The battle of the neighborhoods: where to place a new pharmacy in the Covid era?

Introduction

The client for this project wants to explore the best place for a pharmacy in the Toronto area. The latest spread of COVID 19 rose sales of hand sanitizers, disposable gloves and masks worldwide. It is likely that the sales volume of those items after the pandemic is over will stay at a high level, due to population awareness of virus spread mechanisms, increase fear, and new habits. While most of those items might be found in convenience stores or other retail shops, pharmacies are the most reliable places to get those items. Pharmacies offer prescription and over-the-counter medicines, required by people who got infected with COVID 19 but might not require hospitalization. Therefore, pharmacies have a clear role in the prevention of the disease and the recovery of infected individuals.

The main assumption of this work is that the larger the number of Covid Cases in a given neighborhood, the higher the need for disease prevention and palliation items. In other words, the hypothesis is that the market demand of a given neighborhood increases with the number of Covid Cases in the neighborhood. The number of existing pharmacies in each neighborhood will be used to estimate market supply. A neighborhood with a high number of COVID infections and a low number of pharmacies might be in need of new pharmacies. The number of Covid cases, existing pharmacies and distance to downtown will be analyzed for each neighborhood in Toronto to select the best neighborhoods for opening a new pharmacy.

Data

Zip Codes and a list of Toronto neighborhoods were sourced from Wikipedia. Coordinates of Toronto neighborhoods were loaded from a csv file that has the geographical coordinates of each postal code, available at: http://cocl.us/Geospatial_data. A database for covid cases in Toronto was retrieved from the web, this database is updated weekly on Wednesdays, https://open.toronto.ca/dataset/covid-19-cases-in-toronto/. Foursquare was used to find out the number of existing pharmacies available in a given neighborhood. The distance of each neighborhood to downtown was calculated with a Geopy function, described in https://geopy.readthedocs.io/en/stable/#module-geopy.distance.

From the database with Covid cases in the Toronto area, only confirmed cases were selected. Confirmed cases were grouped by Zip Code, and added to the dataframe with neighborhoods and neighborhoods coordinates. Neighborhoods with with M7Y and M7R postal codes were dropped, because they are mail processing centers, and not actual neighborhoods. Foursquare was used to find pharmacies within a radius of 1000 m. A python program add the number of existing pharmacies in a given neighborhood to each neighborhood in the dataframe. This python program used Foursquare. The distance of each neighborhood to downtown was then calculated and added to the dataframe. Downtown Toronto coordinates, required to calculate the distance to downtown, correspond to the coordinates of Toronto City Hall. In the final

dataframe, each row stored the zip code, Borough, neighborhood name and coordinates, distance to downtown, number of covid cases and number of existing pharmacies, Table 1. All coding was performed on a Jupiter Notebook, through www.kaggle.com website. Using pandas some basic statistical information of the relevant features was obtained, table 2.

Table 1. First 5 lines of the dataframe constructed to solve the problem.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	distance	Covid_Cases	Pharmacies
0	M3A	North York	Parkwoods	43.753259	-79.329656	11.917950	939.0	1.0
1	M4A	North York	Victoria Village	43.725882	-79.315572	9.751236	615.0	0.0
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.881525	1325.0	10.0
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	9.730671	1352.0	0.0
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494	1.077659	0.0	22.0

Table 2. Statistical information of features of the dataframe in table 1, obtained with pandas.

	distance	Covid_Cases	Pharmacies
count	101.000000	101.000000	101.000000
mean	9.510620	979.138614	5.267327
std	6.040528	825.515088	6.248025
min	0.327377	0.000000	0.000000
25%	4.020698	403.000000	1.000000
50%	9.440207	850.000000	3.000000
75%	14.071196	1454.000000	7.000000
max	24.860915	4702.000000	27.000000

Methodology section

Exploratory Data Analysis

To aid visualization of the results, a GeoJson file with neighborhoods of Toronto was used in this project. The link for this file is here. In Figure 1 the color of the neighborhood indicates the number of pharmacies in the given neighborhood, and the red circle has a radius that is proportional to the number of covid cases in the given neighborhood.

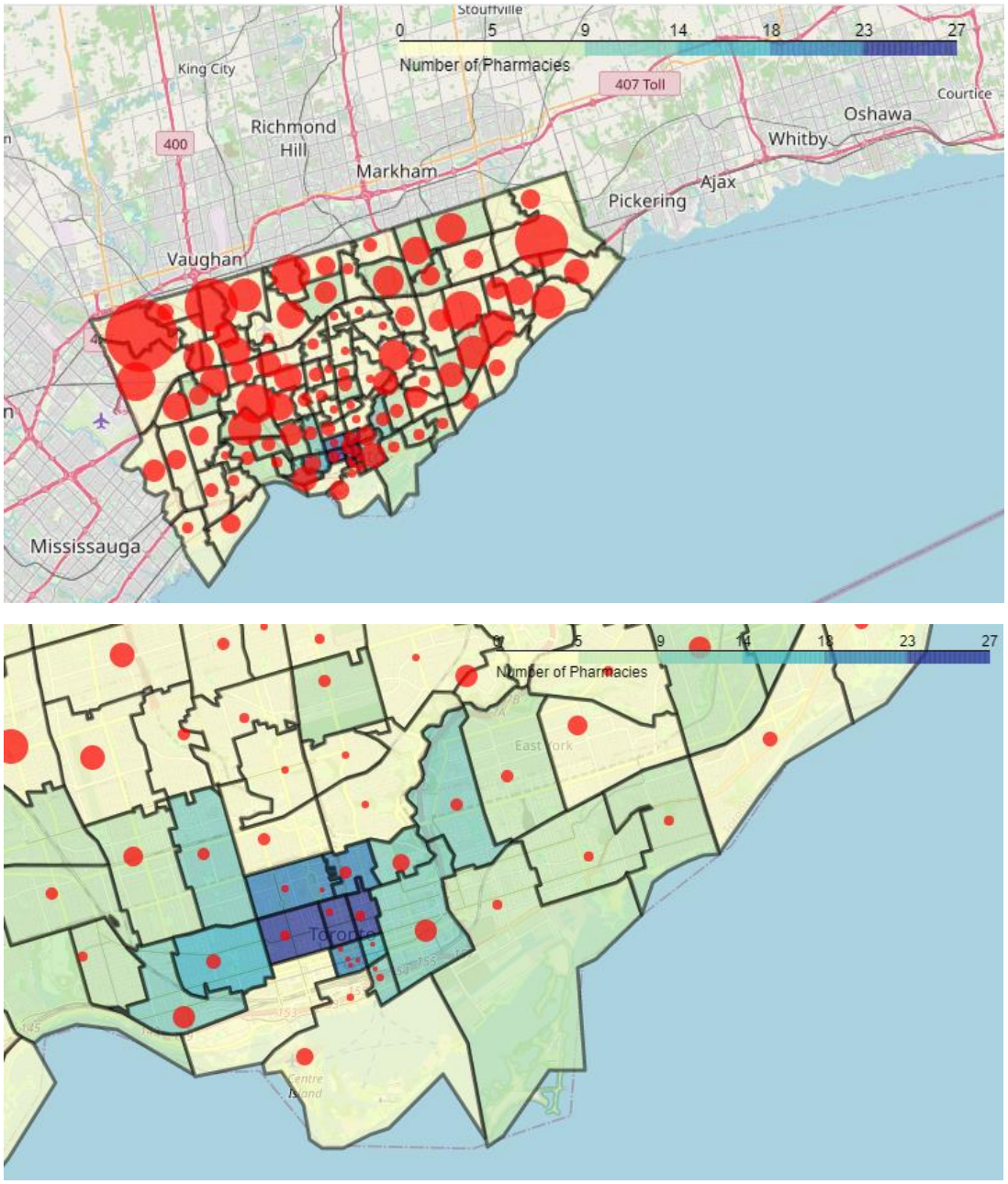

Figure 1. Folium Choropleth maps of Toronto, where the color of the neighborhood indicates the number of existing pharmacies (scale in upper right) and the red circle has a radius that is proportional to the number of Covid cases.

From the figure above it can be concluded that there is not a direct correlation between number of covid cases and number of existing pharmacies. For example, in downtown the figure shows the largest number of pharmacies, but the number of covid cases is low. Far from downtown there are many neighborhoods with the largest number of covid cases, and the number of existing pharmacies is lower than 5. Notice that for very small neighborhoods, or if the neighborhood location falls close to the neighborhood limit, Foursquare might count pharmacies that are out of the neighborhood. Therefore, the number of pharmacies in each neighborhood should be interpreted as the number of pharmacies close to the neighborhood coordinates. The next section uses a machine learning algorithm to divide neighborhoods into clusters.

Neighborhood clustering

An unsupervised Kmeans clustering model was used to separate neighborhoods into 5 different clusters, according to their main features: distance to downtown, number of covid cases and number of existing pharmacies. The features where normalized with an sklearn library, to improve the reliability of the algorithm. A map showing different colors for each cluster in presented in figure 2. The same color code is kept for each cluster in the next figures, for clarity and ease of visualization.

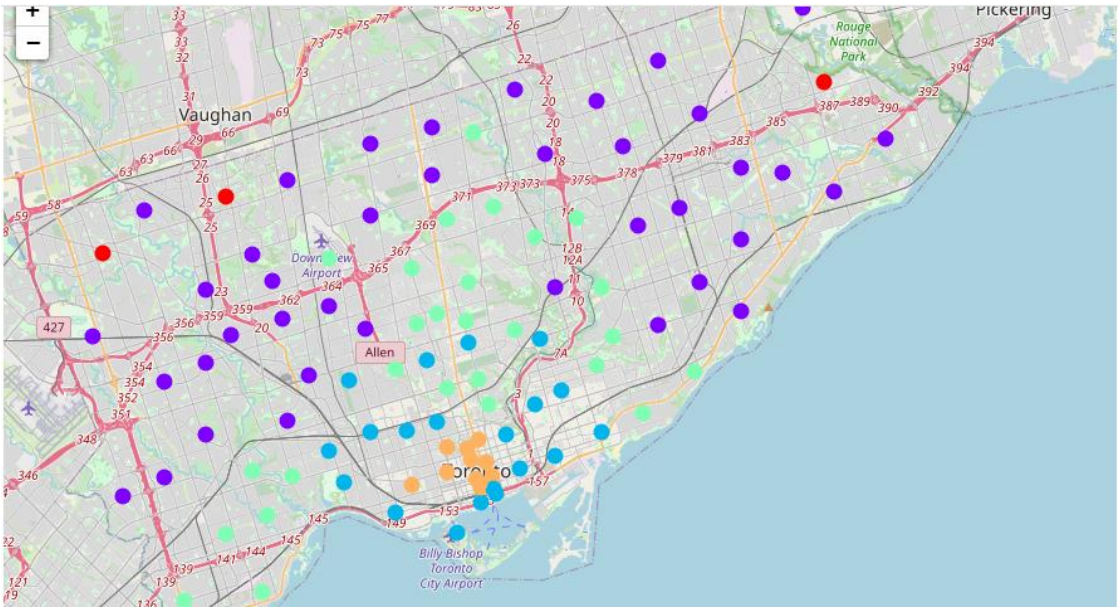

Figure 2. Toronto Neighborhoods, clustered according to the number of existing pharmacies, number of covid cases and distance to downtown.

Results

Figure 3 shows box plots showing normalized feature distribution. Figure 4 shows a bubble plot, where the x and y position of the bubble indicates the number of existing pharmacies and distance to downtown, respectively. The number of covid cases is indicated by the size of the bubble. In figure 3 and figure 4, each cluster is indicated by a different color, following the same color code than in figure 2. Figure 3 clearly indicates that neighborhoods in the red cluster have the largest number of covid cases, are far from downtown, and have a relatively low number of pharmacies. On the other hand, neighborhoods in the yellow cluster have the largest number of pharmacies, are closest to downtown and have covid cases lower than average. Neighborhoods in the purple cluster have a larger than average number of covid cases, figure 3, have a distance to downtown larger than average and have a mixed number of pharmacies, but less than 8, figure 4. Neighborhoods in the light blue cluster have larger than average values of covid cases, and are closer to downtown than the rest of the cluster, except for the yellow cluster. Finally, neighborhoods in the light green cluster have 5 or less existing pharmacies, average or lower than average covid cases and an intermediate distance to downtown.

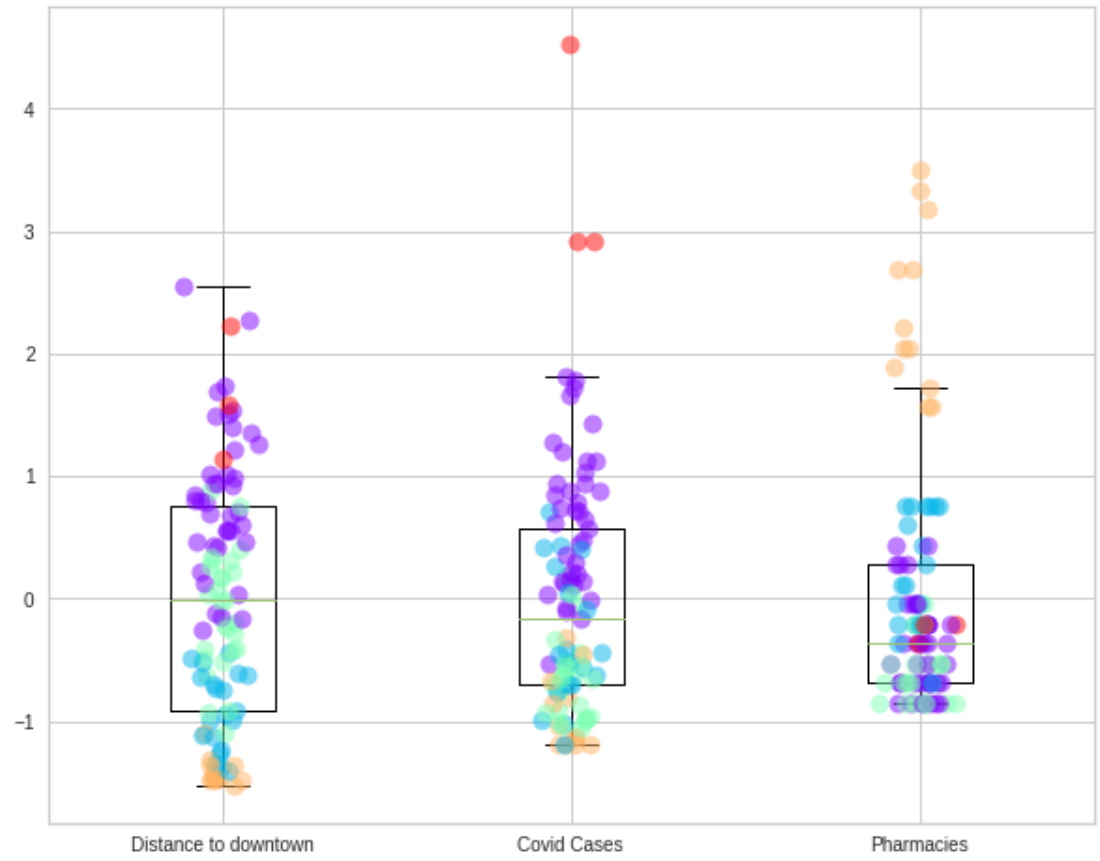

Figure 3. Box plots, showing distribution of distance to downtown, covid cases and existing pharmacies for each neighborhood. The cluster that each neighborhood belongs to is indicated with a color, same color scheme as preceding and next figure.

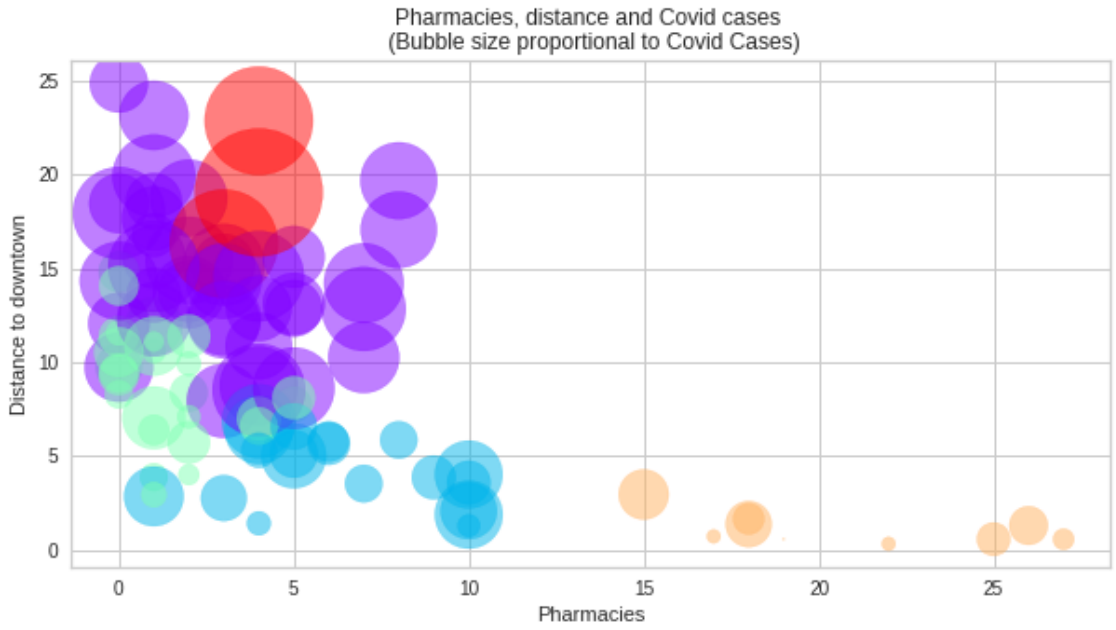

Figure 4. Bubble plots, showing distance to downtown, existing pharmacies and covid cases for each neighborhood, this last one indicated by bubble size. The cluster that each neighborhood belongs to is indicated with a color, same color scheme as preceding two figure.

Using the inverse transformation for standard scaler, the centers of each cluster are obtained, and are shown in the next table, that further highlights their difference.

Table 3. Cluster centers, with color code and cluster label as index.

	distance to downtown	Covid Cases	Existing Pharmacies	color code
0	14.587057	1512.743590	2.871795	Purple
1	4.047961	703.850000	6.550000	Light blue
2	8.886676	435.888889	1.407407	Light Green
3	1.044486	216.583333	20.000000	Yellow
4	19.414714	3817.000000	3.666667	Red

Discussion

Covid cases in neighborhood in the purple cluster are higher than the average number of covid cases for Toronto neighborhoods. The number of pharmacies is 8 or less, most neighborhoods have pharmacies below average. The distance from downtown ranges from 8 to 25 km. Those neighborhood probably need some more pharmacies.

For the light blue cluster, we have neighborhoods close to downtown, with a number of pharmacies lower than 10. There is a mixed amount of covid cases. Those are probably the best neighborhoods to open pharmacies, if the customer decides to prioritize closeness to downtown.

For the light green cluster, we have neighborhoods with a number of pharmacies lower than 5. There is a mixed amount of covid cases. Those are probably good neighborhoods to open pharmacies, but the distance to downtown is intermediate.

In the yellow cluster we find the neighborhoods with the largest amount of pharmacies, and also closest to downtown. Curiously, the number of Covid cases was below average for all neighborhoods. The neighborhoods in this cluster do not need more pharmacies.

In the red cluster, we find the neighborhoods with the largest amount of covid cases, and with a number of pharmacies lower than the average. Despite they are far from downtown, they are in need of more pharmacies.

Conclusions

In this project, Toronto neighborhoods were analyzed according to the number of Covid Cases, existing pharmacies and distance to downtown Toronto. Kmeans clustering was used to sort the neighborhoods into 5 groups, and each group was analyzed to determine which group contains

the best options for opening a new pharmacy. Neighborhoods with postal code M1B, M3N and M9V (cluster 4, red) had the largest number of covid cases, and the number of pharmacies was well below average, so they are the best option if the investor is willing to open it more than 16 km from downtown. The closest to downtown neighborhoods (cluster 3, yellow), had the highest amount of pharmacies, but curiously the number of Covid cases were lower than average in those neighborhoods. The market is likely saturated in those neighborhoods, so they would not be a good option. The rest of the clusters (0,1 and 2) have neighborhoods with good options to open a pharmacy, the choice will depend on whether the investors prioritizes distance to downtown, or market demand.