

The Battle of Neighborhood Report

1. Introduction:

1.1. Background:

The main objective of this project is to help people in exploring better facilities around their neighborhood. It will help people in making a smart and efficient decision on choosing excellent neighborhoods out of numbers of other neighborhoods in Scarborough, Toronto.

Many people are migrating to many different states of Canada and need a lot of research for sound housing prices and reputed schools for their children. 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 intends to create an analysis of features for people who are migrating to Scarborough in search of the best neighborhood as a relative analysis between neighborhoods. The highlights 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.

It will help people to get the 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.

1.2. Problem:

The major goal of this project is to recommend a better neighborhood in a new city for the person who is moving 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.

- 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

1.3. The Location:

Scarborough is a favorite destination for new immigrants in Canada to reside in. As a result, it is one of the most distinct 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 constraints on immigrants and refugees, the general trend of immigration into Canada has been one of the rises.

1.4. Foursquare API:

This project would use Four-square API as its prime data collecting source as it has a database of millions of places, notably their places API which provides the capacity to perform location search, location sharing and details about a business.

1.5. Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to HTTP request restrictions, the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

1.6. Clustering Strategy:

To compare the similarities of two cities, we decided to examine 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: a k-means clustering algorithm

1.7. Libraries Which are Used to Develop the Project:

- Pandas: To create and manipulate data frames.
- Folium: Python visualization library would be used to visualize the neighborhood cluster distribution of using an interactive leaflet map.
- Scikit Learn: To import 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: To Python Plotting Module.

2.Data Acquisition and Cleaning:

2.1. Data Description:

Data link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Will use Scarborough dataset which we scrapped from wikipedia on Week Dataset consisting of latitude and longitude, zip codes.

2.2. 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

2.3.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:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Name of the venue e.g. the name of a store or restaurant
- Venue Latitude
- Venue Longitude
- Venue Category

2.4. Data Cleaning:

The data preparation for each of the three sources of data is done separately. From the Canada crime data, the crimes during the most recent year (2016) are only selected. The major categories of crime are pivoted to get the total crimes per borough as per the category

```
df_2.head(10)
```

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1A\n	Not assigned\n	Not assigned\n	43.64869	-79.38544
1	M1B\n	Scarborough\n	Malvern, Rouge	43.81153	-79.19552
2	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871
3	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.76575	-79.17520
4	M1G\n	Scarborough\n	Woburn	43.76820	-79.21761
5	M1H\n	Scarborough\n	Cedarbrae	43.76969	-79.23944
6	M1J\n	Scarborough\n	Scarborough Village	43.74309	-79.23526
7	M1K\n	Scarborough\n	Kennedy Park, Ionview, East Birchmount Park	43.72861	-79.26367
8	M1L\n	Scarborough\n	Golden Mile, Clairlea, Oakridge	43.71406	-79.28412
9	M1M\n	Scarborough\n	Cliffside, Cliffcrest, Scarborough Village West	43.72360	-79.23496

The second data is scraped from a Wikipedia page using the BeautifulSoup library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form. This is important because we will be merging the two datasets together using the Borough names.

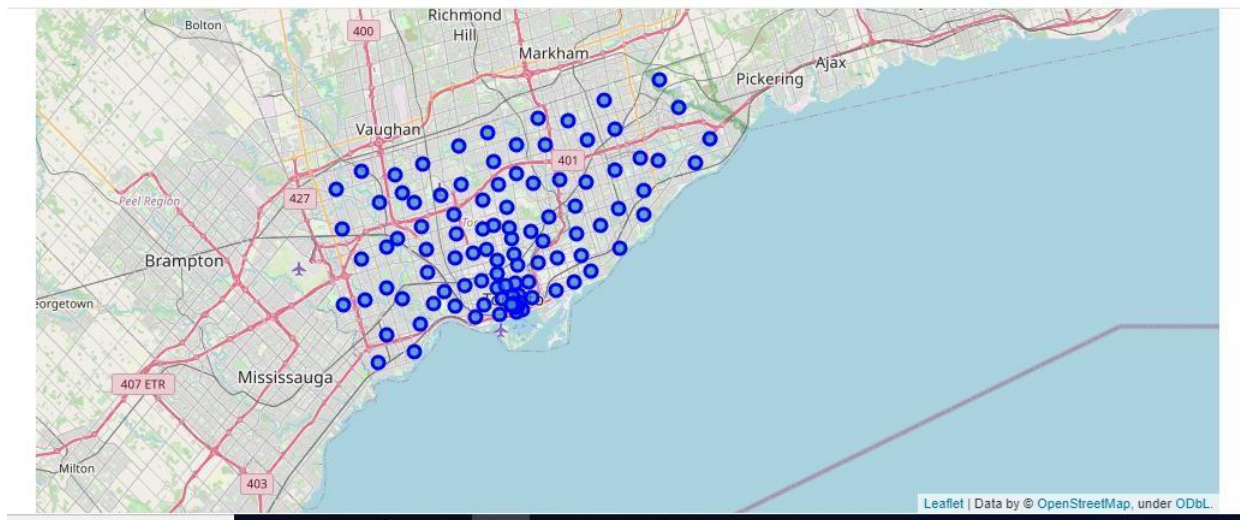
The two datasets are merged on the Borough names to form a new dataset that combines the necessary information in one dataset. The purpose of this dataset is to visualize the crime rates in each borough and identify the borough with the least crimes recorded during the year 2016.

After visualizing the crime in each borough we can find the borough with the lowest crime rate and hence tag that borough as the safest borough. The third source of data is acquired from the list of neighbourhoods in the safest borough on Wikipedia. This dataset is created from scratch, the pandas data frame is created with the names of the neighbourhoods and the name of the borough with the latitude and longitude are left blank

The coordinates of the neighbourhoods is be obtained using Google Maps API geocoding to get the final dataset. The new dataset is used to generate the venues for each neighbourhood using the Foursquare API.

3. Methodology:

Map of Scarborough:



The Geographical Co-ordinate of Neighborhood_1 are 43.773077, - 79.257774.

Nearby Venues/Locations:

Out[80]:

	venue.name	venue.categories	venue.location.lat	venue.location.lng
0	Disney Store	[{"id": "4bf58dd8d48988d1f3941735", "name": "T..."}, {"id": "4bf58dd8d48988d10c951735", "name": "C..."}]	43.775537	-79.256833
1	SEPHORA	[{"id": "4bf58dd8d48988d10c951735", "name": "C..."}]	43.775017	-79.258109
2	American Eagle Outfitters	[{"id": "4bf58dd8d48988d103951735", "name": "C..."}]	43.776012	-79.258334
3	St. Andrews Fish & Chips	[{"id": "4edd64a0c7ddd24ca188df1a", "name": "F..."}]	43.771865	-79.252645
4	Hot Topic	[{"id": "4bf58dd8d48988d103951735", "name": "C..."}]	43.775450	-79.257929

Categories of Nearby Venues/Locations:

	name	categories	lat	lng
0	Disney Store	Toy / Game Store	43.775537	-79.256833
1	SEPHORA	Cosmetics Shop	43.775017	-79.258109
2	American Eagle Outfitters	Clothing Store	43.776012	-79.258334
3	St. Andrews Fish & Chips	Fish & Chips Shop	43.771865	-79.252645
4	Hot Topic	Clothing Store	43.775450	-79.257929

```

Clothing Store    7
Coffee Shop      4
Restaurant       4
Sandwich Place   2
Tea Room         2
Gas Station      2
Pharmacy         2
Supermarket      1
Discount Store   1
Movie Theater    1
Name: categories, dtype: int64
(Count)

```

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Agincourt	28	28	28	28	28	28
Alderwood, Long Branch	10	10	10	10	10	10
Bathurst Manor, Wilson Heights, Downsview North	4	4	4	4	4	4
Bayview Village	5	5	5	5	5	5
Bedford Park, Lawrence Manor East	26	26	26	26	26	26

4. Results:

Most Common venues near neighbourhood:

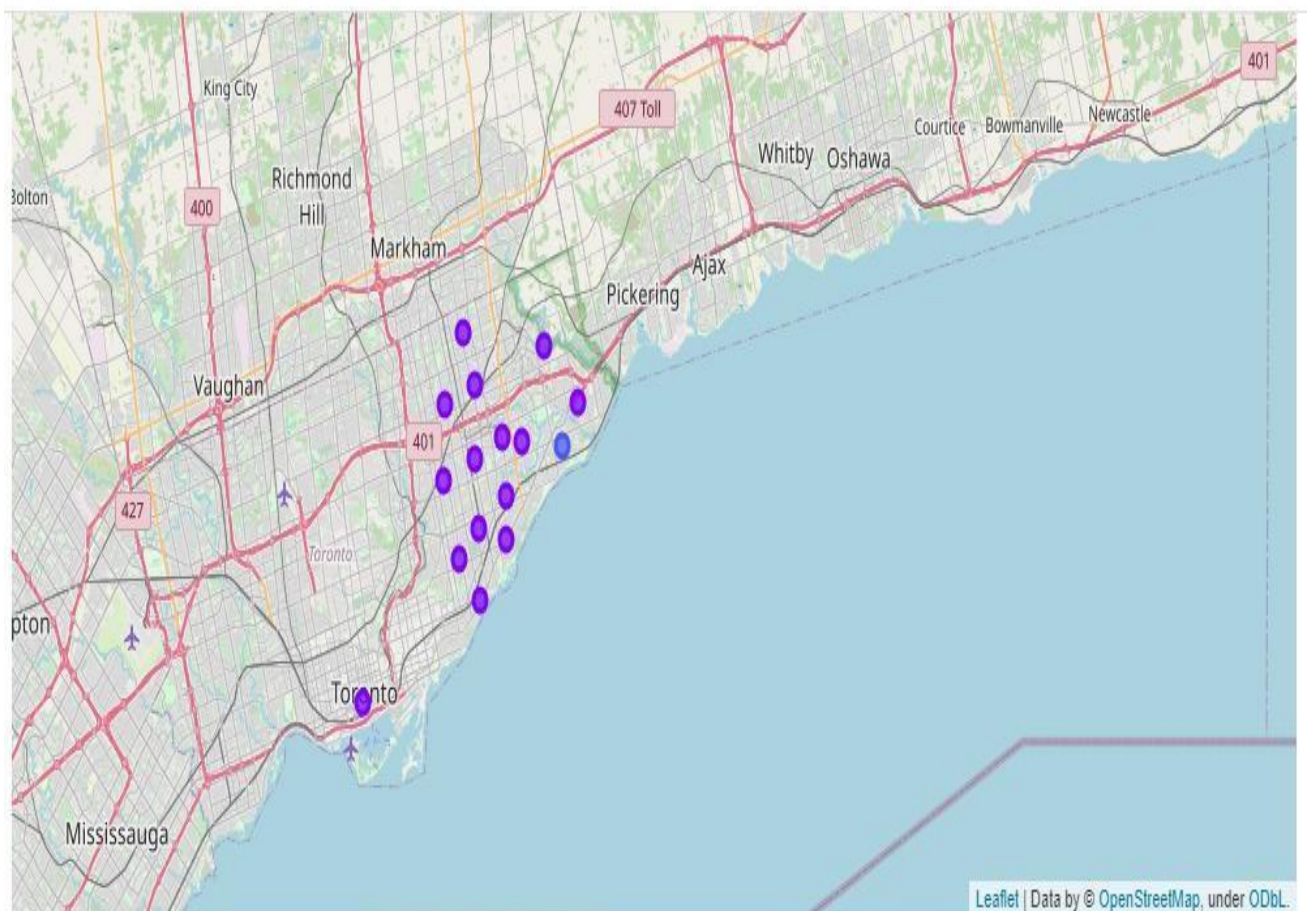
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Chinese Restaurant	Pizza Place	Bank	Pool	Bakery	Japanese Restaurant	Badminton Court	Discount Store	Sandwich Place
1	Alderwood, Long Branch	Sandwich Place	Pub	Dance Studio	Gym	Pharmacy	Coffee Shop	Print Shop	Pizza Place	Convenience Store	Gas Station
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Convenience Store	Other Great Outdoors	Yoga Studio	Event Space	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant
3	Bayview Village	Golf Driving Range	Gas Station	Park	Asian Restaurant	Trail	Yoga Studio	Dog Run	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
4	Bedford Park, Lawrence Manor East	Sandwich Place	Restaurant	Thai Restaurant	Italian Restaurant	Coffee Shop	Pet Store	Pub	Juice Bar	Sports Club	Liquor Store

Clustering approach:

	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	
0	M1A\n	Not assigned\n	Not assigned\n	43.64869	-79.38544	1	Coffee Shop	Hotel	Café	Japanese Restaurant	Bookstore	Restaurant	Be
1	M1B\n	Scarborough\n	Malvern, Rouge	43.81153	-79.19552	1	Zoo Exhibit	Fast Food Restaurant	Farmers Market	Construction & Landscaping	History Museum	Falafel Restaurant	De St
2	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	1	Bar	Golf Course	Fish & Chips Shop	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Ea Et Re
3	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill	43.76575	-79.17520	2	Park	Athletics & Sports	Gym / Fitness Center	Yoga Studio	Doner Restaurant	Dumpling Restaurant	Ea Et Re
4	M1G\n	Scarborough\n	Woburn	43.76820	-79.21761	1	Coffee Shop	Chinese Restaurant	Park	Fast Food Restaurant	Falafel Restaurant	Dumpling Restaurant	Ea Et Re

6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Restaurant	Beer Bar	Movie Theater	Arts & Crafts Store	Monument / Landmark
Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store
Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space
Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	Event Space

Maps of Clustering:



5. Conclusion:

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and longitude 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.