Battle of Neighborhoods, Toronto

(Data Science Capstone Project Report)

Srijan Rudra

1. Introduction

People migrate for a variety of reasons. People move in search of better career and work opportunities or they may move for personal reasons. This migration may be between different cities or within the same city. It may be domestic or international. According to the World Migration Report 2020, there are 164 million migrant workers in the world.

Whatever be the cause of relocation, finding suitable accommodation in a suitable neighborhood is a crucial decision that people who move have to make. The choice of locality is determined by various factors: housing rent prices, safety of neighborhood, availability of preferred cuisine, parks and fitness clubs, leisure opportunities etc. Depending on the individual, some of these factors will carry more weight over the others. However a good decision requires individuals to balance these criteria wisely.

This project aims to help migrants with this decision. Based on above parameters, this project will try and classify neighborhoods in the city of Toronto into distinct categories. Based on those categories, people can make data aware decisions in choosing the right neighborhood for them according to their preferences.

Whom will this analysis be beneficial for?

- Housing service providers
- Rentals/ leasing agencies
- Individuals/ families moving to a new place

2. Data Collection and Usage

2.1 Data Sources

In this project, I will use the following variables to cluster neighborhoods in the city of Toronto: average rent, total crime rate, cuisine index, recreation and fitness index, and entertainment and

nightlife index. Actually data is available by Forward Sortation Area (FSA) and not particularly by each neighborhood. FSA represents the first three letters of the Canadian Postal Code eg:- M2G and may represent an area consisting of one or a few neighborhoods. In this report, I use the terms FSA and neighborhood interchangeably, though they are not identical. I have chosen the following data sources:

1. List of postal codes of Canada: M (List of all Forward Sortation Areas (FSAs) of Toronto)
2. http://cocl.us/Geospatial_data (Coordinates of all FSAs in Toronto)
3. 2016 Census Boundary files (Boundary coordinates of FSAs in Toronto)

   The above three links contain information about - names of all the neighborhoods in Toronto, their geospatial coordinates and their boundary coordinates. These will be used to set up the neighborhoods dataframe.

4. Foursquare API (https://foursquare.com)

   This API will be used to get information about different venues in a neighborhood in Toronto. This will provide data for three of the above variables - cuisine index, recreation and fitness index, and entertainment and nightlife index.

5. https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-

   This website of the Toronto Police department contains incident counts for assault, auto theft, break and enter, robbery, theft over and homicide neighborhood-wise for the city of Toronto. These numbers will be used to set up the total crime rate.

6. https://www.torontorentals.com/toronto

   This website contains listings of over 4000 properties advertised for rent in Toronto. Using web scraping, information about rent, address and neighborhood of the properties will be extracted. This will be used to evaluate average rent for a given neighborhood.

2.2. Definition of Variables

All variables are defined for each FSA/ neighborhood.

- Average Rent: This is defined as the mean rent value of all 1 bedroom apartments and accommodations as advertised on the torontorentals website. To ensure parity in comparison, I have restricted myself to only 1 bedroom accommodations.

- Total Crime Rate: This is defined as the sum of assault, auto theft, break and enter, robbery, theft over and homicide rates as given on the torontopolice website.
- Cuisine Index: This is defined as the number of different types of restaurants available in a neighborhood as returned by the foursquare search. For example, if Italian, American and Indian restaurants are available in a neighborhood, then the cuisine index will be 3.
- Recreation and Fitness Index: This is defined as the number of venues having the keywords - park, fitness, gym or sports in their category name as returned by the foursquare search.
- Entertainment and Nightlife Index: This is defined as the number of venues having the keywords - theater, movie, Karaoke bar, nightclub, strip club, music venue, rock club and other nightlife in their category name as returned by the foursquare search.

2.3. The Actual Dataframe

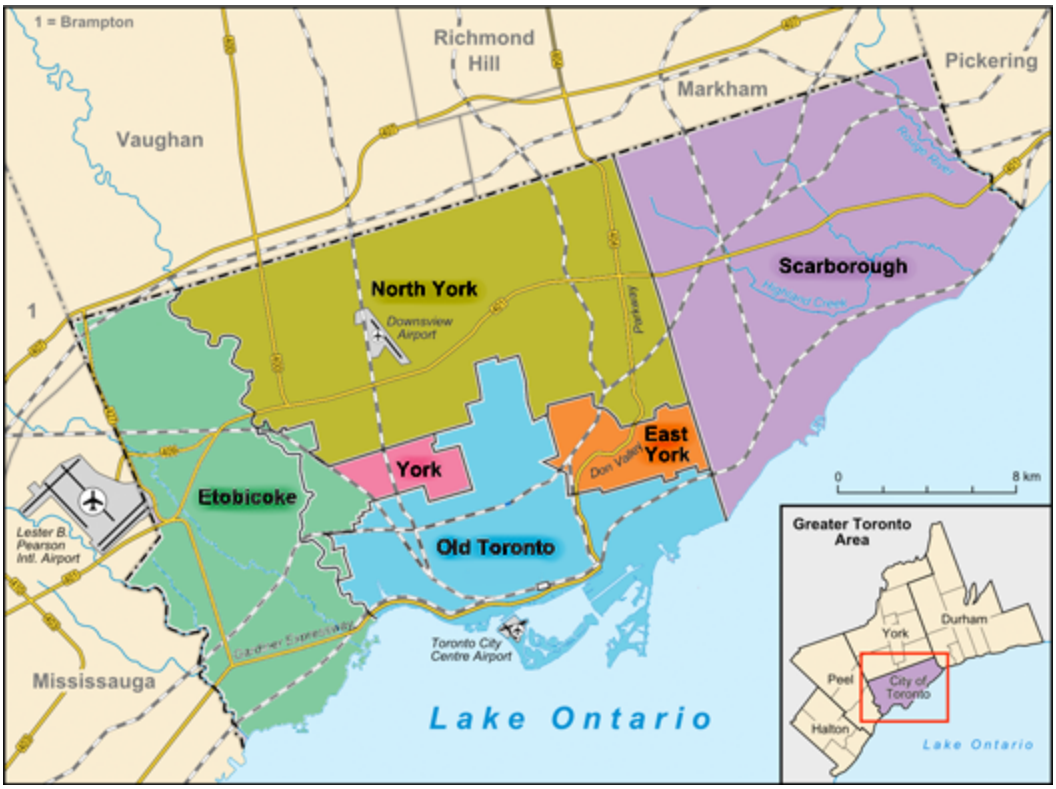
After acquisition of data from all the above sources and data cleaning the dataframe is set up which looks like this:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Average Rent	Total Crime Rate	Cuisine Index	Recreation and Fitness Index	Entertainment and Nightlife Index
0	M3A	North York	Parkwoods	43.753259	-79.329656	1988.071429	862.00000	0	8	0
1	M4A	North York	Victoria Village	43.725882	-79.315572	1725.000000	1313.70000	3	8	0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2108.398693	3853.30000	4	12	0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1862.035714	1300.37619	3	10	0
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	2242.003181	5314.50000	2	36	3
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	1917.307692	884.50000	0	6	0
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	1436.363636	1153.10000	0	4	0
7	M3B	North York	Don Mills	43.745906	-79.352188	1649.666667	798.10000	0	9	0
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	1814.583333	1354.30000	1	3	0
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2052.679245	4537.30000	15	44	9

3. Methodology

3.1. Introduction to Toronto

Toronto is the capital of Ontario in Canada. It is the most populous city of Canada and a major financial hub of the world. The territorial extent of Toronto is divided into 10 boroughs and 103 FSAs. Let us take a look at the map of Toronto, organised by its boroughs.



Source: https://wikitravel.org/en/Toronto

Here Old Toronto is further subdivided into East Toronto, West Toronto, Central Toronto and Downtown Toronto. Downtown Toronto is the part towards the shore of Lake Ontario and is the main commercial area of the city. York and East York constitute the inner suburbs, whereas Etobicoke, North York and Scarborough form the outer suburban areas.

3.2. Exploratory data analysis

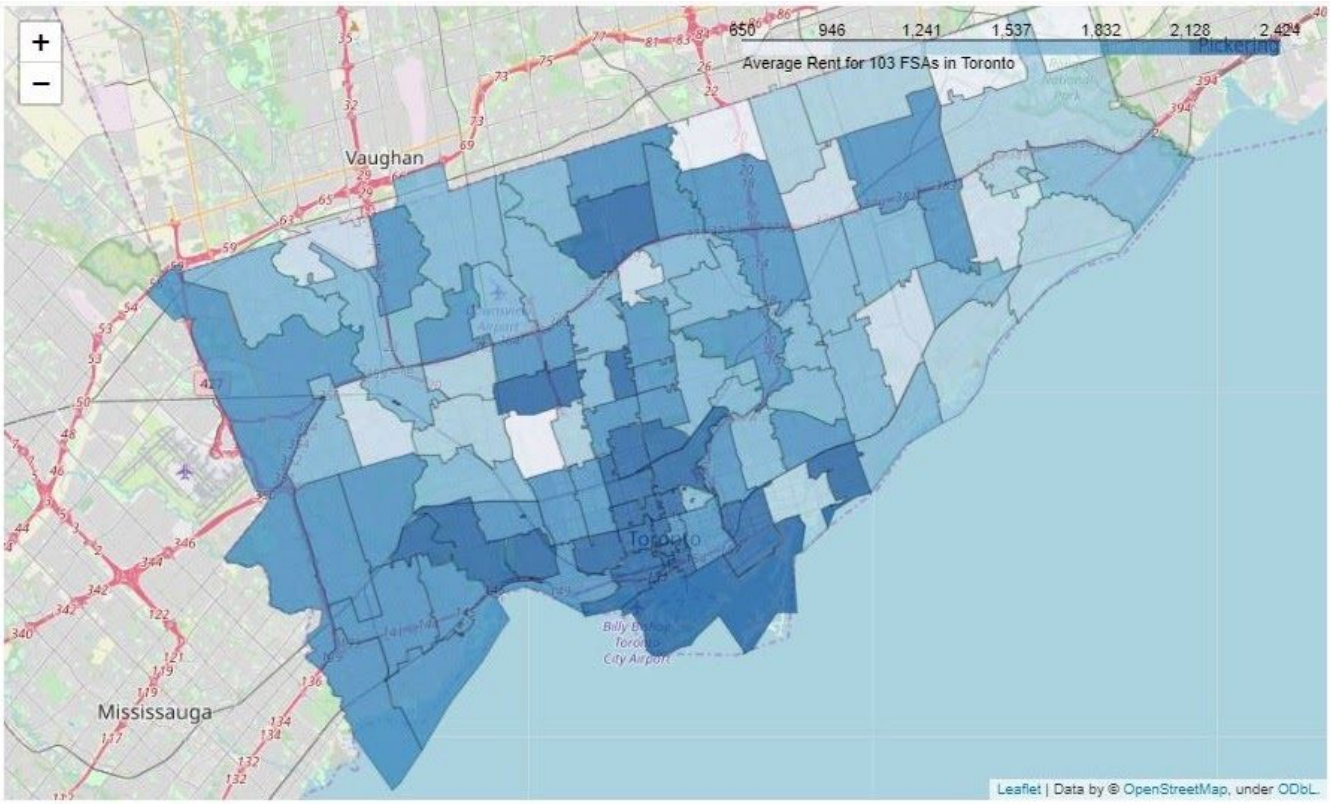
First let us broadly check the statistics for the five variables for the 103 FSAs in Toronto.

	Latitude	Longitude	Radius	Average Rent	Total Crime Rate	Cuisine Index	Recreation and Fitness Index	Entertainment and Nightlife Index
count	103.000000	103.000000	103.000000	102.000000	102.000000	103.000000	103.000000	103.000000
mean	43.704608	-79.397153	834.951456	1859.031545	1580.395572	3.883495	12.883495	0.980583
std	0.052463	0.097146	193.501653	388.499270	927.935567	3.534982	13.664970	2.196167
min	43.602414	-79.615819	500.000000	650.000000	504.200000	0.000000	0.000000	0.000000
25%	43.660567	-79.464763	750.000000	1688.750000	1010.337500	2.000000	5.000000	0.000000
50%	43.696948	-79.388790	1000.000000	1916.653846	1300.376190	3.000000	7.000000	0.000000
75%	43.745320	-79.340923	1000.000000	2146.294492	1727.300000	5.000000	14.500000	1.000000
max	43.836125	-79.160497	1000.000000	2423.739130	5314.500000	16.000000	58.000000	14.000000

Here we see the mean and spread of the variables for the city of Toronto. For example, the total crime rate has a mean value of 1860 and varies from 650 to around 2420. (Crime rate is defined as number of reported crimes per 100,000 population).

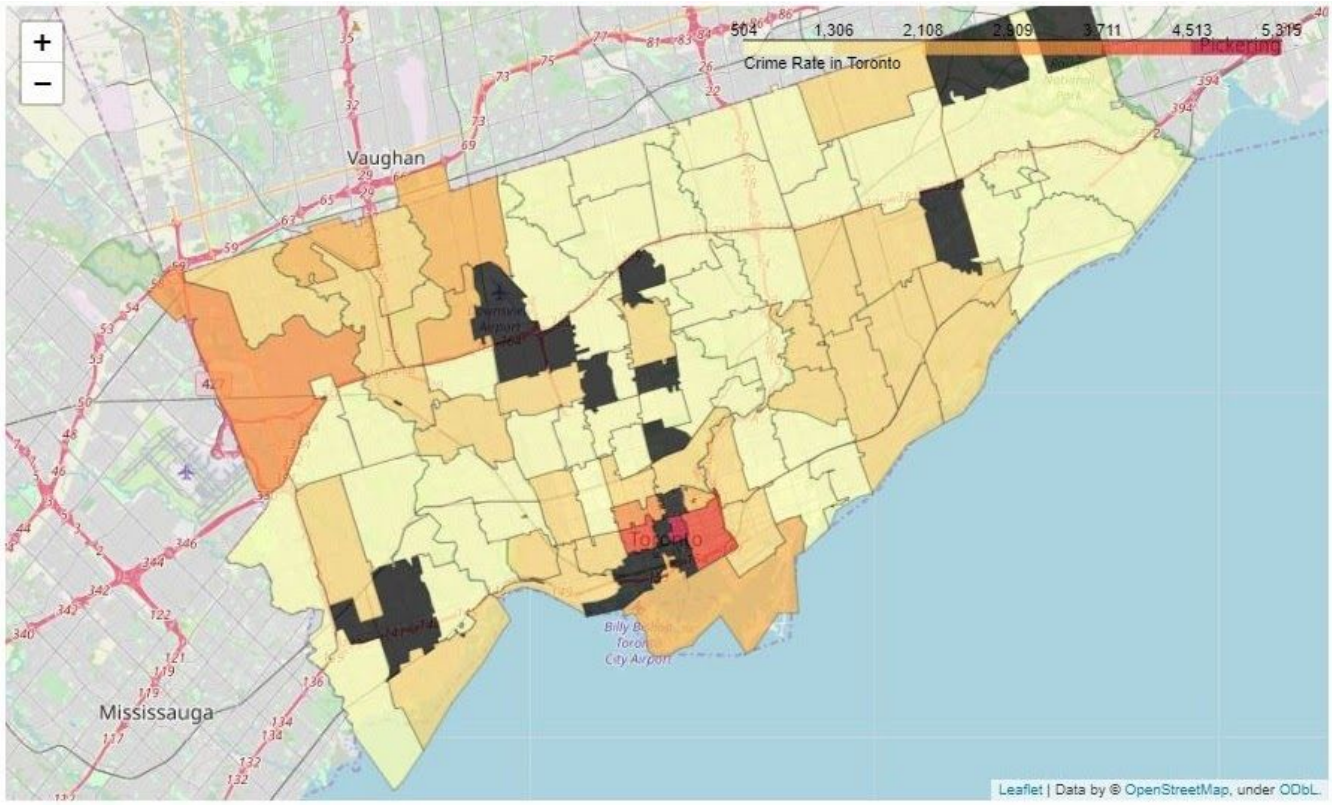
I performed exploratory data analysis for each of the five variables separately. A choropleth map was drawn for each of the five variables. These maps visually depict the spatial change in variable's value across the geography of toronto.

A) Average Rent



Here we see that Old Toronto, especially Downtown Toronto is particularly expensive. Etobicoke, despite being an outer suburb is particularly expensive whereas Scarborough on the eastern front is relatively cheaper.

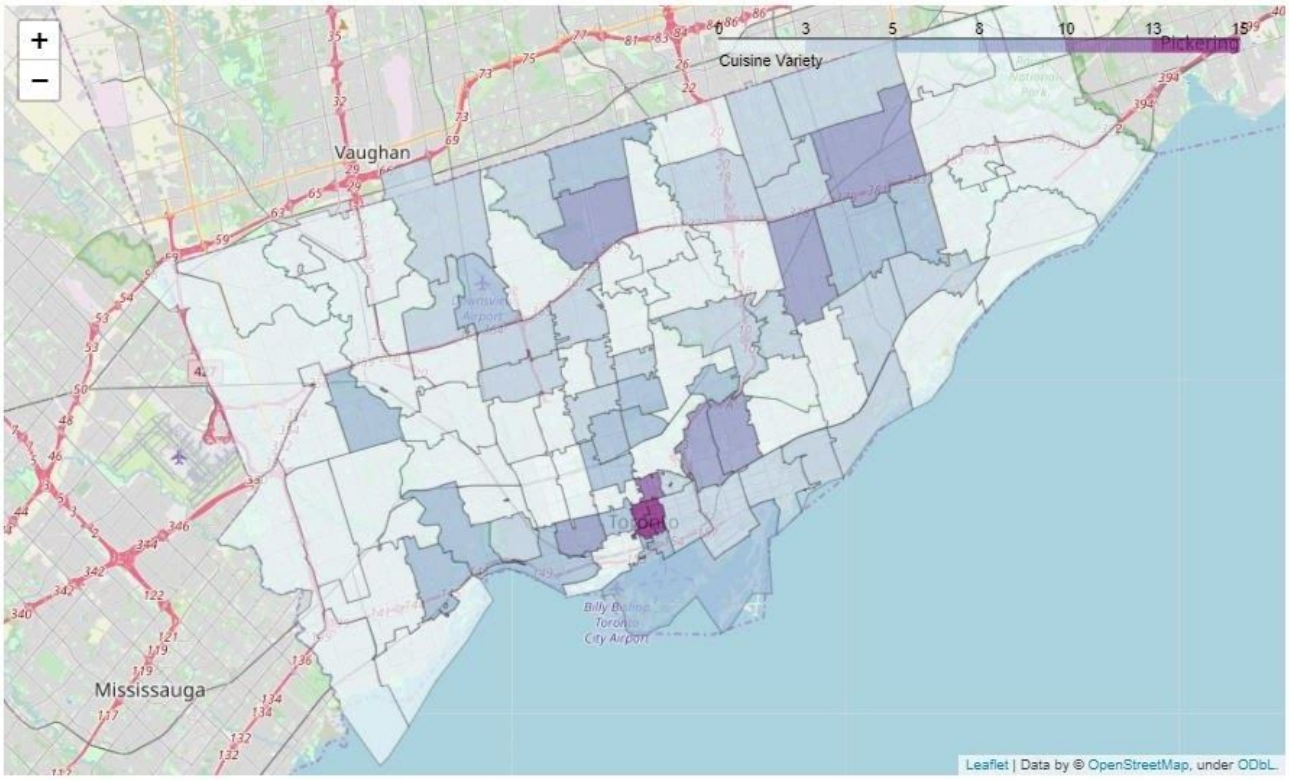
B) Total Crime Rate



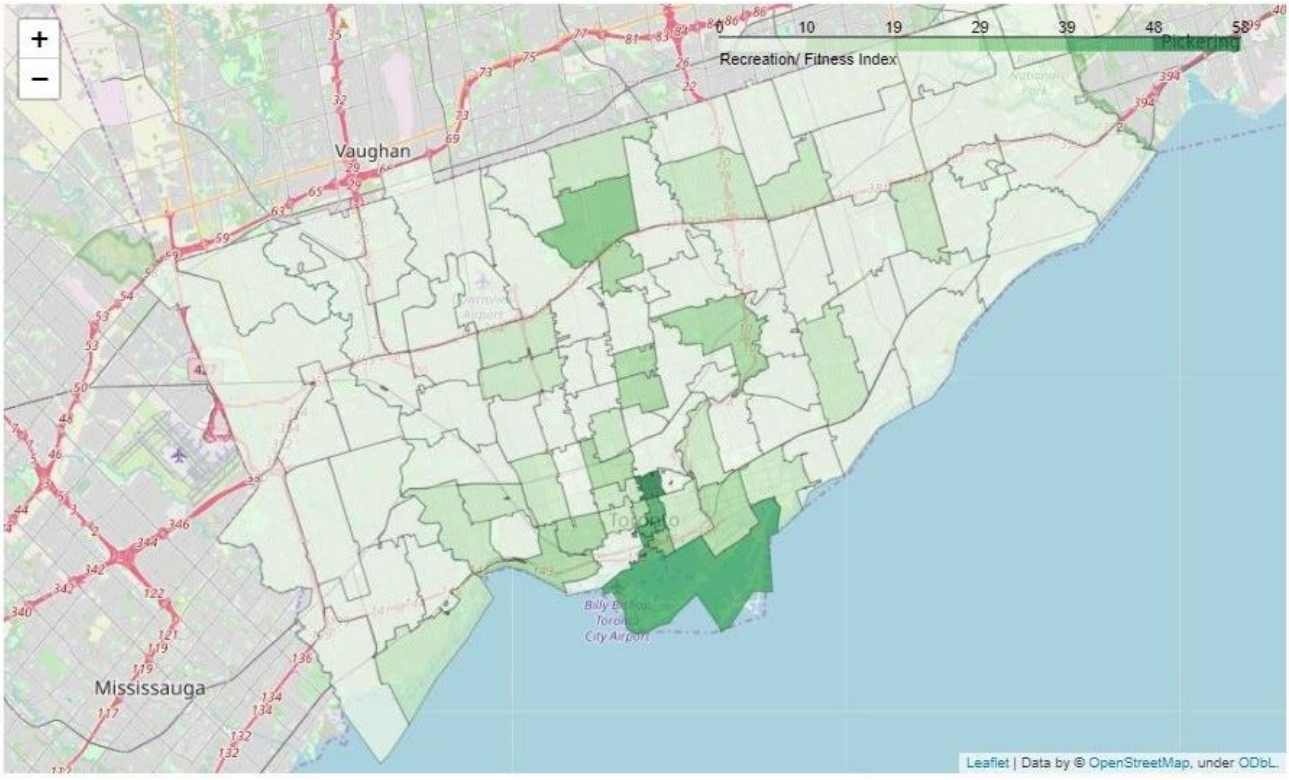
The crime rate for most of the FSAs is less than 1300 per 100,000 population. It is high in some parts of Old Toronto and Northern Etobicoke. (Note black regions depict FSAs for which data is not available)

Similarly, let's see choropleth maps for the other three variables as well.

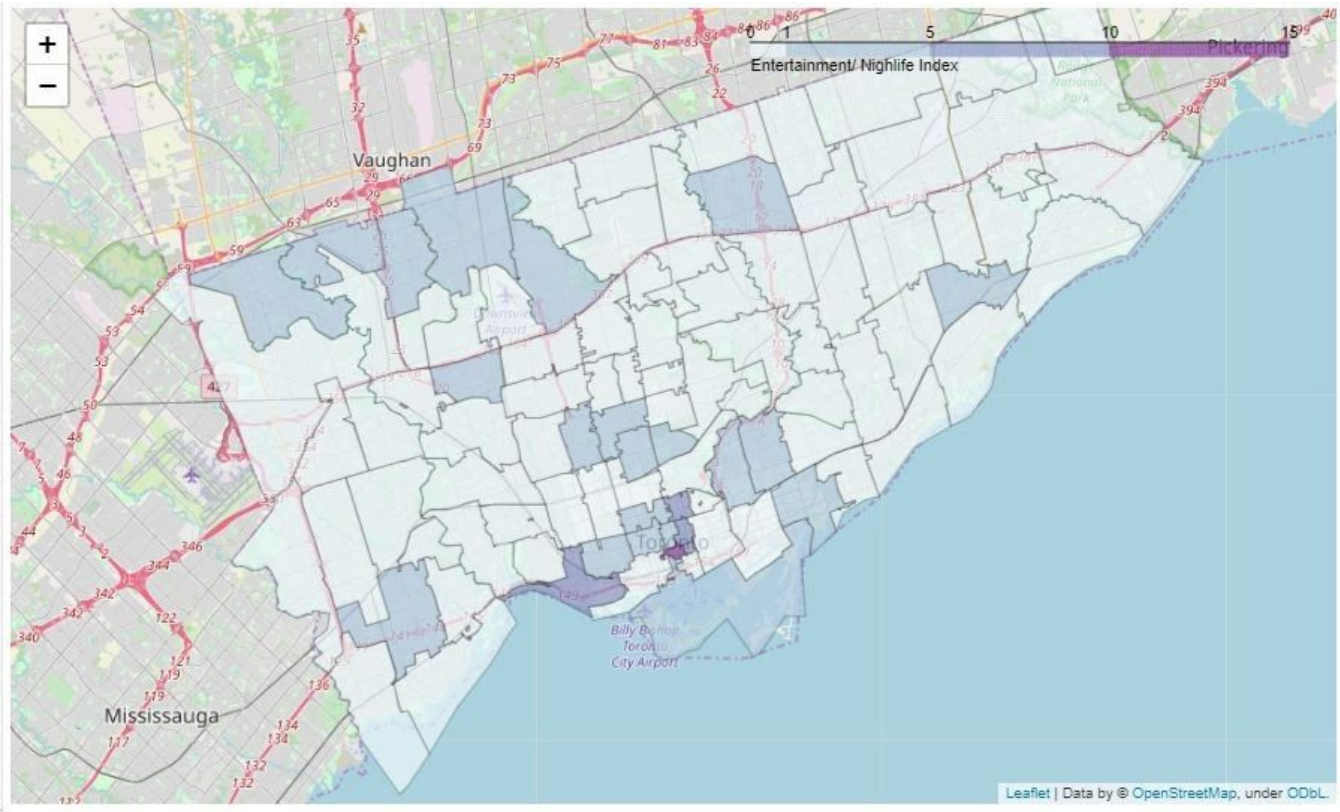
C) Cuisine Index



D) Recreation and Fitness Index



E) Entertainment and Nightlife Index

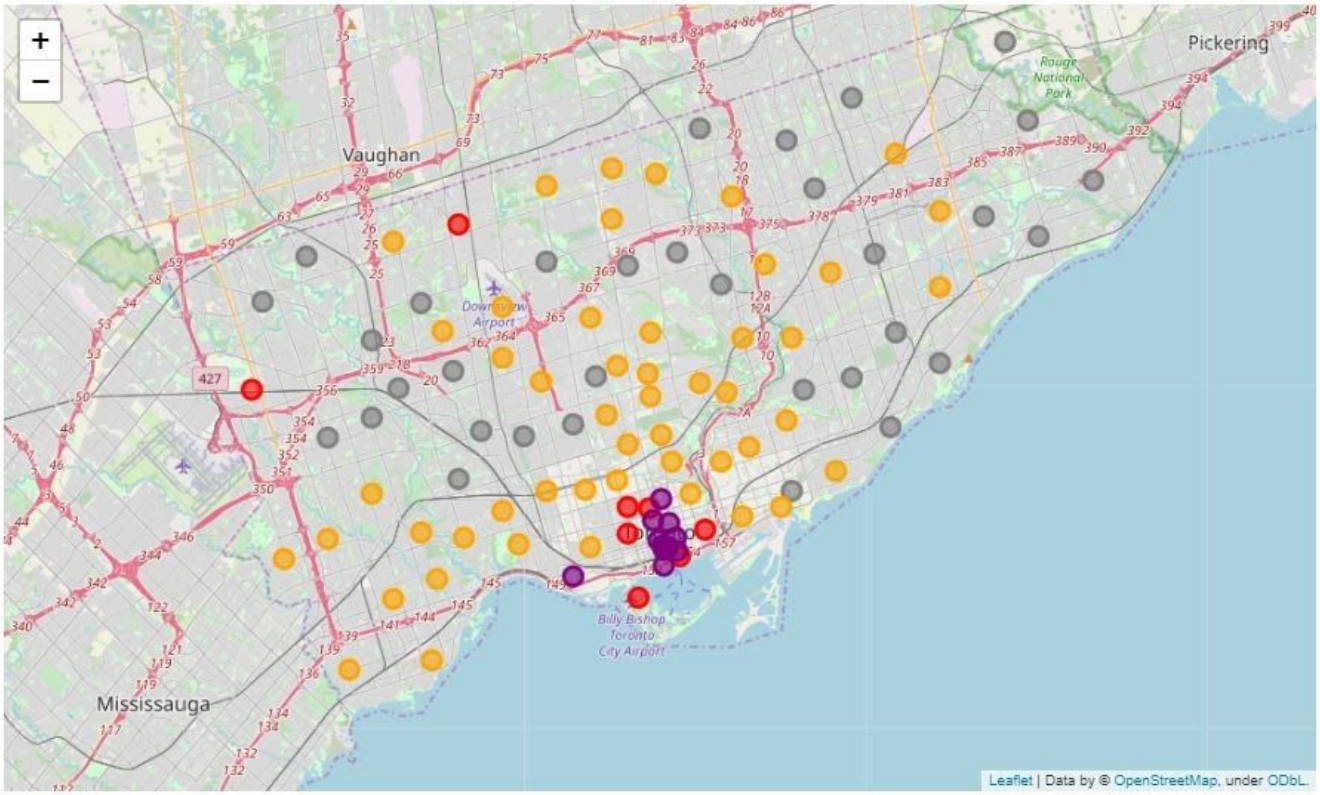


3.3. Clustering Neighborhoods in Toronto

I segmented neighborhoods in Toronto into four clusters using k-means clustering. The following four groups showing the mean values for the five indices were returned by the algorithm-

Label	Average Rent	Total Crime Rate	Cuisine Index	Recreation and Fitness Index	Entertainment and Nightlife Index	Number of FSAs	Color
3	1427.389020	1224.501686	1.757576	6.272727	0.181818	33	gray
0	2007.905410	1188.684076	3.740000	9.980000	0.320000	50	orange
2	2153.139179	3566.976389	3.250000	14.250000	1.250000	8	red
1	2263.363272	2983.797980	11.272727	45.454545	6.000000	11	purple

And here is a map showing the clustered neighborhoods in Toronto.



4. Results

From the above table, we can see that there is generally a positive correlation between average rent and restaurant variety, fitness venues and entertainment options. The more expensive an area is the higher number of commercial venues in that area. The final results can be summarised as follows:-

Cluster	**Brief Description**	**Spatial Extent**
Gray (Poor Man's Friend)	Cheap neighborhoods with low rent; safe; low number of eating and entertainment options	Scattered in the outer suburbs
Orange (Decent and Safe)	Moderate cost of living; safe; moderate number of restaurants and fitness venues but low number of clubs and theaters	Most frequent, prevalent in Central Toronto, East & West Toronto, East York and southern Etobicoke

Red (Always watch out for your purse)	Highest Crime Rate; slightly expensive; moderate number of commercial venues	Downtown Toronto and a few outliers in North western part
Purple (Rich and luxurious)	Very expensive; moderate crime rate; high-end neighborhoods with a wide variety of restaurants, clubs, bars etc.	Concentrated only in Downtown Toronto

5. Conclusion

In this report I segmented and analysed neighborhoods in the city of Toronto in Canada. Four types of neighborhoods were found. Any person who wishes to move to Toronto or is trying to find suitable neighborhood to live in Toronto will find help from this analysis. Though I chose only five variables for consideration, the project can be improved by choosing many more variables and clustering neighborhoods into more groups. Another improvement can be done by improving the quality of data. The Foursquare data though comprehensive misses out on some venues especially those in suburban regions.