# Capstone Project - The Battle of the Neighborhoods -[¶](#Capstone-Project---The-Battle-of-the-Ne)

## Finding a Better Place in North York, Toronto[¶](#Finding-a-Better-Place-in-North-York,-T)

### Applied Data Science Capstone by IBM/Coursera[¶](#Applied-Data-Science-Capstone-by-IBM/Co)

## Table of contents[¶](#Table-of-contents)

* [Introduction: Business Problem](#introduction)
* [Data](#data)
* [Methodology](#methodology)
* [Analysis](#analysis)
* [Results and Discussion](#results)
* [Conclusion](#conclusion)

## Introduction: Business Problem[¶](" \l "Introduction:-Business-Problem)

### Background[¶](" \l "Background)

The purpose of this project is to help people in exploring better facilities around their neighborhood. It will help Asian tourists making a smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in **North York, Toronto.**

Lots of people are migrating to various Asia and needed lots of research for convenient neighborhoods. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

This project aims to create an analysis of features for people migrating to North York to search for 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 an emergency, water resources both fresh and wastewater, and excrement conveyed in sewers and recreational facilities.

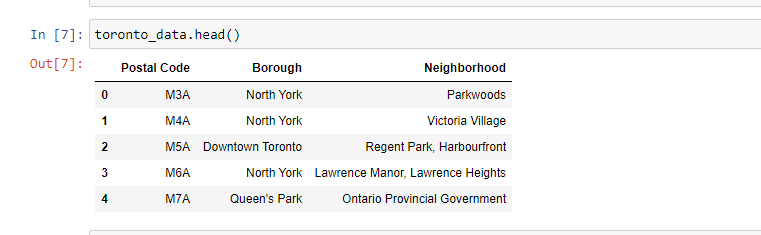
### Problem [¶](" \l "Problem-)

The major purpose of this project, is to suggest a better neighborhood in a North York, Toranto for the person who relocate there. Social presence in society in terms of like minded people. Connectivity to the airport, city center, markets and all daily things nearby.

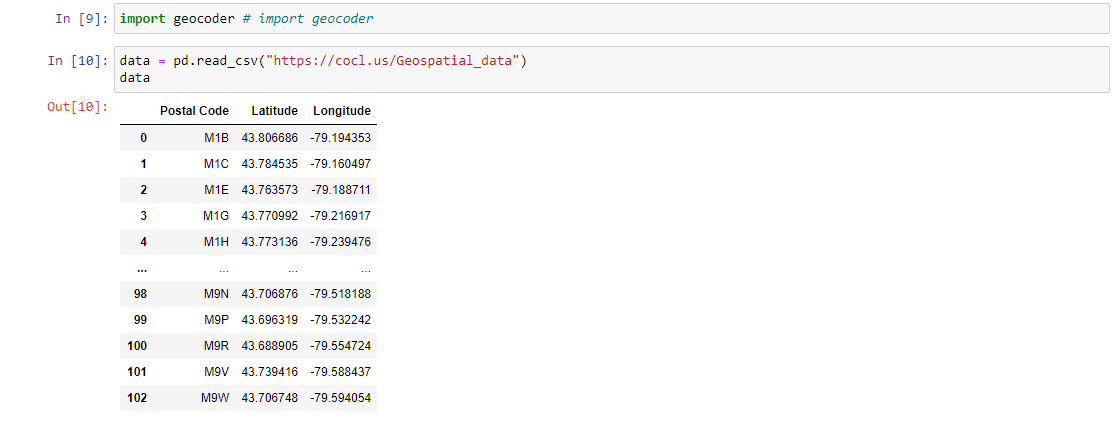
## Data [¶](" \l "Data-)

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

To generate the Postal/Borough/Neighborhood data of Toronto metro area as followed:

**

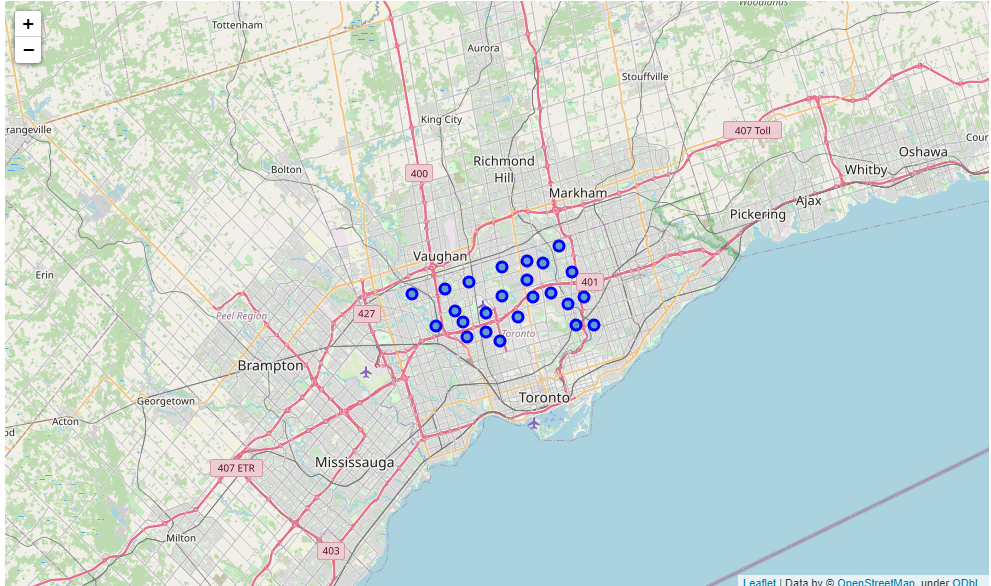
with the library **geocoder** Dataset consisting of latitude and longitude, zip codes.

**

Conbined both information as new dataframe:

**

#### Neighborhood Map of North York, Toronto[¶](#Neighborhood-Map-of-North-York,-Toronto)

**

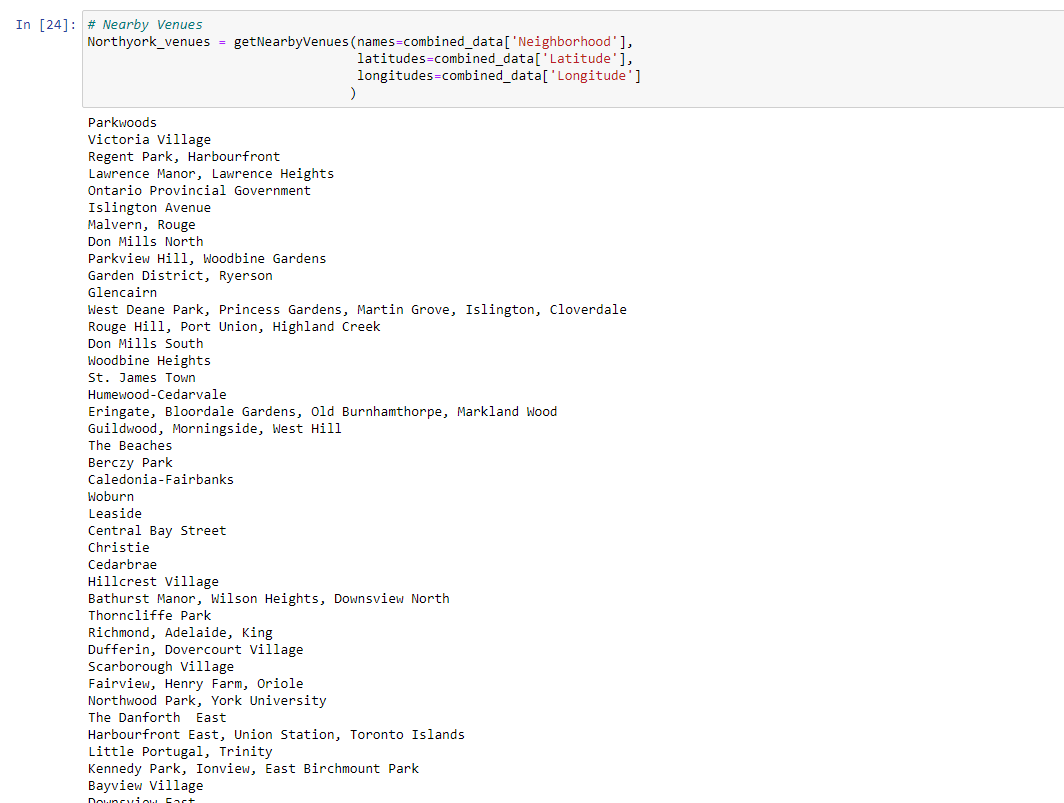
**Foursquare API Data:**

We will need data about different venues in different neighborhoods of **North York** 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, and menus. 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

### Nearby Venues within North York, Toronto[¶](#Nearby-Venu-within-North-York,-Toronto)

OK. Let us now **filter** those locations: we're interested only in **locations with no more than two restaurants in radius of 500 meters**, and **no asian restaurants in radius of 100 meters**, and **more than 10 customers**.

**

Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category and find out the top 10 Venues of each neighborhood in North York.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 | Sandwich Place | Pool Hall | Shanghai Restaurant | Badminton Court | Latin American Restaurant | Breakfast Spot | Lounge | Convenience Store | Dry Cleaner | Dumpling Restaurant |
| 1 | "Alderwood, Long Branch" | Pizza Place | Convenience Store | Pub | Dance Studio | Sandwich Place | Gas Station | Coffee Shop | Pharmacy | Pool | College Stadium |
| 2 | "Bathurst Manor, Wilson Heights, Downsview North" | Coffee Shop | Park | Bank | Ice Cream Shop | Community Center | Sandwich Place | Diner | Shopping Mall | Bridal Shop | Pharmacy |
| 3 | Bayview Village | Bank | Japanese Restaurant | Grocery Store | Intersection | Cafe | Skating Rink | Chinese Restaurant | Playground | Dry Cleaner | Dumpling Restaurant |
| 4 | "Bedford Park, Lawrence Manor East" | Coffee Shop | Restaurant | Sandwich Place | Italian Restaurant | Women's Store | Pub | Cupcake Shop | Frozen Yogurt Shop | Thai Restaurant | Indian Restaurant |

## Methodology [¶](" \l "Methodology-)

### Clustering Approach:[¶](#Clustering-Approach:)

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.

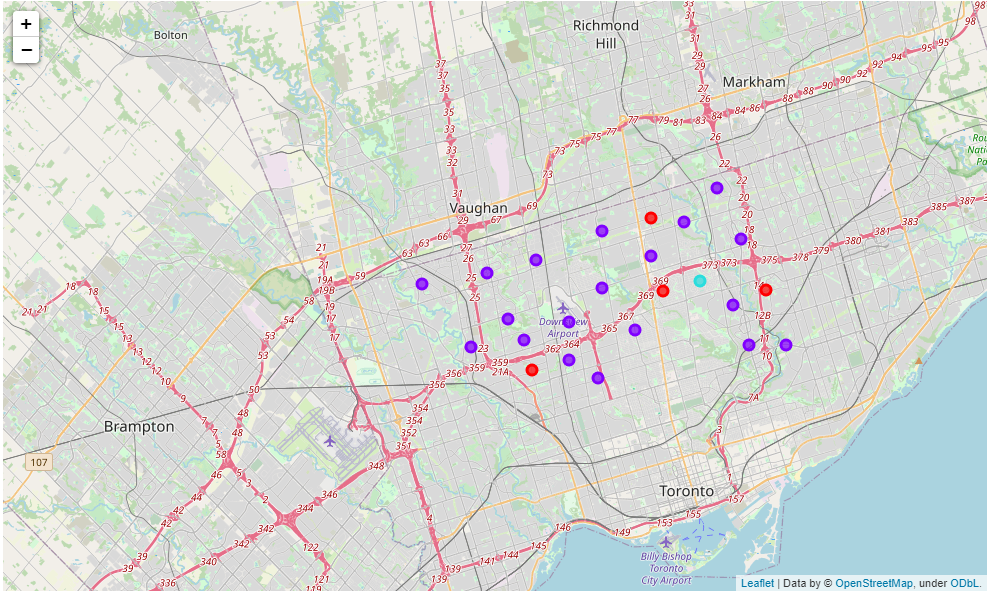
#### K mean clustering[¶](#K-mean-clustering)

Run k-means to cluster the neighborhood into 4 clusters and insert the cluster labes to each neighbor.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Postal Code | 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 | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
| 0 | M3A | North York | Parkwoods | 43.7532586 | -79.3296565 | 0 | Fast Food Restaurant | Pet Store | Park | Food & Drink Shop | Women's Store | Donut Shop | Diner | Discount Store | Distribution Center | Dive Bar |
| 1 | M4A | North York | Victoria Village | 43.7258823 | -79.3155716 | 1 | Pizza Place | Hockey Arena | Sporting Goods Shop | Park | Portuguese Restaurant | Coffee Shop | Playground | College Stadium | Donut Shop | Diner |
| 3 | M6A | North York | "Lawrence Manor, Lawrence Heights" | 43.718518 | -79.4647633 | 1 | Clothing Store | Coffee Shop | Vietnamese Restaurant | Fast Food Restaurant | Paper / Office Supplies Store | Boutique | Bowling Alley | Seafood Restaurant | Cafe | Park |
| 7 | M3B | North York | Don Mills North | 43.7459058 | -79.352188 | 1 | Japanese Restaurant | Cafe | Gym | Paper / Office Supplies Store | Caribbean Restaurant | Doner Restaurant | Dim Sum Restaurant | Diner | Discount Store | Distribution Center |
| 10 | M6B | North York | Glencairn | 43.709577 | -79.4450726 | 1 | Pizza Place | Gas Station | Metro Station | Coffee Shop | Fish Market | Latin American Restaurant | Sandwich Place | Restaurant | Ice Cream Shop | Italian Restaurant |

## Analysis [¶](" \l "Analysis-)

Change the neighthood color with clustered color.

**

## Results and Discussion [¶](" \l "Results-and-Discussion-)

### Examine Clusters[¶](#Examine-Clusters)

#### cluster1[¶](#cluster1)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 | Parkwoods | Park | Food & Drink Shop | Convenience Store | Dessert Shop | Diner | Discount Store | Distribution Center | Dive Bar | Dog Run | Doner Restaurant |
| 49 | "North Park, Maple Leaf Park, Upwood Park" | Home Service | Business Service | Bakery | Construction & Landscaping | Park | Doner Restaurant | Dim Sum Restaurant | Diner | Discount Store | Distribution Center |
| 52 | "Willowdale, Newtonbrook" | Park | Coffee Shop | Trail | Doner Restaurant | Dim Sum Restaurant | Diner | Discount Store | Distribution Center | Dive Bar | Dog Run |
| 66 | York Mills West | Park | Convenience Store | Intersection | Home Service | Bowling Alley | Donut Shop | Diner | Discount Store | Distribution Center | Dive Bar |

#### cluster2[¶](#cluster2)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 |
| 1 | Victoria Village | Coffee Shop | Hockey Arena | Pizza Place | Park | Portuguese Restaurant | Intersection | Playground | College Theater | Dog Run | Dessert Shop |
| 3 | "Lawrence Manor, Lawrence Heights" | Clothing Store | Coffee Shop | Vietnamese Restaurant | Fast Food Restaurant | Furniture / Home Store | Grocery Store | Hobby Shop | Park | Boutique | Paper / Office Supplies Store |
| 7 | Don Mills North | Japanese Restaurant | Gym | Cafe | Caribbean Restaurant | Paper / Office Supplies Store | Athletics & Sports | Dumpling Restaurant | Dry Cleaner | Drugstore | Donut Shop |
| 10 | Glencairn | Pizza Place | Italian Restaurant | Gas Station | Sushi Restaurant | Flower Shop | Sandwich Place | Latin American Restaurant | Fish Market | Asian Restaurant | Ice Cream Shop |
| 13 | Don Mills South | Gym | Restaurant | Clothing Store | Art Gallery | Sporting Goods Shop | Beer Store | Sandwich Place | Discount Store | Supermarket | Dim Sum Restaurant |
| 27 | Hillcrest Village | Sandwich Place | Pharmacy | Fast Food Restaurant | Chinese Restaurant | Restaurant | Tennis Court | Bakery | Distribution Center | Department Store | Dessert Shop |
| 28 | "Bathurst Manor, Wilson Heights, Downsview North" | Coffee Shop | Bank | Park | Pharmacy | Gas Station | Shopping Mall | Bridal Shop | Diner | Sandwich Place | Sushi Restaurant |
| 33 | "Fairview, Henry Farm, Oriole" | Clothing Store | Coffee Shop | Fast Food Restaurant | Baseball Field | Japanese Restaurant | Restaurant | Bank | Bus Station | Shopping Mall | Bar |
| 34 | "Northwood Park, York University" | Coffee Shop | Massage Studio | Pizza Place | Vietnamese Restaurant | Fast Food Restaurant | Japanese Restaurant | Bar | Caribbean Restaurant | Dog Run | Dim Sum Restaurant |
| 39 | Bayview Village | Bank | Japanese Restaurant | Skating Rink | Intersection | Cafe | Chinese Restaurant | Grocery Store | Playground | Dumpling Restaurant | Distribution Center |
| 40 | Downsview East | Coffee Shop | Airport | Bakery | Sandwich Place | Park | Chinese Restaurant | Dumpling Restaurant | Dry Cleaner | Drugstore | Eastern European Restaurant |
| 46 | Downsview West | Pizza Place | Moving Target | Vietnamese Restaurant | Grocery Store | Park | Bank | Shopping Mall | Dive Bar | Dessert Shop | Dim Sum Restaurant |
| 50 | Humber Summit | IT Services | Dance Studio | Home Service | Bakery | Construction & Landscaping | Flower Shop | Women's Store | Discount Store | Distribution Center | Dive Bar |
| 53 | Downsview Central | Business Service | Baseball Field | Vietnamese Restaurant | Middle Eastern Restaurant | Korean Restaurant | Donut Shop | Diner | Discount Store | Distribution Center | Dive Bar |
| 55 | "Bedford Park, Lawrence Manor East" | Coffee Shop | Italian Restaurant | Sandwich Place | Restaurant | Pharmacy | Bagel Shop | Bakery | Bank | Butcher | Cafe |
| 57 | "Humberlea, Emery" | Convenience Store | Discount Store | Gas Station | Auto Garage | Electronics Store | Eastern European Restaurant | Dumpling Restaurant | Dry Cleaner | Drugstore | Donut Shop |
| 59 | Willowdale South | Coffee Shop | Ramen Restaurant | Korean Restaurant | Japanese Restaurant | Middle Eastern Restaurant | Pizza Place | Fast Food Restaurant | Sandwich Place | Bank | Dessert Shop |
| 60 | Downsview Northwest | Gas Station | Grocery Store | Pizza Place | Falafel Restaurant | Fast Food Restaurant | Shopping Mall | Discount Store | Sandwich Place | Athletics & Sports | Pharmacy |
| 72 | Willowdale West | Coffee Shop | Grocery Store | Pharmacy | Park | Supermarket | Pizza Place | Greek Restaurant | Gourmet Shop | Dance Studio | Deli / Bodega |

#### cluster3[¶](#cluster3)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |  |
| 45 | "York Mills, Silver Hills" | Cafeteria | Women's Store | Doner Restaurant | Dim Sum Restaurant | Diner | Discount Store | Distribution Center | Dive Bar | Dog Run | Donut Shop |

## Conclusion [¶](" \l "Conclusion-)

In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 5 ) different clusters and for 102 different lattitude and logitude from dataset, which have very-similar neighborhoods around them.

As the model's selection, there are most venues, there are most Asian related restaurant in the cluster1. The scope covers from **Far Eastern to Middle Eastern Asia**. It reflects our expectation for tourists and new immigrants.

| **Items related to Asians** | **Counts** |
| --- | --- |
| Asian Restaurant | 1 |
| Chinese Restaurant | 3 |
| Dim Sum Restaurant | 3 |
| Dumpling Restaurant | 4 |
| Japanese Restaurant | 5 |
| Korean Restaurant | 2 |
| Middle Eastern Restaurant | 2 |
| Ramen Restaurant | 1 |
| Sushi Restaurant | 2 |
| Vietnamese Restaurant | 4 |

This project can be reused for other cities, just think about changing clustering size to adapt to your city.

Also, I have created/modify a huge quantity of function in order to adapt.

It's very far from being perfect, a lot of work can be done, other source of data can be found, but in the end the result seams to correlate with the real world, when we know the city, the area predicted seams correct.

### Libraries Which are Used to Develope the Project:[¶](#Libraries-Which-are-Used-to-Develope-th)

**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.