

# Capstone Project - Battle Of Neighbourhoods

Simka Janevska  
May,2021

## 1. Introduction

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in North York, Toronto. The aim of this project is to demonstrate how Foursquare data can be used to explore and compare neighbourhoods or cities of choice and which problems can it solve.

Lots of people are migrating to various states of Canada and needed lots of research for good place to live. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, etc.

This Project aim to create an analysis of features for a people migrating to North York to search a best neighborhood as a comparative analysis between neighborhoods.

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. Problem description

For this project, I chose a hypothetical business problem. Imagine that I have a friend, who got job offer in North York, Toronto, so he and his family are moving to North York. It is only natural, that he wants to move to a neighbourhood, which has easy access to School, Cafe, Super market, Medical shops, Grocery shops, mall, theatre, hospital, etc..

He asked me to assist him in creating a detailed analysis of the areas in North York, and choose which is best for him.

### **3. Target audience**

In the future, this approach could be used as a service, helping people moving homes to find neighbourhoods similar to the ones they have been living in and used to.

### **4. Data Description**

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)

I used North York dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

We need data about different venues in different neighborhoods of that specific borough. In order to gain that information we used "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 is 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 500 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. Venue Latitude
6. Venue Longitude
7. Venue Category

## 5. Methodology

I used Toronto dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes. I sliced the original dataframe and create a new dataframe of the North York.

All the analysis was done using Python-programming language with the following Python's libraries:

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.

## 6. Results

Folium library was used to visualize map of North York and neighbourhoods were superimposed on top of it as in the picture below:



Then I use Foursquare API to explore areas around North York and find 10 most common venues in each neighbourhood. I've set the radius of search to 500 meters around the centre of each neighbourhood and limit to 100 venues. We create dataframe and display the top 10 venues for each neighborhood:

	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	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Gas Station	Restaurant	Fried Chicken Joint	Diner	Deli / Bodega	Ice Cream Shop	Mobile Phone Shop	Bridal Shop
1	Bayview Village	Japanese Restaurant	Chinese Restaurant	Café	Bank	Discount Store	Construction & Landscaping	Convenience Store	Cosmetics Shop	Deli / Bodega	Department Store
2	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Restaurant	Sandwich Place	Butcher	Grocery Store	Juice Bar	Comfort Food Restaurant	Café	Fast Food Restaurant
3	Don Mills North	Caribbean Restaurant	Café	Gym	Japanese Restaurant	Women's Store	Diner	Comfort Food Restaurant	Construction & Landscaping	Convenience Store	Cosmetics Shop
4	Don Mills South	Coffee Shop	Clothing Store	Restaurant	Gym	Italian Restaurant	Discount Store	Bike Shop	Beer Store	Sandwich Place	Grocery Store

I've decided to use K-means clustering algorithm to group neighbourhoods into 5 distinct groups, based on top 10 venues in each neighbourhood.

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
Parkwoods	43.753259	-79.329656	3.0	Park	Food & Drink Shop	Diner	Clothing Store	Coffee Shop	Comfort Food Restaurant	Construction & Landscaping	Convenience Store	Cosmetics Shop
Victoria Village	43.725882	-79.315572	1.0	Coffee Shop	Pizza Place	Hockey Arena	Portuguese Restaurant	Women's Store	Dim Sum Restaurant	Comfort Food Restaurant	Construction & Landscaping	Convenience Store
Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Furniture / Home Store	Women's Store	Miscellaneous Shop	Boutique	Coffee Shop	Vietnamese Restaurant	Accessories Store	Sporting Goods Shop
Don Mills North	43.745906	-79.352188	1.0	Caribbean Restaurant	Café	Gym	Japanese Restaurant	Women's Store	Diner	Comfort Food Restaurant	Construction & Landscaping	Convenience Store
Glencairn	43.709577	-79.445073	1.0	Sushi Restaurant	Asian Restaurant	Pizza Place	Bakery	Japanese Restaurant	Women's Store	Diner	Comfort Food Restaurant	Construction & Landscaping

Most of the neighborhoods fall into Cluster 2 which has most variety of venues like: coffee shop, clothing store, differen kinds of restaurants, park etc. Cluster 1 has most common venue fabric shop, Cluster 3 has one most common venue as food truck, and Cluster 4 and five has park as most common venue. Now we can conclude that best place to live in North York will be cluster 2.

## 7. Conclusion

Foursquare data and K-means algorithm, alongside with folium mapping tools can be powerful tools helping people living in large cities to make educated decisions. This approach can be replicated for various other tasks, e.g. choosing a restaurant or a store location.

## 8. References

[https://github.com/Simbo632/Coursera\\_Capstone/blob/main/Final%20notebook%20week5.ipynb](https://github.com/Simbo632/Coursera_Capstone/blob/main/Final%20notebook%20week5.ipynb)