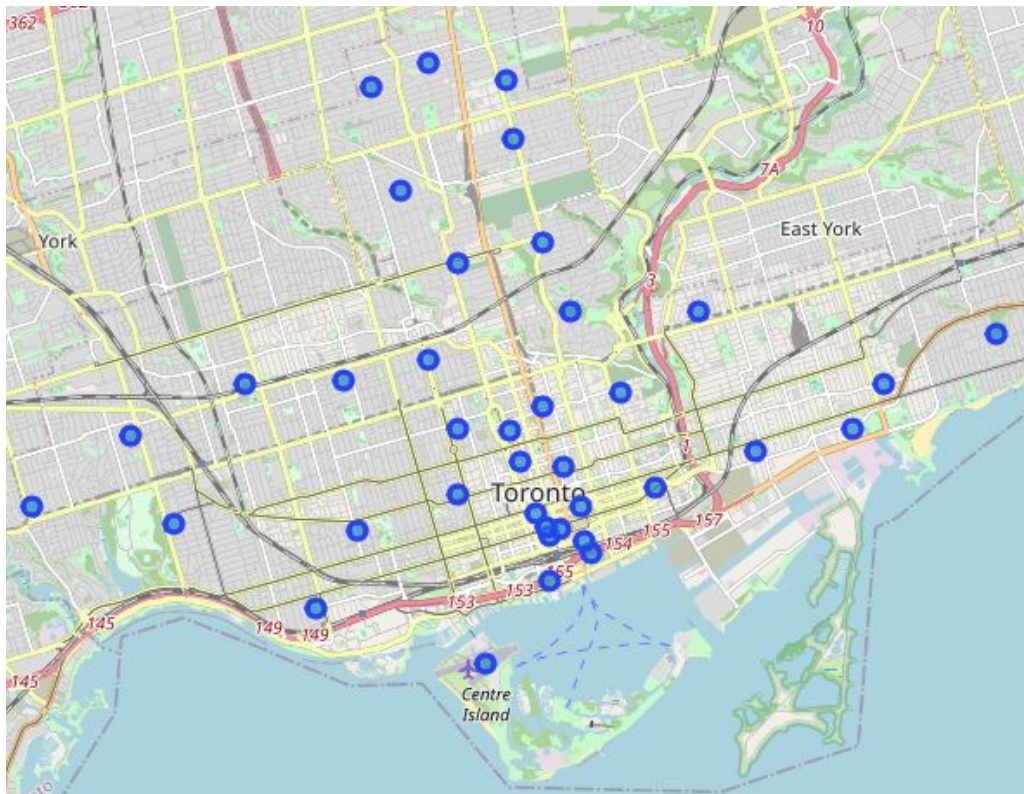
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

**Final Borough's include:**

North York	24
Downtown Toronto	19
Scarborough	17
Etobicoke	12
Central Toronto	9
West Toronto	6
East Toronto	5
York	5
East York	5
Mississauga	1

Name: Borough, dtype: int64

**Map of Scarborough**

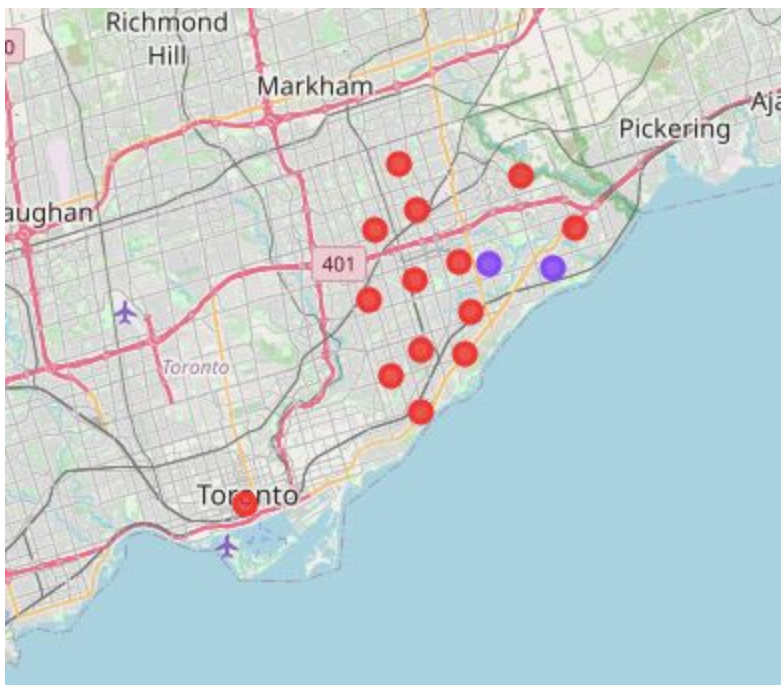


### 3. Methodology Section

#### 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 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



#### Workflow:

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 80 and the radius parameter would be set to 120.

### 4. Results Section

#### Map of Clusters in Scarborough



Toronto is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas, 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 has used the 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 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.

## **6. Conclusion Section**

In this project, using k-means cluster algorithm I separated the neighborhood into 10 different clusters and for 69 different latitude and longitude from the 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.