Battle of Neighborhood

Vishesh Khandelwal

22/8/2020

Introduction: Business Problem

The aim of this project is to find a safe and secure location for opening of commercial establishments in Toronto, Canada. Specifically, this report will be targeted to stakeholders interested in opening any business place like Grocery Store in Toronto City, Canada.

The first task would be to choose the safest borough by analyzing crime data for opening a grocery store and short listing a neighborhood, where grocery stores are not amongst the most common venues, and yet as close to the city as possible.

We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood where grocery store is not amongst the most common venue can be selected.

### Data

#### Based on definition of our problem, factors that will influence our decision are:

#### Finding the safest borough based on crime statistics

#### Finding the most common venues

##### Choosing the right neighborhood within the borough

##### We will be using the geographical coordinates of Toronto to plot neighborhoods in a borough that is safe, and finally cluster our neighborhoods and present our findings.

#### Following data sources will be needed to extract/generate the required information:

#### Part 1 - Using the real world Data set of Toronto Crime Data from the website of Toronto Police Department. - A dataset consisting of the crime statistics of each Neighborhood in Toronto along with type of crime. (<https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-/data?geometry=-80.289%2C43.557%2C-78.436%2C43.904>)

#### Part 2 - Gathering additional information about the borough from the wikipedia about the borough in Toronto. ([https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Toronto)¶](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto)%C2%B6)

#### Part 3 - Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighborhood's co-ordinates. This data will be fetched using Open Cage Geocoder to find the safest borough and explore the neighborhood by plotting it on maps using Folium and perform exploratory data analysis.¶

#### Part 4 - Creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighborhood along with co-ordinates. This data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

### Methodology

#### Categorization of methodology section into two parts:

##### Exploratory Data Analysis: Visualize the crime reports in different Toronto boroughs to identity the safest borough and normalize the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.

##### Modelling: To help stakeholders choose the right neighborhood within a borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use K-Means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.

##### We are finding the lowest assault, robbery, theftover, autotheft cases in each borough. Following are the findings of in each type of cases.

##### Average of Lowest Assault Cases in last 5 years

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Name | Average Assault Rate | Borough | |
| 1 | Yonge-St.Clair | 31.0 | Old Toronto |
| 34 | Forest Hill North | 30.8 | Old Toronto |
| 117 | Bridle Path-Sunnybrook-York Mills | 27.8 | North York |
| 128 | Lawrence Park North | 27.7 | Old Toronto |
| 65 | Edenbridge-Humber Valley | 27.2 | Etobicoke |
| 53 | Kingsway South | 25.8 | Etobicoke |
| 31 | Princess-Rosethorn | 24.0 | Etobicoke |
| 120 | Lawrence Park South | 22.7 | Old Toronto |
| 126 | Markland Wood | 20.7 | Etobicoke |
| 55 | Forest Hill South | 18.5 | Old Toronto |

##### Average of Lowest Autotheft Cases in last 5 years

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Name | Average Autotheft Rate | Borough |
| 51 | Casa Loma | 6.8 | Old Toronto |
| 103 | Blake-Jones | 6.5 | Old Toronto |
| 28 | Ionview | 6.5 | Scarborough |
| 13 | Danforth | 6.2 | Old Toronto |
| 110 | Playter Estates-Danforth | 6.2 | Old Toronto |
| 30 | Broadview North | 5.3 | East York |
| 50 | Old East York | 4.3 | East York |
| 1 | Yonge-St.Clair | 4.3 | Old Toronto |
| 113 | Woodbine-Lumsden | 3.0 | East York |
| 37 | Guildwood | 2.7 | Scarborough |

##### Lowest of average robbery rate in last 5 years

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Name | Average Robbery Rate | Borough |
| 120 | Lawrence Park South | 5.7 | Old Toronto |
| 59 | Humber Heights-Westmount | 5.7 | Etobicoke |
| 1 | Yonge-St.Clair | 5.7 | Old Toronto |
| 51 | Casa Loma | 5.3 | Old Toronto |
| 93 | Lambton Baby Point | 5.3 | York |
| 125 | Maple Leaf | 5.2 | North York |
| 113 | Woodbine-Lumsden | 5.0 | East York |
| 128 | Lawrence Park North | 3.8 | Old Toronto |
| 46 | Bayview Woods-Steeles | 3.7 | North York |
| 117 | Bridle Path-Sunnybrook-York Mills | 3.3 | North York |

Lowest of Average Theftover in last 5 year0

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Name | Average Theftover Rate | Borough |
| 58 | Caledonia-Fairbank | 1.8 | York |
| 37 | Guildwood | 1.7 | Scarborough |
| 131 | Morningside | 1.7 | Scarborough |
| 104 | Rexdale-Kipling | 1.7 | Etobicoke |
| 14 | Rustic | 1.5 | North York |
| 103 | Blake-Jones | 1.5 | Old Toronto |
| 79 | Elms-Old Rexdale | 1.5 | Etobicoke |
| 30 | Broadview North | 1.3 | East York |
| 93 | Lambton Baby Point | 1.2 | York |
| 54 | Runnymede-Bloor West Village | 1.2 | Old Toronto |

### Based on the above exploratory data analysis we are able to find the borough which has more crime and the borough which records less crime

### Old Toronto in also known as 'DownTown' has some area (Lawrence South Park, Younge St. Claire) which records the lowest crime in each type where as there were some area(Bay street corridor, Church Yonge corridor) which records the highest crime.

### To start a grocery shop we need to find the place where there is more residency and low crime rate

### On that fact Scarborough and Old Toronto both becomes the best for the location and safety.

### Best Neighborhood venue where the average crime rate is less than 60.

| **Neighborhood** | **No.** |
| --- | --- |
| **Casa Loma** | 23 |
| **Centennial Scarborough** | 2 |
| **Forest Hill North** | 4 |
| **Forest Hill South** | 4 |
| **Guildwood** | 4 |
| **Highland Creek** | 4 |
| **Lawrence Park North** | 53 |
| **Lawrence Park South** | 53 |
| **Mount Pleasant East** | 42 |
| **Runnymede-Bloor West Village** | 40 |
| **Steeles** | 1 |
| **Yonge-St.Clair** | 57 |

### Results and Discussion

###### Based on all the analysis and everything we found that the casa loma and centennial scarborough is best for the grocery store.

###### The objective of the business problem was to help stakeholders identify one of the safest borough in Toronto, and an appropriate neighborhood within the borough to set up a commercial establishment especially a Grocery store. This has been achieved by first making use of Toronto crime data to identify a safe borugh with considerable number of neighborhood for any business to be viable. After selecting the borough it was imperative to choose the right neighborhood where grocery shops were not among venues in a close proximity to each other. We achieved this by grouping the neighborhoods into clusters to assist the stakeholders by providing them with relavent data about venues and safety of a given neighborhood.

### Conclusion

###### I have explored the crime data to understand different types of crimes in all neighborhoods of Toronto and later categorized them into different boroughs, this helped me in grouping the neighborhoods into boroughs and choose the safest borough first. Once we confirmed the borough the number of neighborhoods for consideration also comes down, we further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.