COVID-19 Toronto Report
Systematic Issues of COVID-19 in Toronto

Jonathan (Po-hsuan) Lai

July 1, 2020

1. Introduction

The year 2020 would be marked in history book by the emergence of COVID-19 (Novel Coronavirus) and its impact spared no corner of the world. Half a million of people has surrendered their lives to the virus globally with more than 10 million cases infected. In Toronto - Canada's largest city - almost 15000 cases of infection were found, and more than 1000 deaths have been reported so far.

Those lost lives will never be returned, and we must strive to improve our response to next pandemic. Governments need to evaluate the spread of pandemic in their communities in order to better prepare for the next inevitable event. This report will assess the cases of infection in 140 neighborhoods in Toronto along with their respective neighborhood profiles to uncover systematic issues in the city of Toronto related thru the spread of COVID-19. Key insights and recommended measures discussed will help improve the quality of decision making to protects residents against future waves of infectious diseases.

2. Data Description

Open Data Toronto is an excellent source for high quality data and provided the perfect data to perform analysis in our report. Additionally, we have access to COVID-19 cases by neighborhood thanks to the transparency of City of Toronto in response to the pandemic.

All the data used in this report were listed out as follows:

COVID-19 Toronto Neighborhood Maps – City of Toronto has given public access to number of cases for all 140 neighborhoods. For this part, we will subset to sporadic cases as the outbreak related cases were all well understood and Canadian government has already committed to improve in that aspect.

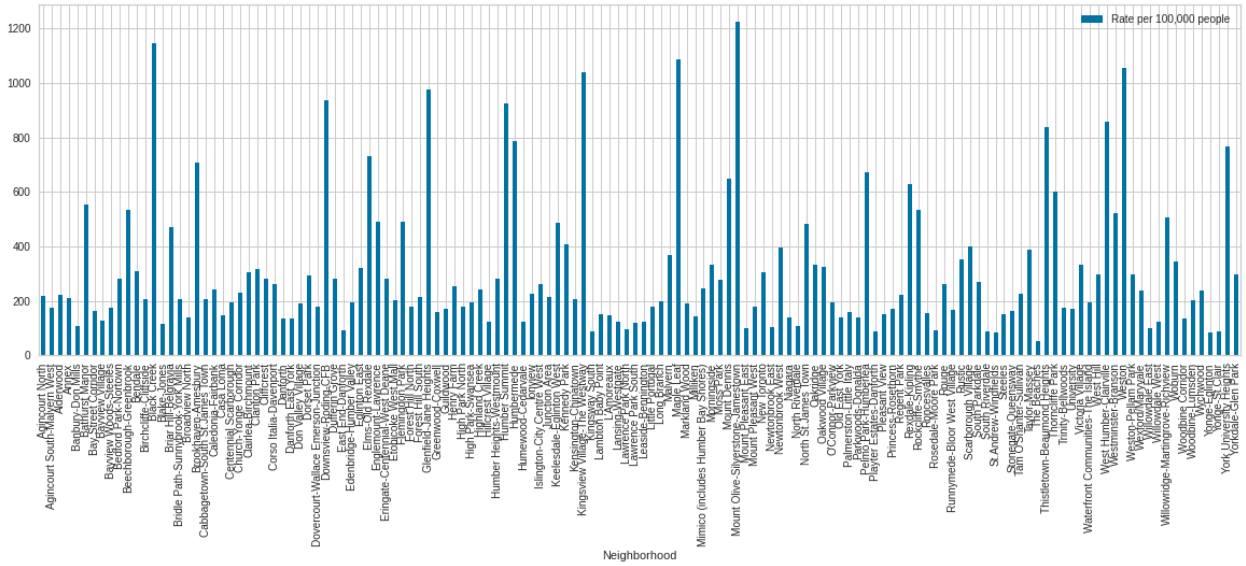
Neighborhood Profiles – 2016 Census of Population documented all keys demographic information needed and is available on Open data Toronto. Given the topic of this report, Only income, occupation, ethnicities and commuting related data were included as these were the main characteristics which would best reflect the systematic issues interested.

Neighborhoods Geospatial – Open data Toronto provided the geojson data required to locate our neighborhoods and their boundaries. This data was last updated at June 29, 2020.

Foursquare venues – Foursquare API was used to explore and extract venues in each of the 140 neighborhoods. Venues data will allow the report to investigate neighborhood construct and its potential link to an infectious disease.

3. Methodology and Analysis

As using the number of cases wouldn't be fair given the population difference among each neighborhood, Rate per 100,000 people was selected instead for all the analyses regarding COVID-19 number.
By mapping out the COVID-19 Rate per 100,000 people against neighborhoods, a raw distribution of cases in the city of Toronto was available. However, this plot was not informative enough and neighbor profile data should be able to support what is missing.



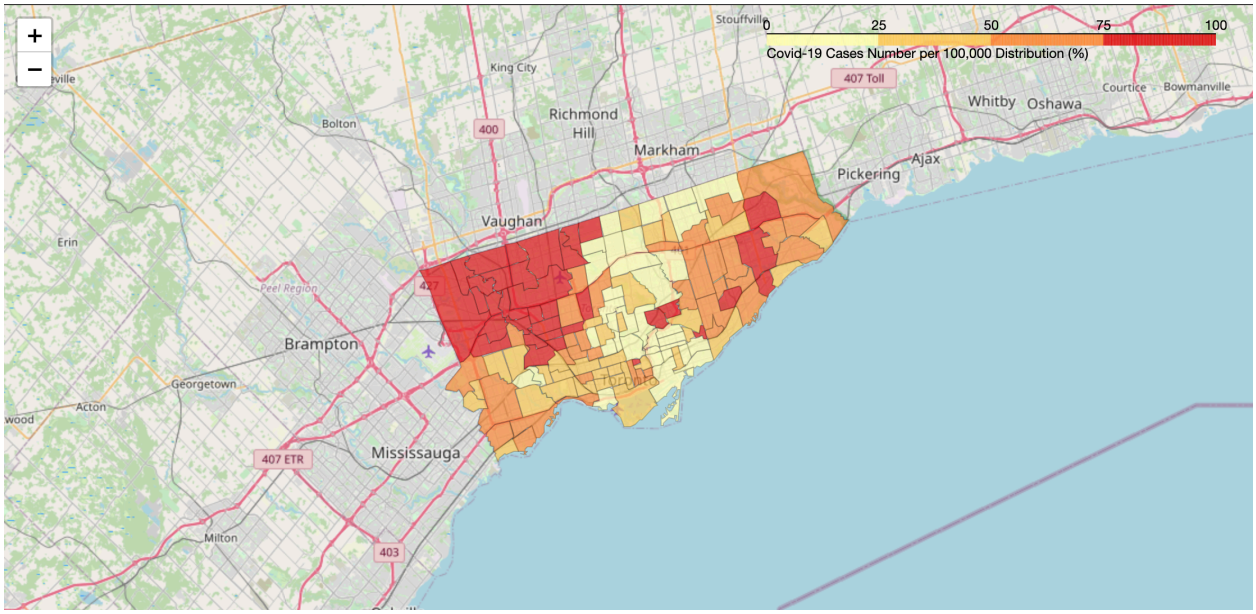
All parameters related to Neighborhood Profile were transformed into percentage term to facilitate initial data exploration and neighborhoods were categorized into 4 groups: "Bottom 25%","25% to 50%","50% to 75%" and "Top 25%" in terms of COVID-19 Rate per 100,000 people. Among all neighborhood profiles data, some interesting findings were found and listed below.

covid_cat	Total visible minority population	Latin American	Black	Southeast Asian	6 Sales and service occupations	7 Trades, transport and equipment operators and related occupations	9 Occupations in manufacturing and utilities	Public transit	Between 5 a.m. and 5:59 a.m.	Between 12 p.m. and 4:59 a.m.	Total income: Average amount ($)
Bottom 25%	0.393706	0.014338	0.039586	0.007334	0.194078	0.049571	0.019370	0.383061	0.025760	0.112190	76610.028571
25% to 50%	0.347969	0.020182	0.043191	0.010764	0.206506	0.067783	0.021055	0.353799	0.034634	0.129043	68738.000000
50% to 75%	0.518294	0.027850	0.095275	0.012821	0.258017	0.106261	0.044375	0.379313	0.053712	0.166873	42400.057143
Top 25%	0.626352	0.056894	0.172441	0.030905	0.276187	0.149152	0.078692	0.381981	0.070070	0.203235	33245.885714

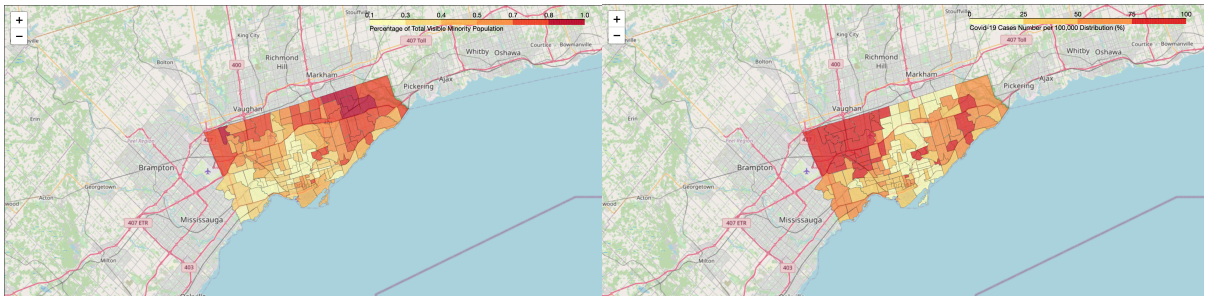
Most of the features included above showed interesting insight to the COVID-19 status in Toronto. For instance, taking public transit as the main method to work actually showed little to

no link with COVID-19 cases. More data visualizations were presented to better showcases some other connections.

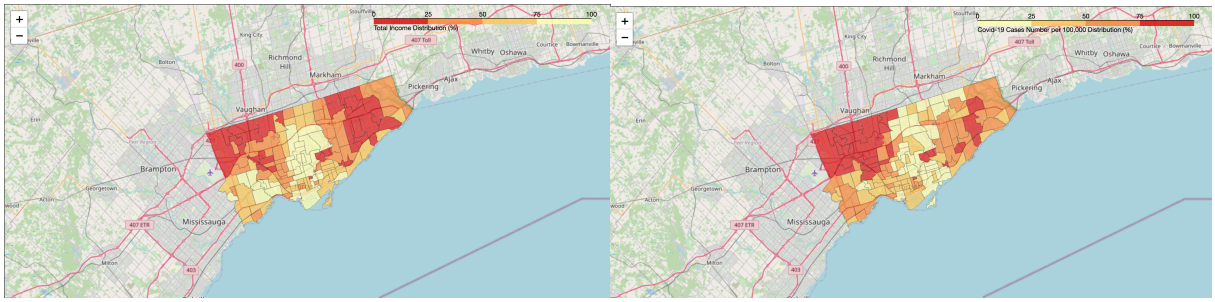
The most crowded downtown toronto was not the hardest hit neighborhood surprisingly; If an infectious disease wasn't most rampant in the overcrowded area, other factors must be investigated in an effort to explain for the distribution of COVID-19 cases.



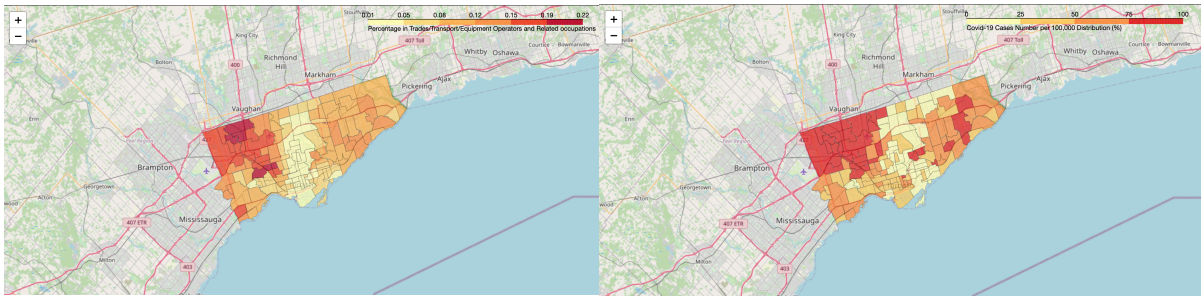
The percentage of visible minority was one possible factor. As mapped out in the choropleth, neighborhoods with more visible minority tended to have higher cases of COVID-19; even thou the link was not exceptionally distinctive.



Using the distribution of total income in Toronto, a more obvious link with COVID-19 cases can be clearly observed. Lower total income almost covered all regions of higher COVID-19 cases.



Much like total income, occupations also showed a link to COVID-19 cases. As the example given in the graph, Neighborhoods with higher Trade, transport and equipment operators also mirrored the COVID-19 distribution to an extent.



Based on the initial data exploration, further inferential statistics on Spearman correlation were listed as followings for some key features:

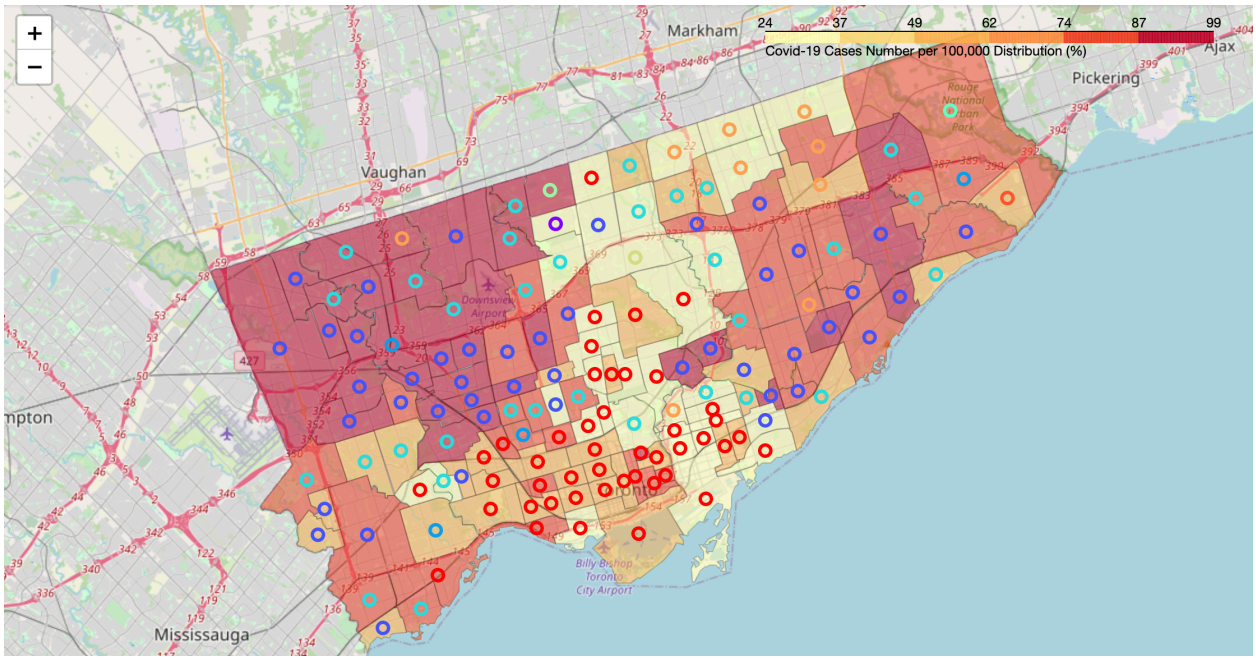
Correlation of COVID-19 & Income	-0.6927598346854432
Correlation of COVID-19 & Black population	0.725389779361921
Correlation of COVID-19 & Trades/Transport/Equipment occupations	0.7562135094356127

More workers in trade/transport/equipment and larger black population all suggested higher COVID-19 case number in a given neighborhood. Additionally, lower income neighborhoods were linked to higher COVID-19 cases as demonstrated by the negative correlation. While inferential statistics were useful for investigating individual features, a more comprehensive model was needed to better assess various features simultaneously.

K-Means clustering were used to group all 140 neighborhoods based on their features as the distinct characteristic of certain clusters could unveil the systematic issues in City of Toronto. Two types of K-Means clustering were employed to deal with two potential kinds of systematic issues: one in neighborhood planning associated with venue profiles and the other one in neighborhood demographic profiles. The cluster number of 10 and 5 were used respectively for the first and the second K-Means.

4. Results

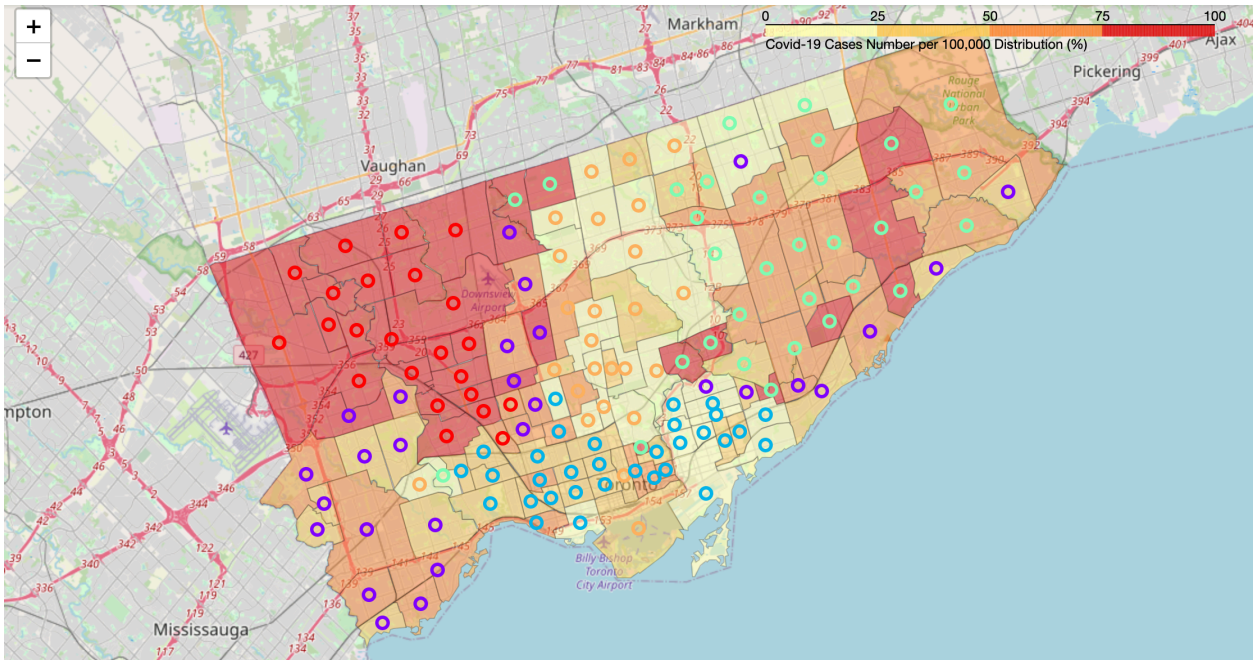
Clustering by venues' categories, Cluster 2, 4 and 6 stood out (blue, light blue and light green circles).



Looking at the five most common venue categories in each cluster, cluster 2, 4 and 6 didn't really differ too much from the rest outside the inclusion of Chinese restaurant, bus line and gym.

	Cluster Group	First	Second	Third	Fourth	Fifth
0	0	Coffee Shop	Park	Café	Italian Restaurant	Bank
1	1	Coffee Shop	Fast Food Restaurant	Sandwich Place	Bank	Grocery Store
2	2	Park	Pizza Place	Chinese Restaurant	Pharmacy	Coffee Shop
3	3	Pizza Place	Park	Coffee Shop	Grocery Store	Furniture / Home Store
4	4	Park	Coffee Shop	Café	Pizza Place	Italian Restaurant
5	5	Park	Coffee Shop	Italian Restaurant	Bakery	Sushi Restaurant
6	6	Café	Bus Line	Gym / Fitness Center	Coffee Shop	College Gym
7	7	Zoo Exhibit	Zoo	Other Great Outdoors	Dessert Shop	Theme Park Ride / Attraction
8	8	Park	Baseball Field	Pool	Zoo Exhibit	Fish & Chips Shop
9	9	Park	Coffee Shop	Electronics Store	Health & Beauty Service	Event Service

Clustering by demographic characteristics, cluster 0 (red circle) stood out the most.



Looking at cluster 0, an aberration could be seen as Black/Latin American/Southeast Asian population were exceptionally high.

Cluster Group	Latin American	Total visible minority population	South Asian	Chinese	Black	Filipino	Arab	Southeast Asian	West Asian	Korean	Japanese	Visible minority; n.i.e.
0	0.078275	0.609397	0.124614	0.021077	0.209834	0.056234	0.012892	0.044514	0.013404	0.004091	0.001721	0.021907
1	0.028954	0.315900	0.059340	0.037334	0.066167	0.059202	0.006617	0.009828	0.008271	0.011191	0.004842	0.010248
2	0.021288	0.317965	0.055730	0.090988	0.052838	0.024835	0.007727	0.014277	0.007048	0.009899	0.007977	0.009834
3	0.015705	0.729333	0.237859	0.168827	0.101576	0.094221	0.018736	0.008134	0.031145	0.011167	0.003276	0.017887
4	0.015177	0.385855	0.049711	0.149319	0.025885	0.028848	0.011120	0.005909	0.040789	0.036049	0.006998	0.003609

Cluster also saw high percentages in trade, transport and equipment operators and related occupation along with occupations in manufacturing and utilities.

0 Management occupations	1 Business, finance and administration occupations	2 Natural and applied sciences and related occupations	3 Health occupations	4 Occupations in education, law and social, community and government services	5 Occupations in art, culture, recreation and sport	6 Sales and service occupations	7 Trades, transport and equipment operators and related occupations	8 Natural resources, agriculture and related production occupations	9 Occupations in manufacturing and utilities
0.061596	0.141954	0.042196	0.049998	0.089367	0.021949	0.277449	0.177815	0.008086	0.091374
0.117232	0.171669	0.072586	0.053326	0.137473	0.047925	0.226381	0.108718	0.007345	0.031307
0.135842	0.162758	0.075829	0.047560	0.160321	0.108772	0.212930	0.052957	0.004823	0.016144
0.073359	0.165604	0.082528	0.059224	0.097978	0.025676	0.280005	0.105796	0.004228	0.061423
0.160418	0.207693	0.097351	0.074545	0.158260	0.060008	0.168947	0.033316	0.003399	0.010476

No significant aberration was found in commuting method as opposed to commuting hours, which those leaving before 6 am and after 12 pm both showed higher percentages in Cluster 0. The total income for cluster 0 was also significantly low compared to the entire city.

Car, truck, van - as a driver	Car, truck, van - as a passenger	Public transit	Walked	Bicycle	Other method	Between 5 a.m. and 5:59 a.m.	Between 6 a.m. and 6:59 a.m.	Between 7 a.m. and 7:59 a.m.	Between 8 a.m. and 8:59 a.m.	Between 9 a.m. and 11:59 a.m.	Between 12 p.m. and 4:59 a.m.	Total income: Average amount ($)	Rate per 100,000 people
0.536117	0.067767	0.351216	0.029823	0.006556	0.008734	0.081533	0.200434	0.229378	0.172928	0.111536	0.204139	32463.391304	759.386196
0.557646	0.046295	0.337659	0.033456	0.014473	0.010409	0.052278	0.144746	0.277059	0.247890	0.132149	0.145202	52733.750000	267.880594
0.309944	0.030011	0.416270	0.141926	0.088384	0.013188	0.029018	0.093876	0.228351	0.322589	0.190444	0.136037	56832.093750	168.706131
0.501417	0.054888	0.392246	0.036897	0.005577	0.008662	0.052261	0.149940	0.238163	0.213573	0.159224	0.186859	33114.781250	302.011272
0.475145	0.033877	0.361219	0.099580	0.016310	0.013567	0.020248	0.082151	0.265001	0.352940	0.187175	0.092136	105331.440000	138.074272

The most common venues for this clustering method was listed as followings;

	Cluster Group	First	Second	Third	Fourth	Fifth
0	0	Coffee Shop	Pizza Place	Grocery Store	Fast Food Restaurant	Gas Station
1	1	Park	Coffee Shop	Pizza Place	Convenience Store	Grocery Store
2	2	Café	Coffee Shop	Park	Bakery	Italian Restaurant
3	3	Park	Coffee Shop	Chinese Restaurant	Pizza Place	Pharmacy
4	4	Park	Coffee Shop	Café	Italian Restaurant	Pizza Place

5. Discussion

Clustering based on venue characteristics showed that the clusters with most COVID-19 cases didn't differ too much from the rest of Toronto in terms of most common venues. The clusters did point out venues like Chinese restaurant, bus line and gym but the first two categories wouldn't fit with the demographics profiling as more Chinese or commuting population didn't provide much difference in COVID-19 cases. Gym is a possible source of infection due to its nature of being an indoor venue with high chance of transmission, it would still be rash to make assumption considering the magnitude of lockdown in Toronto. Most likely the lockdown measure worked, and no imperative insights were available as a result.

.

Clustering based on demographic profiles confirmed many insights seen in exploratory analysis. K-Means clustering grouped together the hardest hit neighborhoods into the top left corner of City of Toronto. These neighborhoods clearly consisted mostly of visible minorities such as black, Latin American or Southeast Asian population. While it is possible to attribute such connection with genetic difference among races, the link could actually be well explained by other parts of the analysis instead. Visible minorities were taking up more dangerous and unstable occupations which would expose them to higher risks; The less regular commute hours and higher proportion of labor work population in these neighborhoods corroborated this viewpoint. To make matter worse, their average total incomes also ranked way lower than the rest of the City of Toronto. It is surprising that labor work occupations showed more significant link to the COVID-19 cases than healthcare occupation, who intuitively should be exposed directly to higher infectious risks. Government has to oversee the improvement in these sectors to better protect their workers; otherwise the wide spread of COVID-19 compounded with low income in these areas would unfortunately produce disastrous outcome. This would be the most important step in dealing with a potential second wave in the future as other factors like modes of commutes or neighborhood venues profiles didn't stand out in our analysis.

6. Conclusion

The situation of COVID-19 pandemic in Toronto was evaluated in this report. As no significant inference could be made based on the clustering of neighborhood venue profiles, Toronto lockdown measured likely succeeded in containing the spread of the pandemic. On the other hand, clustering on neighborhood demographics elucidated the improvement could be done to better protect blue collar workers in Toronto. The imbalance of cases between this population and the rest of the city of Toronto was truly alarming.

This report was an attempt to investigate and explain the systematic bias existed in Toronto for COVID-19 cases. As all data were analyzed on neighborhood level, future analyses on patient level will be needed to further confirm any insights and findings while avoiding the potential bias associated with neighborhood level data.

Reference:

City of Toronto COVID-19 Summary
https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/
Neighborhood profiles
https://open.toronto.ca/dataset/neighbourhood-profiles/
Neighbourhoods
https://open.toronto.ca/dataset/neighbourhoods/