Faculty of Engineering & Applied Science

Project Title: Analyzing Toronto's Neighborhood

Group 36

Due Date: 04/11/2022

GitHub: https://github.com/abdulbhutta/Analyzing-Toronto-

First Name	Last Name	Student ID
Abdul	Bhutta	100785884
Shreya	Sharma	100725334
Disha	Panday	100725039

Table Of Contents:

Introduction

We took the city of Toronto and segmented it into separate neighborhoods based on their geological coordinates. The neighborhoods will then be grouped into clusters using a combination of location data from the Foursquare API and artificial intelligence (clustering).

A. **Problem Statement :** The goal of this project is to better comprehend and visualize Toronto's neighborhood for new-comers. We wanted to visualize each segment of the city to determine which neighborhood we should reside in since we would be graduating next year and going to Toronto for employment.

   Researching and moving might be difficult for newcomers. After graduation, students will be extremely busy seeking jobs, and finding a new home will be an additional stress, which we hope to alleviate.

B. **Target Audience :** The primary target audience for these tools will be people who are actively planning a move and looking for an apartment after graduation and/or a job in Toronto, Ontario, but they can also be extended to an indirect target audience of people who are thinking about moving but are unsure whether their personality/interests will match what Toronto has to offer.

Data Sources

A. **Types of Data Required** : The objective of this report is to examine the neighborhoods of Toronto city using various data sets in order to discover the ideal location. The project will make use of the following datasets :
   1. Toronto's neighborhoods classified by postal code
   2. Geospatial Coordinates of each neighborhood with postal code
   3. Shape data to obtain the Latitude, Longitude locations and the geometry of the city
   4. Each neighborhood's venue (Police Station) type
   5. Income and Crime Rate Statistics Corresponding to the city of Toronto

B. Data Explanation

I. Toronto's Postal Codes, Boroughs and Neighborhoods:

The list of boroughs, neighborhoods, and postal codes for Toronto was scraped from the Wikipedia site using Python's “pandas” package.We wrangled and cleaned up the data after WebScraping, and then read it into our pandas data frame. To read data from a csv file into a pandas DataFrame, use the read_csv() method from pandas. It also comes with support for a variety of file types and data sources (csv, excel, sql, json, parquet, etc.) with the prefix read_*.

Many postal codes had a "not assigned" indication, and duplicate postal codes were also discovered, necessitating more data cleansing.

	Postal Code	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
...	...	...	...
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...
102	M9W	Etobicoke	Northwest

Figure 1: Toronto’s Postal code set with boroughs and neighborhoods scraped from Wikipedia

II. Toronto’s Neighbourhood Coordinates :

Neighborhoods latitude and longitude information was obtained from geospatial coordinates(csv File)

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476
...	...	...	...
98	M9N	43.706876	-79.518188
99	M9P	43.696319	-79.532242
100	M9R	43.688905	-79.554724
101	M9V	43.739416	-79.588437
102	M9W	43.706748	-79.594054

103 rows × 3 columns

Figure 2: Postal Code Latitude Longitude from geospatial coordinates (CSV FILE)

III. **Combined data :** We combined the data from the scraped wikipedia page as well as the csv file.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

103 rows × 5 columns

Figure 3: Combined data from the wikipedia page and the csv file

IV. **City of Toronto’s corresponding statistics:** The information was gathered from websites covering Toronto's neighborhoods. The information was about the area's crime rate, income, and police stations. It should be noted that the data was gathered in 2016, implying that it may differ from current numbers (2022)

i) Sample from Total Household Incomes and Income Brackets :

	Neighbourhood	Total - Household total income groups in 2015 for private households - 100% data	Under $5,000	$5,000 to $9,999	$10,000 to $14,999	$20,000 to $24,999	$25,000 to $29,999	$30,000 to $34,999	$35,000 to $39,999	$40,000 to $44,999
0	Agincourt North	9,120	155	105	160	320	540	420	455	420
1	Agincourt South-Malvern West	8,135	315	140	195	315	400	370	385	370
2	Alderwood	4,620	55	45	80	145	150	155	170	160
3	Annex	15,935	850	485	655	620	530	525	555	540
4	Banbury-Don Mills	12,125	265	155	235	395	445	405	470	460
...	...	...	...	...	...	...	...	...	...	...
135	Wychwood	5,885	120	120	185	325	270	225	225	225
136	Yonge-Eglinton	5,680	205	105	145	175	170	190	210	195
137	Yonge-St.Clair	7,015	215	120	185	225	220	215	235	260
138	York University Heights	10,165	345	230	340	525	525	590	570	530
139	Yorkdale-Glen Park	5,345	100	65	120	245	240	290	275	280

140 rows × 19 columns

$40,000 to $44,999	$45,000 to $49,999	$50,000 to $59,999	$60,000 to $69,999	$70,000 to $79,999	$80,000 to $89,999	$90,000 to $99,999	$100,000 and over	$200,000 and over	Population
420	435	800	700	635	525	515	2,505	325	29,113
370	415	770	645	595	510	405	2,030	285	23,757
160	165	335	300	320	275	250	1,915	360	12,054
540	505	1,000	900	795	715	605	5,895	2,670	30,526
460	400	930	885	780	655	605	4,615	1,750	27,695
...	...	...	...	...	...	...	...	...	...
225	225	400	390	315	300	275	1,895	610	14,349
195	180	370	375	325	275	245	2,340	1,110	11,817
260	260	505	445	415	340	325	2,865	1,380	12,528
530	530	950	845	615	605	515	1,955	195	27,593
280	230	470	370	360	295	235	1,550	325	14,804

Figure 4&5: Toronto Neighborhoods Total Household Incomes and Income Brackets

	Neighbourhood	Total - Household total income groups in 2015 for private households - 100% data	Population	Total People Working in Household	Average Total Income Groups in the Neighbourhood
0	Agincourt North	9,120	29113	9120	0.313262
1	Agincourt South-Malvern West	8,135	23757	8135	0.342425
2	Alderwood	4,620	12054	4620	0.383275
3	Annex	15,935	30526	15935	0.522014
4	Banbury-Don Mills	12,125	27695	12125	0.437805
...	...	...	...	...	...
135	Wychwood	5,885	14349	5885	0.410133
136	Yonge-Eglinton	5,680	11817	5680	0.480663
137	Yonge-St.Clair	7,015	12528	7015	0.559946
138	York University Heights	10,165	27593	10165	0.368391
139	Yorkdale-Glen Park	5,345	14804	5345	0.361051

140 rows × 5 columns

Figure 6: Final Household Data

ii) Crimes :

	Neighbourhood	Assault_AVG	AutoTheft_AVG	BreakandEnter_AVG	Homicide_AVG	Robbery_AVG	TheftOver_AVG	Average Crime Rate
0	Agincourt North	74.8	29.7	53.5	0.2	30.2	4.7	193.1
1	Agincourt South-Malvern West	117.8	36.7	79.8	0.2	27.3	13.3	275.1
2	Alderwood	36.3	16.2	24.7	0.2	6.8	6.8	91.0
3	Annex	246.3	22.0	147.5	0.5	40.8	29.5	486.6
4	Banbury-Don Mills	80.5	21.8	73.2	0.0	15.0	10.3	200.8
...	...	...	...	...	...	...	...	...
135	Wychwood	70.2	13.2	34.0	0.3	13.8	2.3	133.8
136	Yonge-Eglinton	75.8	9.0	28.0	0.3	19.5	4.8	137.4
137	Yonge-St.Clair	31.0	4.3	23.3	0.0	5.7	4.3	68.6
138	York University Heights	333.2	106.3	113.2	0.8	75.8	36.3	665.6
139	Yorkdale-Glen Park	160.2	55.5	63.3	1.2	31.5	22.5	334.2

140 rows × 8 columns

Figure 7: Crime Rate Data

V. Latitude Longitude and Geometry Dataset for each Neighborhood

	Neighbourhood	Latitude	Longitude	geometry
0	Agincourt North	43.805441	-79.266712	POLYGON ((-79.24213 43.80247, -79.24319 43.802...
1	Agincourt South-Malvern West	43.788658	-79.265612	POLYGON ((-79.25498 43.78122, -79.25797 43.780...
2	Alderwood	43.604937	-79.541611	POLYGON ((-79.54866 43.59022, -79.54876 43.590...
3	Annex	43.671585	-79.404001	POLYGON ((-79.39414 43.66872, -79.39588 43.668...
4	Banbury-Don Mills	43.737657	-79.349718	POLYGON ((-79.33055 43.73979, -79.33044 43.739...
...	...	...	...	...
135	Wychwood	43.676919	-79.425515	POLYGON ((-79.43592 43.68015, -79.43492 43.680...
136	Yonge-Eglinton	43.704689	-79.403590	POLYGON ((-79.41096 43.70408, -79.40962 43.704...
137	Yonge-St.Clair	43.687859	-79.397871	POLYGON ((-79.39119 43.68108, -79.39141 43.680...
138	York University Heights	43.765736	-79.488883	POLYGON ((-79.50529 43.75987, -79.50488 43.759...
139	Yorkdale-Glen Park	43.714672	-79.457108	POLYGON ((-79.43969 43.70561, -79.44011 43.705...

140 rows × 4 columns

Figure 8: Total Police Station for each Neighborhood

VI. **Venue :** We used FourSquare to find the sort of venue (in this instance, police stations) that surrounded the coordinates of each neighborhood. The query contained each Toronto neighborhood's latitude and longitude coordinates, with a limit of 50 results per neighborhood based on a 10km radius within the neighborhood.

	Neighbourhood	Latitude	Longitude	Station	Station Lat	Station Long
0	Agincourt North	43.805441	-79.266712	Toronto Police Service, Divisions, 42 Division	43.789177	-79.240020
1	Agincourt North	43.805441	-79.266712	Toronto Police Property Unit	43.774605	-79.267090
2	Agincourt North	43.805441	-79.266712	York Regional Police District 5 HQ	43.876477	-79.287645
3	Agincourt North	43.805441	-79.266712	Toronto Police Service - 43 Division	43.770656	-79.173907
4	Agincourt North	43.805441	-79.266712	Toronto Police Service, Divisions, 41 Division	43.730515	-79.277079
...	...	...	...	...	...	...
3616	Yorkdale-Glen Park	43.714672	-79.457108	Toronto Police Service	43.661197	-79.384973
3617	Yorkdale-Glen Park	43.714672	-79.457108	Toronto Police Svc	43.756491	-79.551813
3618	Yorkdale-Glen Park	43.714672	-79.457108	Toronto Police Service, Divisions, 23 Division	43.718423	-79.569583
3619	Yorkdale-Glen Park	43.714672	-79.457108	Cogient Corp	43.650135	-79.381155
3620	Yorkdale-Glen Park	43.714672	-79.457108	Earnscliffe Ontario Inc	43.650090	-79.381007

3621 rows × 6 columns

Figure 9: Police Stations Near Each Neighborhood

	Neighbourhood	Police Stations
0	Mount Pleasant West	44
1	Leaside-Bennington	43
2	Lawrence Park North	43
3	Mount Pleasant East	43
4	Lawrence Park South	43
...	...	...
135	West Hill	7
136	Morningside	7
137	Rouge	5
138	Centennial Scarborough	5
139	Highland Creek	4

140 rows × 2 columns

Figure 10: Total Police Station for each Neighborhood

Methodology

1. **Foursquare API:** We signed up for a foursquare API developer account and learned how to access the API with Python. Longitude, Latitude and venues all of them can be found using Foursquare. We obtained the project's data from a third-party source. And finally, we utilized Python's folium package to illustrate the map. On an interactive leaflet map, folium makes it simple to see data that's been altered in Python. It allows data to be bound to a map for choropleth visualization, as well as complex vector/raster/HTML infographics to be sent as markers on the map. It takes advantage of the Python ecosystem's data manipulation capabilities as well as the mapping capabilities of the leaflet.js module.

2. **Preparing data:** This stage was all about gathering information from all of the previously stated sources. We figured out what our project's key contribution was and what we intended to accomplish. We also combined all of the obtained datasets into one to begin clustering.

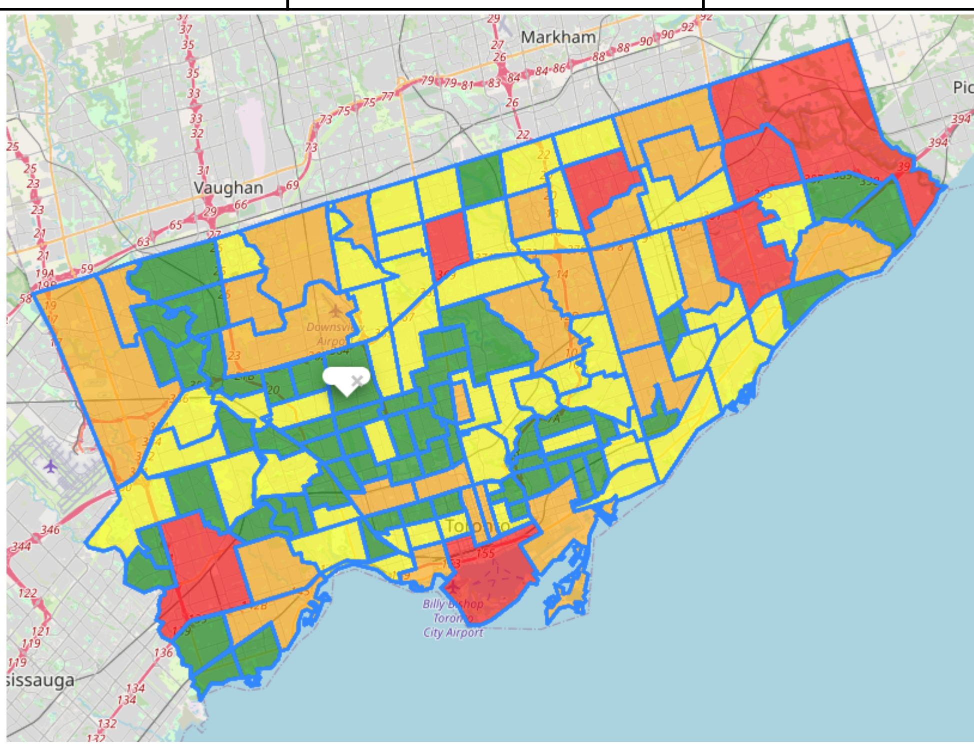
3. **Neighborhood segmentation and clustering(K-means Clustering):** We learned clustering using Python for segmenting and clustering the neighborhoods in this stage. The cluster criteria had to be in line with our project's objectives. We next chose the crime rate as well as the income source and tax, and explored the relationship between the two. We divided the region into clusters based on the priorities that produce the desired outcomes, assigned a score to each neighborhood, and displayed the results. **K-means Clustering :** k-means clustering is a vector quantization method derived from signal processing that aims to divide n observations into k clusters, with each observation belonging to the cluster with the closest mean, which serves as the cluster's prototype[1]. The unsupervised machine learning approach k-means clustering is used to find clusters of data items in a dataset. There are a variety of clustering algorithms available, but k-means is one of the most popular and accessible. These characteristics make k-means clustering in Python relatively simple to implement, especially for beginner programmers and data scientists.

Results and Conclusion :

After clustering, the data was divided into four clusters. The geometry coordinates for each neighborhood can be used to visualize these four clusters on the map. To represent the ideal location, each cluster is assigned a color. According to the data, the green area is the best place to live in Toronto because it has the lowest crime rate and population while also having the most police stations.

	Population	Average Crime Rate	Total People Working in Household	Average Total Income Groups in the Neighbourhood	Police Stations
Cluster Labels					
0	19405.040000	227.890000	7814.900000	0.403285	24.780000
1	29705.120000	460.956000	12338.000000	0.414668	21.920000
2	49725.714286	554.885714	20385.000000	0.397379	15.142857
3	11562.258621	132.339655	4672.931034	0.404700	29.793103

Cluster	Color	Ideal Location
0	Yellow	Ideal
1	Orange	Less than Ideal
2	Red	**NOT** Ideal
3	Green	Most Ideal



References :

[1]: "K-means clustering," *Wikipedia*, 20-Mar-2022. [Online]. Available: https://en.wikipedia.org/wiki/K-means_clustering. [Accessed: 11-Apr-2022].

[2]"List of postal codes of Canada: M," *Wikipedia*, Mar. 11, 2022. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (accessed Apr. 12, 2022).

[3]"Neighbourhood Crime Rates (Boundary File)," *data.torontopolice.on.ca*. https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-

[4]"Open Data Dataset." https://open.toronto.ca/dataset/neighbourhoods/

Appendix :

Final Data Set Used

	Neighbourhood	Population	Average Crime Rate	Cluster Labels	Total People Working in Household	Average Total Income Groups in the Neighbourhood	Police Stations	geometry
0	Agincourt North	29113	193.1	1	9120	0.313262	9	POLYGON ((-79.24213 43.80247, -79.24319 43.802...
1	Agincourt South-Malvern West	23757	275.1	0	8135	0.342425	10	POLYGON ((-79.25498 43.78122, -79.25797 43.780...
2	Alderwood	12054	91.0	3	4620	0.383275	15	POLYGON ((-79.54866 43.59022, -79.54876 43.590...
3	Annex	30526	486.6	1	15935	0.522014	38	POLYGON ((-79.39414 43.66872, -79.39588 43.668...
4	Banbury-Don Mills	27695	200.8	1	12125	0.437805	31	POLYGON ((-79.33055 43.73979, -79.33044 43.739...
...	...	...	...	...	...	...	...	...
135	Wychwood	14349	133.8	3	5885	0.410133	37	POLYGON ((-79.43592 43.68015, -79.43492 43.680...
136	Yonge-Eglinton	11817	137.4	3	5680	0.480663	41	POLYGON ((-79.41096 43.70408, -79.40962 43.704...
137	Yonge-St.Clair	12528	68.6	3	7015	0.559946	41	POLYGON ((-79.39119 43.68108, -79.39141 43.680...
138	York University Heights	27593	665.6	1	10165	0.368391	18	POLYGON ((-79.50529 43.75987, -79.50488 43.759...
139	Yorkdale-Glen Park	14804	334.2	3	5345	0.361051	43	POLYGON ((-79.43969 43.70561, -79.44011 43.705...

140 rows × 8 columns

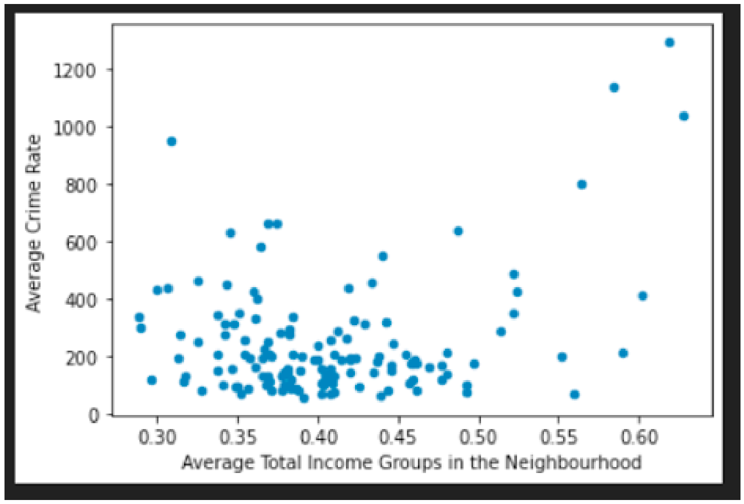
Statistics on Data

	Population	Average Crime Rate	Cluster Labels	Total People Working in Household	Average Total Income Groups in the Neighbourhood	Police Stations
count	140.000000	140.000000	140.000000	140.000000	140.000000	140.000000
mean	19511.221429	246.273571	1.521429	7949.428571	0.405608	25.864286
std	10033.589222	206.329192	1.343612	4794.749042	0.068062	11.723561
min	6577.000000	58.800000	0.000000	2650.000000	0.288197	4.000000
25%	12019.500000	119.175000	0.000000	5137.500000	0.361737	15.000000
50%	16749.500000	188.400000	1.000000	6572.500000	0.389001	26.000000
75%	23854.500000	289.700000	3.000000	9537.500000	0.439355	37.000000
max	65913.000000	1292.200000	3.000000	40750.000000	0.627951	44.000000

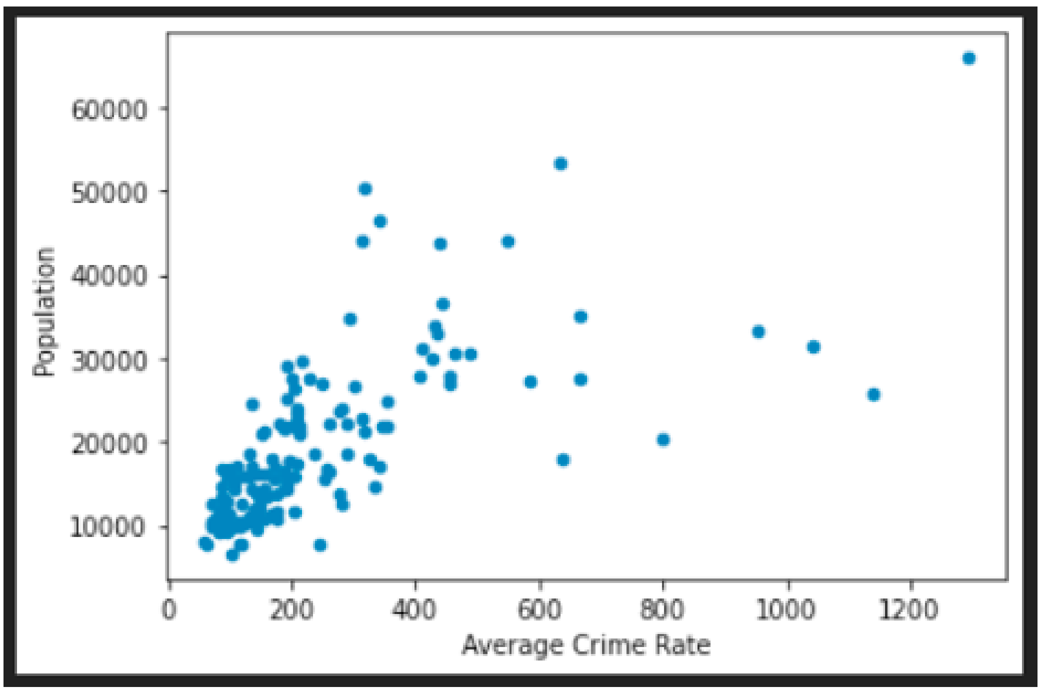
Clusters Data

	Population	Average Crime Rate	Total People Working in Household	Average Total Income Groups in the Neighbourhood	Police Stations
Cluster Labels					
0	19405.040000	227.890000	7814.900000	0.403285	24.780000
1	29705.120000	460.956000	12338.000000	0.414668	21.920000
2	49725.714286	554.885714	20385.000000	0.397379	15.142857
3	11562.258621	132.339655	4672.931034	0.404700	29.793103

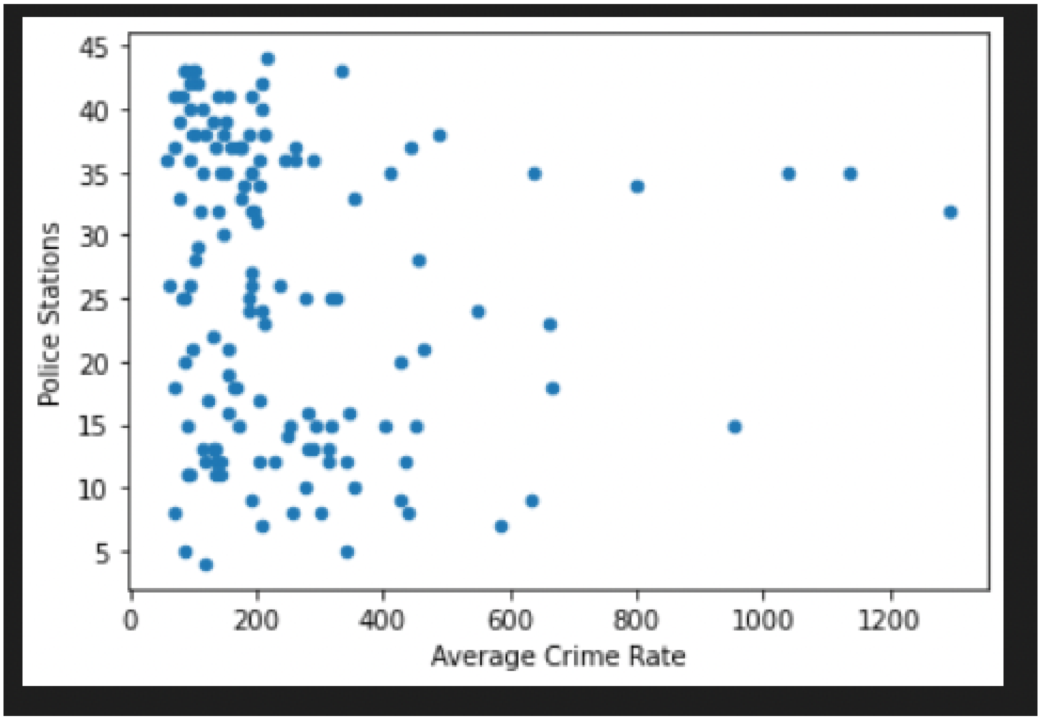
Average Crime Rate vs Average Total Income Groups in The Neighborhood



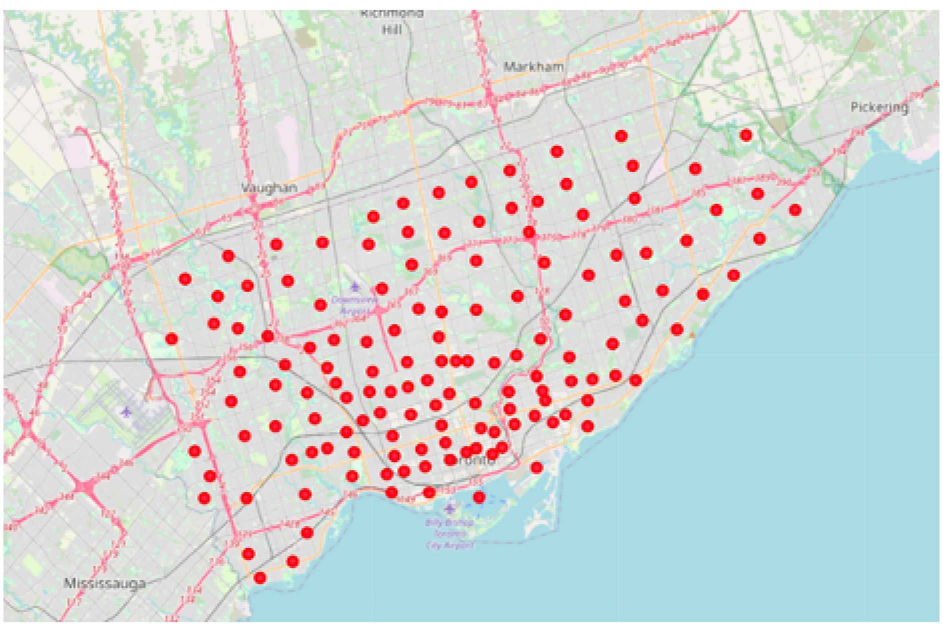
Population in the Neighborhood vs Average Crime Rate



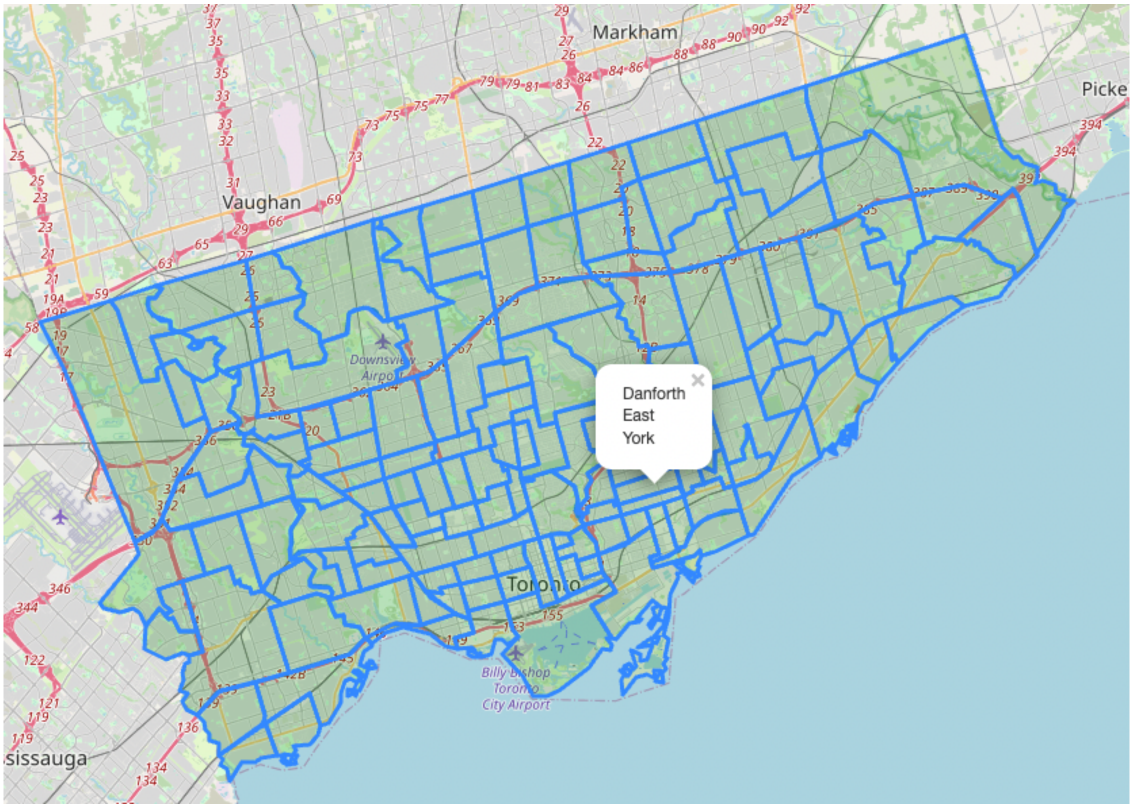
Average Crime Rate vs Police Stations



All the Neighborhoods visualized on the maps



Neighborhoods Visualization with Geometry


Markham
Vaughan
Picke
Downsview Airport
Danforth East York
Toronto
Billy Bishop Toronto City Airport
sissauga