Understanding COVID-19 in Toronto Neighbourhoods – A Data-driven Approach For Vaccination Prioritization

By Chris Blackwood

Introduction

COVID-19 has been a major health and economic issue worldwide since early 2020. The virus was first identified in humans in Wuhan, China in December of 2019, and the World Health Organization recognized a global pandemic in March of 2020 [1]. As of late January 2021, over 100 million cases have been identified worldwide, and over 2 million people have died [2].

Many organizations and countries initiated vaccine development promptly. The Pfizer-BioNTech vaccine was the first vaccine approved for use in Canada, and distribution began in late December of 2020 in that country [3]. Vaccine production is complex, worldwide demand is extreme, and distribution logistics and vaccine storage is non-trivial. As a result, it will take many months before everyone receives a vaccine. In Canada, the priority recipients of the vaccine are generally the elderly and healthcare workers, and individual provinces determine the exact distribution order [4]. In the province of Ontario for example, health care workers in hospitals, long-term care homes, and retirement homes were amongst the first to receive vaccination [5].

Business Problem

Data scientists can play a valuable role in understanding the geographic and demographic distributions of COVID-19 infections at a detailed level. Data can be used to show whether COIVD-19 infections are more prevalent in certain areas or populations rather than others.

I examine COVID-19 infections in the city of Toronto, the capital of the Canadian province of Ontario. I am interested in the rates of COVID-19 infection and I will sort these data by geographic area within the city to see if certain neighbourhoods have more infections than others, or if the distribution is random. I will generate choropleth maps to categorize the distribution of COVID-19 cases. To enhance this analysis, I will query Foursquare for neighbourhoods of interest to see the types and abundances of venues in these neighbourhoods, and determine if there are any correlations between the types of venues and the COVID-19 infection rates.

Secondly, I will attempt to understand the different regions of Toronto in terms of demographic characteristics. I will compare COVID-19 infection rates to a variety of census data, including population statistics, household income, and education level. If there are regions of the city that have particularly high or low COVID-19 cases, I will attempt to identify any demographic factors of significance in these areas. I will use an assortment of statistical and mapping analysis techniques to determine relevance of demographic factors.

The target audience for the study is government officials responsible for public health measures designed to limit the spread of COVID-19 and to coordinate vaccination distribution. My hope is that trends will become apparent, such as particular geographic

regions, common venues, or demographic factors correlate with high COVID-19 case rates. If I am successful in identifying factors that correlate with high COVID-19 infection rates, then these correlations could be used to help prioritize neighbourhoods for heath measures and vaccine distributions.

Data

The city of Toronto maintains a strong online database of city-specific information. I will reference three key databases from the city of Toronto and augment with data obtained from Foursquare. The main data sources are listed in Table 1.

Item	Description	Location
COVID-19 data for city of Toronto	Daily update of COVID-19 cases in Toronto broken down by neighbourhood in Excel and PDF formats	https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/
City of Toronto neighbourhood profiles	Toronto census data in CSV format	https://open.toronto.ca/dataset/neighbourhood-profiles/
Toronto neighbourhood spatial information	Geojson file of 140 Toronto neighbourhoods	https://open.toronto.ca/dataset/neighbourhoods/
Foursquare queries	Query results of venue information for specific neighbourhoods	Results returned from Foursquare, converted to dataframe in Python

Table 1. Key data sources utilized in project.

COVID-19 infection data for Toronto is updated daily. The COVID-19 data I employ was retrieved on January 15, 2021. The dataset consists of neighbourhood name and ID number, cumulative case count per neighbourhood, and cumulative infection rate per 100,000 people per neighbourhood. An example is shown in Figure 1.

Neighbourhood ID	Neighbourhood Name	Rate per 100,000 people	Case Count
138	Eglinton East	3464.172813	789
47	Don Valley Village	1829.876899	495
38	Lansing-Westgate	1336.302895	216
9	Edenbridge-Humber Valley	2471.837786	384
44	Flemingdon Park	3720.421283	816
59	Danforth East York	1222.351572	210
129	Agincourt North	2109.02346	614
99	Mount Pleasant East	846.4977645	142
137	Woburn	3791.717304	2028
102	Forest Hill North	2045.915977	262
111	Rockcliffe-Smythe	3205.070574	713
130	Milliken	2009.634201	534
55	Thorncliffe Park	4368.012128	922
7	Willowridge-Martingrove-Richview	2735.150749	606

Figure 1. Sample of COVID-19 data for neighbourhoods in the city of Toronto.

The Toronto neighbourhood profile dataset is an extensive collection of demographic data collected as part of a 2016 census. This is the most recent census data available for the city. One assumption of the project is that demographic trends have not changed in a significant manner for Toronto neighbourhoods since 2016. An example of the dataset is shown in Figure 2. The dataset may be indexed by either Neighbourhood ID or name, which makes it possible to compare with the COVID-19 data. The data contains over 2000 demographic classifications, and not all will be considered in this study. Broad categories of classification include education level, household income, age and gender, housing density, mother tongue, method of transport to employment, and citizenship.

Id	Category	Topic	Data Source	Characteristi	City of Toron	Agincourt Nc	Agincourt So	Alderwood	Annex	Banbury-Dor	Bathurst Mai	Bay Street Cc	Bayview Villa	Bayview Woo	Bedford Park	Beechboroug	Bendale	Birchcliffe-Cl	Black Creek
1	Neighbourhc	Neighbourhood Information	City of Toron	Neighbourhood	Number	129	128	20	95	42	34	76	52	49	39	112	127	122	24
2	Neighbourhc	Neighbourhood Information	City of Toron	TSNS2020	Designation	No Designati	No Designati	No Designati	No Designati	No Designati	No Designati	No Designati	No Designati	No Designati	No Designati	NIA	No Designati	No Designati	NIA
3	Population	Population and dwellings	Census Profil	Population, 2	2,731,571	29,113	23,757	12,054	30,526	27,695	15,873	25,797	21,396	13,154	23,236	6,577	29,960	22,291	21,737
4	Population	Population and dwellings	Census Profil	Population, 2	2,615,060	30,279	21,988	11,904	29,177	26,918	15,434	19,348	17,671	13,530	23,185	6,488	27,876	21,856	22,057
5	Population	Population and dwellings	Census Profil	Population C	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	2.80%	33.30%	21.10%	-2.80%	0.20%	1.40%	7.50%	2.00%	-1.50%
6	Population	Population and dwellings	Census Profil	Total private	1,179,057	9,371	8,535	4,732	18,109	12,473	6,418	18,436	10,111	4,895	9,052	2,796	11,293	9,637	7,671
7	Population	Population and dwellings	Census Profil	Private dwell	1,112,929	9,120	8,136	4,616	15,934	12,124	6,089	15,074	9,532	4,698	8,607	2,650	10,766	9,198	7,324
8	Population	Population and dwellings	Census Profil	Population d	4,334	3,929	3,034	2,435	10,863	2,775	3,377	14,097	4,195	3,240	4,209	3,614	4,011	3,765	6,282
9	Population	Population and dwellings	Census Profil	Land area in	630.2	7.41	7.83	4.95	2.81	9.98	4.7	1.83	5.1	4.06	5.52	1.82	7.47	5.92	3.46
10	Population	Age characteristics	Census Profil	Children (0-1	398,135	3,840	3,075	1,760	2,360	3,605	2,325	1,695	2,415	1,515	4,555	1,120	4,550	3,345	4,600
11	Population	Age characteristics	Census Profil	Youth [15-24	340,270	3,705	3,360	1,235	3,750	2,730	1,940	6,860	2,505	1,635	3,210	855	4,605	2,440	3,290
12	Population	Age characteristics	Census Profil	Working Age	1,229,555	11,305	9,965	5,220	15,040	10,810	6,655	13,065	10,310	4,490	8,410	2,750	12,050	9,075	8,525
13	Population	Age characteristics	Census Profil	Pre-retireme	336,670	4,230	3,265	1,825	3,480	3,555	2,030	1,760	2,540	1,825	3,075	885	3,535	3,520	2,425
14	Population	Age characteristics	Census Profil	Seniors (65+	426,945	6,045	4,105	2,015	5,910	6,975	2,940	2,420	3,615	3,685	3,980	965	5,250	3,910	2,870
15	Population	Age characteristics	Census Profil	Older Senior	66,000	925	555	320	1,040	1,640	710	330	610	740	660	145	900	630	370
16	Population	Age characteristics	Census Profil	Male: 0 to 04	69,895	660	575	360	445	570	435	470	455	205	675	180	770	570	745
17	Population	Age characteristics	Census Profil	Male: 05 to (	69,350	695	540	270	365	660	355	230	395	260	795	195	790	580	800
18	Population	Age characteristics	Census Profil	Male: 10 to 1	64,945	660	460	225	325	675	415	130	410	320	880	170	775	535	785
19	Population	Age characteristics	Census Profil	Male: 15 to 1	74,240	840	780	285	465	715	490	585	520	385	880	210	955	585	765
20	Population	Age characteristics	Census Profil	Male: 20 to 2	97,415	1015	1000	355	1215	700	530	2485	735	445	765	215	1370	650	840
21	Population	Age characteristics	Census Profil	Male: 25 to 2	113,905	1015	1045	355	2080	645	465	2115	1075	405	405	225	1135	530	725
22	Population	Age characteristics	Census Profil	Male: 30 to 3	108,895	835	820	410	1610	735	485	1695	1040	285	450	200	995	600	635
23	Population	Age characteristics	Census Profil	Male: 35 to 3	94,070	680	625	455	1055	735	580	1010	805	230	590	205	835	675	580
24	Population	Age characteristics	Census Profil	Male: 40 to 4	86,535	760	610	420	835	815	435	560	685	310	665	170	840	740	585
25	Population	Age characteristics	Census Profil	Male: 45 to 4	90,860	890	760	440	850	1010	535	500	605	390	760	225	950	835	660
26	Population	Age characteristics	Census Profil	Male: 50 to 5	98,735	1160	970	515	920	1110	605	540	620	430	845	220	1010	1020	680
27	Population	Age characteristics	Census Profil	Male: 55 to 5	88,145	1060	850	540	855	930	565	475	560	450	790	210	945	915	640
28	Population	Age characteristics	Census Profil	Male: 60 to 6	72,270	925	710	390	755	730	385	375	555	375	670	230	765	810	465
29	Population	Age characteristics	Census Profil	Male: 65 to 6	60,360	925	630	300	780	715	380	295	475	405	560	120	650	640	380
30	Population	Age characteristics	Census Profil	Male: 70 to 7	42,320	590	425	205	640	570	225	285	345	315	400	125	495	420	285
31	Population	Age characteristics	Census Profil	Male: 75 to 7	32,730	490	350	155	485	505	200	200	265	310	270	65	435	245	245
32	Population	Age characteristics	Census Profil	Female: 10 to	62,165	595	475	255	355	600	330	105	350	270	820	180	765	575	730
33	Population	Age characteristics	Census Profil	Male: 80 to 8	25,670	380	240	105	335	435	165	150	230	255	250	90	355	200	175
34	Population	Age characteristics	Census Profil	Male: 85 to 8	15,665	210	155	65	230	340	110	80	150	165	145	25	230	125	95
35	Population	Age characteristics	Census Profil	Male: 90 to 9	6,185	100	45	35	80	175	80	35	75	55	60	10	80	55	40
36	Population	Age characteristics	Census Profil	Male: 95 to 9	1,280	25	20	10	25	20	20	0	10	20	15	0	10	25	5
37	Population	Age characteristics	Census Profil	Male: 100 ye	125	0	0	0	5	5	5	0	0	0	0	0	10	5	0
38	Population	Age characteristics	Census Profil	Female: 0 to	66,105	620	530	320	480	500	400	500	405	185	605	190	735	510	755
39	Population	Age characteristics	Census Profil	Female: 05 to	65,680	620	495	325	400	600	365	235	405	265	780	200	710	570	785
40	Population	Age characteristics	Census Profil	Female: 15 to	71,285	865	690	290	550	660	405	790	425	365	840	205	900	555	820
41	Population	Age characteristics	Census Profil	Female: 20 to	97,330	975	895	310	1520	650	510	3005	815	435	725	225	1385	635	860
42	Population	Age characteristics	Census Profil	Female: 25 to	119,040	1005	975	350	2265	745	480	2440	1145	410	475	235	1155	555	845
43	Population	Age characteristics	Census Profil	Female: 30 to	115,675	935	835	430	1675	860	575	1645	1170	285	645	240	1130	635	805
44	Population	Age characteristics	Census Profil	Female: 35 to	102,240	775	715	450	1040	895	615	825	860	310	785	260	925	715	730
45	Population	Age characteristics	Census Profil	Female: 40 to	95,860	910	745	445	915	975	575	530	760	385	835	225	945	870	735
46	Population	Age characteristics	Census Profil	Female: 45 to	100,065	1060	885	430	855	1170	650	555	780	535	985	270	1030	895	780
47	Population	Age characteristics	Census Profil	Female: 50 to	103,670	1290	975	515	935	1115	630	665	750	520	985	280	1090	1045	795
48	Population	Age characteristics	Census Profil	Female: 55 to	94,660	1165	915	485	915	1005	605	500	775	525	885	250	990	990	730
49	Population	Age characteristics	Census Profil	Female: 60 to	81,600	1070	795	400	940	895	470	425	660	475	730	185	830	810	610
50	Population	Age characteristics	Census Profil	Female: 65 to	70,180	985	690	325	950	955	415	445	585	520	660	155	795	690	470
51	Population	Age characteristics	Census Profil	Female: 70 to	51,285	690	450	210	700	790	275	320	440	400	505	110	580	450	355
52	Population	Age characteristics	Census Profil	Female: 75 to	43,430	575	405	180	565	730	280	250	355	400	390	90	545	345	330
53	Population	Age characteristics	Census Profil	Female: 80 to	34,965	485	350	210	425	650	285	170	315	330	305	60	500	290	235
54	Population	Age characteristics	Census Profil	Female: 85 to	25,135	350	205	130	345	615	265	135	230	265	260	45	350	250	130
55	Population	Age characteristics	Census Profil	Female: 90 to	13,500	160	100	70	260	360	165	50	115	180	140	60	155	140	80
56	Population	Age characteristics	Census Profil	Female: 95 to	3,475	60	20	5	90	105	45	5	25	55	35	0	65	40	15
57	Population	Age characteristics	Census Profil	Female: 100	650	10	0	5	25	20	10	0	0	10	10	0	10	5	15
58	Families, hou	Household and dwelling characteristi	Census Profil	Occupied pri	1,112,930	9125	8135	4620	15935	12125	6080	15075	9535	4700	8610	2655	10765	9205	7320
59	Families, hou	Household and dwelling characteristi	Census Profil	Single-detac	269,675	3345	2790	2840	645	3485	2105	5	2025	1565	4820	665	3870	4900	680
60	Families, hou	Household and dwelling characteristi	Census Profil	Apartment	493,275	2120	3145	255	8165	6270	2055	14800	5900	1775	1995	665	4850	945	4065
61	Families, hou	Household and dwelling characteristi	Census Profil	Other attac	349,880	3645	2200	1525	7125	2370	1925	280	1610	1365	1790	1325	2040	3350	2585
62	Families, hou	Household and dwelling characteristi	Census Profil	Semi-detac	71,225	805	330	545	1185	285	165	0	120	495	45	80	290	425	660
63	Families, hou	Household and dwelling characteristi	Census Profil	Row house	61,630	1440	515	85	595	740	45	15	880	200	65	15	325	40	1140
64	Families, hou	Household and dwelling characteristi	Census Profil	Apartment	48,540	645	695	330	455	40	640	0	75	100	315	90	1075	880	155
65	Families, hou	Household and dwelling characteristi	Census Profil	Apartment	165,630	735	610	560	4795	1315	1070	240	520	570	1260	1130	340	1960	630
66	Families, hou	Household and dwelling characteristi	Census Profil	Other sing	2,860	15	65	0	90	0	15	10	5	0	115	0	20	40	5
67	Families, hou	Household and dwelling characteristi	Census Profil	Movable dw	95	5	0	0	0	0	0	0	0	0	0	0	5	0	0
68	Families, hou	Household and dwelling characteristi	Census Profil	Private hous	1,112,930	9125	8140	4620	15925	12125	6095	15075	9530	4695	8610	2650	10760	9200	7330
69	Families, hou	Household and dwelling characteristi	Census Profil	1 person	359,955	1350	1610	1105	7885	4360	1660	7605	3195	975	2390	870	2805	2960	1660
70	Families, hou	Household and dwelling characteristi	Census Profil	2 persons	333,425	2370	2325	1440	5220	3820	1790	5330	3260	1505	2415	765	3190	2860	1820
71	Families, hou	Household and dwelling characteristi	Census Profil	3 persons	175,720	1995	1680	885	1540	1755	1095	1410	1610	935	1190	475	1850	1460	1355
72	Families, hou	Household and dwelling characteristi	Census Profil	4 persons	146,580	1750	1335	795	885	1515	970	575	1000	820	1515	320	1465	1320	1215
73	Families, hou	Household and dwelling characteristi	Census Profil	5 or more p	97,245	1645	1175	390	390	675	575	145	470	480	1110	225	1465	605	1360

Figure 2. Sample of demographic data for the City of Toronto.

A Geojson dataset broken down by neighbourhoods is also available for the City of Toronto. This will allow trends identified for specific neighbourhoods to be identified in map view. A map view of the Toronto neighbourhoods from this file is shown in Figure 3.

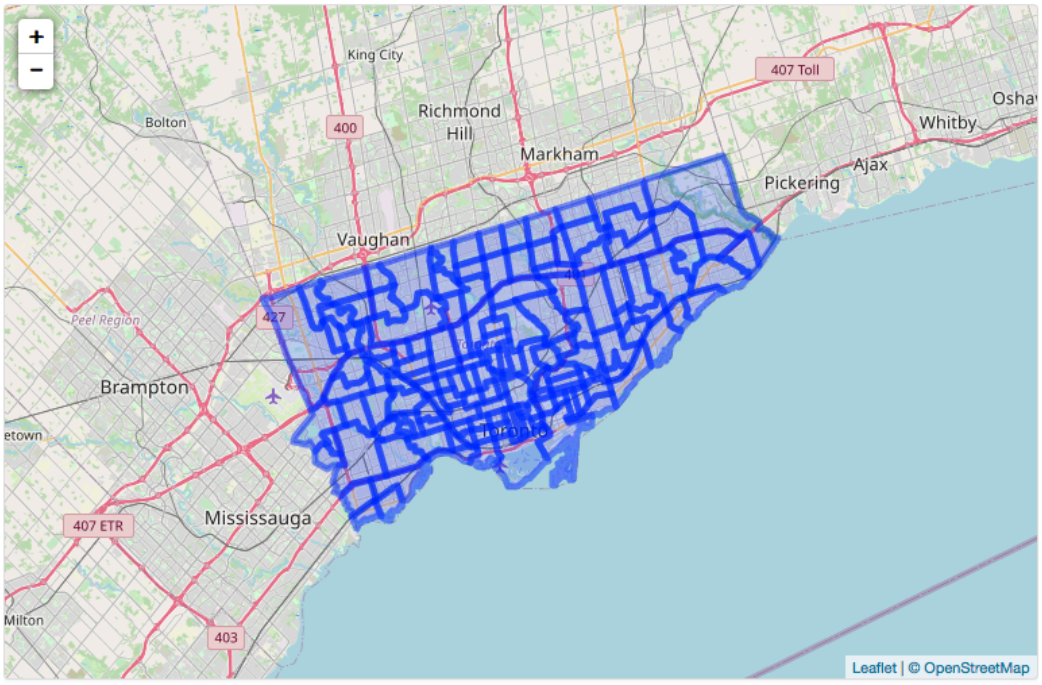

Figure 3. Map view of Geojson file used in this study.

Once I identify neighbourhoods of high or low rates of COVID-19 infection, I will run queries in Foursquare to understand the types and abundances of venues in each neighbourhood. Figure 4 is a sample of a Foursquare result converted to a data frame for analysis.

[62]:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	1	Black Creek	Home Service	Hotel	Fast Food Restaurant	Intersection	Electronics Store	Coffee Shop	Italian Restaurant
1	1	Forest Hill South	Pizza Place	Trail	Gift Shop	Intersection	Israeli Restaurant	Coffee Shop	Korean Restaurant
2	4	Humbermede	Golf Course	Furniture / Home Store	Coffee Shop	Department Store	Dance Studio	Gastropub	Food & Drink Shop
3	3	Maple Leaf	Park	Trail	Convenience Store	Furniture / Home Store	Food & Drink Shop	Fish & Chips Shop	Fast Food Restaurant
4	0	Mount Olive-Silverstone-Jamestown	Park	Baseball Field	Caribbean Restaurant	Trail	Dance Studio	Furniture / Home Store	Food & Drink Shop
5	1	Rosedale-Moore Park	Coffee Shop	Fish & Chips Shop	Breakfast Spot	Pizza Place	Fast Food Restaurant	Intersection	Trail
6	1	Runnymede-Bloor West Village	Asian Restaurant	Café	Food & Drink Shop	Diner	Market	Coffee Shop	Sushi Restaurant
7	1	The Beaches	Fast Food Restaurant	Italian Restaurant	Ice Cream Shop	Burrito Place	Liquor Store	Food & Drink Shop	Fish & Chips Shop
8	2	Thistletown-Beaumond Heights	Convenience Store	Home Service	Construction & Landscaping	Dance Studio	Furniture / Home Store	Food & Drink Shop	Fish & Chips Shop
9	1	Woodbine Corridor	Fast Food Restaurant	Italian Restaurant	Ice Cream Shop	Burrito Place	Liquor Store	Food & Drink Shop	Fish & Chips Shop

Figure 4. Example of data frame created from Foursquare data query.

Methodology

Three datasets – COVID-19 data, demographic data, and a Geojson file - were downloaded from the City of Toronto website. The COVID-19 and demographic data each contained columns of neighbourhood name and neighbourhood number (1-140), so merging the datasets was possible. The Geojson file referenced the same neighbourhood number, allowing the statistical data to be referenced to the spatial data. Data cleaning and manipulation were completed in both Excel and Python. For many of the figures included in this report, abbreviations for data categories are employed and will be explained as required. The most important is 'Covid19Rate,' representing the number of confirmed COVID-19 infections per 100,000 residents of each neighbourhood. Covid19CaseCount is the total cumulative cases of COVID-19 for the given neighbourhood.

The most fundamental question to answer is if certain neighbourhoods have higher rates of COVID-19 infections than others. Figures 5 and 6 show the neighbourhoods with the lowest and highest number of COVID-19 infections per 100,000 residents in table form. There are significant differences, as the neighbourhood with the highest infection rate (Thistletown-Beaumond Heights) has a rate of more than ten times the neighbourhood with the lowest infection rate (The Beaches).

Neighbourhood	Covid19Rate
Bridle Path-Sunnybrook-York Mills	938.916469
Newtonbrook East	938.062993
Lawrence Park South	935.502998
Willowdale East	854.582226
Mount Pleasant East	846.497764
Rosedale-Moore Park	831.620704
Woodbine Corridor	813.332270
Forest Hill South	810.659709
Runnymede-Bloor West Village	645.481629
The Beaches	625.956322

Figure 5. Neighbourhoods with lowest rates of COVID-19 infection.

Neighbourhood	Covid19Rate
Thistletown-Beaumond Heights	6650.579151
Mount Olive-Silverstone-Jamestown	6463.555259
Maple Leaf	6201.167046
Black Creek	6095.597369
Humbermede	5969.765198
West Humber-Clairville	5586.575408
Humber Summit	5565.399485
Glenfield-Jane Heights	5559.017415
Weston	5491.329480
Downsview-Roding-CFB	5232.226406

Figure 6. Neighbourhoods with highest rates of COVID-19 infection.

Next Foursquare was used to understand the number and variety of venues present within each neighbourhood of interest. Foursquare queries were analyzed for the five neighbourhoods with the highest infection rates and the five neighbourhoods with the lowest infection rates. The number of venues per neighbourhood, and the types and abundance were considered. It was difficult to identify clear trends from the data. This will be expanded upon in the Results section.

The demographic data were examined for factors correlating with COVID-19 infection rates. A subset of the full demographic suite was chosen for investigation, specifically data related to visible minority status, age, household income, employment, education, and commute to work.

For each demographic category, a similar workflow was employed. A correlation matrix was generated and plotted in tables and as heatmaps. The critical information is the correlation between Cvod19Rate and the demographic feature of interest. Regression plots were also examined to visualize correlations that warrant further attention. For brevity, not all figures are captured here.

The correlation matrices for the visible minority data were split for ease of display and shown in figures 7 and 8. A correlation of 1.0 represents a perfect positive correlation, where the y-axis variable increases as the x-axis variable increases. Conversely, a value of -1.0 means the y-axis variable decreases at a proportional rate as the x-axis increases. A value of zero represents no correlation. Of primary interest is the row Covid19Rate, showing the correlation between the infection rate and the visible minority populations.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	VMAboriginal	VMLatinAmerica	VMBlack	VMArab	VMNotSpecified	VMMultiple	VMNot
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.260068	-0.221717	0.022347	-0.070637	0.083444	0.183580	-0.073978
Covid19Rate	-0.160640	1.000000	0.730923	-0.096396	0.533566	0.654493	0.166010	0.493298	0.311711	-0.334786
Covid19CaseCount	-0.027007	0.730923	1.000000	0.223191	0.602199	0.851278	0.462577	0.778155	0.776954	0.040880
VMAboriginal	0.260068	-0.096396	0.223191	1.000000	0.173971	0.322458	0.206579	0.381450	0.435464	0.530763
VMLatinAmerica	-0.221717	0.533566	0.602199	0.173971	1.000000	0.649863	0.198034	0.431688	0.488615	0.192868
VMBlack	0.022347	0.654493	0.851278	0.322458	0.649863	1.000000	0.395387	0.806073	0.736919	-0.062478
VMArab	-0.070637	0.166010	0.462577	0.206579	0.198034	0.395387	1.000000	0.364912	0.524611	0.165470
VMNotSpecified	0.083444	0.493298	0.778155	0.381450	0.431688	0.806073	0.364912	1.000000	0.725973	-0.009871
VMMultiple	0.183580	0.311711	0.776954	0.435464	0.488615	0.736919	0.524611	0.725973	1.000000	0.229005
VMNot	-0.073978	-0.334786	0.040880	0.530763	0.192868	-0.062478	0.165470	-0.009871	0.229005	1.000000

Figure 7. Correlation matrix for a portion of visible minority demographic data. The categories shown are visible minority Aboriginal, Latin American, Black, Arab, unspecified, multiple visible minorities, or not a visible minority.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	VMSouthAsian	VMChinese	VMFilipino	VMSoutheastAsian	VMWestAsian	VMKorean	VMJapanese
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.238937	0.218038	0.172835	-0.171758	-0.120980	-0.179417	0.087104
Covid19Rate	-0.160640	1.000000	0.730923	0.377711	-0.194829	0.410632	0.458677	-0.013907	-0.216567	-0.423823
Covid19CaseCount	-0.027007	0.730923	1.000000	0.740276	0.083278	0.666958	0.519743	0.214637	-0.015151	-0.060593
VMSouthAsian	0.238937	0.377711	0.740276	1.000000	0.182574	0.537870	0.059744	0.278101	-0.021373	0.025877
VMChinese	0.218038	-0.194829	0.083278	0.182574	1.000000	0.083815	-0.014296	0.413517	0.408836	0.410311
VMFilipino	0.172835	0.410632	0.666958	0.537870	0.083815	1.000000	0.159245	0.234661	0.070573	-0.028514
VMSoutheastAsian	-0.171758	0.458677	0.519743	0.059744	-0.014296	0.159245	1.000000	0.024845	-0.035139	-0.044588
VMWestAsian	-0.120980	-0.013907	0.214637	0.278101	0.413517	0.234661	0.024845	1.000000	0.834930	0.494523
VMKorean	-0.179417	-0.216567	-0.015151	-0.021373	0.408836	0.070573	-0.035139	0.834930	1.000000	0.651388
VMJapanese	0.087104	-0.423823	-0.060593	0.025877	0.410311	-0.028514	-0.044588	0.494523	0.651388	1.000000

Figure 8. Correlation matrix for a second portion of visible minority demographic data. The categories shown are visible minority South Asian, Chinese, Filipino, Southeast Asian, West Asian, Korean, and Japanese.

The visible minorities with the highest correlation to COVID-19 infections per 100,000 are Black (0.65), and Latin American (0.53). Japanese has the highest negative correlation at -0.45. Regression plots for visible minority Black and visible minority Japanese are shown in figures 9 and 10.

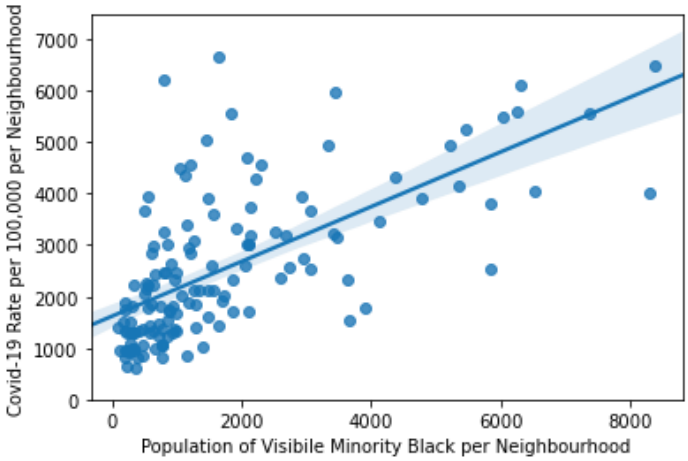

Figure 9. Regression plot of COVID-19 cases per 100,000 people for visible minority Black populations showing a strong positive correlation.

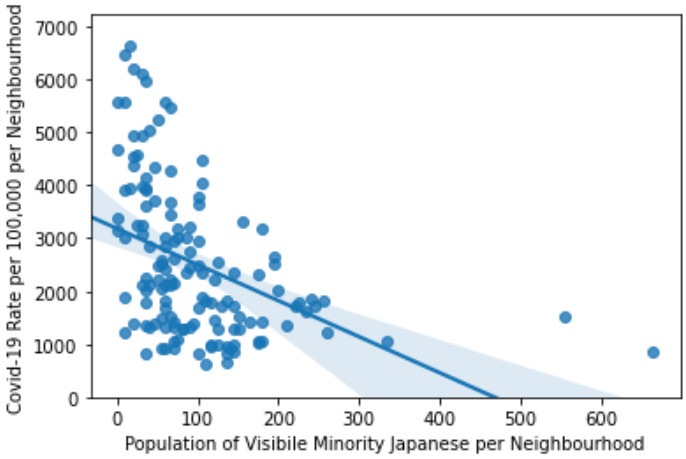

Figure 10. Regression plot of COVID-19 cases per 100,000 people for visible minority Japanese populations showing a negative correlation.

Total household income was separated into bins. The heat map of the correlation matrix for total household income and COVID-19 infections is shown in Figure 11. The bins chosen for household income are below $25,000, $25,000-$40,000, $40,000-60,000, $60,000-$80,000, and greater than $80,000. The most important row to consider is Covid19Rate. Neither of the income bins has a very high correlation with Covid19Rate, with the max value 0.20 for $25,000-$40,000. But the highest income bin ($80,000 plus) has the largest negative correlation at -0.26.

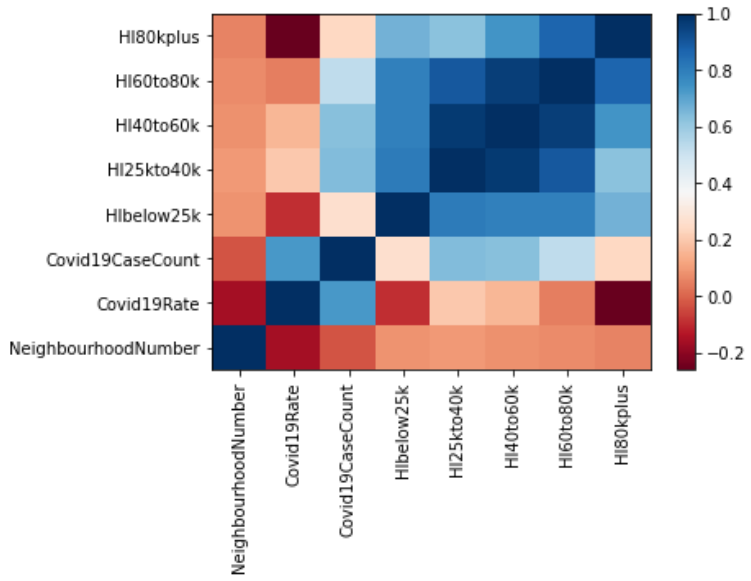

Figure 11. Heatmap of correlation matrix for total household income data.

Age of residents was grouped in bins of 0-14 years, 15-25 years, 25-54 years, 55-64 years, 65 plus, and 85 plus, the latter two subject to some overlap due to lack of specific differentiation. The heat map of the correlation matrix for age and COVID-19 infections is shown in Figure 12. Correlations do not differ greatly by age bin, but the highest positive correlations are for the two youngest bins, representing the range of infants, to school age, to young professionals.

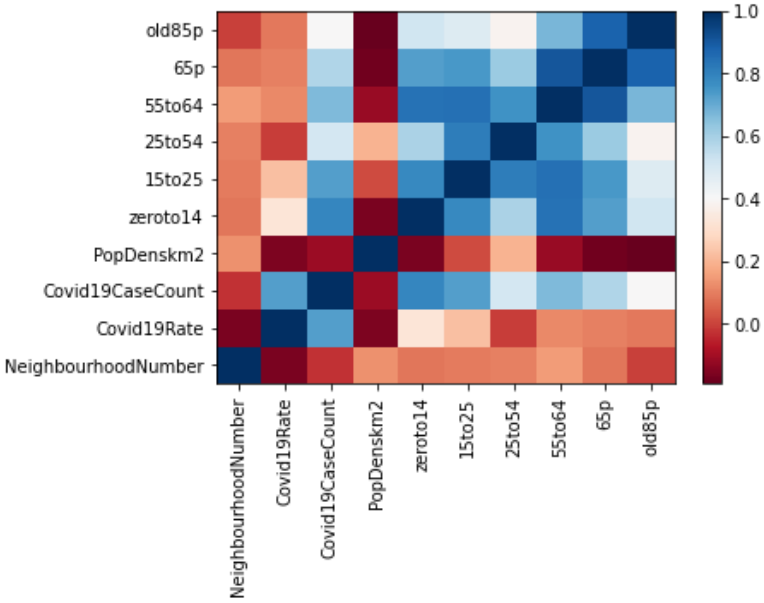

Figure 12. Heatmap of correlation matrix for age data.

An assortment of data describing commuting to work was analyzed, including duration of commute and method of transportation. There are high positive correlations between COVID-19 infection rates and both long commutes to work and commuting as a passenger

in a private vehicle, the second of which is shown in Figure 13. The strongest negative correlation is with those who cycle to work. There is no significant correlation between taking public transit and COVID-19 infections, as shown in Figure 14.

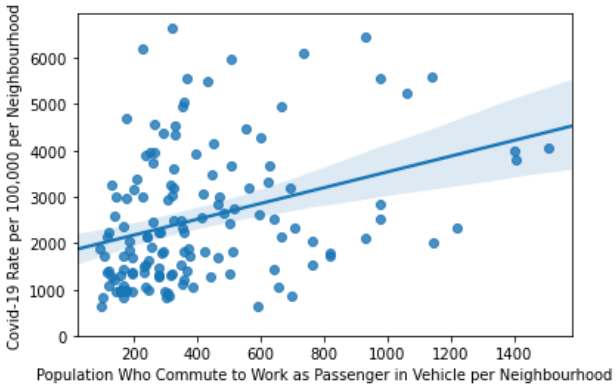

Figure 13. Regression plot showing a positive correlation between commuting to work as a passenger in a private vehicle and COVID-19 infections.

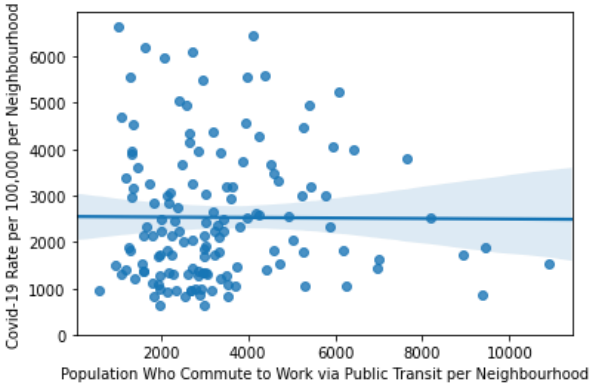

Figure 14. Regression plot showing no correlation between commuting to work via public transit and COVID-19 infections.

Employment data were also examined, and the correlation matrix is shown in table form in Figure 15. COVID-19 infection rate has a strong positive correlation with unemployment, and a strong negative correlation with employment. The trend for unemployment is shown in Figure 16.

	NeighbourhoodNumber	Covid19Rate	Covid19CaseCount	EmploymentRate	UnemploymentRate
NeighbourhoodNumber	1.000000	-0.160640	-0.027007	0.044850	0.091673
Covid19Rate	-0.160640	1.000000	0.730923	-0.518157	0.605126
Covid19CaseCount	-0.027007	0.730923	1.000000	-0.376483	0.486920
EmploymentRate	0.044850	-0.518157	-0.376483	1.000000	-0.765290
UnemploymentRate	0.091673	0.605126	0.486920	-0.765290	1.000000

Figure 15. Correlation matrix for employment and unemployment rates.

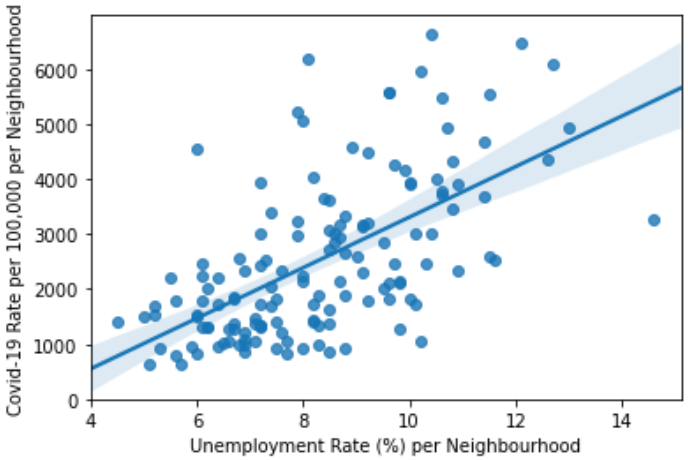

Figure 16. Regression plot of COVID-19 cases per 100,000 people vs. unemployment rate, showing a strong positive correlation.

The final demographic category considered was education level. Education was broken down into several bins, considering only adults aged 25-64. The categories and abbreviations are: no high school diploma (A25to64NoCert), high school diploma (A25to64HSdip), trade certificate or diploma (A25to64Trade), diploma from a non-university college or CEGEP (A25to64non-uniDip), university certificate or diploma below bachelor level (A25to64UniDipbelowBach), and university diploma of bachelor level or greater (A25to64UniBachorhigher). The heatmap of the correlation matrix is shown in Figure 17. Considering the row of Covid19rate, the highest positive correlations are associated with education less than high school and certified trades. University education bachelor or higher, has a negative correlation.

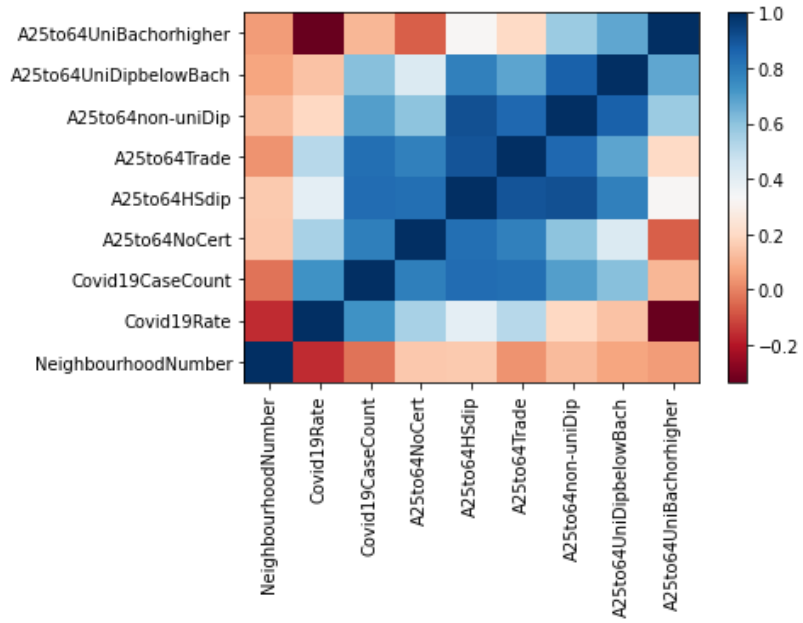

Figure 17. Heatmap of correlation matrix for education levels

Results

Neighbourhoods in Toronto with the highest and lowest number of COVID-19 infections per 100,000 people were identified in the Methodology section and shown in table form. These data are repeated in bar charts and are shown below. Figure 18 shows the neighbourhoods with the lowest number of infections per 100,000 people, while figure 19 shows the neighbourhoods with the highest number of infections. This is better visualized in the choropleth map shown in Figure 20. The highest rates are observed in neighbourhoods to the northwest, at the city limits. Case rates are also high in eastern neighbourhoods, and lowest in the centre and in the lakeside downtown core.

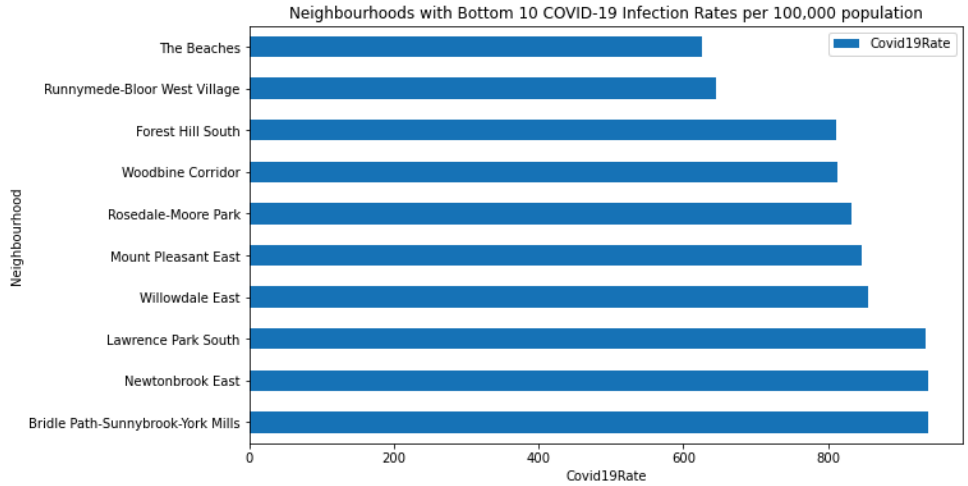

Figure 18. Neighbourhoods with lowest number of COVID-19 infections per 100,000 residents.

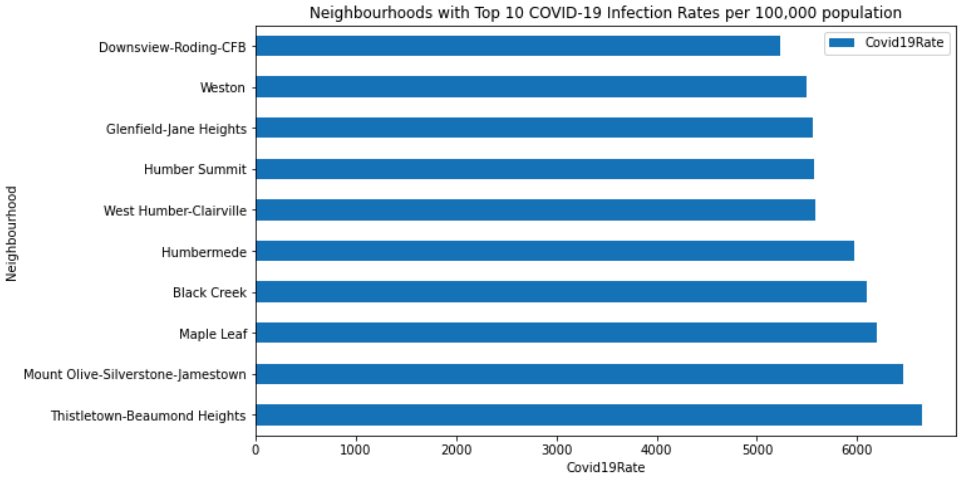

Figure 19. Neighbourhoods with highest number of COVID-19 infections per 100,000 residents.

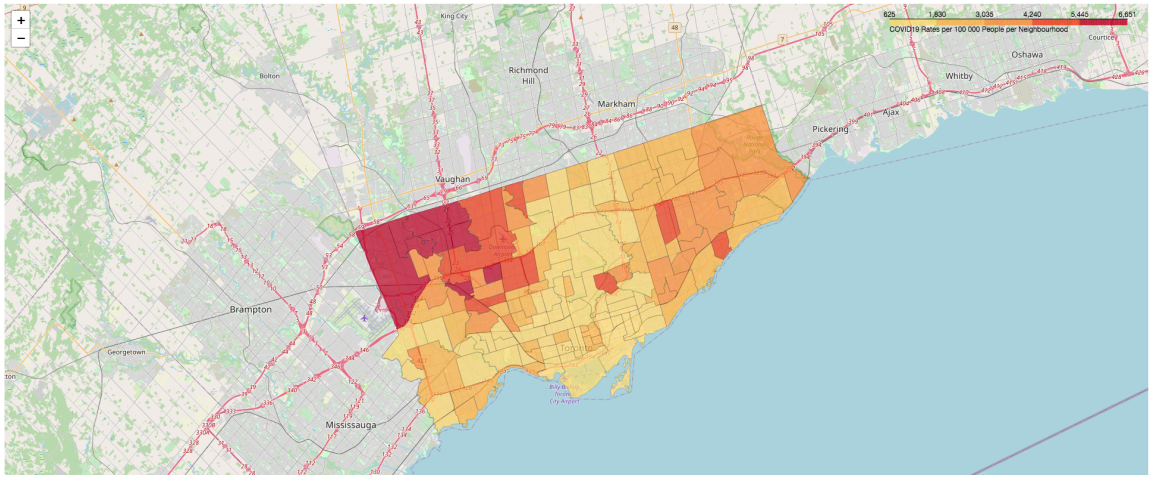

Figure 20. Choropleth map of COVID-19 infections per 100,000 residents of each Toronto neighbourhood.

The neighbourhoods with the highest and lowest rates of infection were tested with queries in Foursquare. Figure 21 shows the number of venues (of all categories) for the neighbourhoods with the top five and bottom five infections per 100,000 residents. It is apparent that the five neighbourhoods with the highest number of venues are the five neighbourhoods with the lowest rate of infection per 100,000 residents. The neighbourhoods with fewer venues have the highest infection rates.

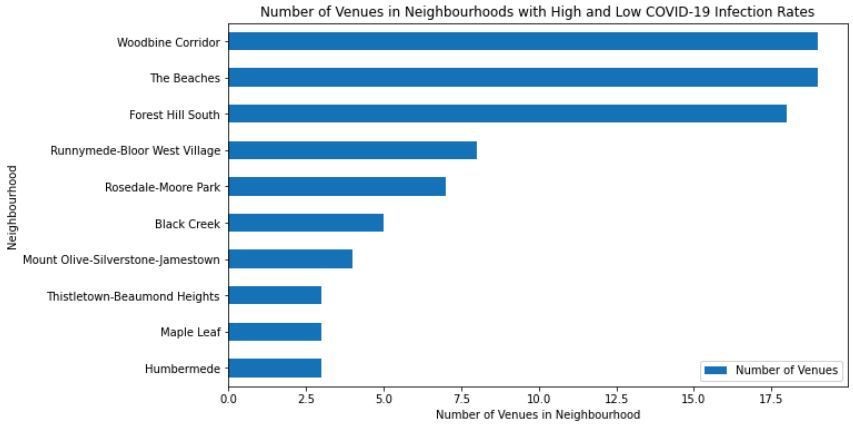

Figure 21. Number of venues (uncategorized) for the neighbourhoods with top five and bottom five COVID-19 infection rates.

Data returned from Foursquare were further examined and categorized by the top five venues per neighbourhood. This breakdown is shown in Figure 22. It is difficult to make strong classifications of the results. Four of the neighbourhoods with the lowest rates of COVID-19 infection have at least one outdoor venue in the top five venues. Runnymede-Bloor West Village is the exception. Similarly, ignoring intersection as a venue, four of the five neighbourhoods with the highest rates of COVID-19 infection have at least one outdoor venue in the top five venues. Of the indoor venues, most are an assortment of recreational facilities or food services. Since clear relationships between COVID-19 infection rates and venues returned from Foursquare could not be established, demographic data were analyzed to understand the geographic distribution of infection rates.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Black Creek	Hotel	Intersection	Fast Food Restaurant	Coffee Shop	Home Service
Forest Hill South	Bagel Shop	Restaurant	Pizza Place	Trail	Optical Shop
Humbermede	Golf Course	Coffee Shop	Department Store	Trail	Dance Studio
Maple Leaf	Tennis Court	Pharmacy	Park	Convenience Store	Golf Course
Mount Olive-Silverstone-Jamestown	Park	Baseball Field	Caribbean Restaurant	Dance Studio	Golf Course
Rosedale-Moore Park	Coffee Shop	Intersection	Breakfast Spot	Pizza Place	Fish & Chips Shop
Runnymede-Bloor West Village	Asian Restaurant	Sushi Restaurant	Food & Drink Shop	Pizza Place	Café
The Beaches	Fast Food Restaurant	Pizza Place	Park	Steakhouse	Sandwich Place
Thistletown-Beaumond Heights	Convenience Store	Construction & Landscaping	Home Service	Bank	Baseball Field
Woodbine Corridor	Fast Food Restaurant	Pizza Place	Park	Steakhouse	Sandwich Place

Figure 22. Most common venue types per neighbourhood, using the subset of neighbourhoods with the top five and bottom five rates of infection.

Figures 23-27 are a series of choropleth maps constructed from the Toronto demographic data. The chosen figures help to describe the demographic characteristics associated (or not associated) with high rates of COVID-19 infections. These maps should be directly compared with Figure 20, the map of infection rates by neighbourhood. The demographic data explain COVID-19 prevalence better than the venue data obtained from Foursquare.

Figure 23 shows the population of people who identify as visible minority Black for each Toronto neighbourhood. As was suggested in the statistical analysis, neighbourhoods with high populations of people who identify as visible minority Black generally have high rates of COVID-19 infection per 100,000 residents.

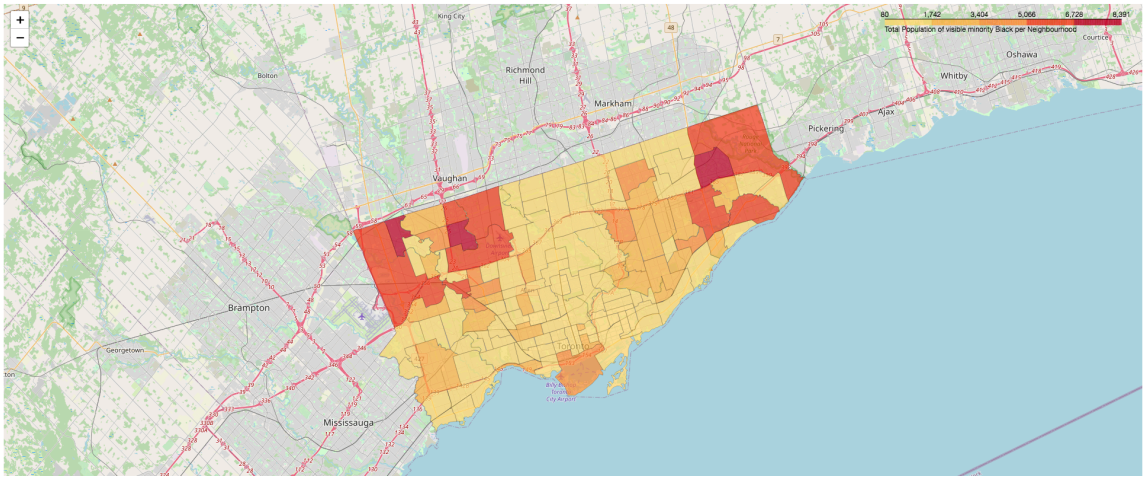

Figure 23. Choropleth map of population of people who identify as visible minority Black for each Toronto neighbourhood.

Figure 24 is a choropleth map for the population with total household income of $80,000 or greater for each Toronto neighbourhood. Neighbourhoods with high populations of wealthy people generally do not have high rates of COVID-19 infection. The number of people per neighbourhood who commute to work as a vehicle passenger (not public transit) is presented in Figure 25. This map resembles the COVID-19 infection rate map in Figure 20. Figure 26 shows the unemployment rate by neighbourhood. Many areas with high COVID-19 infection rates (Figure 20) are also regions of high unemployment.

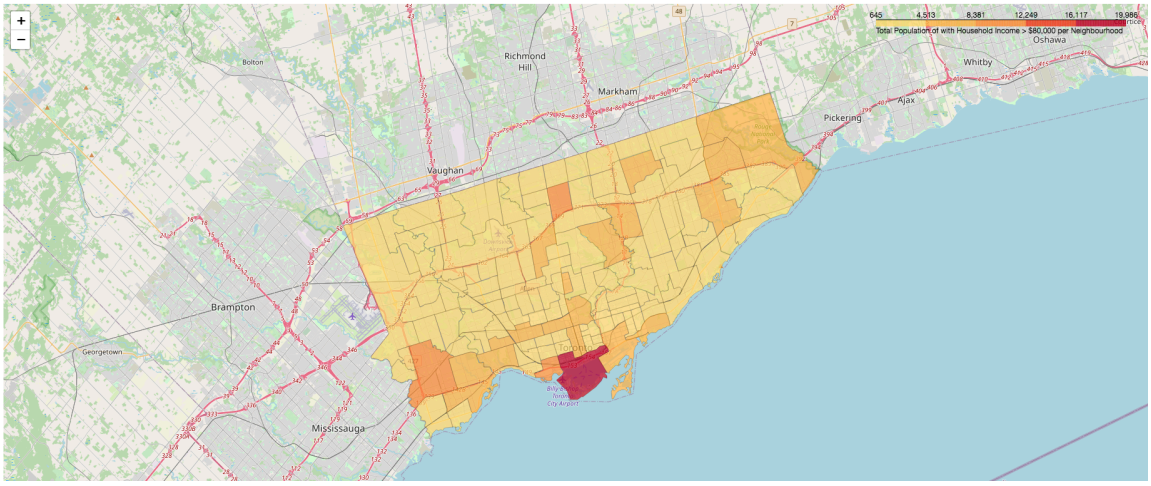

Figure 24. Choropleth map of population of people with total household income $80,000 and greater for each Toronto neighbourhood.

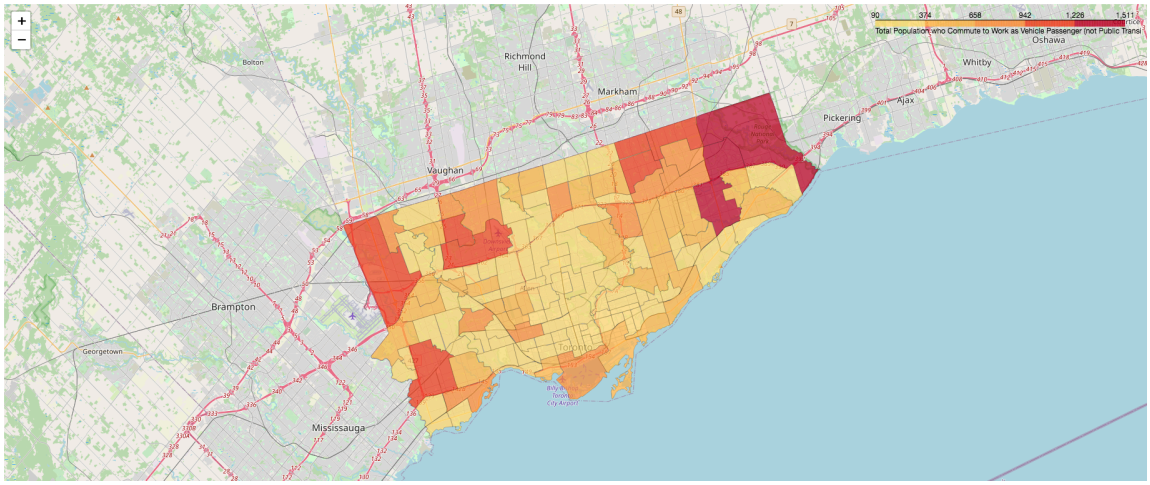

Figure 25. Choropleth map of population of people who commute to work in a private vehicle (not public transit) as a passenger for each Toronto neighbourhood.

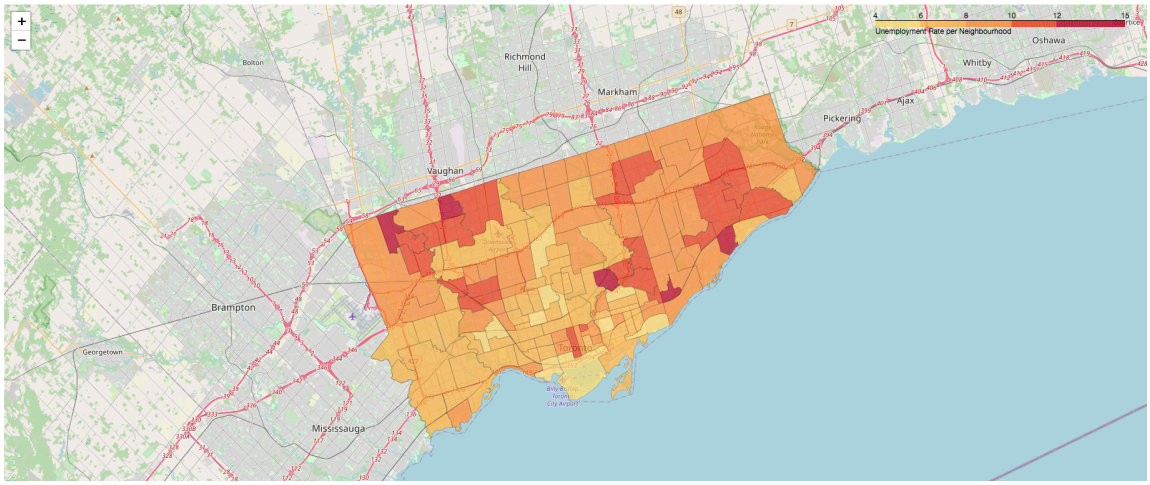

Figure 26. Choropleth map of unemployment rate for each Toronto neighbourhood.

Lastly, the number of people per neighbourhood with education level of Bachelor's Degree or higher is shown in Figure 27. The region of the most highly educated people corresponds with neighbourhoods of low rates of COVID-19 infection.

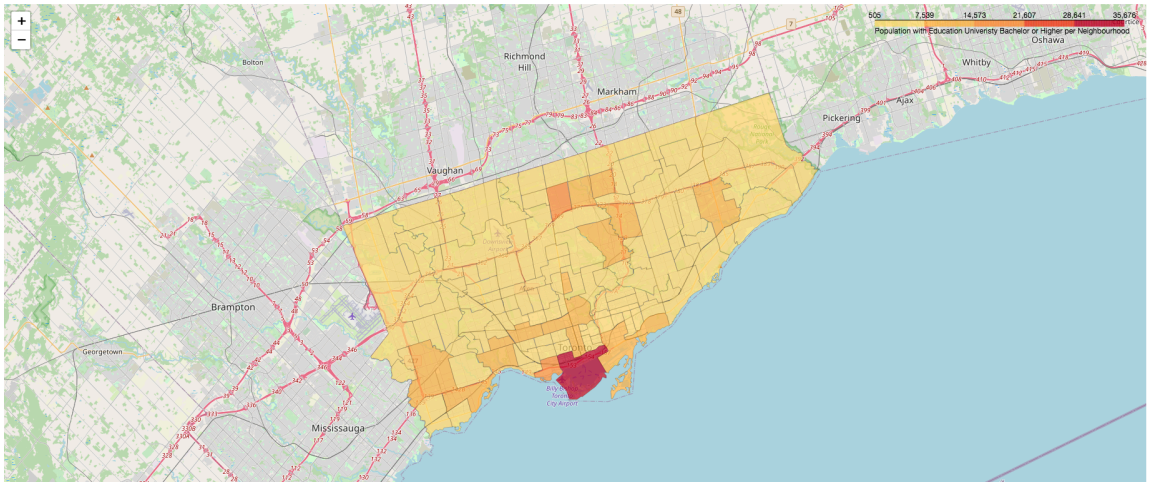

Figure 27. Choropleth map of number of people with university degrees at bachelor level or higher for each Toronto neighbourhood.

Discussion

A rich database was analyzed to understand the distribution of COVID-19 cases in Toronto. It was relatively easy to determine which neighbourhoods had recorded the highest number of COVID-19 infections. The top five neighbourhoods in terms of infection rates are Thistletown-Beaumond Heights, Mount Olive-Silverstone-Jamestown, Maple Leaf, Black Creek, and Humbermede. These have been presented in table and map form.

It was not immediately obvious why infection rates were highest in these regions. Foursquare queries of the neighbourhoods with highest and lowest rates of infections showed that neighbourhoods with the most public venues had the lowest rates of infections. It could perhaps be concluded that access to public venues is helpful in reducing COVID-19 infection rates. Perhaps access to public venues reduces gatherings in private homes for example. But the data at hand could not be used to verify this, and quarantine factors such as restricted access hours or outright closures of these venues were not clear. Further breakdowns into the nature of venues, be they indoor or outdoor in nature, did not add further clarification.

Thus demographic data were analyzed, and proved helpful in understanding the characteristics of individuals most or least prone to COVID-19 infections. Factors that strongly positively correlate with COVID-19 infections were common in neighbourhoods where infection rates are high. These factors include, but are not limited to, identifying as a visible minority black, commuting to work as a passenger in a private vehicle, and being unemployed. Conversely, factors that have strong negative correlations with infection rates were common in neighbourhoods where infection rates are low. These factors include high household income, and high education.

Conclusions

The goal of this study was to help inform policy makers responsible for public heath measures. Regions with high infection rates of COVID-19 should be prioritized for measures aimed at reducing infections and for prioritized vaccinations. Based on this study, it is reasonable to conclude that neighbourhoods with the highest infection rates (Thistletown-Beaumond Heights, Mount Olive-Silverstone-Jamestown, Maple Leaf, Black Creek, and Humbermede) should be subjected to new health measures and prioritized for vaccinations.

Fortunately, the richness of the Toronto dataset allows for a more nuanced vaccination prioritization. This study suggests demographic groups that should be prioritized for vaccination and those who should wait. *Vaccination priority groups* include those who identify as visible minority Black or Latin American, occupants of households with low to moderate incomes, people aged 25 or lower, those who commute to work as a passenger or for a long duration, the unemployed, and those with low education or who work in trades. Conversely, groups that should be *low vaccination priority* include visible minority Japanese or those who are not a visible minority, occupants of households with high income, and individuals who are highly educated.

References

[1] Wikipedia, retrieved January 21, 2021. *COVID-19 Pandemic.* https://en.wikipedia.org/wiki/COVID-19_pandemic

[2] World Health Organization, retrieved January 21, 2021. *WHO Coronavirus Disease (COVID-19) Dashboard*. https://covid19.who.int/

[3] CBC, retrieved January 21, 2021. *Here's the COVID-10 vaccine rollout plan, province by province. https://www.cbc.ca/news/canada/covid19-vaccine-rollout-plans-canada-1.5836262*

[4] Government of Canada, retrieved January 21, 2021. *Vaccines and treatments for COVID-19: Vaccine rollout. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/prevention-risks/covid-19-vaccine-treatment/vaccine-rollout.html*

[5] Pelly, Lauren, retrieved January 21, 2021. *Health-care workers lining up for COVID-19 vaccine, but some warn of 'real troubles' with hesitancy. https://www.cbc.ca/news/canada/toronto/health-care-vaccines-covid-19-hesitancy-1.5852119*