The Battle of Neighbourhood Report

1. Introduction:

1.1. Background:

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort?

To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move you'll end up regretting. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

1.2. Problem:

The crime statistics dataset of Canada found on Kaggle has crimes in each Boroughs of Canada from 2008 to 2016. The year 2016 being the latest we will be considering the data of that year which is actually old information as of now. The crime rates in each borough may have changed over time.

This project aims to select the safest borough in Canada based on the total crimes, explore the neighbourhoods of that borough to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering

1.3. Interest:

Expats who are considering to relocate to Canada will be interested to identify the safest borough and explore its neighbourhoods and common venues around each neighbourhood.

2.Data Acquisition and Cleaning:

2.1. Data Acquisition:

The data acquired for this project is a combination of data from three sources. The first data source of the project uses a Canada crime data from Wikipedia that shows the crime per borough in Canada.

The dataset contains the following columns:

- lsoa_code: code for Lower Super Output.
- Borough: Common name for borough.
- major category : High level categorization of crime
- Minor category: Low level categorization of crime within major category.
- value : monthly reported count of categorical crime in given borough
- year : Year of reported counts, 2008-2016
- month : Month of reported counts, 1-12

The second source of data is scraped from a Wikipedia page that contains the list of boroughs. This page contains additional information about the boroughs, the following are the columns:

- Borough: The names of the all boroughs.
- Inner : Categorizing the borough as an borough
- Status: Categorizing the borough as Royal, City or other borough.
- Local authority: The local authority assigned to the borough.
- Political control: The political party that control the borough.
- Headquarters: Headquarters of the Boroughs.
- Area (sq. mi): Area of the borough in square miles.
- Population (2013 EST.): The population in the borough recorded during the year 2013.
- Co-ordinates: The latitude and longitude of the boroughs.
- Nr. in map: The number assigned to each borough to represent visually on a map.

The third data source is the list of neighbourhoods of the place. This dataset is created from scratch using the list of neighbourhood available on the site.

- Neighbourhood: Name of the neighbourhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

2.2. 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

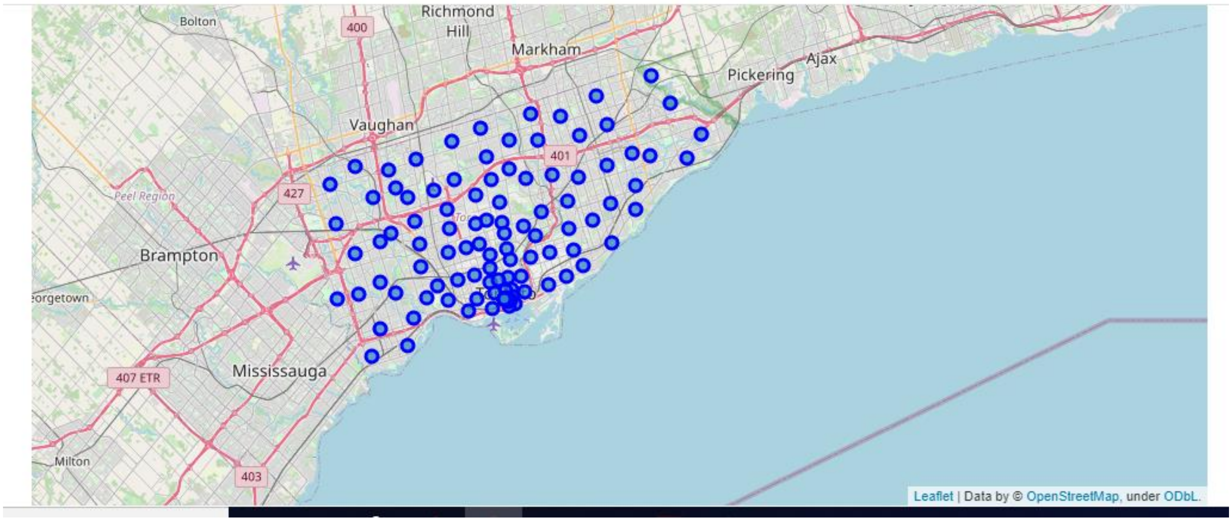
The second data is scraped from a Wikipedia page using the Beautiful Soup library in python. Using this library we can extract the data in the tabular format as shown in the website. After the web scraping, string manipulation is required to get the names of the boroughs in the correct form. This is 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...	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:

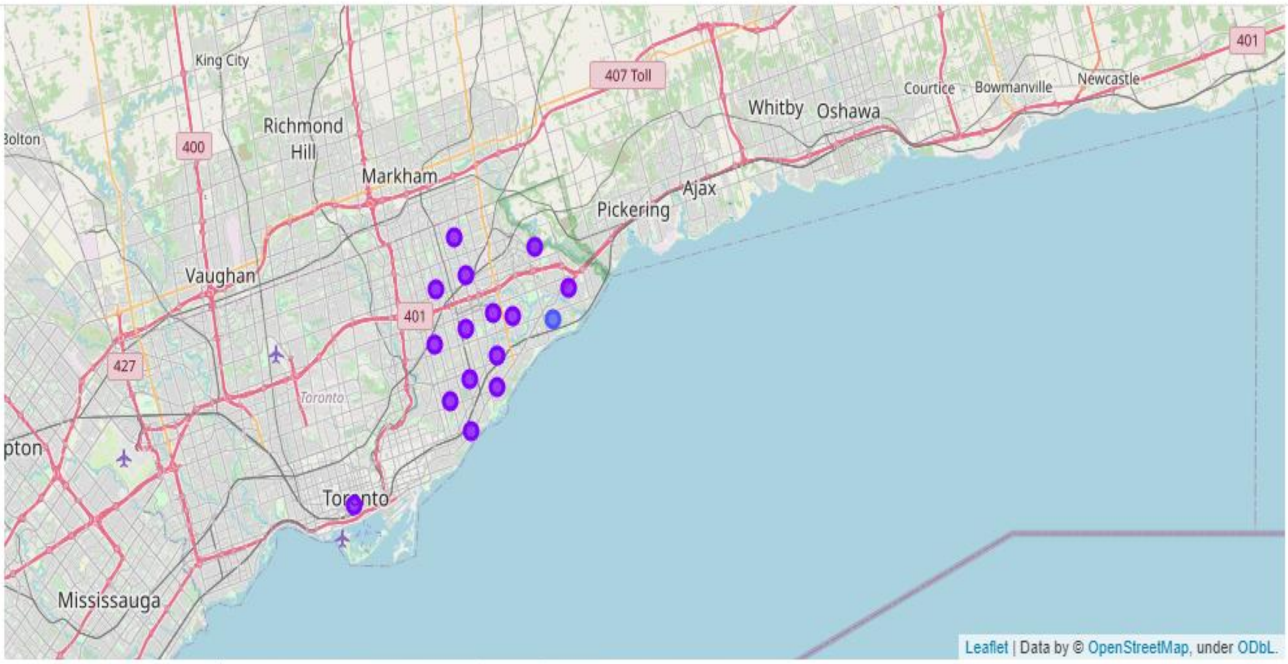
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Chinese Restaurant	Pizza Place	Bank	Pool	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	C
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	Do Sh
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 Eu 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 Eu 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 Eu 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 neighbourhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighbourhoods around them. Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made.