

Coursera Capstone Report

Introduction and Business Problem

Through this project I aim to select the safest borough in Toronto based on the number of crimes, and explore the different venues located in each neighborhood that will be clustered overall using k-mean clustering. This report will primarily target immigrants that arrive in Toronto every year in an attempt to relocate themselves and start a new life. Safety and amenities have been selected as primary goals when selecting a neighborhood and during the research I will focus on the safest borough within Toronto through examining crime statistics and explore the 10 most common venues in each neighborhood so that the best neighborhood suited to the client's need can be selected. As a student living in Toronto, Canada, I was able to observe thousands of immigrants settle down in various places across Toronto every year. Hopefully, this project will aid individuals or families when deciding on a neighborhood to select when looking for an apartment.

Data

The factors that will be put into consideration for this problem are:

- The total number of crimes that have been committed in each neighbourhood during 2004 (the dataset is based on crime rates that occurred in 2004)
- The most commonly sighted venues/amenities within the neighbourhood that was selected to be the safest

The dataset will be acquired through the following steps:

- Preprocessing the 2004 Toronto Police Homicide Data that was accessed from Kaggle
- Creating a new dataset of the selected neighbourhood in Toronto and generating their coordinates using Google Maps API Geocoding

1) Preprocessing the 2004 Toronto Police Data that was accessed from Kaggle

Toronto Crime Data

	Crime_Number	Year	Division	Crime_Type	Hood_ID	Neighbourhood
0	111878	2004	D53	Other	98	Rosedale-Moore Park (98)
1	125755	2004	D42	Shooting	137	Woburn (137)
2	136086	2004	D42	Shooting	132	Malvern (132)
3	148623	2004	D13	Shooting	93	Dovercourt-Wallace Emerson-Junction (93)
4	148619	2004	D42	Shooting	131	Rouge (131)

Figure 1 Raw Data

The data preparation was done through several steps. From the Toronto crime data provided by the Toronto police department (uploaded in Kaggle), the crimes during the year 2004 were selected. The major categories of homicide crimes were then selected to be pivoted into a new table after being processed once again as shown below in Figure 2.

	Crime_Number	Year	Division	Crime_Type	Hood_ID	Neighbourhood	No_of_Crime
0	111878	2004	D53	Other	98	Rosedale-Moore Park (98)	1
1	125755	2004	D42	Shooting	137	Woburn (137)	1
2	136086	2004	D42	Shooting	132	Malvern (132)	1
3	148623	2004	D13	Shooting	93	Dovercourt-Wallace Emerson-Junction (93)	1
4	148619	2004	D42	Shooting	131	Rouge (131)	1

Figure 2 After first preprocessing

The second data was scraped from a Wikipedia page containing additional information on the various neighbourhoods located within Toronto. Using the Beautiful Soup Library from python, a table containing extra information such the location (latitude and longitude) was extracted to be merged with the processed data earlier on.

	Neighbourhood	No_of_CrimeOther	No_of_CrimeShooting	No_of_CrimeStabbing	Total
0	Agincourt North (129)	3	1	1	5
1	Agincourt South-Malvern West (128)	1	5	1	7
2	Alderwood (20)	0	5	1	6
3	Annex (95)	1	3	4	8
4	Banbury-Don Mills (42)	1	3	2	6

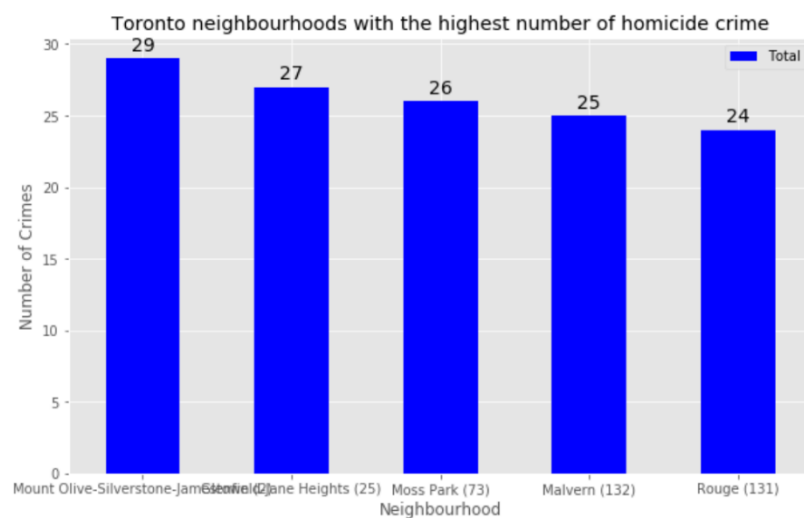
Figure 3 Pivot Table

The total number of crimes that occurred within each neighbourhood was then summed up as shown in Figure 3 to get an accurate visualization representation as shown later on in the report.

	No_of_CrimeOther	No_of_CrimeShooting	No_of_CrimeStabbing	Total
count	130.000000	130.000000	130.000000	130.000000
mean	2.023077	4.046154	1.738462	7.807692
std	2.140294	4.332787	1.654469	6.277925
min	0.000000	0.000000	0.000000	1.000000
25%	1.000000	1.000000	0.250000	3.000000
50%	1.000000	2.000000	1.000000	6.000000
75%	2.000000	6.000000	3.000000	11.750000
max	11.000000	21.000000	8.000000	29.000000

Figure 4 Description

The two datasets were merged based on the neighbourhoods and then were visualized using bar graphs to express the data in the most simplistic yet clear way (See figure 5 and 6)



The purpose of this dataset was to visualize the crime rates in each neighbourhood and identify the neighbourhood with the most and least crimes recorded during the year 2004.

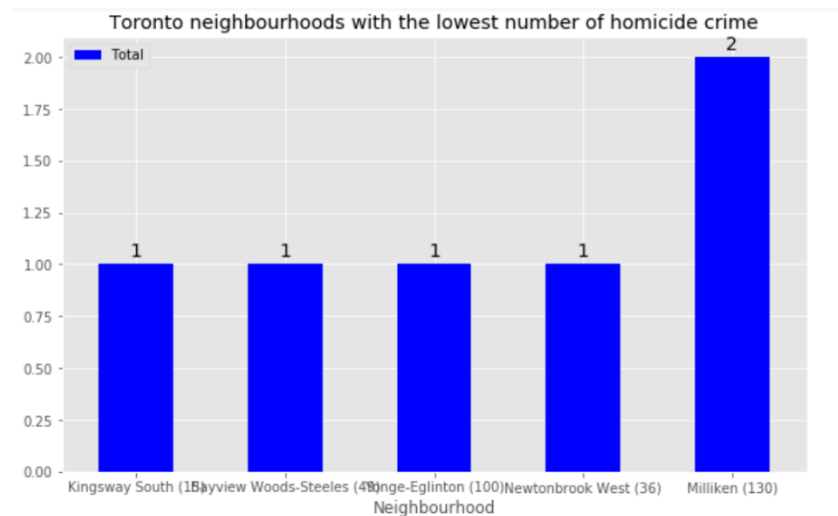


Figure 5 Final - Bottom 5 Crime Rate

After visualization, it was clear that Kingsway South, Bayview Woods-Steeles, Yonge-Eglinton, Newtonbrook West were the safest neighbourhoods with 1 or 2 homicide crimes happening in a year. On the contrary, Mount Olive-Silverstone-Jamestown, Glenfield-Jane, Heights, Moss, ParkMalvern and Rogue were the neighbourhoods classified to be the most dangerous, showing crimes rates that went up as high as 27 homicides per year.

Results and Discussion

The aim of this project was to provide information to those (immigrants) who want to relocate to the safest neighbourhood in Toronto. Through observing the 2004 homicide data provided by the Toronto police department (in Kaggle) I was able to lead to the conclusion that Kingsway South, Bayview Woods-Steeles, Yonge-Eglinton, Newtonbrook West were the safest neighbourhoods within Toronto. Data showing results of 5 neighbourhoods with the highest crime rates have also been dealt with; Mount Olive-Silverstone-Jamestown, Glenfield-Jane, Heights, Moss, ParkMalvern and

Rogue being the neighbourhoods. New immigrants arriving in Toronto will hopefully be able to choose neighbourhoods that suit their safety needs through these results

Conclusion

This project helps individuals and families get a better grasp at understanding Toronto neighborhoods with respect to crime rates as well as venues that are located within the neighborhood. It is always helpful to make use of technology to stay one step ahead, for instance, sufficiently examining and understanding a particular neighbourhood before moving in. Further progress of this project can be achieved by taking multiple other factors, such as cost of living in areas, into consideration in order to provide a much detailed and beneficial report to those in need.