IBM Capstone Project
Battle of the Neighbourhoods

1. Introduction

1.1. Problem Statement

This project aims to analyse if there is a relation between the different neighbourhoods in Toronto and their homicide (murder) rates. It clusters similar neighbourhoods based on venues called by the Foursquare API by using K-Means Clustering. It then narrows down the focus of to check whether neighbourhoods with higher homicide rates have more medical facilities present to check if the rise of homicide rates tends to lead to the rise in medical facilities as well.

1.2. Interested Parties

By detecting a pattern of similarity between neighbourhoods where homicides take place one can identify whether their neighbourhood is safe or not. Oftentimes this can help potential house buyers as well as people who want to start businesses in a particular locality.It can also help the police identify higher risk neighbourhoods thus helping them keep a closer watch on them.

2. Data Acquisition and Analysis

2.1. Data Acquisition

- Homicide Records of Toronto: Acquired from the public safety data portal of the Toronto Police Service. It contains 1093 rows and 13 columns as listed below.

X	Y	Index_	Event_Unique_ID	Occurrence_year	Division	Homicide_Type	Occurrence_Date	Hood_ID	Neighbourhood	Lat	Long
-79.386620	43.648640	201	1787320	2006	D52	Shooting	2006-10-27T04:00:00.000Z	77	Waterfront Communities-The Island (77)	43.648640	-79.386620
-79.304939	43.715988	202	1470621	2006	D54	Stabbing	2006-11-05T05:00:00.000Z	43	Victoria Village (43)	43.715988	-79.304939
-79.304939	43.715988	203	1470621	2006	D54	Stabbing	2006-11-05T05:00:00.000Z	43	Victoria Village (43)	43.715988	-79.304939
-79.228691	43.809391	204	1448992	2006	D42	Shooting	2006-11-14T05:00:00.000Z	132	Malvern (132)	43.809391	-79.228691
-79.484482	43.649334	205	1945800	2006	D11	Other	2006-11-25T05:00:00.000Z	87	High Park-Swansea (87)	43.649334	-79.484482

- Foursquare API: To retrieve venues in each neighbourhood listed above (2471 rows x 7 columns) as well as specially searching and retrieving medical centres in Toronto (250 rows x 4 columns).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Agincourt North (129)	43.807407	-79.281151	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
1	Agincourt North (129)	43.807407	-79.281151	Dolly's Crafts & Sports Cards	43.809942	-79.279811	Hobby Shop
2	Agincourt North (129)	43.807407	-79.281151	Rosewood Plaza	43.810271	-79.279965	Shop & Service
3	Agincourt North (129)	43.807407	-79.281151	Reginos Pizza	43.810560	-79.280120	Pizza Place
4	Agincourt South-Malvern West (128)	43.785169	-79.257304	The Roti Hut	43.787277	-79.258724	Caribbean Restaurant

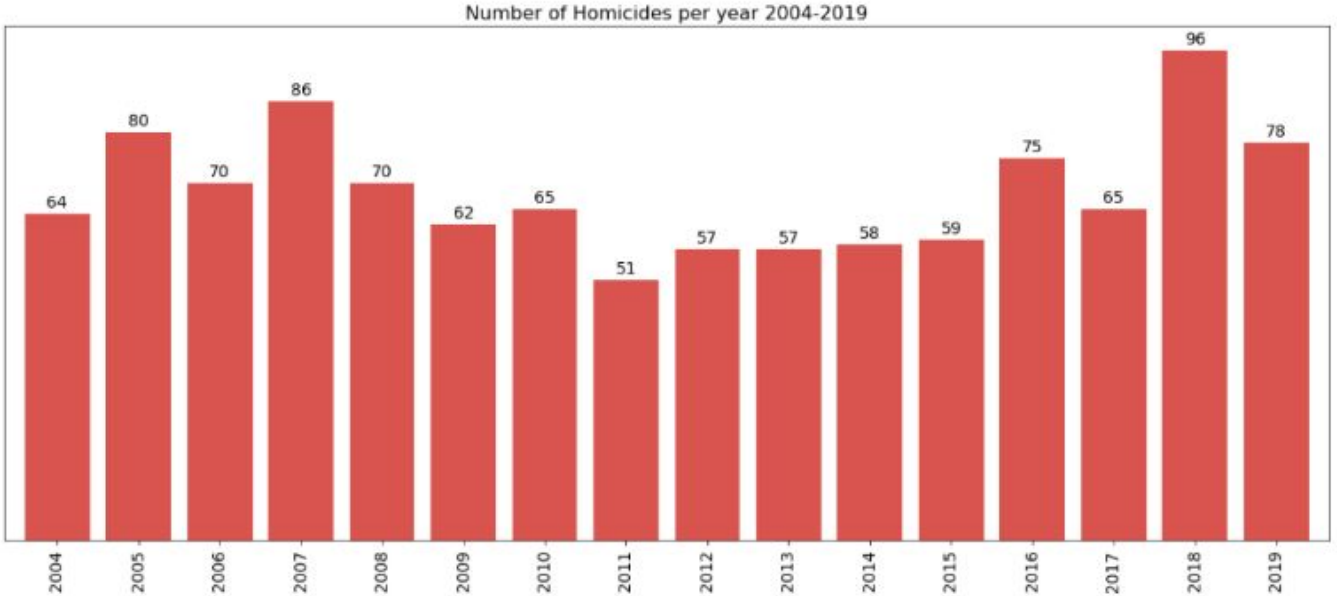
:

	name	categories	lat	lng
0	CIRA Medical Centre	Medical Center	43.655395	-79.386459
1	St Michael's Hospital Medical Imaging	Medical Center	43.653684	-79.378706
2	The Toronto Centre For Medical Imaging	Medical Center	43.649604	-79.384442
3	Oxford Medical Imaging	Medical Lab	43.648542	-79.378560
4	Carlton + Church Medical Clinic	Medical Center	43.661912	-79.379798
5	Maple Leaf Medical Clinic	Medical Center	43.661448	-79.383711
6	Medical Sciences Building	Medical School	43.661362	-79.393900
7	Queen Spadina Medical Centre	Doctor's Office	43.648578	-79.397088
8	491 Medical Centre	None	43.665475	-79.380216
9	Medical Building	Doctor's Office	43.655813	-79.386844

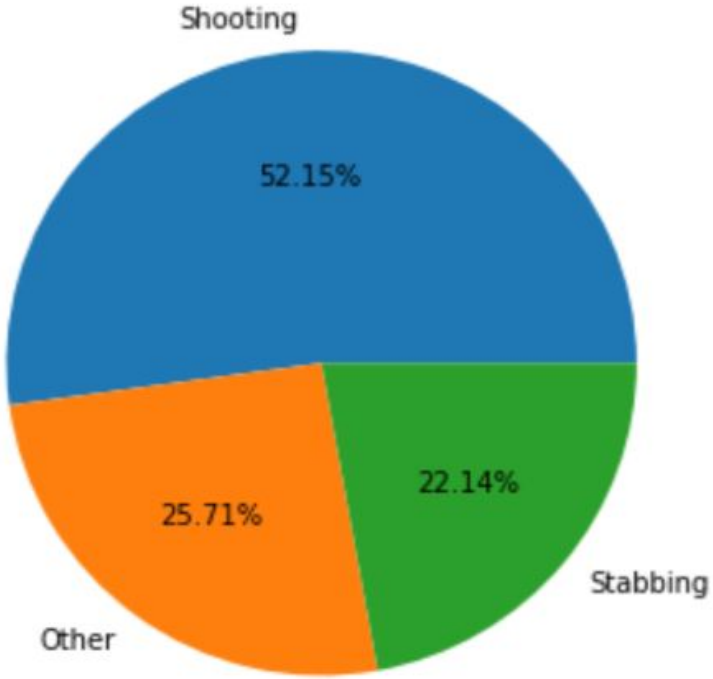
3. Methodology

3.1. Exploratory Data Analysis

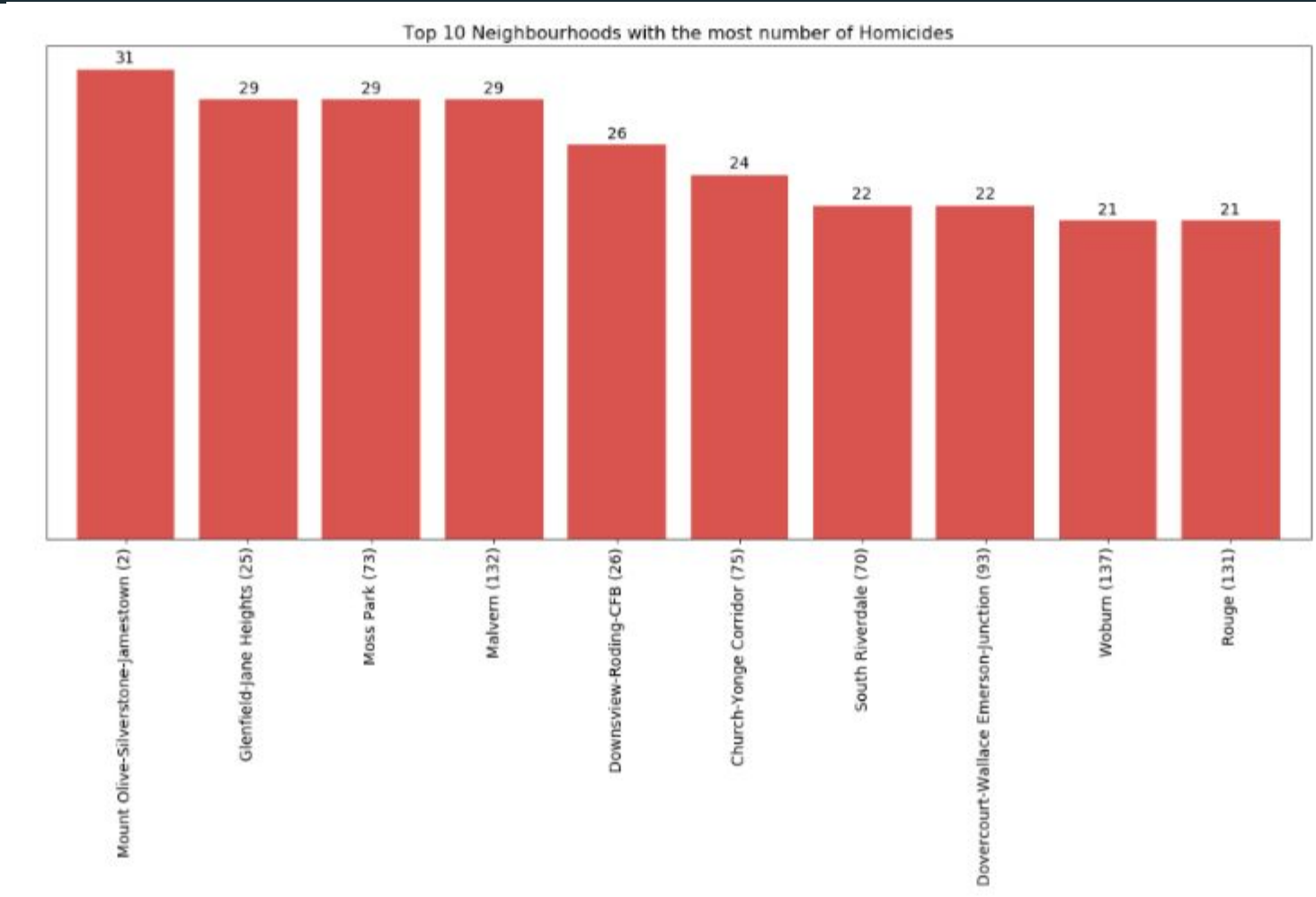
- **Occurence Years:** To visualise the total number of homicides in each year a bar graph was made. We can see that 2018 had the most number of homicides (96) while 2011 had the least (51)



- **Type of Homicide:** A pie chart was used to depict the type of homicide. The most common cause was Shooting (52.15%)



- **Neighbourhoods:** A map to plot all the neighbourhoods where a homicide has occurred using Folium. A bar graph was used to depict the 10 neighbourhoods with the highest number of homicides till date.


Top 10 Neighbourhoods with the most number of Homicides
31
29
29
29
26
24
22
22
21
21
Mount Olive-Silverstone-Jamestown (2)
Glenfield-Jane Heights (25)
Moss Park (73)
Malvern (132)
Downsview-Roding-CFB (26)
Church-Yonge Corridor (75)
South Riverdale (70)
Dovercourt-Wallace Emerson-Junction (93)
Woburn (137)
Rouge (131)

- **Different venues present in the neighbourhoods:** The number of venues retrieved for each neighbourhood was checked along with the 3 most common venues in each neighbourhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Agincourt North (129)	Caribbean Restaurant	Pizza Place	Hobby Shop
1	Agincourt South-Malvern West (128)	Gym / Fitness Center	Breakfast Spot	Sporting Goods Shop
2	Alderwood (20)	Pizza Place	Skating Rink	Pharmacy
3	Annex (95)	Coffee Shop	Italian Restaurant	Restaurant
4	Banbury-Don Mills (42)	Coffee Shop	Shoe Store	Gourmet Shop

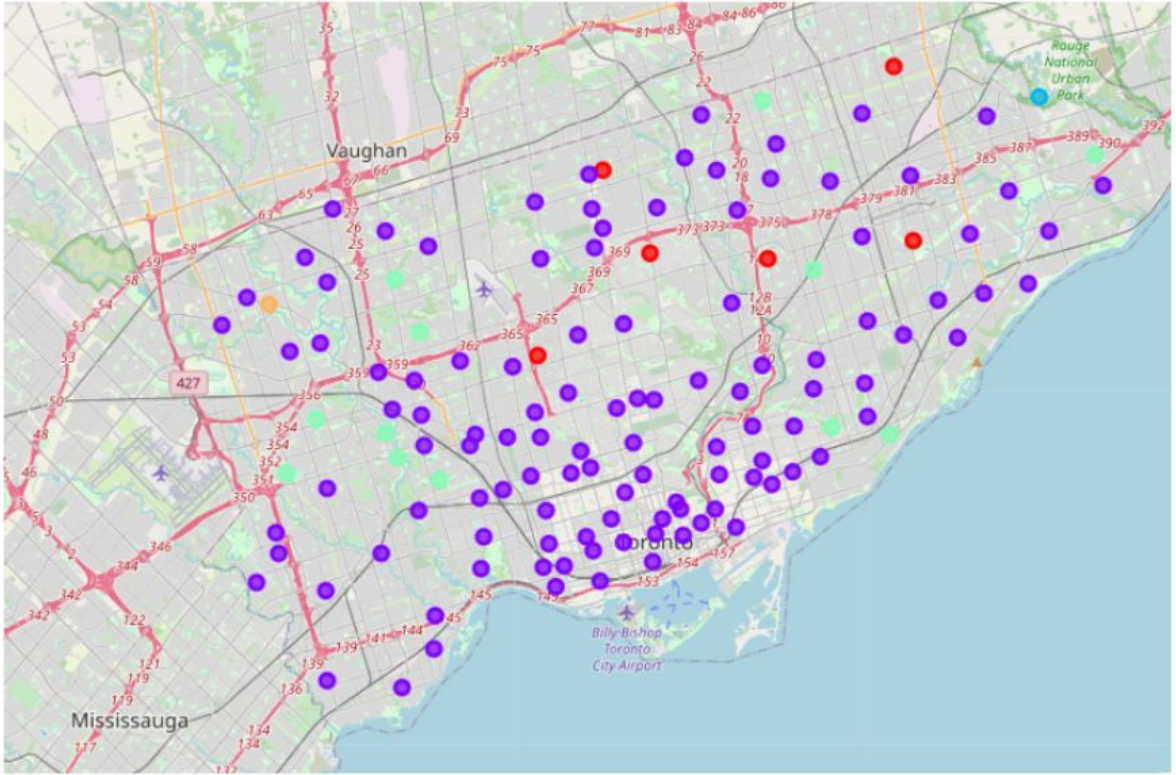
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt North (129)	4	4	4	4	4	4
Agincourt South-Malvern West (128)	13	13	13	13	13	13
Alderwood (20)	10	10	10	10	10	10
Annex (95)	61	61	61	61	61	61
Banbury-Don Mills (42)	23	23	23	23	23	23
Bathurst Manor (34)	21	21	21	21	21	21
Bay Street Corridor (76)	72	72	72	72	72	72
Bayview Village (52)	11	11	11	11	11	11
Bayview Woods-Steeles (49)	18	18	18	18	18	18
Bedford Park-Nortown (39)	22	22	22	22	22	22
Beechborough-Greenbrook (112)	4	4	4	4	4	4
Bendale (127)	4	4	4	4	4	4
Birchcliffe-Cliffside (122)	4	4	4	4	4	4
Black Creek (24)	5	5	5	5	5	5
Blake-Jones (69)	14	14	14	14	14	14
Briar Hill-Belgravia (108)	12	12	12	12	12	12
Broadview North (57)	10	10	10	10	10	10
Brookhaven-Amesbury (30)	6	6	6	6	6	6
Cabbagetown-South St.James Town (71)	55	55	55	55	55	55
Caledonia-Fairbank (109)	5	5	5	5	5	5
Casa Loma (96)	22	22	22	22	22	22
Centennial Scarborough (133)	3	3	3	3	3	3
Church-Yonge Corridor (75)	73	73	73	73	73	73
Clairlea-Birchmount (120)	7	7	7	7	7	7
Cliffcrest (123)	10	10	10	10	10	10
Corso Italia-Davenport (92)	23	23	23	23	23	23
Danforth (66)	28	28	28	28	28	28

3.2. Modelling

One hot encoding is done on the venues data. (One hot encoding is the process by which categorical variables are converted into a form that could be provided to ML algorithms for prediction.) The venues are grouped by neighbourhoods and the mean of the venues are calculated.

To find a similarity between the neighbourhoods K-Means is used to cluster similar neighbourhoods. K-Means is an unsupervised machine learning algorithm that clusters data based into a pre-defined number of clusters. Data points within a cluster are similar to each other while data points in different clusters are not. Here we decide to divide the data into 5 different clusters and visualise it on the map.

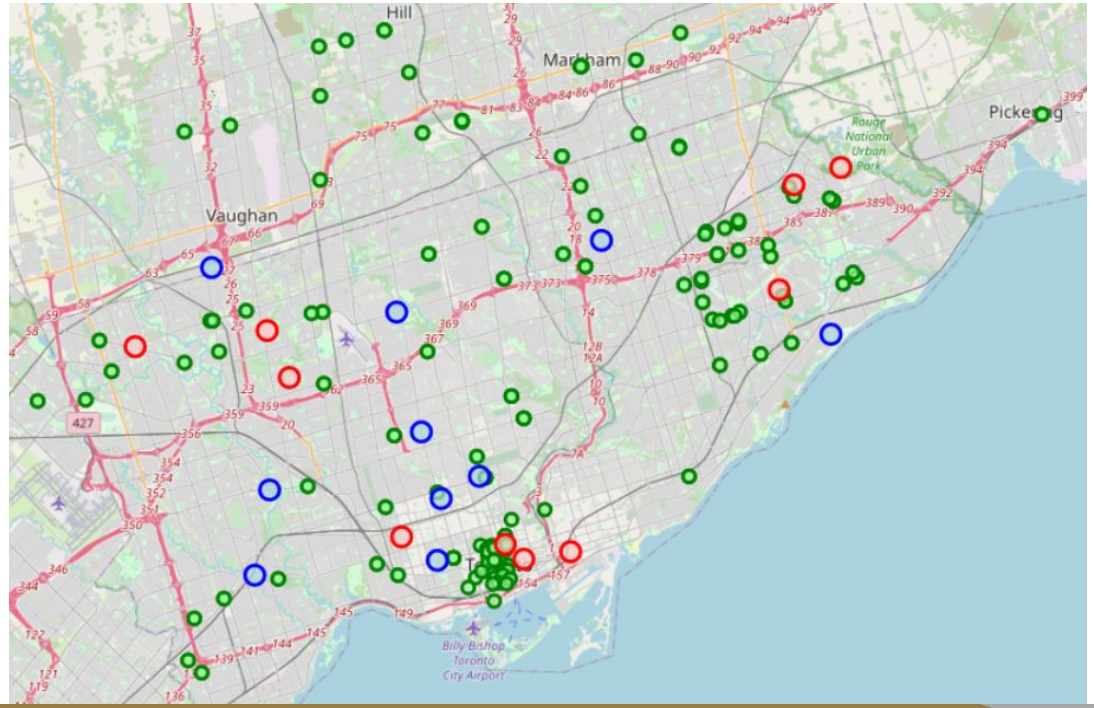
4. Results



From the map above we can see that the majority of the neighbourhoods where homicides take place belong to cluster 2 depicted in purple. Cluster 2 mostly consists of grocery stores, restaurants and pharmacies.

	Total_Homicides	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	31	Grocery Store	Discount Store	Pizza Place	Pharmacy	Beer Store	Sandwich Place	Fast Food Restaurant	Hardware Store	Gym Pool	Fried Chicken Joint
2	29	Italian Restaurant	Diner	Coffee Shop	Sandwich Place	Thai Restaurant	Café	Yoga Studio	Bus Stop	Food Truck	Breakfast Spot
3	29	Pizza Place	Fast Food Restaurant	Gym / Fitness Center	Pharmacy	Bubble Tea Shop	Restaurant	Sandwich Place	Skating Rink	Convenience Store	Grocery Store
5	24	Coffee Shop	Japanese Restaurant	Spa	Diner	Burger Joint	Thai Restaurant	Falafel Restaurant	Bar	Sandwich Place	Ramen Restaurant
6	22	Café	Bar	American Restaurant	Coffee Shop	Sandwich Place	Bakery	Seafood Restaurant	Taco Place	Falafel Restaurant	Italian Restaurant
7	22	Park	Pharmacy	Art Gallery	Bus Line	Brazilian Restaurant	Smoke Shop	Café	Bank	Grocery Store	Gym
8	21	Indian Restaurant	Bank	Coffee Shop	Bakery	Vietnamese Restaurant	Discount Store	Supermarket	Fast Food Restaurant	Ethiopian Restaurant	Event Service

We then plotted all the medical centers in the Toronto region (depicted in green) with the neighbourhoods with the most homicides (red) and the neighbourhoods with 1 homicide case(blue). The area where the density of medical centers are the highest is Toronto City which has the highest population density as well.



5. Discussion

We can observe from the clustering that the majority of the clusters have been sorted into one cluster thus depicting that neighbourhoods where Homicides have occurred are similar in nature although further analysis in different cities are required . They belong to cluster 2 which mainly consists of restaurants and pharmacies as compared to the other cluster which contains more of parks and public spaces such as museums, etc.

From the second map we cannot conclude if there are more medical facilities present near areas with higher homicides as the distribution of them seems fairly even. In most cases we do see that neighbourhoods where homicides have occured do have a fair share of medical facilities nearby but nothing concrete enough to draw up a conclusion.

6. Conclusion

Thus through this we do find a pattern between neighbourhoods where homicides have occurred based on the venues in that area.Further analysis in cities with more homicides must be done.