Where is a right place to move in Toronto?

Jahyun Kim

Feb, 2020

1 Introduction

1.1 Background

How many times have you moved in your life? I moved six times in Korea and one more time to Canada in 2009. Meanwhile, my children have been grown and they are staying and working at Toronto area. They are spending over 2.5 hours for commutation and I spend about an hour.
As you know, moving is very hard for all family to be satisfied. I like current place, but they asked me to considering move to Toronto. Hence, I am trying to find a right place of Toronto by using my all analyzing skills and tools I have learned through this course.

1.2 The factor of consideration

I am going to consider the place as an order of priority below.

- Safety
- Neighborhood
- Time of commutation
- Housing price (?)

1.3 Expected Problems

In fact, generally I know what to do before buying a house, but mostly I relied on the realtors' advice and it was not able for me to get the detail data of safety from them. Therefore, I am going to analyze the safety, neighborhood and the market trend of housing price in Toronto as well. However, I was not sure, where I could get relevant raw data for each section and how I could manipulate the data to get the result and could use the right tools I learned for plotting location and charts. These things were my challenge.

2 Data Description

2.1 Data Sources

In order to research safety and neighborhood, the open data of Crime Rate of Toronto was should be collected and manipulated. The data files are csv format for numbers and geojson files were used to plot charts or maps to present locations. Basically, MCI_2014_2018.csv, Toronto_CRB.csv are used for numeric data and TorontoDivision.geojson, Toronto_CRB.geojson files are geometry data by referencing from the Toronto Police Service Public Safety

Data Portal (http://data.torontopolice.on.ca/search?q=crime) and used for plotting the boundary of Division and Neighborhoods. The Toronto_CRB.csv file is a summery version from the MCI_2014_2018. The location data of Police Division was referenced from the Toronto Open Data for developers (https://open.toronto.ca/dataset/police-boundaries/)

2.2 Data cleansing

The data has been grouped and cleansed by Police Boundaries such as D11, D51 in Toronto. The year of data has been used for the last two years (2017-2018) for the sampling but full dataset has been applied for creating Choropleth map of Major Crime Indicator 2014-2018. The datasets such as Assault, Break and Enter, Auto Theft and Robbery etc. have been manipulated and merged for presenting charts and incidents boundaries of maps.

2.3 Data Table

The data has been referenced from MCI_2014_2018.csv that includes 167,525 incidents of rows and selected the data only from the 2017-2018. There were 70,646 rows of incidents for 2017-2018 but the first 1,000 incidents of rows was used in order to reduce the time and catch the concept of dataset and create the map of Police Boundaries of Division.

The datasets have been reconfigured as below, and the data was grouped by Division. There are Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide data separated by columns from 2014 to 2018 but 2017-2018 data has been used, and each incident-year column has been renamed by year and merged in this table. One of incidents data is as below and the other tables of incident are same structure. The data has been manipulated and declared.

Table1. Full table of MCI_2014_2018

	ucr_code	ucr_ext	offence	reportedyear	...	MCI	Division	Hood_ID	Neighbourhood	Lat	Long
7004	1430	100	Assault	2017	...	Assault	D22	17	Mimico (includes Humber Bay Shores) (17)	43.614082	-79.492294
7008	1420	100	Assault With Weapon	2017	...	Assault	D22	17	Mimico (includes Humber Bay Shores) (17)	43.614082	-79.492294
7012	1430	100	Assault	2017	...	Assault	D33	53	Henry Farm (53)	43.773651	-79.348854
7013	1430	100	Assault	2017	...	Assault	D31	27	York University Heights (27)	43.763363	-79.500862

Table2. Assault Data Table, df_AS201718

	Division	Hood_ID	Neighbourhood	Assault_2017	Assault_2018	Total
0	D53	97	Yonge-St.Clair	46	61	107
1	D31	27	York University Heights	120	138	258
2	D32	38	Lansing-Westgate	226	197	423
3	D32	31	Yorkdale-Glen Park	124	127	251
4	D22	16	Stonegate-Queensway	112	128	240
...	...	...	...	...	...	...
135	D33	46	Pleasant View	252	260	512
136	D13	94	Wychwood	74	86	160
137	D53	56	Leaside-Bennington	134	102	236
138	D13	108	Briar Hill-Belgravia	82	90	172
139	D22	17	Mimico	25	26	51

Table3. Location Data of Police Boundaries

	AREA_NAME	LATITUDE	LONGITUDE
0	D00	43.633336	-79.368454
1	D11	43.654151	-79.465607
2	D12	43.695261	-79.493046
3	D13	43.692443	-79.439954
4	D14	43.649990	-79.417189
...	...	...	...
16	D54	43.700513	-79.325294
17	D55	43.665352	-79.323255
18	D58	43.622809	-79.381290

Table4. Grouped by Division Table

	Division	Count
0	D11	6788
1	D12	7371
2	D13	5968
3	D14	12285
4	D22	10260

Table5. Merged Data Table from six types of Incidents

	Division	Hood_ID	Neighbourhood	2017	2018	Total
0	D53	97	Yonge-St.Clair	46	61.0	107.0
1	D31	27	York University Heights	120	138.0	258.0
2	D32	38	Lansing-Westgate	226	197.0	423.0
3	D32	31	Yorkdale-Glen Park	124	127.0	251.0
...	...	...	...	...	...	...
836	D13	94	Wychwood	0	0.0	0.0
837	D53	56	Leaside-Bennington	0	0.0	0.0
838	D13	108	Briar Hill-Belgravia	1	0.0	1.0
839	D22	17	Mimico	1	0.0	1.0

Table6. Tables for Bar chart

	2017	2018	Total
Division			
D52	831	931.0	1762.0
D51	849	919.0	1768.0
D11	1252	1261.0	2513.0
D33	1383	1428.0	2811.0
D55	1654	1757.0	3411
…	…	…	…

3 Methodology

In order to get a right place to move, mostly I followed up the steps of IBM Data Science Methodology. First, my business requirements were 1. the safest place, 2. Good neighborhood 3. Short commute distance.

Second, Data was collected from the Toronto Police Service Public Safety Data Portal, and Toronto Open Data for developers.

Third, data was analyzed and transferred to Pandas Dataframe and to prepare for the next steps.

Fourth, data was manipulated to make a test model. At first, data was extracted from the year 2017 and 2018 and created proper Dataframes to create various maps. 1000 of sample data and geometry data were used to plot the police division map (Fig 1. and 1000 incidents were plotted on the map (Fig 2.) with markers and marker clustering (Fig 3-1/2.) as well. After then, Choropleth map was also created. As for selecting proper Choropleth map, it was also created by a Hood_ID instead of Division (Fig 4.)

Finally, the range was expanded to year 2014 to 2018 and created a bar chart (Fig 5.) and Choropleth map (Fig 6, 7.) the same format as sample drawing.

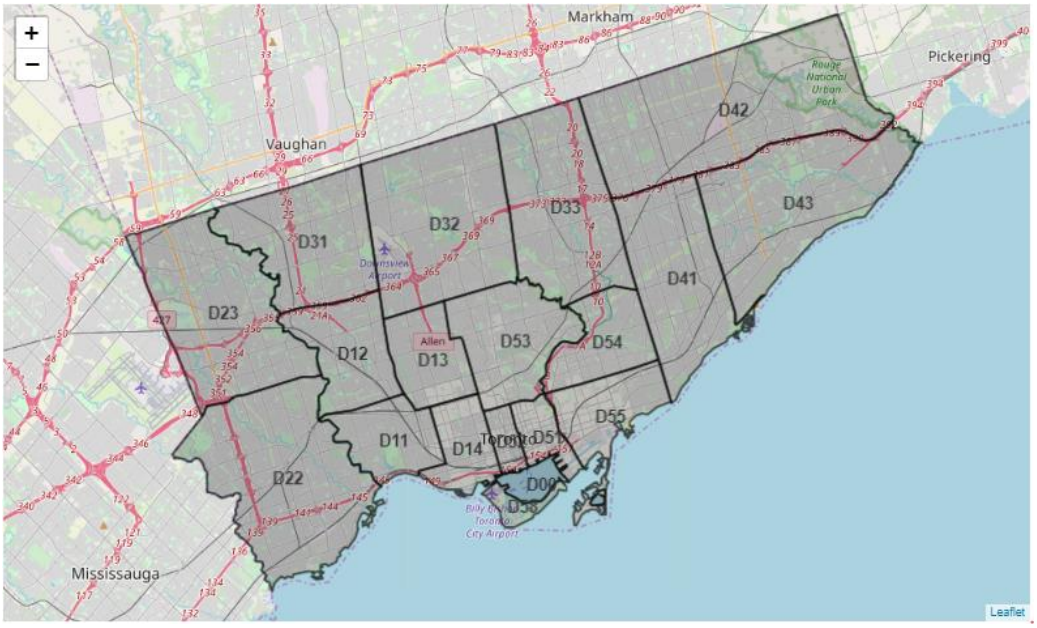

Figure 1. Police Boundary Map in Toronto

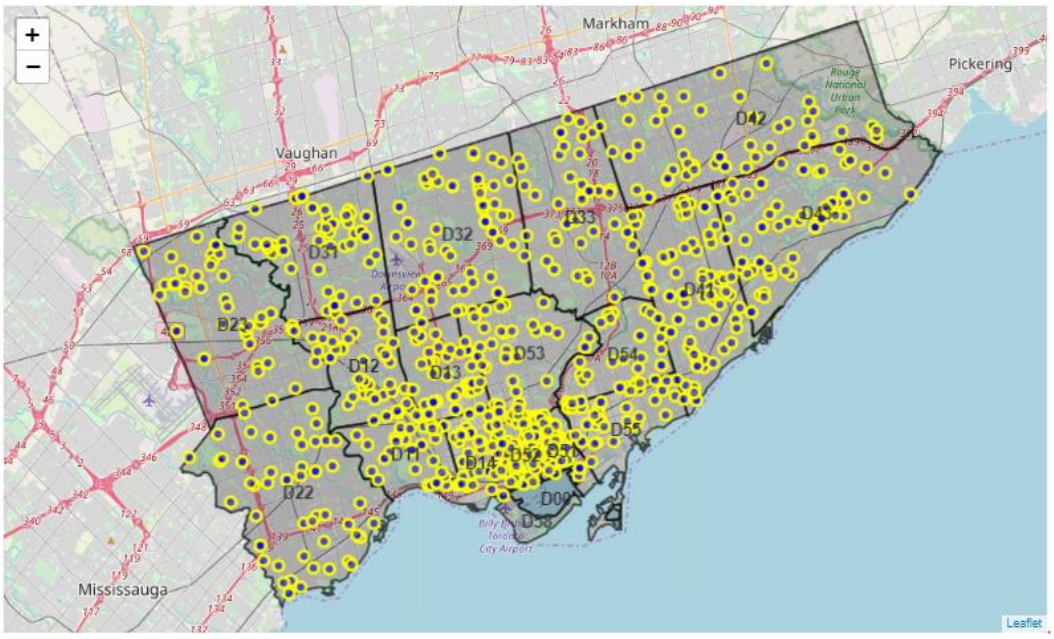

Figure 2. 1000 incidents in Toronto

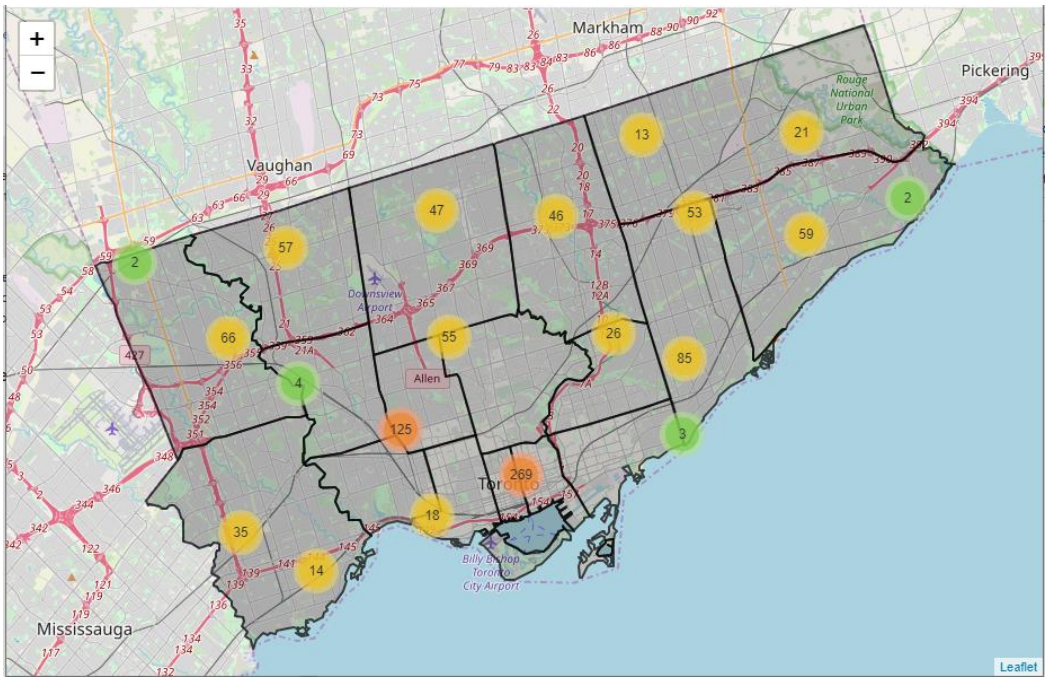

Figure 3-1. Markers Clustering in Toronto

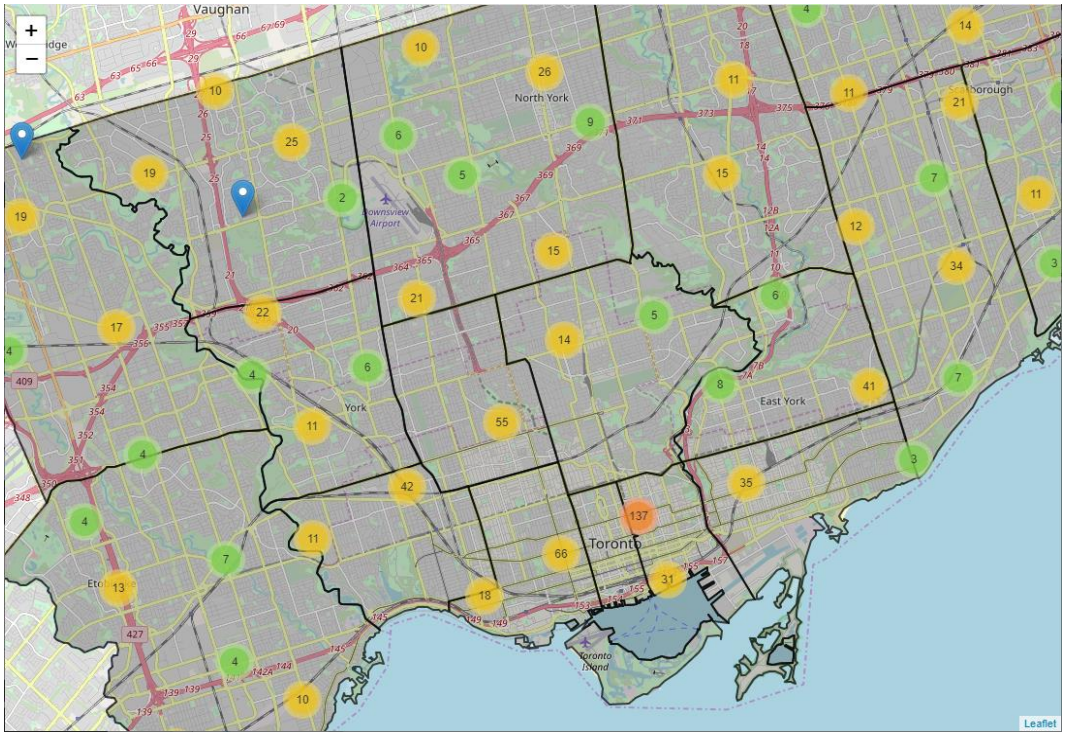

Figure 3-2. Zoom-In Markers Clustering in Toronto

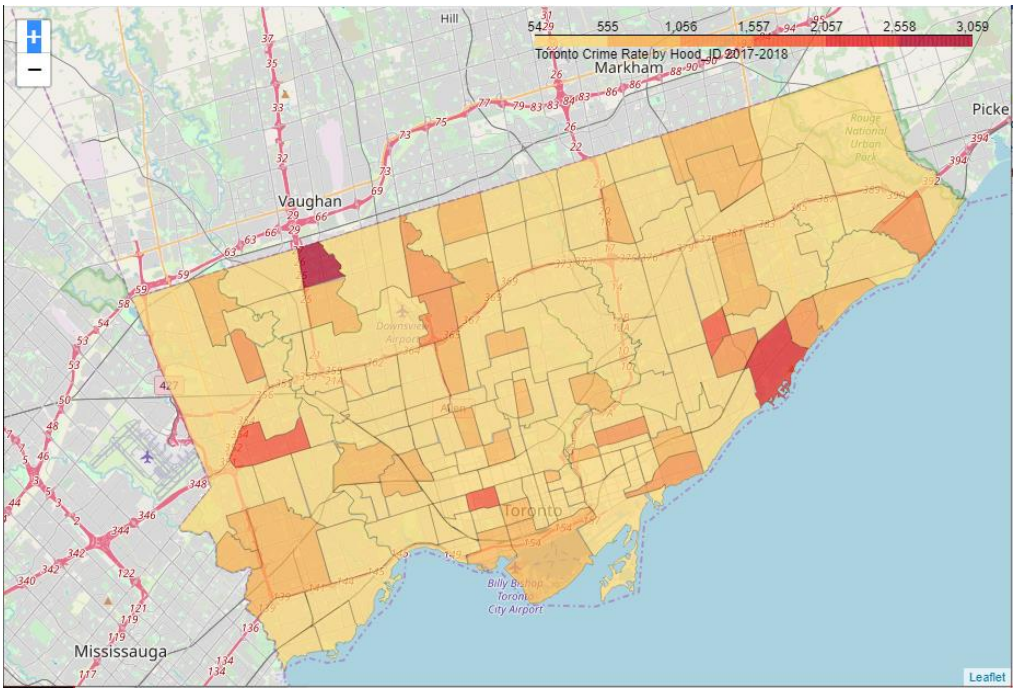

Figure 4. Toronto_Crime_by_HoodID 2017-2018 in Toronto

4 Result

The bar chart in 2017-2018 shows the Top3 low incidents in D52, D51, D11 and the Choropleth maps in 2017-2018 and 2014-2018 apparently shows the density through full divisions. It shows D52, D11, D13 look lower incidents than other divisions. I selected D11 because it is residence area that markets, and parks are close as well as my commute distance for the west is closer than D52.

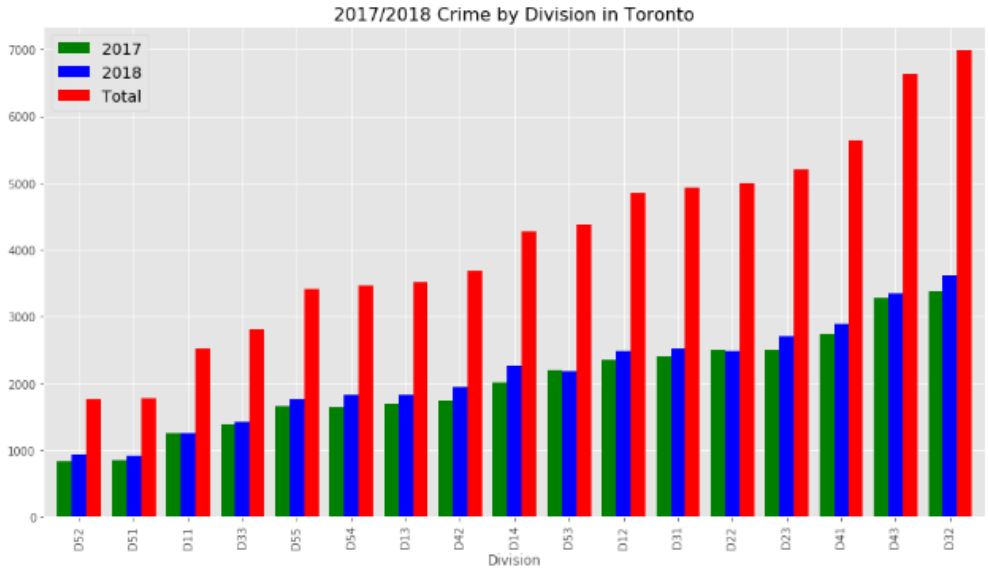

Figure 5. 2017-2018 Incidents Chart in Toronto

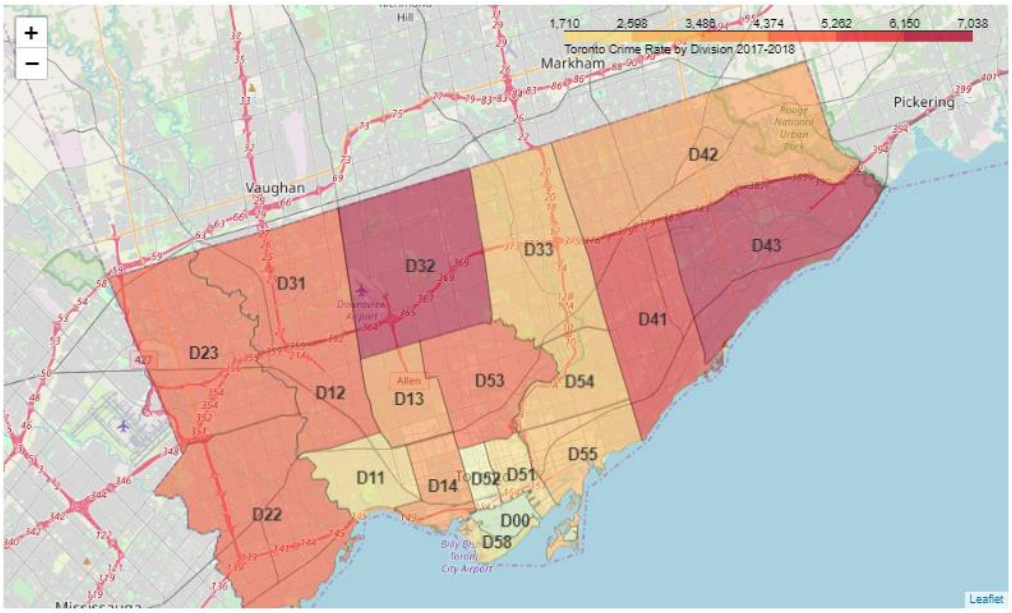

Figure 6. 2017-2018 Incidents Choropleth Map in Toronto

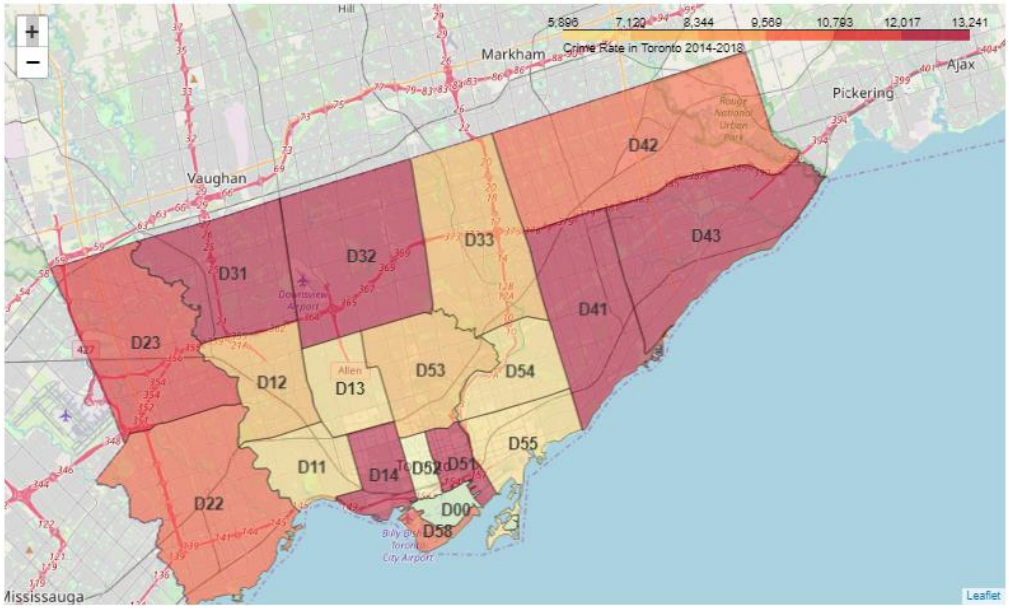

Figure 7. 2014 -2018 Incidents Choropleth Map in Toronto

The neighborhood is also important. Hence, the popular spots around the center of D11 like parks (Fig 8.) and venues (Fig 9.) are explored by using Foursquare API.

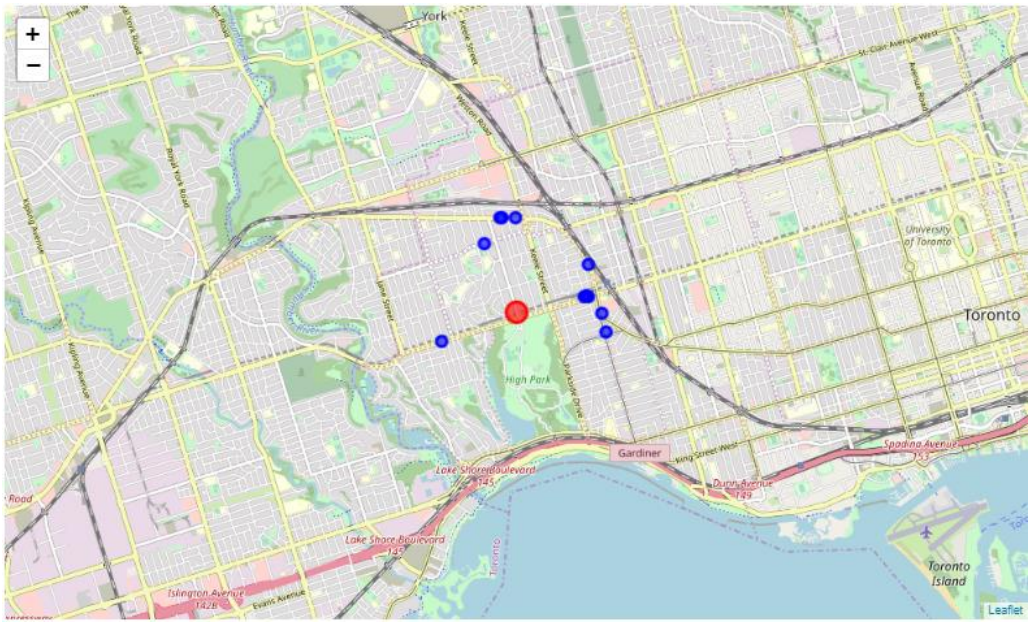

Figure 8. The Parks around the center of D11

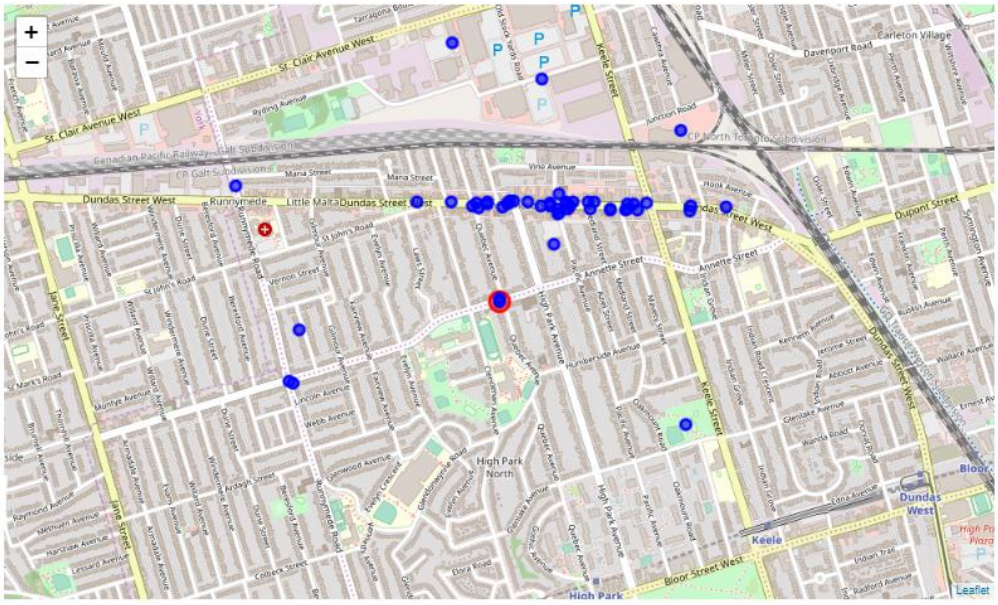

Figure 9. Venues around Annette Food Market

5 Discussion

We usually consider somethings before moving to the place you want.

- whether the place is safe
- how is neighborhood, community
- commute distance to workplace
- whether public transit is convenient

- the type of house, size, number of rooms and bathrooms
- house price etc.

However, the priority will depend on your family and your life. As for me, I considered safety first. There were 167,525 incidents in Toronto from 2014 to 2018 and tend to increase every year and spreading to over all divisions. The numbers of incidents won't easy for you to recognize the safe places. Fortunately, we learn how to plot the map by using Choropleth and it makes me to distinguish divisions and select the right places. What is your priority and how about is my approach?

6 Conclusion

In this project, I tried to get proper data for analyzing Safety and Neighborhoods in Toronto. I was able to get the data of Major Crime Indicator 2014-2018 in Toronto and its geometry data as well. Finally I got Top 3 Divisions of the lowest Crime that D52, D51 and D11 in the year 2017-2018 and the year 2014-2018 by using Bar chart and Choropleth maps. I selected the D11 from the study because it is residence area and there are many parks and markets, especially, it is easier to commute to the west than other divisions.

I used most of all my skills and tools that I learned through this course. It was challenge for me to get and manipulate the data to make proper data frames. Besides plotting maps for Choropleth was not easy as a novice like me. I had to analyze the csv and geometry data to match each other. Eventually, I made it and I got the result, where I should move in Toronto. This approaching method could be very helpful for anyone who want to move to somewhere.