Study of Variables Influencing Whether a COVID-19 patient in Toronto is Hospitalized

Zeyao Li

December 21, 2020

Abstract

This report studies the factors that affect whether a COVID-19 patient in Toronto should be hospitalized. Logistic regression model is fitted by using independent variables of age, neighbourhood, source of infection, and gender. The results show it is more likely that older patients and male patients are admitted into hospitals to treat COVID-19.

Keywords

COVID-19, Toronto, Logistic Regression, Hospitalization, Age, Neighbourhood, Source of Infection, Gender

Introduction

In 2020, COVID-19 is the biggest challenge. It has a large impact around the world by disturbing daily life negatively. After the outbreak of COVID-19 at the start of 2020, it was not long before City of Toronto has been affected. As confirmed by the National Microbiology Lab in Winnipeg in January 27, 2020 , "a man in quarantine in Sunnybrook Hospital is Canada's first documented case of the new coronavirus"(CBC News), which was also the first positive case in Toronto. Over the year, the number of confirmed cases climbs up significantly. As of December 18, 2020, there are total of 52,191 recorded cases in Toronto(D'andrea). During the outbreak of COVID-19, health care system of Toronto faces great pressures to combat the virus. "With GTA hospitals beginning to overflow, health officials and medical professionals are sounding the alarm"(Staff). With limiting spaces for hospitalizations, people with light symptoms should stay at home for self-isolation and leave spaces of hospitalizations for people with serious symptoms. Thus, it is crucial to identify the factors that contributing to whether a patient is hospitalized.

One common way to analyze relatinships by variables is by using logistic regression. Logistic regression performs a predictive analysis when the dependent variable is binary. Also, similar to other regression analysis, logistic regression could be used to explain the relationship between independent and dependent variables(What is Logistic Regression). The main goal of this report is to predict the whether a patient should be admitted into hospital, which is a binary outcome. Thus, logistic regression is an appropriate choice to use.

In this report, the data set is obtained from Toronto Open Data Portal. The data would be explained more specifically in the Methodology section(Section 2) as well as the model used to predict probability of hospitalization of patients in Toronto. In the Results section(Section 3), summary of data and model's output would be provided. Lastly, interpretations of results and discussions of weaknesses and next steps from this study would be included in the Discussion section(Section 4).

Methodology

Data

As mentioned earlier in Section 1, the data set about information of COVID-19 patients from Toronto is used to conduct study in this report. All cases that are tested positive before December 10, 2020 are included in this data set. In other words, the data set used in this report is a census of all positive cases in Toronto before December 10, 2020. The data is obtained from Toronto Open Data Portal, which is a website with "digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime and anywhere"(City of Toronto). The contributor for COVID-19 data is Toronto Public Health. Since Toronto Public Health is an authoritative organization, the data should be accurate and reliable to perform analysis.

Some data cleanings were performed before using it in this report. In the cleaning process, the outcome variable, "is the patient ever hospitalized", is assigned a factor of 0 and 1 for each case, where 0 means "no" and 1 means "yes". In this case, "ever hospitalized" includes COVID-19 patients who were admitted into hospitals before and those who are currently hospitalized. Meanwhile, 4 predictors are chosen for this report to predict the probability of admitting into hospital. Cases with missing value for at least one predictor are removed to prevent errors. All four predictors are categorical, and here are lists of categories for each predictor.

1. age of patients: 19 and younger, 20 to 29 Years, 30 to 39 Years, 40 to 49 Years, 50 to 59 Years, 60 to 69 Years, 70 to 79 Years, 80 to 89 Years, 90 and older
2. neighbourhood: Agincourt North, Agincourt South-Malvern West, Alderwood, Annex, Banbury-Don Mills (These are the neighbourhoods where people are test positived. There are 140 neighbourhoods in total, and a complete list would be provided in Appendix)
3. source of infection: Close contact, Community, Healthcare, Institutional, N/A - Outbreak associated, Pending, Travel
4. gender: female, male, other, transgender

Model

In this section, the chosen model of logistic regression would be introduced. Since all independent variables are categorical, "dummy" variable coding would be used for each independent variable in logistic regression. The set up of the model is the following:

$$log(\frac{\hat{p}}{1-\hat{p}}) = \beta_0 + \beta_1 x_{age} + \beta_2 x_{neighbourhood} + \beta_3 x_{infection\ source} + \beta_4 x_{gender}$$

$log(\frac{\hat{p}}{1-\hat{p}})$ represents the log odds of admitting into hospitals.

$\hat{p}$ is the probability of admitting into hospital.

β_{0j} represents the intercept parameter, which is the log odds of admitting into hospitals when age is in the range of 19 and younger; neighbourhood is Agincourt North; source of infection is close contact; gender is female.

For β_1, β_2, β_3, and β_4, those are the slope parameters. For each β, it is the change in log odds of admitting into hospitals comparing to reference level when a single category in the variables is 1 in dummy variable coding.

Originally, two models were built for this study. Besides the model introduced above, a second model had explanatory variables of age, source of infection, and gender. However, AIC is 14455.2 for the second model

and 14351.57 for the model introduced above. Thus, since better model usually has lower AIC, the final model in this report consists of age, neighbourhood, source of infection, and gender as explanatory variables.

In the next section, the logistic regression model would be fitted on R and results would be presented.

Results

Table 1: Table 1: number of patients hospitalized in different age groups

age group	number
19 and younger	32
20 to 29 Years	66
30 to 39 Years	115
40 to 49 Years	187
50 to 59 Years	341
60 to 69 Years	460
70 to 79 Years	532
80 to 89 Years	614
90 and older	311

Table 1 shows the number of patients who are admitted into hospitals for different age groups. There is a trend that for people who are older, number of hospitalized patients increases. However, the number of patients drop from 80 to 89 years to 90 and older.

Table 2: Table 2: number of patients hospitalized with different genders

gender	number
FEMALE	1271
MALE	1387

Table 2 summarizes hospitalized patients by genders. As shown from the table, there is no other or transgender patients admitted into hospitals. There are more male patients than female patients, but these two numbers do not differ by much.

Table 3: Table 3: summary of logistic regression model

term	estimate	std.error	statistic	p.value
(Intercept)	-4.9065064	0.2996828	-16.3723349	0.0000000
age_group20 to 29 Years	0.3196724	0.2177575	1.4680199	0.1420988
age_group30 to 39 Years	1.0148334	0.2028747	5.0022673	0.0000006
age_group40 to 49 Years	1.7201832	0.1940465	8.8647978	0.0000000
age_group50 to 59 Years	2.3144520	0.1875610	12.3397322	0.0000000
age_group60 to 69 Years	3.1585019	0.1862351	16.9597582	0.0000000
age_group70 to 79 Years	4.2025549	0.1881995	22.3303156	0.0000000
age_group80 to 89 Years	4.2983576	0.1896609	22.6633810	0.0000000
age_group90 and older	4.0209282	0.1962790	20.4857805	0.0000000
neighbourhoodAgincourt South-Malvern West	0.0487135	0.3905041	0.1247451	0.9007253
neighbourhoodAlderwood	-0.1896730	0.5004695	-0.3789900	0.7046953

term	estimate	std.error	statistic	p.value
neighbourhoodAnnex	-0.3238166	0.3223204	-1.0046420	0.3150693
neighbourhoodBanbury-Don Mills	-0.5023265	0.3690104	-1.3612802	0.1734252
neighbourhoodBathurst Manor	-0.3132553	0.3092127	-1.0130741	0.3110248
neighbourhoodBay Street Corridor	-0.1065385	0.4609308	-0.2311378	0.8172078
neighbourhoodBayview Village	-0.0799474	0.4809618	-0.1662240	0.8679807
neighbourhoodBayview Woods-Steeles	-1.8819694	0.4417065	-4.2606783	0.0000204
neighbourhoodBedford Park-Nortown	-0.0399746	0.3440983	-0.1161721	0.9075162
neighbourhoodBeechborough-Greenbrook	-0.7034030	0.3699704	-1.9012413	0.0572704
neighbourhoodBendale	-0.6607077	0.2925053	-2.2587892	0.0238965
neighbourhoodBirchcliffe-Cliffside	-0.7060756	0.2888088	-2.4447854	0.0144938
neighbourhoodBlack Creek	-0.0824815	0.2913409	-0.2831098	0.7770926
neighbourhoodBlake-Jones	-1.1571720	1.0821952	-1.0692821	0.2849426
neighbourhoodBriar Hill - Belgravia	-0.1213854	0.3841866	-0.3159544	0.7520371
neighbourhoodBridle Path-Sunnybrook-York Mills	-1.6980733	0.7815712	-2.1726404	0.0298074
neighbourhoodBroadview North	-0.3629477	0.6601000	-0.5498375	0.5824308
neighbourhoodBrookhaven-Amesbury	-0.3534065	0.3461344	-1.0210095	0.3072499
neighbourhoodCabbagetown-South St. James Town	-0.8797899	0.3340532	-2.6336822	0.0084464
neighbourhoodCaledonia-Fairbank	-0.6676191	0.6077559	-1.0984988	0.2719868
neighbourhoodCasa Loma	-0.0425504	0.4512326	-0.0942983	0.9248722
neighbourhoodCentennial Scarborough	0.0261752	0.4825562	0.0542428	0.9567417
neighbourhoodChurch-Yonge Corridor	-0.7188779	0.4515150	-1.5921460	0.1113519
neighbourhoodClairlea-Birchmount	-0.9516684	0.3166216	-3.0056964	0.0026497
neighbourhoodClanton Park	-0.7573165	0.5106600	-1.4830150	0.1380704
neighbourhoodCliffcrest	0.1489536	0.3697610	0.4028376	0.6870677
neighbourhoodCorso Italia-Davenport	0.1933764	0.4427063	0.4368051	0.6622527
neighbourhoodDanforth	0.6715265	0.4961376	1.3535086	0.1758932
neighbourhoodDanforth-East York	-0.4663267	0.5290729	-0.8814034	0.3780995
neighbourhoodDon Valley Village	-0.0685878	0.3756117	-0.1826030	0.8551095
neighbourhoodDorset Park	-0.4713348	0.2816797	-1.6733007	0.0942681
neighbourhoodDovercourt-Wallace Emerson-Junction	0.0741154	0.3453790	0.2145914	0.8300859
neighbourhoodDownsview-Roding-CFB	-0.4359639	0.2959725	-1.4729878	0.1407543
neighbourhoodDufferin Grove	-0.2178902	0.4024106	-0.5414624	0.5881889
neighbourhoodEast End-Danforth	-1.1557691	0.3770820	-3.0650336	0.0021765
neighbourhoodEdenbridge-Humber Valley	-1.0886690	0.3494982	-3.1149490	0.0018398
neighbourhoodEglinton East	-0.2483980	0.3284594	-0.7562518	0.4494983
neighbourhoodElms-Old Rexdale	-1.1891180	0.5300258	-2.2435097	0.0248640
neighbourhoodEnglemount-Lawrence	-0.4044396	0.3335213	-1.2126350	0.2252694
neighbourhoodEringate-Centennial-West Deane	-0.7021965	0.4487379	-1.5648255	0.1176238
neighbourhoodEtobicoke West Mall	-1.1494467	0.3697428	-3.1087740	0.0018787
neighbourhoodFlemingdon Park	-0.3186072	0.3610172	-0.8825264	0.3774922
neighbourhoodForest Hill North	0.0121521	0.3981015	0.0305252	0.9756482
neighbourhoodForest Hill South	-0.6642556	0.6948373	-0.9559872	0.3390787
neighbourhoodGlenfield-Jane Heights	-0.3396147	0.2726903	-1.2454227	0.2129764
neighbourhoodGreenwood-Coxwell	-0.5670961	0.6585231	-0.8611635	0.3891480
neighbourhoodGuildwood	-1.3685677	0.3614564	-3.7862595	0.0001529
neighbourhoodHenry Farm	-0.1193742	0.4280099	-0.2789053	0.7803175
neighbourhoodHigh Park North	-0.1288158	0.4222864	-0.3050436	0.7603330
neighbourhoodHigh Park-Swansea	0.5859984	0.3661826	1.6002903	0.1095342
neighbourhoodHighland Creek	-1.4285183	0.5867005	-2.4348340	0.0148986

term	estimate	std.error	statistic	p.value
neighbourhoodHillcrest Village	-0.9919166	0.4143026	-2.3941838	0.0166574
neighbourhoodHumber Heights-Westmount	-1.3812693	0.3446478	-4.0077708	0.0000613
neighbourhoodHumber Summit	-0.9436211	0.3933596	-2.3988763	0.0164455
neighbourhoodHumbermede	-0.8959479	0.3749441	-2.3895509	0.0168690
neighbourhoodHumewood-Cedarvale	0.2311116	0.5101801	0.4530001	0.6505487
neighbourhoodIonview	-0.1490114	0.4880518	-0.3053188	0.7601233
neighbourhoodIslington-City Centre West	-0.7151233	0.2807970	-2.5467623	0.0108727
neighbourhoodJunction Area	-0.2697231	0.4513749	-0.5975589	0.5501343
neighbourhoodKeelesdale-Eglinton West	-0.5643274	0.5112718	-1.1037718	0.2696921
neighbourhoodKennedy Park	-0.4714129	0.3721149	-1.2668478	0.2052097
neighbourhoodKensington-Chinatown	-0.7127151	0.3643367	-1.9561989	0.0504417
neighbourhoodKingsview Village-The Westway	-0.1902297	0.3083244	-0.6169789	0.5372486
neighbourhoodKingsway South	0.0465584	0.4326930	0.1076014	0.9143119
neighbourhoodL'Amoreaux	-0.1195882	0.3179313	-0.3761448	0.7068093
neighbourhoodLambton Baby Point	0.1352678	0.5899125	0.2293014	0.8186347
neighbourhoodLansing-Westgate	-0.7696953	0.5219594	-1.4746269	0.1403129
neighbourhoodLawrence Park North	0.0366581	0.5904266	0.0620875	0.9504931
neighbourhoodLawrence Park South	-0.7945326	0.6630490	-1.1983015	0.2307997
neighbourhoodLeaside-Bennington	-0.6263141	0.4163761	-1.5042027	0.1325292
neighbourhoodLittle Portugal	-0.5014531	0.3625806	-1.3830113	0.1666614
neighbourhoodLong Branch	0.5530908	0.5334274	1.0368624	0.2998000
neighbourhoodMalvern	-0.5054539	0.3000370	-1.6846386	0.0920583
neighbourhoodMaple Leaf	-0.4663552	0.3159970	-1.4758214	0.1399919
neighbourhoodMarkland Wood	-1.0659619	0.5581019	-1.9099773	0.0561361
neighbourhoodMilliken	-0.0536572	0.3944174	-0.1360418	0.8917883
neighbourhoodMimico (includes Humber Bay Shores)	-0.1846529	0.3244567	-0.5691140	0.5692788
neighbourhoodMorningside	-1.3671511	0.3536575	-3.8657493	0.0001107
neighbourhoodMoss Park	0.4901344	0.2986840	1.6409799	0.1008016
neighbourhoodMount Dennis	-0.1985171	0.3023140	-0.6566586	0.5114005
neighbourhoodMount Olive-Silverstone-Jamestown	-0.6501962	0.2892087	-2.2481903	0.0245641
neighbourhoodMount Pleasant East	-0.8399671	0.6752633	-1.2439105	0.2135325
neighbourhoodMount Pleasant West	-1.0735370	0.3427319	-3.1322934	0.0017345
neighbourhoodNew Toronto	-0.7548067	0.3970468	-1.9010523	0.0572952
neighbourhoodNewtonbrook East	-0.3818509	0.5292567	-0.7214852	0.4706110
neighbourhoodNewtonbrook West	0.3482285	0.3054719	1.1399690	0.2542992
neighbourhoodNiagara	-0.1545783	0.3839002	-0.4026522	0.6872041
neighbourhoodNorth Riverdale	-0.4548911	0.6168227	-0.7374747	0.4608337
neighbourhoodNorth St. James Town	-0.2894590	0.3099908	-0.9337664	0.3504244
neighbourhoodO'Connor-Parkview	-1.0671328	0.4994243	-2.1367257	0.0326203
neighbourhoodOakridge	0.0286139	0.3860579	0.0741182	0.9409163
neighbourhoodOakwood Village	-0.2783228	0.3497457	-0.7957862	0.4261563
neighbourhoodOld East York	-0.3080868	0.5344583	-0.5764468	0.5643132
neighbourhoodPalmerston-Little Italy	0.1500529	0.4952492	0.3029846	0.7619016
neighbourhoodParkwoods-Donalda	-0.1400961	0.3543307	-0.3953824	0.6925607
neighbourhoodPelmo Park-Humberlea	-0.5188128	0.4148166	-1.2507041	0.2110425
neighbourhoodPlayter Estates-Danforth	-1.9318505	0.5724651	-3.3746172	0.0007392
neighbourhoodPleasant View	-1.0401509	0.5583867	-1.8627786	0.0624934
neighbourhoodPrincess-Rosethorn	-0.2395788	0.4690565	-0.5107675	0.6095139
neighbourhoodRegent Park	-0.3333795	0.5440386	-0.6127865	0.5400175
neighbourhoodRexdale-Kipling	-1.2820161	0.4024132	-3.1858203	0.0014434

term	estimate	std.error	statistic	p.value
neighbourhoodRockcliffe-Smythe	-0.4558106	0.3571776	-1.2761457	0.2019040
neighbourhoodRoncesvalles	0.4003400	0.4801294	0.8338168	0.4043842
neighbourhoodRosedale-Moore Park	-0.3355018	0.4605187	-0.7285301	0.4662892
neighbourhoodRouge	-0.9319433	0.2805358	-3.3220119	0.0008937
neighbourhoodRunnymede-Bloor West Village	0.5345229	0.5605235	0.9536136	0.3402793
neighbourhoodRustic	-0.2699795	0.3339828	-0.8083637	0.4188813
neighbourhoodScarborough Village	-0.0124494	0.3231541	-0.0385245	0.9692695
neighbourhoodSouth Parkdale	-0.1784434	0.2815550	-0.6337782	0.5262256
neighbourhoodSouth Riverdale	-0.1347280	0.4867579	-0.2767865	0.7819441
neighbourhoodSt.Andrew-Windfields	-0.9994711	0.6292821	-1.5882719	0.1122249
neighbourhoodSteeles	0.2464446	0.4180381	0.5895266	0.5555081
neighbourhoodStonegate-Queensway	-1.2378152	0.5835790	-2.1210757	0.0339154
neighbourhoodTam O'Shanter-Sullivan	-0.7125210	0.3454316	-2.0626977	0.0391414
neighbourhoodTaylor-Massey	-0.0270741	0.3844918	-0.0704154	0.9438630
neighbourhoodThe Beaches	-0.7747915	0.6854710	-1.1303053	0.2583476
neighbourhoodThistletown-Beaumond Heights	-1.1240459	0.3375895	-3.3296236	0.0008696
neighbourhoodThorncliffe Park	-0.2528028	0.3218712	-0.7854159	0.4322098
neighbourhoodTrinity-Bellwoods	0.2617519	0.4267908	0.6133026	0.5396763
neighbourhoodUniversity	-0.3857026	0.4971521	-0.7758240	0.4378529
neighbourhoodVictoria Village	-1.4261469	0.4148853	-3.4374485	0.0005872
neighbourhoodWaterfront Communities-The Island	-0.9849839	0.3893336	-2.5299228	0.0114088
neighbourhoodWest Hill	-0.4871555	0.3696502	-1.3178823	0.1875430
neighbourhoodWest Humber-Clairville	-0.4916640	0.2760021	-1.7813778	0.0748508
neighbourhoodWestminster-Branson	-0.8827945	0.3450908	-2.5581513	0.0105230
neighbourhoodWeston	0.1987458	0.2757720	0.7206887	0.4711011
neighbourhoodWeston-Pellam Park	-0.0836564	0.4325352	-0.1934095	0.8466383
neighbourhoodWexford/Maryvale	-0.3193179	0.3753354	-0.8507536	0.3949062
neighbourhoodWillowdale East	-0.9416358	0.4744670	-1.9846181	0.0471870
neighbourhoodWillowdale West	-0.0140270	0.4045089	-0.0346765	0.9723377
neighbourhoodWillowridge-Martingrove-Richview	-0.1007952	0.3361821	-0.2998233	0.7643119
neighbourhoodWoburn	-0.4402107	0.2816861	-1.5627705	0.1181066
neighbourhoodWoodbine Corridor	-0.4716737	0.7866988	-0.5995607	0.5487990
neighbourhoodWoodbine-Lumsden	0.4078251	0.5689745	0.7167722	0.4735147
neighbourhoodWychwood	-0.3220899	0.3888172	-0.8283839	0.4074531
neighbourhoodYonge-Eglinton	-0.4780266	0.6999031	-0.6829898	0.4946133
neighbourhoodYonge-St.Clair	-0.7085112	0.5181971	-1.3672620	0.1715432
neighbourhoodYork University Heights	-0.0623421	0.2712080	-0.2298682	0.8181942
neighbourhoodYorkdale-Glen Park	-0.6267610	0.2928337	-2.1403312	0.0323280
source_of_infectionCommunity	0.3784125	0.0724531	5.2228633	0.0000002
source_of_infectionHealthcare	1.4246055	0.0854524	16.6713288	0.0000000
source_of_infectionInstitutional	0.2760098	0.1550007	1.7807002	0.0749614
source_of_infectionN/A - Outbreak associated	-0.0806850	0.0690311	-1.1688213	0.2424757
source_of_infectionPending	0.7371385	0.6259683	1.1775972	0.2389572
source_of_infectionTravel	0.6438293	0.1236817	5.2055347	0.0000002
genderMALE	0.4643565	0.0458963	10.1175249	0.0000000
genderOTHER	-10.8028968	184.3861364	-0.0585884	0.9532799
genderTRANSGENDER	-9.9272375	236.9730225	-0.0418918	0.9665849

Table 3 is a summary of outputs from fitting data into logistic regression model. The estimate of intercept is around -4.9065. As stated in the model section, the estimate of intercept is the log odds of admitting into

hospitals when age is in the range of 19 and younger; neighbourhood is Agincourt North; source of infection is close contact; gender is female.

In order to interpret the estimates of β, reference category should be included. For the variable age group, the reference category is 19 and younger. Thus, for instance, the estimate of β for patients who fall in the age group of 20 to 29 years is around 0.3197. It means that keeping other variables constant, in comparison to patients from age group of 19 and younger, the log odds of admitting into hospitals for patients in age group of 20 to 29 years are higher by 0.3197 on average. For other variables in age group, the interpretations of estimates of β are similar based on the reference group of 19 and younger. Likewise, since all variables are categorical, similar types of interpretations could be applied for estimates of β of other independent variables. The only difference here is the reference category of each independent variable. Agincourt North is the reference category in the variable neighbourhood; Close contact is the reference category in the variable source of infection; Female is the reference category in the variable gender.

Discussion

Summary

In previous sections, after introducing the background and main objective of this report, data and model are thoroughly discussed. In the Result section, two tables are used to show number of hospitalized patients with respect to age and gender. A third table is created to present the summaries of outputs from logistic regression model. For the rest sections, findings from results would be concluded as well as weaknesses and potential improvements of the report.

Conclusions

By using a benchmark of 0.05, it is possible to examine the significance of categories of each variable for predicting probability of hospitalization. As shown from Table 3, for the age group, the estimates of β for all categories except age group of 20 to 29 years have a p-value less than 0.05, which suggests these categories are useful predictors. It is worth-noticing that the estimates of β of age group increase as age increases, but it drops from age group of 80 to 89 years towards 90 and older. This finding is consistent with results from Table 1. Thus, for people who tested positive, there is an increasing chance of hospitalization if they are older.

For variables neighbourhood and source of infection, there are only a few p-values for the estimate of β that are less than the significance level of 0.05. This indicates that there might be multicollinearity within each category such that they are highly correlated, so it is not useful to formally conclude from these neighbourhood and source of infection.

The last independent variable is gender. Table 3 shows that p-value for estimate of β for male is less than 0.05, but p-values are close to 1 for transgender and other. This finding is consistent with Table 2 since there are no other or transgender patients admitted to hospitals, so only categories of male and female are useful predictors. In general, since the estimate of β for male patients is positive, the result suggests that male patients have a higher probability of hospitalization.

Weakness & Next Steps

This report uses Toronto COVID-19 data set from Toronto Public Health, which is trustworthy, but there are still weaknesses and limitations. On December 10, 2020, this data set was downloaded from Toronto Open Data Portal. Nevertheless, there is an update to the data weekly, so this report is only valid for analyzing COVID-19 cases before December 10, 2020. The data set used in this report is incomplete since the COVID-19 crisis is still lasting, so there could be potential errors, such as positive cases are instead tested negative for COVID-19. Another drawback is that there are too many categories for the variable

neighbourhood, and most of them are not significant. Thus, it is hard to draw any useful conclusion from the variable neighbourhood.

Some steps could be taken to improve this report in future. Firstly, it is important to update the logistic regression model in this report with newly updated data. With the most recent data, the model could produce more accurate estimates of β and better results. Another possible improvement is combining categories from the variable neighbourhood. For instance, the neighbourhoods "Bayview Village" and "Bayview Woods-Steeles" are very close, so they could be combined to form a single neighbourhood "Bayview". When there are less neighbourhoods but larger size for each of them, a more significant pattern of hospitalized patients might be found. Lastly, the key factor determining whether a patient should be hospitalized is the health of patient. Thus, this report could be paired with another logistic regression model that uses independent variables in medicals' perspectives.

Appendix

A) Neighbourhoods: "Agincourt North", "Agincourt South-Malvern West", "Alderwood", "Annex", "Banbury-Don Mills", "Bathurst Manor", "Bay Street Corridor", "Bayview Village", "Bayview Woods-Steeles", "Bedford Park-Nortown", "Beechborough-Greenbrook", "Bendale", "Birchcliffe-Cliffside", "Black Creek", "Blake-Jones", "Briar Hill - Belgravia", "Bridle Path-Sunnybrook-York Mills", "Broadview North", "Brookhaven-Amesbury", "Cabbagetown-South St. James Town", "Caledonia-Fairbank", "Casa Loma""Centennial Scarborough", "Church-Yonge Corridor", "Cliffcrest", "Corso Italia-Davenport", "Danforth", "Danforth-East York", "Don Valley Village", "Dorset Park", "Dovercourt-Wallace Emerson-Junction", "Downsview-Roding-CFB", "Dufferin Grove", "East End-Danforth", "Edenbridge-Humber Valley", "Eglinton East", "Elms-Old Rexdale", "Englemount-Lawrence", "Eringate-Centennial-West Deane", "Etobicoke West Mall" "Flemingdon Park", "Forest Hill North", "Forest Hill South", "Glenfield-Jane Heights", "Greenwood-Coxwell", "Guildwood" "Henry Farm", "High Park North", "High Park-Swansea", "Highland Creek", "Hillcrest Village", "Humber Heights-Westmount", "Humber Summit", "Humbermede" "Humewood-Cedarvale", "Ionview" "Islington-City Centre West", "Junction Area", "Keelesdale-Eglinton West", "Kennedy Park" "Kensington-Chinatown", "Kingsview Village-The Westway" "Kingsway South", "L'Amoreaux", "Lambton Baby Point", "Lansing-Westgate" "Lawrence Park North", "Lawrence Park South", "Leaside-Bennington", "Little Portugal", "Long Branch", "Malvern", "Maple Leaf", "Markland Wood", "Milliken", "Mimico (includes Humber Bay Shores)", "Morningside", "Moss Park", "Mount Dennis", "Mount Olive-Silverstone-Jamestown" "Mount Pleasant East", "Mount Pleasant West", "New Toronto", "Newtonbrook East", "Newtonbrook West", "Niagara", "North Riverdale", "North St. James Town", "O'Connor-Parkview", "Oakridge", "Oakwood Village", "Old East York", "Palmerston-Little Italy", "Parkwoods-Donalda", "Pelmo Park-Humberlea", "Playter Estates-Danforth" "Pleasant View", "Princess-Rosethorn", "Regent Park", "Rexdale-Kipling" "Rockcliffe-Smythe", "Roncesvalles", "Rosedale-Moore Park", "Rouge" "Runnymede-Bloor West Village", "Rustic", "Scarborough Village", "South Parkdale", "South Riverdale", "St.Andrew-Windfields", "Steeles", "Stonegate-Queensway", "Tam O'Shanter-Sullivan", "Taylor-Massey", "The Beaches", "Thistletown-Beaumond Heights", "Thorncliffe Park", "Trinity-Bellwoods", "University", "Victoria Village", "Waterfront Communities-The Island", "West Hill", "West Humber-Clairville" "Westminster-Branson", "Weston", "Weston-Pellam Park", "Wexford/Maryvale", "Willowdale East", "Willowdale West", "Willowridge-Martingrove-Richview", "Woburn", "Woodbine Corridor", "Woodbine-Lumsden" "Wychwood", "Yonge-Eglinton", "Yonge-St.Clair", "York University Heights", "Yorkdale-Glen Park"

B) Code and data supporting this analysis is available at: https://github.com/zeyao-li/STA304_Final_Project_ZL

References

1. Wu, Changbao, and Mary E. Thompson. "Basic Concepts in Survey Sampling." Sampling Theory and Practice. Springer, Cham, 2020. 3-15.
2. Timeline of COVID-19 cases across Canada | CBC News. (2020, March 14). Retrieved from https://www.cbc.ca/news/health/canada-coronavirus-timeline-1.5482310
3. D'andrea, A., & Lavoie, J. (2020, December 21). How many cases of coronavirus are there in Toronto? Retrieved from https://www.toronto.com/news-story/9873970-how-many-cases-of-coronavirus-are-there-in-toronto-/
4. Staff, N. (2020, December 02). More COVID-19 patients in the ICU, GTA hospitals feeling the pressure. Retrieved from https://toronto.citynews.ca/2020/12/02/more-covid-19-patients-in-the-icu-gta-hospitals-feeling-the-pressure/
5. City of Toronto. (2018, January 11). What is Open Data? Retrieved from https://www.toronto.ca/city-government/data-research-maps/open-data/what-is-open-data/
6. Tidyverse. (n.d.). Retrieved from https://www.tidyverse.org/
7. Xie, Y. (2020, September 22). A General-Purpose Package for Dynamic Report Generation in R [R package knitr version 1.30]. Retrieved from https://cran.r-project.org/web/packages/knitr/index.html
8. Open source & professional software for data science teams. (2020, December 02). Retrieved from https://rstudio.com/
9. Open Data Dataset. (n.d.). Retrieved from https://open.toronto.ca/dataset/covid-19-cases-in-toronto/
10. What is Logistic Regression? (2020, March 09). Retrieved from https://www.statisticssolutions.com/what-is-logistic-regression/