City of Toronto Neighbourhoods comparative study (COVID-19)

Applied Data Science by IBM/Coursera

Lamis Kusto

June 27, 2020

1. Introduction

As COVID-19 hits everywhere in the world, I decided to take Toronto as a case study, the study will characterize Toronto neighbourhoods by features include people testing positive for COVID-19, health units, populations, income, housing and way of transportation.

The purpose of this project aims to create an analysis of features for the City of Toronto as a comparative analysis between neighborhoods to show the effective of COVID-19 on each neighborhood. According to the Government of Toronto report on June 26th 2020 the number of cases are 14,134 with 1,072 deaths.

2. Data Sources

We will be collecting data from the following sources:

- City of Toronto Neighbourhood Profiles to get the populations, income, education, housing and labour. Source of Data: **Statistics Canada, Toronto Open Data portal, the file in CSV format.**

  In these profiles, "neighbourhood" refers to the City of Toronto's 140 social planning neighbourhoods. These social planning neighbourhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these

social planning neighbourhoods are consistent over time, allowing for comparison between Census years, the last being in 2016.

Each data point in this file is presented for the City's 140 neighbourhoods, as well as for the City of Toronto as a whole. The data is sourced from a number of Census tables released by Statistics Canada. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information.

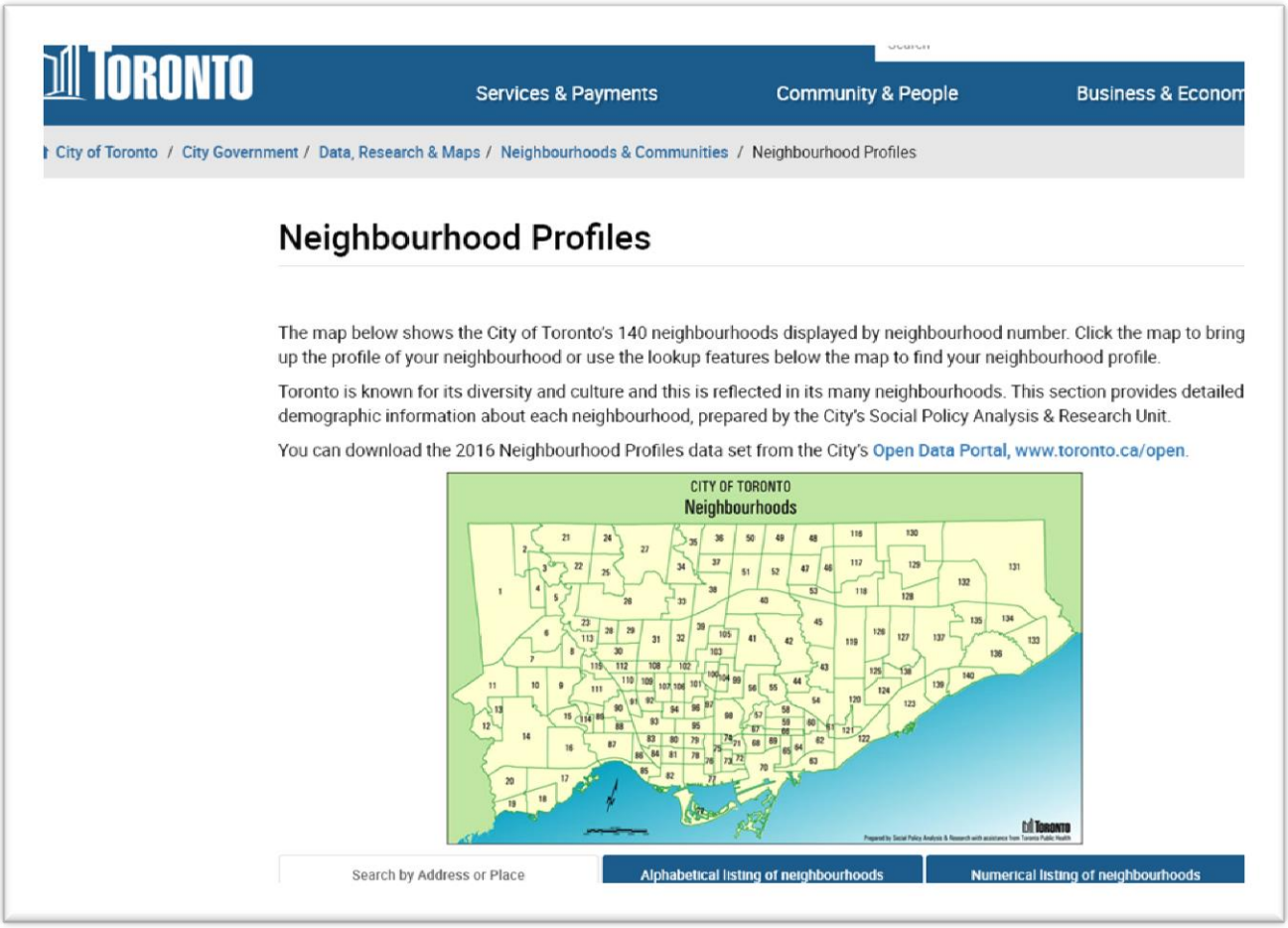

Figure (1) City of Toronto Neighbourhood Profile Website

- City of Toronto Neighbourhoods Boundaries for city shape to get the area name and location. Source of Data: **Toronto Open Data, the file in GeoJSON format.**
  This file contains the City's 140 neighbourhood's geo data like latitude, longitude and boundary shape.

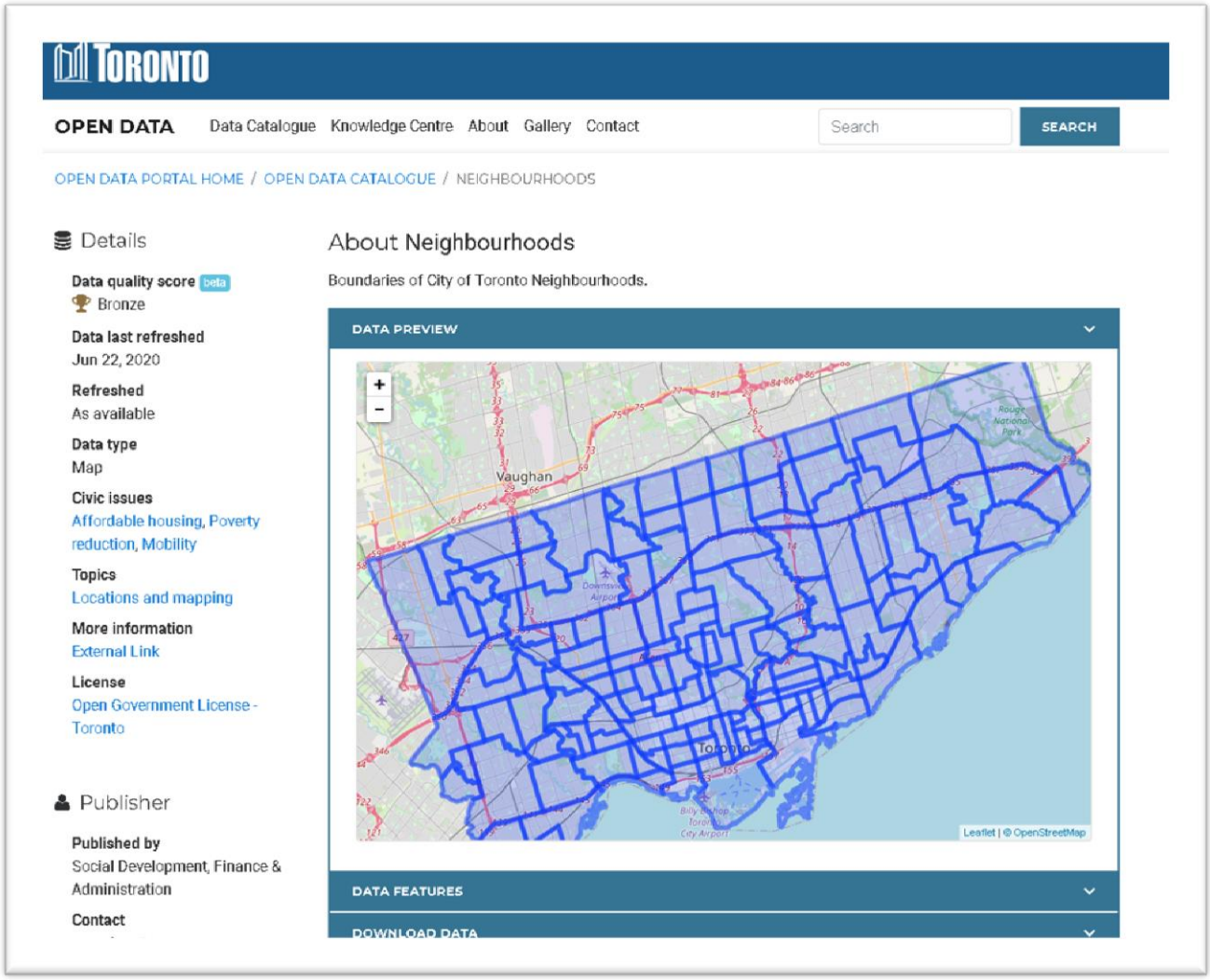

Figure (2) Boundaries of City of Toronto Neighbourhoods Website

- City of Toronto Cases of COVID-19 for each neighbourhood and the number of infected of COVID-19. Source of Data: **Ontario Ministry of Health, integrated Public Health Information System (iPHIS), the file in CSV format.**

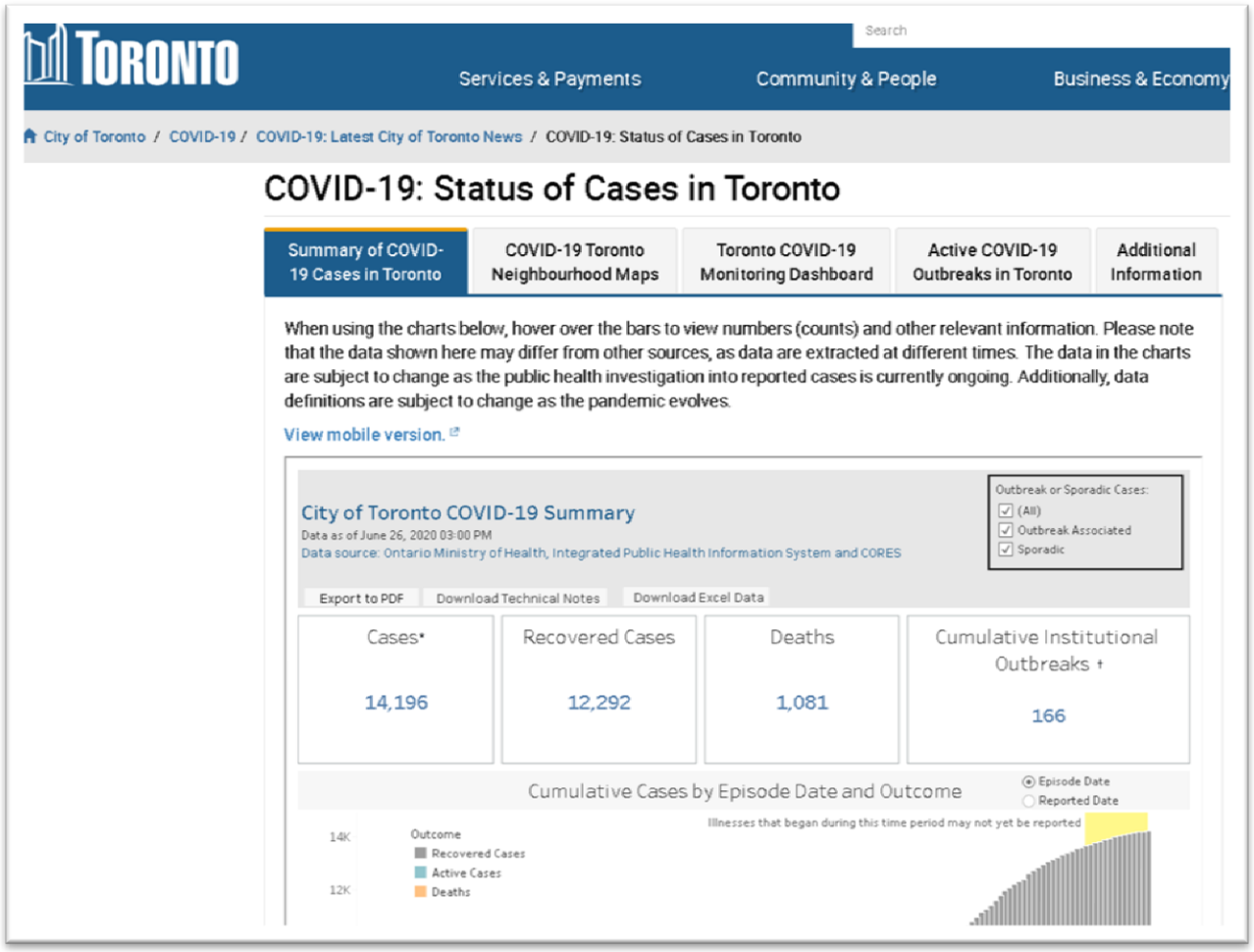

Figure (3) COVID-19 Status of Cases in Toronto

- Foursquare API to get the Hospitals information. Source of Data: **Foursquare API**

  The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the center of the City of Toronto. The information obtained per venue as follows:

  1. Neighborhood
  2. Neighborhood Latitude

3. Neighborhood Longitude
4. Venue
5. Name of the venue (the name of the Hospital)
6. Venue Latitude
7. Venue Longitude
8. Venue Category

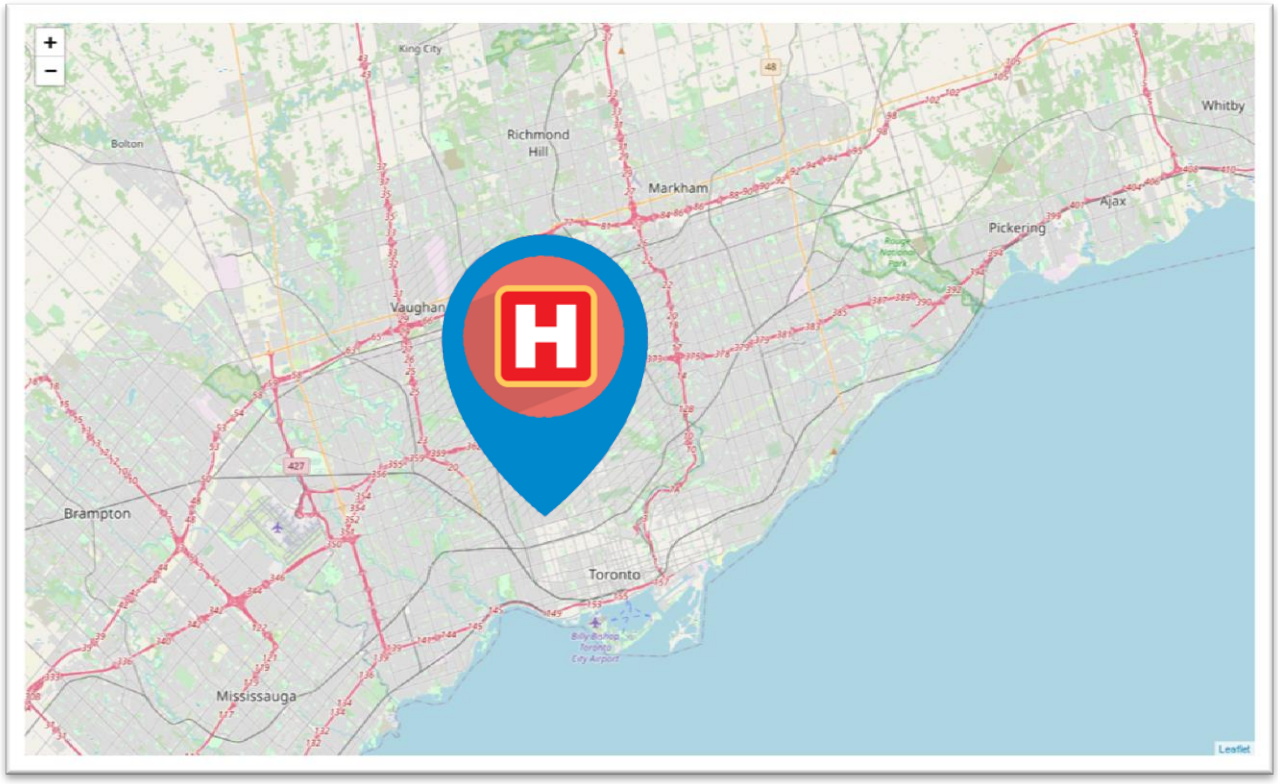

Figure (4) Use Foursquare API to get Hospitals in the City of Toronto

Note: This summary provides an overview of COVID-19 cases among residents of Toronto. Please note that the data shown here may differ from other sources, as data are extracted at different times. The data in the charts are subject to change as the public health investigation into reported cases is currently ongoing. It can take up to two weeks for symptomatic individuals to seek care, get tested, and for Toronto Public Health to receive the results. Additionally, data definitions are subject to change as the pandemic evolves.

3. Methodology

The purpose of this project aims to compare the relationship between people infected by COVID-19 and their living and economic circumstances and create an analysis of features for the City of Toronto as a comparative analysis between neighborhoods to show the effective of COVID-19 on each neighborhood.

As mentioned in the previous section, we will collect the city of Toronto neighbourhood data from different data sources.

3.1 Data Group A: Cases of COVID-19

1. Data was pulled into from the City of Toronto as a CSV File to create a dataframe.
2. The dataframe contains all the COVID-19 cases data of the 140 neighborhoods of Toronto.
3. Data is filtered to remove the unnecessary columns

	No	Neighbourhood	Total_Cases	Rate_People	Y	X
0	129	Agincourt North	69	237.007522	43.805441	-79.266712
1	128	Agincourt South-Malvern West	51	214.673570	43.788658	-79.265612
2	20	Alderwood	36	298.656048	43.604937	-79.541611
3	95	Annex	80	262.071677	43.671585	-79.404001
4	42	Banbury-Don Mills	33	119.155082	43.737657	-79.349718
...	...	...	...	...	...	...
135	94	Wychwood	80	557.530142	43.676919	-79.425515
136	100	Yonge-Eglinton	13	110.011001	43.704689	-79.403590
137	79	Yonge-St.Clair	26	207.535121	43.687859	-79.397871
138	27	York University Heights	390	1413.401950	43.765736	-79.488883
139	31	Yorkdale-Glen Park	163	1101.053769	43.714672	-79.457108

140 rows × 6 columns

4. Plot a Bar Chart of the dataframe

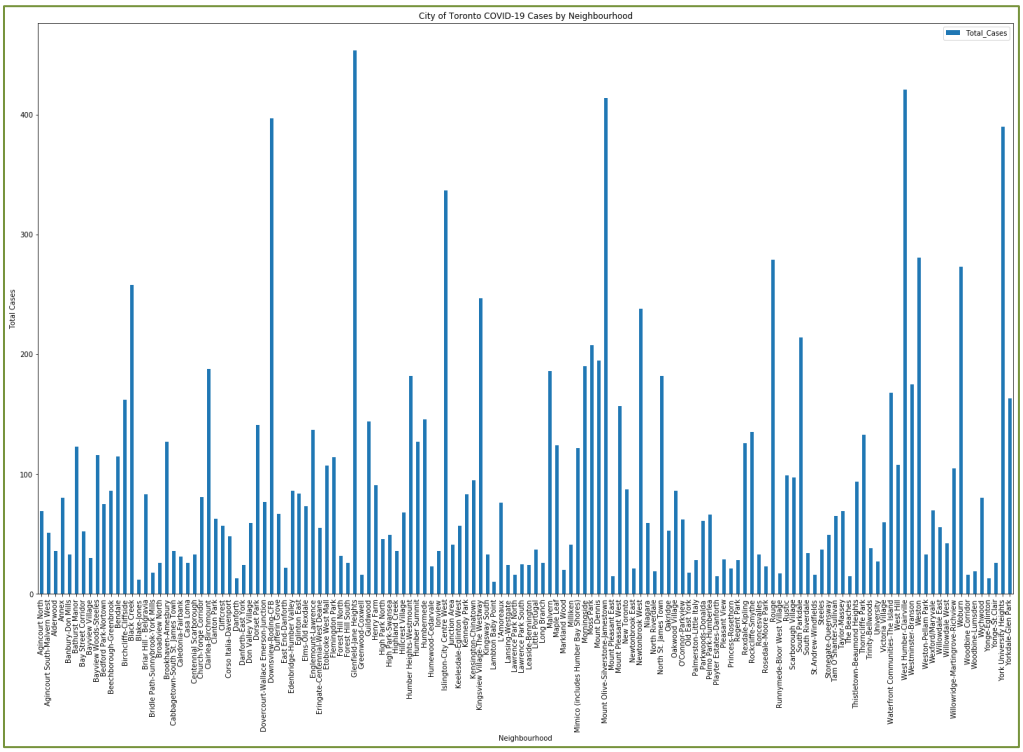

Figure (5) City of Toronto COVID-19 Cases by Neighbourhood

5. Sort the data from high to low then cut the first 10 neighbourhoods with the most cases and plot the Bar Chart again to get a clear view of the situation.

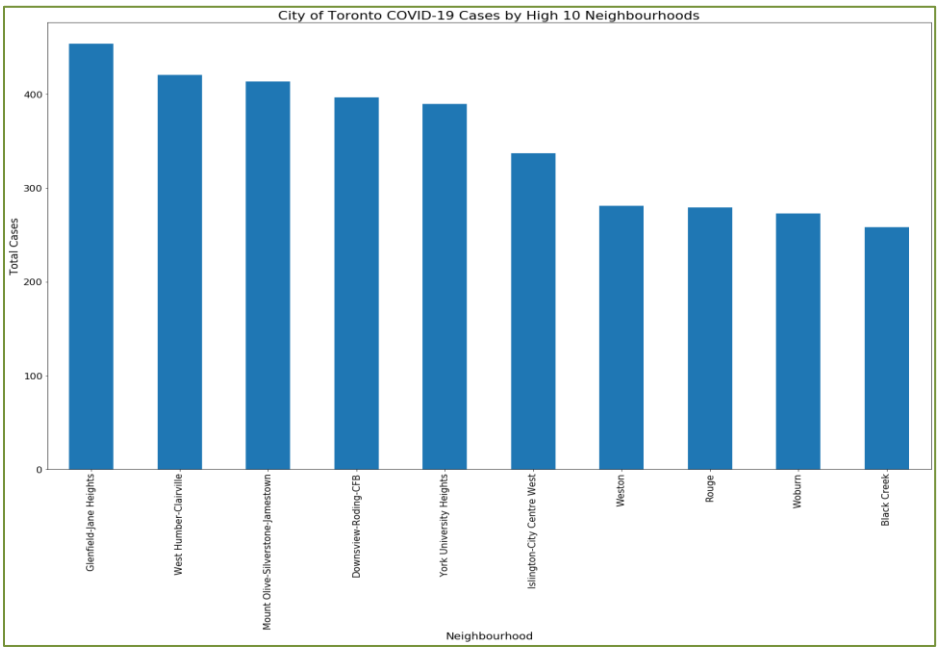

Figure (6) 10 Neighbourhoods that have the Highest Number of Cases

3.2 Data Group B: Neighbourhood Profiles

1. Data was pulled into from the City of Toronto as a CSV File to create a dataframe.

2. This dataframe contains all the census data (2016) of the neighborhoods of Toronto.

3. Data is filtered to remove the unnecessary columns

Neighbourhood	Unsuitable_Housing	Public_Transit	Low_income
Glenfield-Jane Heights	23.2	3965	21.8
West Humber-Clairville	17.4	4380	15.3
Mount Olive-Silverstone-Jamestown	30.8	4110	24.6
Downsview-Roding-CFB	18.8	6085	18.1
York University Heights	21.2	5405	23.8
...	...	...	...
The Beaches	3.7	2995	10.0
Danforth	5.7	2290	12.2
Yonge-Eglinton	6.1	2935	13.2
Blake-Jones	7.4	1605	22.0
Lambton Baby Point	7.1	1400	16.8

140 rows × 3 columns

4. Plot a Bar Chart of the dataframe

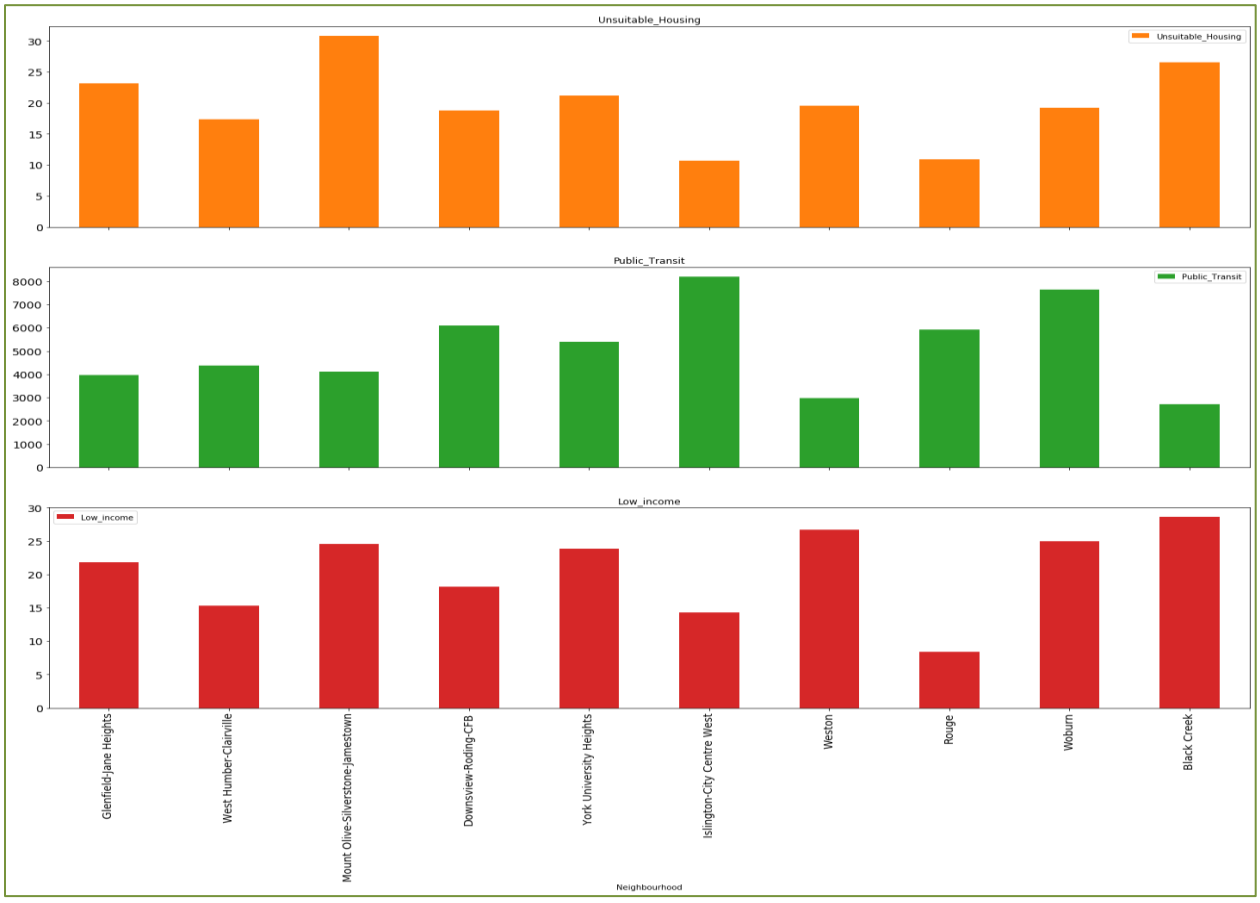

Figure (7) Neighbourhoods by Percentage of Unsuitable Housing, Public Transit and Percentage of Low Income

3.3 Data Group C: Neighbourhood Profiles

1. Data was pulled into from the City of Toronto neighbourhoods' boundaries as a GeoJSON File.
2. I used a python folium library to generate a map centred on the City of Toronto and visualize the neighborhood's boundaries.

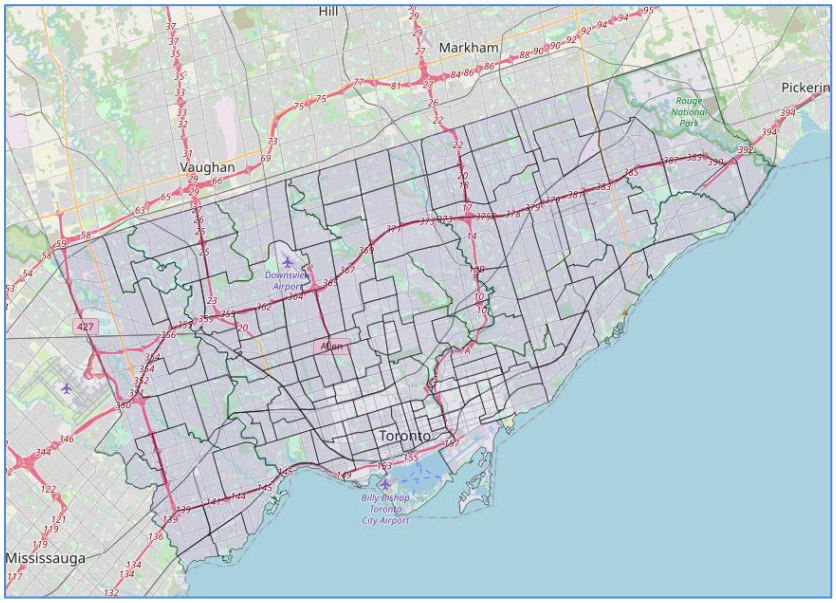

Figure (8) Toronto Neighbourhood Boundaries Extracted from GeoJSON File

3.4 Data Group D: Foursquare API

1. Retrieved from Foursquare a data contained information of venues within a specified distance of the longitude and latitude of the center of the City of Toronto, the data contains Hospitals in the city then transform venues into a dataframe.
2. The dataframe contains all kind of hospitals so it filtered to exclude veterinary hospitals from the dataset
3. Clean the dataframe by keeping only the category, name and location of the hospitals.

	name	categories	address	crossStreet	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress	neighborhood	id
0	Michael Garron Hospital	Hospital	825 Coxwell Ave	Mortimer Ave	43.689573	-79.326173	[{'label': 'display', 'lat': 43.689573, 'lng':...	7918	M4C 3E7	CA	East York	ON	Canada	[825 Coxwell Ave (Mortimer Ave), East York ON ...	NaN	4adb8e3bf964a520f22821e3
3	North York General Hospital	Hospital	4001 Leslie St	at Sheppard Ave E	43.768974	-79.363209	[{'label': 'display', 'lat': 43.76897368914205...	2116	M2K 1E1	CA	Toronto	ON	Canada	[4001 Leslie St (at Sheppard Ave E), Toronto O...	NaN	4b6c7ba2f964a520bb3d2ce3
4	Rouge Valley Centenary Hospital	Hospital	2867 Ellesmere Rd.	at Neilson Rd.	43.780614	-79.205151	[{'label': 'display', 'lat': 43.78061431275554...	11080	M1E 4B9	CA	Toronto	ON	Canada	[2867 Ellesmere Rd. (at Neilson Rd.), Toronto ...	NaN	4b828d15f964a520efd730e3
5	Mount Sinai Hospital Women's and Infants' Depa...	Hospital	700 University Ave., 3rd floor, Ontario Power ...	at College St.	43.659612	-79.390761	[{'label': 'display', 'lat': 43.65961215020 49,...	11898	NaN	CA	Toronto	ON	Canada	[700 University Ave., 3rd floor, Ontario Power...	NaN	4b1fbe8af964a5209e2824e3

3.5 Analysis

In our analysis we will use a Correlation as a technique for investigating the relationship between two quantitative, continuous variables. The correlation coefficient is a measure of the strength of the association between the two variables.

The first step in studying the relationship between two continuous variables is to draw a scatter plot of the variables to check for linearity.

The nearer the scatter of points is to a straight line, the higher the strength of association between the variables.

- I Used correlation to summarize the strength of the linear relationship between data

	Total_Cases	Unsuitable_Housing	Public_Transit	Low_income
Total_Cases	1.000000	0.519884	0.373505	0.253532
Unsuitable_Housing	0.519884	1.000000	0.159268	0.704397
Public_Transit	0.373505	0.159268	1.000000	0.221457
Low_income	0.253532	0.704397	0.221457	1.000000

- Calculate and plot the correlation between cases of COVID and percentage of unsuitable housing, low income and public transit in each Neighbourhood.

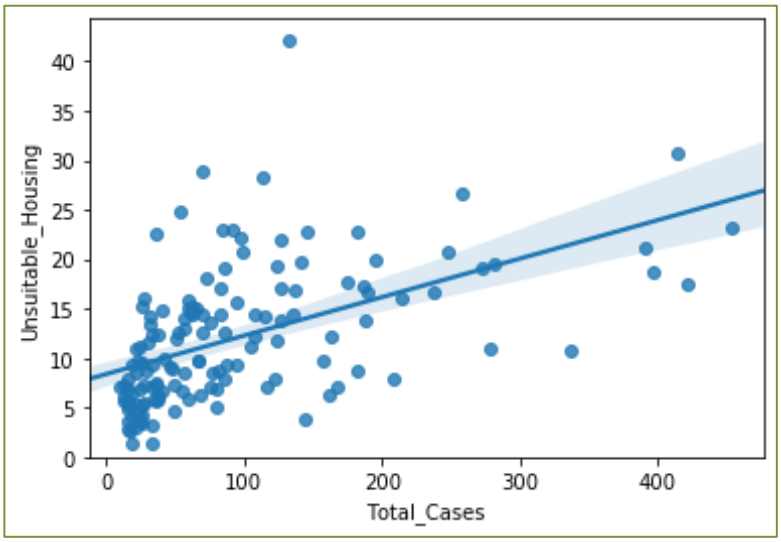

Figure (9) Scatter Plot of Cases of COVID-19 and Percentage of Unsuitable Housing

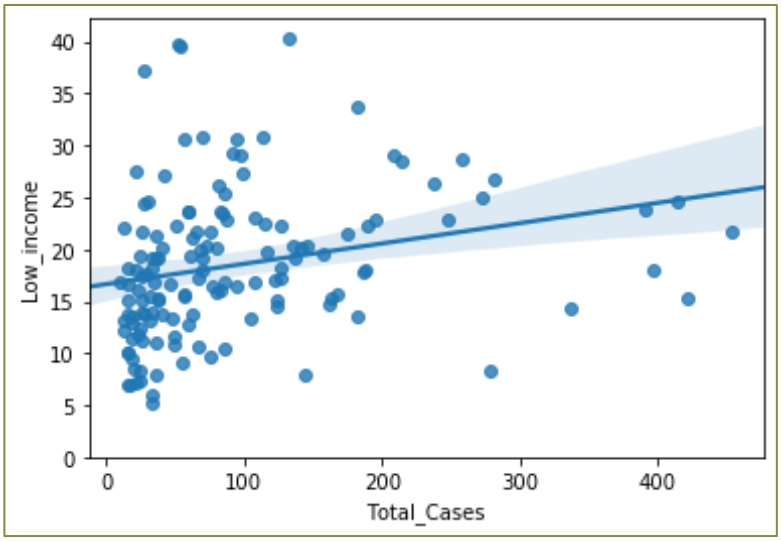

Figure (10) Scatter Plot of Cases of COVID-19 and Percentage of Low Income

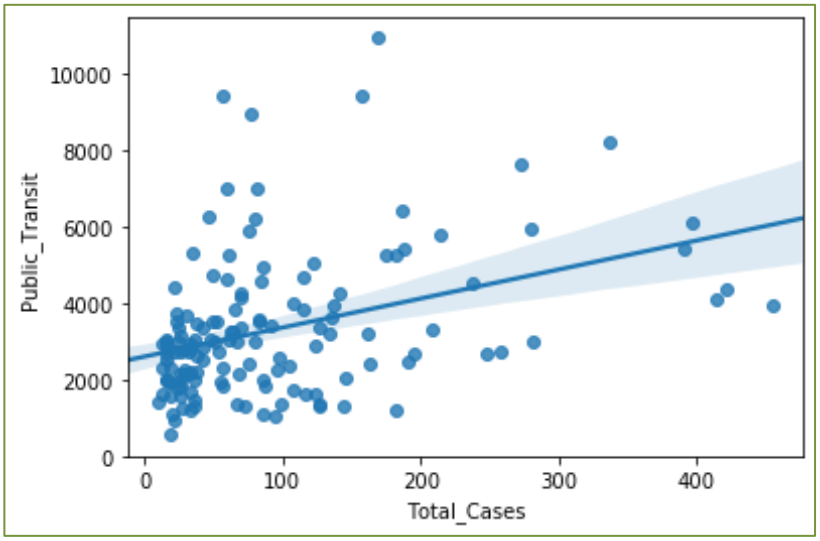

Figure (11) Scatter Plot of Cases of COVID-19 and People Use Public Transit

4. Results and Discussion

Our analysis shows that the neighbourhoods that must affected by COVID-19 the people there live in an unsuitable housing and a high percentage of them use the public transit for their movement. Although the analysis shows that there is a relationship between the infected people and the income but I believe it doesn't count as the result of the correlation doesn't shows a strong relation between the number of infected against their income.

One of my aims was also visualize all Toronto neighbourhoods and focus on the high 10 neighborhood that affected by the virus and show the hospitals in the city with python folium map.

- The Blue circles represent the Hospitals in the city.
- The small Yellow circles represent Toronto neighbourhoods.
- The large Yellow circles represent the most 10 neighbourhoods hardest-hit by the virus.

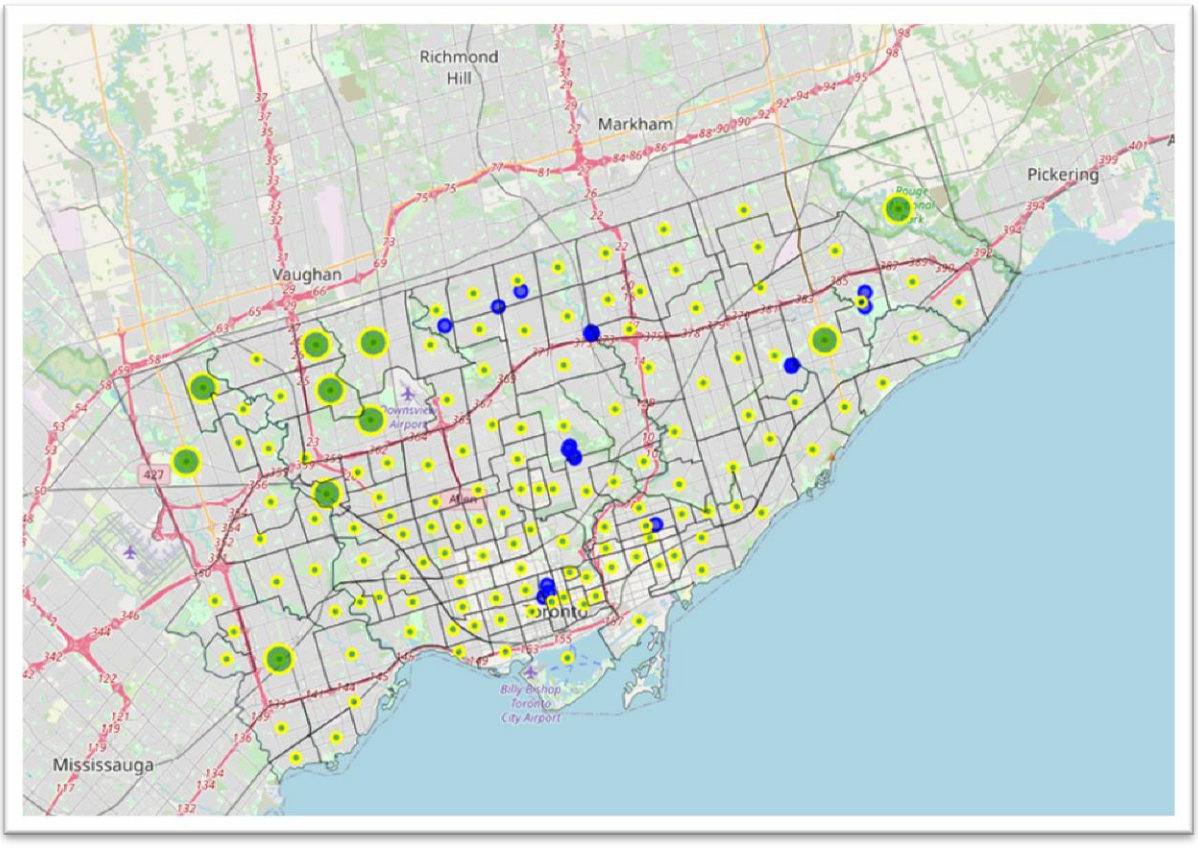

Figure (12) City of Toronto Map

5. Conclusion

As a result, Toronto's northeast and northwest are the city's hardest-hit area for COVID-19, according to the result visualized map. Knowing which areas of the city are hardest-hit is valuable from a public health point of view as Ontarians testing positive for COVID-19 are more likely to live in neighbourhoods characterized by precarious housing and lower income status.

The results could point to other factors: crowded living and working conditions that many people with low income experience, living in denser housing may provide fewer opportunities to properly follow distancing recommendations.

Note: By all means, the reports here should not be used as a measuring tool, because in reality the situation is different.