The Battle of the Neighborhoods

Yago Matiolli

December 3, 2019

1. Introduction

1.1 Background

Moving to another country is a difficult task. Trying to discover where to live, where is safe and have all to offer is no simple task. Imagine that a company is moving to another country to make a new company headquarters and many of the employers are moving with. How can we simplify the search?

First let's define what we are considering a good neighbor:

- Safe place
- Enough attributes of interest for each one (Type of restaurants, Parks and other)

2. Data acquisition and cleaning

2.1 Data sources

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

- The total number of crimes committed in each of the borough during the last year.
- The most common venues in each of the neighborhood in the safest borough selected.

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

- crimes table: http://data.torontopolice.on.ca/datasets/robbery-2014-to-2018
- Toronto Neighborhoods:https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2.2 Crimes tables: Data cleaning and feature selection

Because we are only looking for where is committed more crimes, we don't need all information that is present in the crime dataset. So, first was reduced the table for just the column that was needed.

- X, Lat: Latitude coordinates
- Y, Long: Longitude coordinates
- Reportedyear: the year when the crime was committed
- Offence and MCI: description of the crime
- Neighborhood: the neighborhood where the crime was committed.

After that the crime table was reduced for just the most recent crimes, 2018.

Table 1 - First 5 rows of the crimes table

offence	MCI	Neighbourhood	Lat	Long
Robbery - Business	Robbery	St.Andrew-Windfields (40)	43.759922	-79.389244
Robbery - Purse Snatch	Robbery	Bendale (127)	43.767086	-79.268257
Robbery - Home Invasion	Robbery	L'Amoreaux (117)	43.799000	-79.305984
Robbery - Armoured Car	Robbery	Bedford Park-Nortown (39)	43.726246	-79.417221
Robbery - Home Invasion	Robbery	Bay Street Corridor (76)	43.650280	-79.382401

2.3 Toronto Neighborhoods: Data cleaning and feature selection

For the second table we are more interested in looking for the borough than for isolated neighborhoods. So, for that the table provided via wikia was reorganized groping by the Postal code.

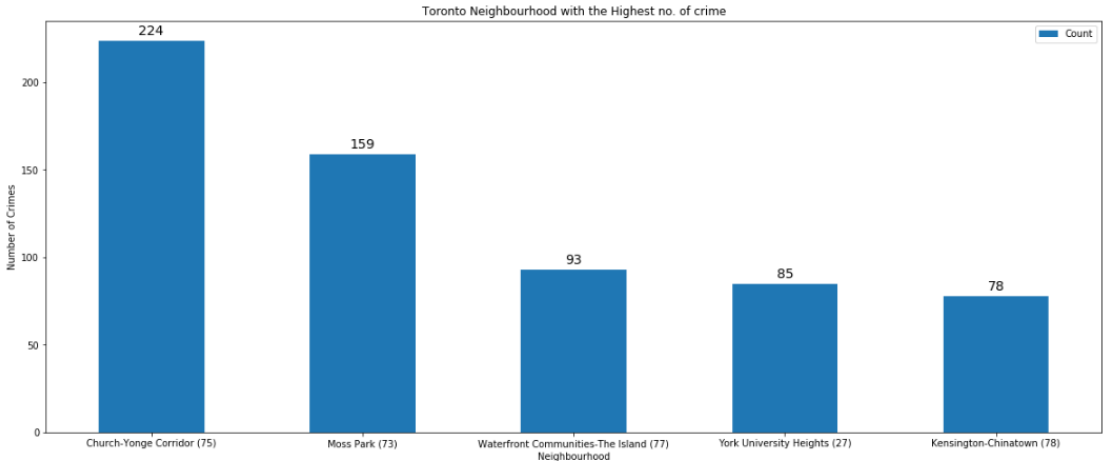
Table 2- First 10 rows of the Borough table

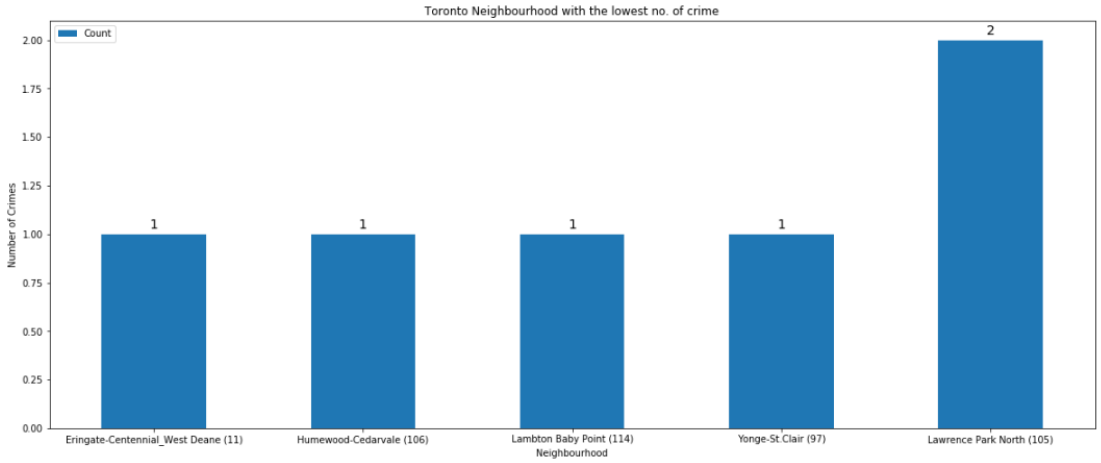
PostalCode	Borough	Neighborhood
M1B	Scarborough	Rouge,Malvern
M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
M1E	Scarborough	Guildwood,Morningside,West Hill
M1G	Scarborough	Woburn
M1H	Scarborough	Cedarbrae
M1J	Scarborough	Scarborough Village
M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park
M1L	Scarborough	Clairlea,Golden Mile,Oakridge
M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West
M1N	Scarborough	Birch Cliff,Cliffside West

3. Exploratory Data Analysis

3.1 Discovering which Neighborhood is the safest

To discover that was looked the crimes table and reorganizing it in a way that was possible to see the crimes committed during all 2018 grouped by neighborhood. After that was looked in the top 5 and the lowest 5 for more information.





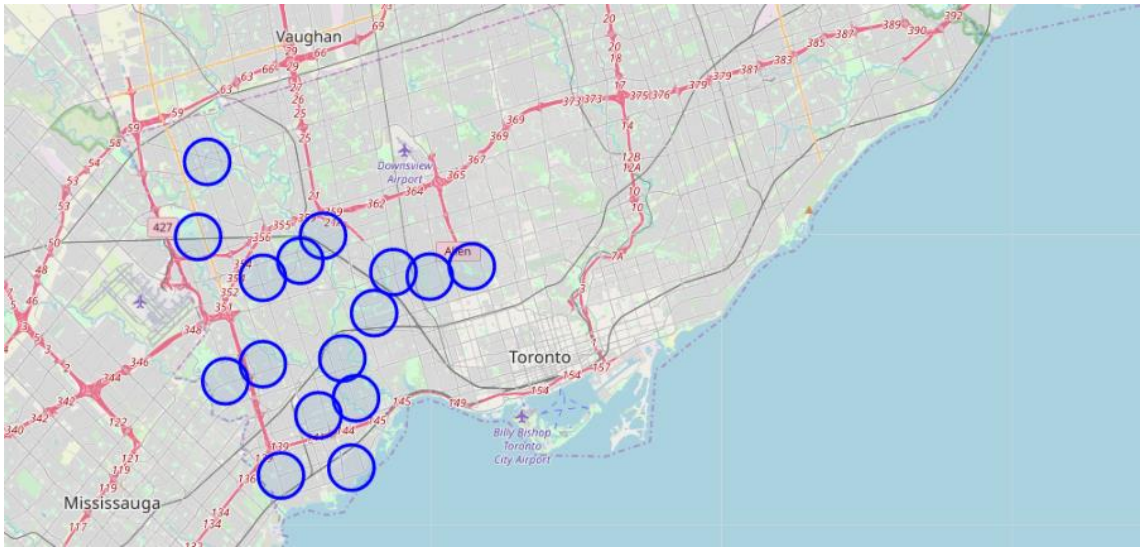
As we can see the first graph show the 5 neighborhoods with more crime in 2018, but in the second graph we have very safe neighborhoods with a total of 6 crimes in 2018 in comparison with 639 from the first.

Using the 2 safest neighborhoods was studied in which Boroughs they are part of. And was compiled all of that neighborhoods in another table.

PostalCode	Borough	Neighborhood
M6C	York	Humewood-Cedarvale
M6E	York	Caledonia-Fairbanks
M6M	York	Del Ray,Keelesdale,Mount Dennis,Silverthorn
M6N	York	The Junction North,Runnymede
M8V	Etobicoke	Humber Bay Shores,Mimico South,New Toronto
M8W	Etobicoke	Alderwood,Long Branch
M8X	Etobicoke	The Kingsway,Montgomery Road,Old Mill North
M8Y	Etobicoke	Humber Bay,King's Mill Park,Kingsway Park Sout...

M8Z	Etobicoke	Kingsway Park South West,Mimico NW,The Queensw...
M9B	Etobicoke	Cloverdale,Islington,Martin Grove,Princess Gar...
M9C	Etobicoke	Bloordale Gardens,Eringate,Markland Wood,Old B...
M9N	York	Weston
M9P	Etobicoke	Westmount
M9R	Etobicoke	Kingsview Village,Martin Grove Gardens,Richvie...
M9V	Etobicoke	Albion Gardens,Beaumond Heights,Humbergate,Jam...
M9W	Etobicoke	Northwes

Based on the table was created a map with all neighborhoods from Etobicoke and York.



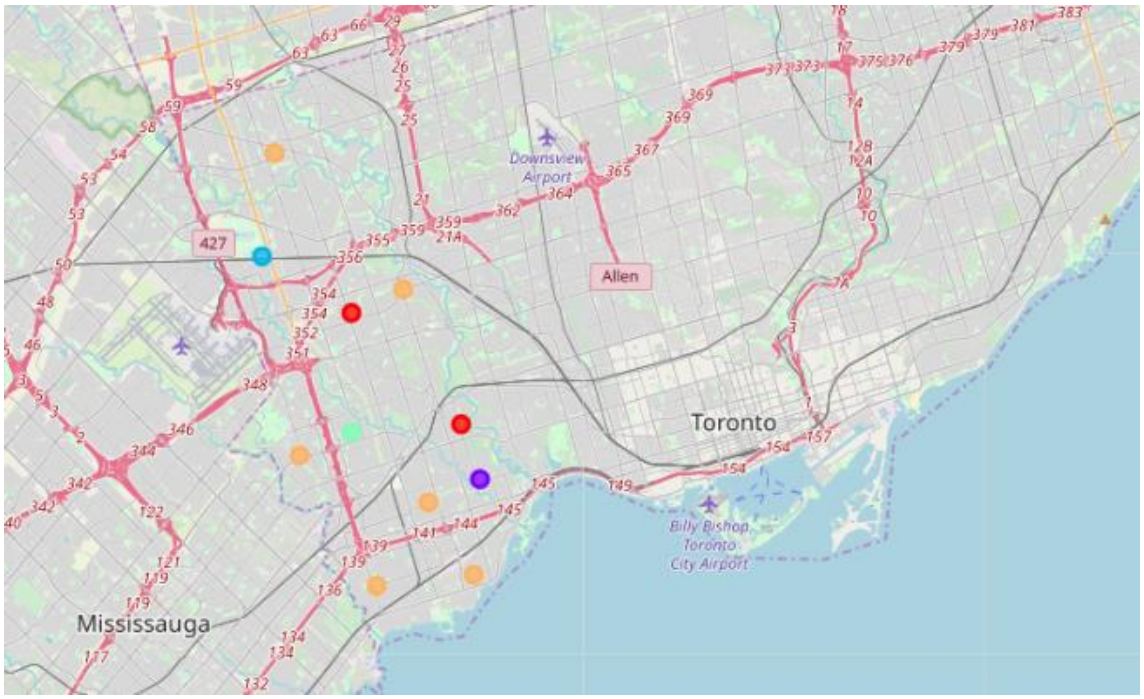
3.2 The most common venues in each of the neighborhood in the safest borough selected

Now with the borough selected is possible to see what are the most common venues in each borough. For that was used foursquare API and filtered around the neighborhood selected above.

Last the most common venues was grouped in a new dataframe and then was clustered in 5 points for Etobicoke and 4 points for York.

Here is a list of the clustered points. Notice that the venues are listed in order of most present.

3.2.1. Etobicoke



Cluster 1 (Red)

1. River
2. Park
3. Wings Joint
4. Coffee Shop
5. Fried Chicken Joint
6. Filipino Restaurant
7. Fast Food Restaurant
8. Drugstore
9. Discount Store
10. Convenience Store

1. Pizza Place
2. Mobile Phone Shop
3. Park
4. Grocery Store
5. Filipino Restaurant
6. Fast Food Restaurant
7. Drugstore
8. Discount Store
9. Convenience Store
10. Coffee Shop

Cluster 2 (Purple)

1. Locksmith
2. Baseball Field
3. Wings Joint
4. Coffee Shop
5. Fried Chicken Joint
6. Filipino Restaurant
7. Fast Food Restaurant
8. Drugstore
9. Discount Store
10. Convenience Store

Cluster 3 (Blue)

1. Drugstore
2. Rental Car Location
3. Wings Joint
4. Chinese Restaurant
5. Fried Chicken Joint
6. Filipino Restaurant
7. Fast Food Restaurant
8. Discount Store
9. Convenience Store
10. Coffee Shop

Cluster 4 (Green)

1. Filipino Restaurant
2. Wings Joint
3. Coffee Shop

4. Grocery Store
5. Fried Chicken Joint
6. Fast Food Restaurant
7. Drugstore
8. Discount Store
9. Convenience Store
10. Chinese Restaurant

Cluster 5 (orange)

1. American Restaurant
2. Pizza Place
3. Gym
4. Fast Food Restaurant
5. Liquor Store
6. Café
7. Fried Chicken Joint
8. Pharmacy
9. Restaurant
10. Sandwich Place

1. Pizza Place
2. Pharmacy
3. Athletics & Sports
4. Skating Rink
5. Sandwich Place
6. Gym
7. Pub
8. Pool
9. Coffee Shop
10. Café

1. Wings Joint
2. Tanning Salon
3. Bakery
4. Burger Joint
5. Burrito Place
6. Convenience Store
7. Discount Store
8. Fast Food Restaurant
9. Grocery Store
10. Gym

1. Pizza Place
2. Pet Store
3. Liquor Store
4. Beer Store
5. Café
6. Coffee Shop
7. Park
8. Pharmacy
9. Fast Food Restaurant
10. Drugstore

1. Pizza Place
2. Chinese Restaurant
3. Intersection
4. Sandwich Place
5. Middle Eastern Restaurant
6. Discount Store
7. Coffee Shop
8. Filipino Restaurant
9. Fast Food Restaurant
10. Drugstore

1. Pizza Place
2. Beer Store
3. Fast Food Restaurant
4. Pharmacy
5. Grocery Store
6. Fried Chicken Joint
7. Sandwich Place
8. Baseball Field
9. Burger Joint
10. Burrito Place

3.2.2. York

Cluster 1 (red)

1. Park
2. Women's Store
3. Market
4. Fast Food Restaurant
5. Trail
6. Tennis Court
7. Sandwich Place
8. Hockey Arena
9. Grocery Store
10. Field

Cluster 2 (Purple)

1. Grocery Store
2. Convenience Store
3. Bus Line
4. Breakfast Spot
5. Women's Store
6. Trail
7. Tennis Court
8. Sandwich Place
9. Park
10. Market

Cluster 3 (Blue)

1. Trail
2. Tennis Court
3. Hockey Arena
4. Field
5. Women's Store
6. Sandwich Place
7. Park
8. Market
9. Grocery Store
10. Fast Food Restaurant

Cluster 4 (Yellow)

1. Fast Food Restaurant
2. Sandwich Place
3. Discount Store
4. Check Cashing Service
5. Women's Store
6. Trail
7. Tennis Court
8. Park
9. Market
10. Hockey Arena

4. Results

Now if all that data collected an analyzed is possible to see what are the most common venues in each clustered.

Example 1: In Etobicoke cluster 3 the most common venues is a drugstore. That information could be useful for an employee that needs medicine with regular basics for him or his family.

Example 2: Imagine that an employee loves pizza and want to live nearby some place that will have a large variety of store in that regards. Looking to the clusters the best place for him would be Cluster 5 in Etobicoke.

5. Discussion

This project could use several layers with the proper datasets. With access of rental prices for example, could be add in the project the mean value of the rental in the borough mention.

6. Conclusion

It's possible to see which neighborhood is safer and where each employee will find a place that will accommodate his/her needs.

That's a simple example of the potential in data science. With a good data, and some insight is possible to understand and expand in some many aspects of our lives. Reducing cost, reducing problems, being more efficient.