

## Final Report – Capstone Project

### The Battle of Neighborhoods Finding a Better Place in North York, Toronto

#### 1. Introduction

The goal 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. Many people are migrating to various states of Canada and needed lots of research for good housing prices. This project is for those people who are looking for better neighborhoods for example ease of accessing to cafe, school, supermarket, 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. 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 waste water and excrement conveyed in sewers and recreational facilities.

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

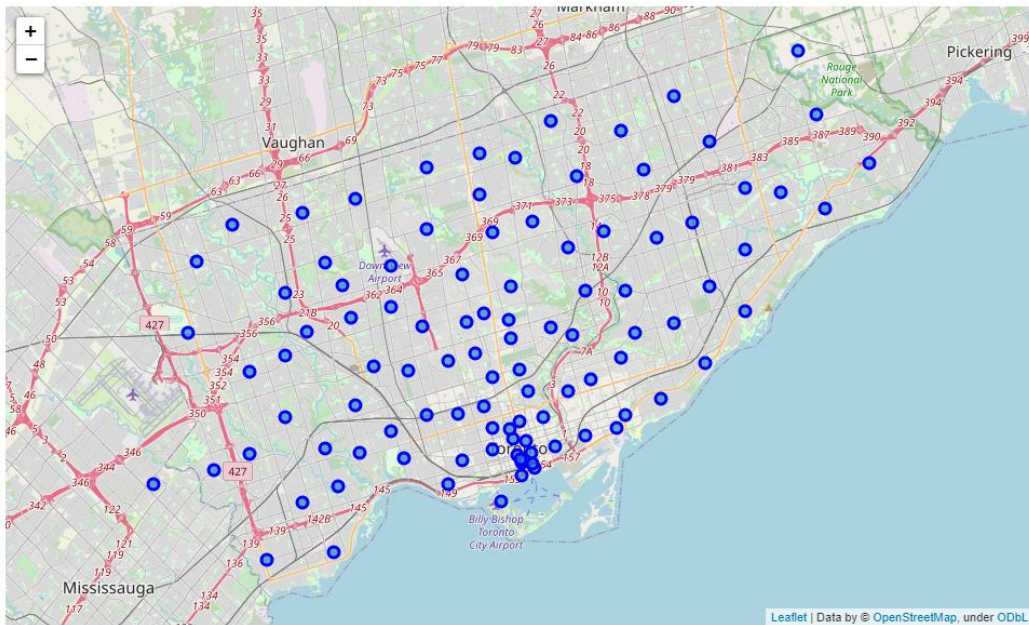
Data source: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

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

## Map of North York, Toronto



### 3. Methodology

#### 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

## K-Means Clustering Approach

```
In [81]: # Using K-Means to cluster neighborhood into 3 clusters
northyork_grouped_clustering = northyork_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters = 3, random_state = 0).fit(northyork_grouped_clustering)
kmeans.labels_
```

```
Out[81]: array([2, 2, 2, 1, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 1, 0, 0, 2, 2, 0,
0])
```

```
In [82]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

northyork_merged = neighborhoods_ny.iloc[:,:]

# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
northyork_merged = northyork_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

northyork_merged.dropna(inplace = True)
northyork_merged.head()# check the last columns!
```

```
Out[82]:
```

	PostalCode	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
0	M2H	North York	Hillcrest Village	43.803762	-79.363452	2.0	Fast Food Restaurant	Chinese Restaurant	Sandwich Place	Restaurant	Tennis Court	Bakery	Pharmacy
1	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	2.0	Clothing Store	Coffee Shop	Fast Food Restaurant	Baseball Field	Women's Store	Juice Bar	Japanese Restaurant
2	M2K	North York	Bayview Village	43.786947	-79.385975	2.0	Bank	Playground	Grocery Store	Café	Japanese Restaurant	Chinese Restaurant	Intersection
4	M2M	North York	Willowdale, Newtonbrook	43.789053	-79.408493	0.0	Park	Trail	Pizza Place	Coffee Shop	Ice Cream Shop	Hotel	Home Service
5	M2N	North York	Willowdale South	43.770120	-79.408493	2.0	Coffee Shop	Korean Restaurant	Ramen Restaurant	Pizza Place	Dessert Shop	Café	Shopping Mall

## Most Common venues near Neighborhood

### Most Common venues near neighborhood

```
In [80]: num_top_venues = 10

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

columns = ['Neighborhood']
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))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = northyork_grouped['Neighborhood']

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

neighborhoods_venues_sorted.head()
```

```
Out[80]:
```

	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	Park	Coffee Shop	Bank	Bridal Shop	Deli / Bodega	Diner	Pharmacy	Pizza Place	Convenience Store	Community Center
1	Bayview Village	Bank	Playground	Grocery Store	Café	Japanese Restaurant	Chinese Restaurant	Intersection	Nightclub	Metro Station	Outdoor Supply Store
2	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Sandwich Place	Sushi Restaurant	Boutique	Restaurant	Pub	Café	Butcher	Cupcake Shop
3	Don Mills North	Japanese Restaurant	Gym	Caribbean Restaurant	Café	Accessories Store	Medical Supply Store	Park	Outdoor Supply Store	Nightclub	Moving Target
4	Don Mills South	Coffee Shop	Restaurant	Gym	Italian Restaurant	Supermarket	Dance Studio	Discount Store	Nightclub	Intersection	Clothing Store

## 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 and Discussion

### The Location:

North York is an eclectic, multicultural district home to the hands-on Ontario Science Centre and the Aga Khan Museum, with exhibits on Islamic culture in a striking modern building. In the area's north, Black Creek Pioneer Village is an 1800s living museum. Sprawling Downsview Park includes a lake, event spaces, and a flea and farmers' market, while Edwards Gardens has a greenhouse, fountains, and botanic gardens.

North York is one of the most diverse and multicultural areas in the Greater Toronto Area, 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.

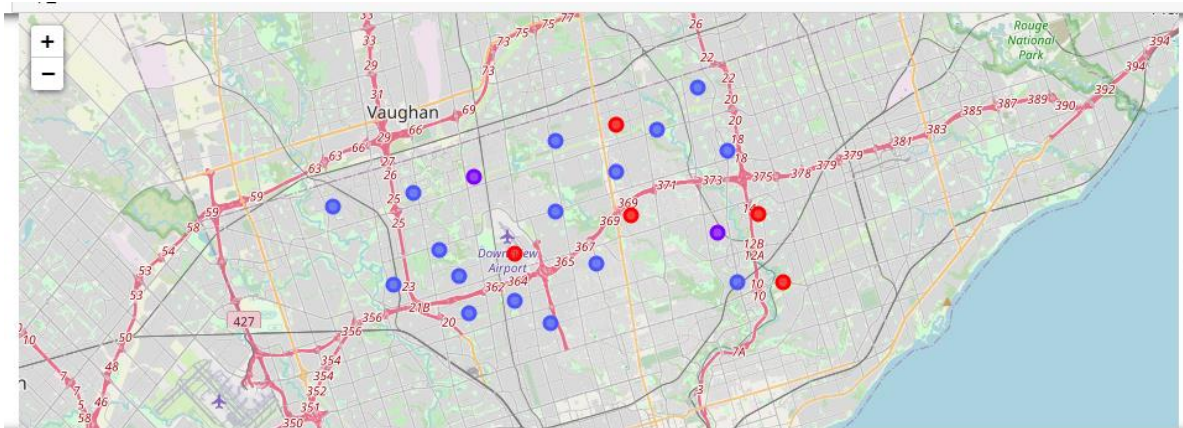
### Foursquare API:

This project has used 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.

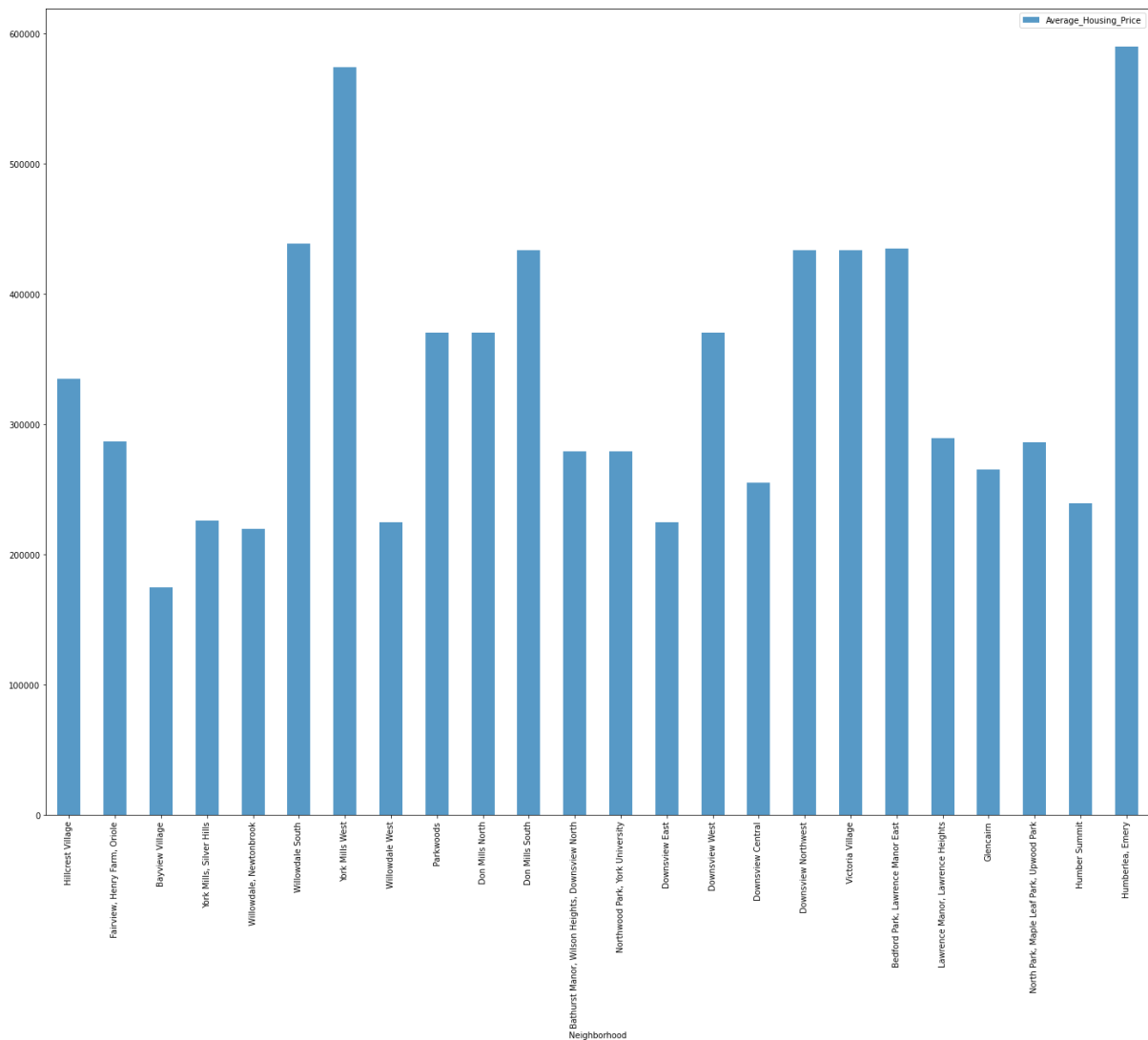
The major purpose of this project is to suggest a better neighborhood in a new city for the person who are shifting there. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

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

### Map of Clusters in North York



## Average Housing Price by Clusters in North York



## 5. Conclusion

In this project, using k-means cluster algorithm I separated the neighborhood into 10 different clusters and for 24 different coordinates 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.