

Capstone Project - The Battle of Neighborhoods

Health Facility Distributions in Neighbourhood of Toronto amid Corvid-19

By Jieun Kim
17 April 2020

Table of contents

1. Introduction

- Motivation
- Problem

2. Methodology

- Data collection
- Data analysis process

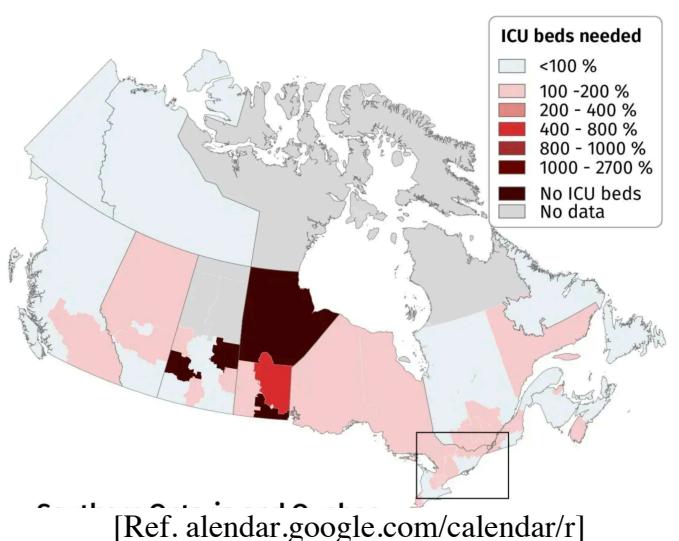
3. Results

4. Discussions and conclusions

1. Introduction

1.1 Motivation

This project is completed as part of IBM's Data Science Professional Certificate course offered by Coursera.org. We focus the issues about health facilities' distribution in Toronto area. Toronto has been heavily hit by Corvid-19 compared to any other large cities in Canada. In fact, some hospital regions across the country would have to boost their ICU beds by at least 1½ times their capacity. Ten hospital regions in Ontario (Toronto is included here), eight in Quebec, three in Saskatchewan and two in Alberta and Manitoba would all face ICU bed capacity challenges, according to the data.



[Ref. calendar.google.com/calendar/r]

Toronto is also marked as a vulnerable region where a patient transfer to a large capacity of hopstical in different regions. In Toronto, there are currently 2,881 confirmed and probable cases of the virus in Toronto, including 147 death as of 17 April. To give some figures about Toronto, Toronto is the largest city, however, recorded relatively lower density of population compared to Vancouver for instance. Many scientists are wondering why Toronto becomes most victim in this pandemic. In this regard, the project looks for the number of hospital facilities and their distributions in the neighbourhood. This attempts will help diagnose the current overcrowding situation and support health-related decision makers concerning how to improve health system in Toronto, in particular.

1.2 Problem

The object of this project is to analyse neighbourhood of Toronto, where health facilities has snot sufficiently reached, or overly populated. With the help of Data Science Methodology and machine learning techniques, we can certainly build an analysis to raise issues about following question
(*Are health facilities in Toronto equally distributed through the neighborhood?*)

2. Methodology

2.1 Data collection

Location & Neighbourhood	Wikipedia Canada Ontario-Toronto postcode data	https://en.wikipedia.org/wiki/ List_of_postal_codes_of_Canada:_M
Venue	Foursquare	Search query (Hospital, Pharmacy)
Support data	Statistics of Corvid-19 Canada	https://www.worldometers.info/coronavirus/

2.2 Data analysis process

Data analysis process consists of 5 steps: Analytic Approach, Data requirement, data collection, data understanding and preparation. The following list illusgtrates details about each step.

1. The analytic approach for this problem is to perform unsupervised learning technique such as K-means Clustering. This will help to identify various patterns based on neighbourhoods in Tronto.
2. •We would require data such as list of Boroughs and Neighbourhood of Toronto, also the Corvid data for Toronto area)
3. •Once we have noted the data requirements, the next step is data collection. We need to scrape data from the online websites using libraries such as Beautiful Soup. Also, using Foursquare.
4. The data understanding, and preparation part would be the most difficult as the collected data would not be clean. Goal here is to clean the data.
5. We would use K-Means algorithm to create K clusters and utilize Foursquare dataset to examine whether there exist any distinct characteristics of neighbourhoods.

3. Results

3.1 Hospital and Pharmacy Information in Toronto

Neighborhood firstly collected through Wikipedia dataset. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. We then call a search to Foursquare to extract a list of hospital and Pharmacy information in relation to latitude and longitude.

The tables enlist a total of 14 hospitals and pharmacies in the neighborhood of Toronto.

	name	categories	address	lat	lng	postalCode	state	
0	The Hospital for Sick Children (SickKids)	Hospital	555 University Ave.	43.657499	-79.386512	M5G 1X8	ON	
1	Toronto General Hospital	Hospital	190 Elizabeth St	43.658762	-79.388292	M5G 2C4	ON	
2	St. Michael's Hospital	Hospital	30 Bond St	43.653784	-79.377809	M5B 1W8	ON	
6	Inpatient Lounge - St. Michael's Hospital	Hospital		Nan	43.653428	-79.379383	Nan	Nan
7	MRI Department Sick Kids Hospital	Hospital	555 University Ave.	43.657149	-79.384323	Nan	ON	

Table 1 Screenshot of hospitals in Toronto (partial list)

	name	categories	address	lat	lng	postalCode	state	ne
0	I.D.A. - Pharmacy By The Grange	Pharmacy	275 Dundas St W	43.654158	-79.390490	M5T 1G1	ON	
1	On Care Pharmacy	Pharmacy	481 Dundas West	43.653188	-79.397056	M5T 1H1	ON	
2	U & C Pharmacy	Pharmacy	700 University Ave,112	43.658808	-79.390989	M5G 1Z5	ON	
3	RK Pharmacy	Pharmacy	302 Spadina Avenue	43.653262	-79.398402	Nan	Nan	
4	Leslie L. Dan Pharmacy Building	College Academic Building	144 College St.	43.659963	-79.391329	M5S	ON	
5	Guardian - Morelli's Pharmacy	Pharmacy	15 York St	43.642418	-79.380937	M5J 0A3	ON	
6	I.D.A. - The Palmerston Pharmacy & HomeCare	Pharmacy	499 College St	43.655889	-79.409718	M6G 1A5	ON	
7	Pharmacy PMH	Pharmacy		Nan	43.656351	-79.391775	Nan	ON
8	Guardian Pharmacy	Pharmacy	755 Dundas St. W	43.655781	-79.383859	Nan	ON	
9	Victor Pharmacy	Pharmacy	123 Edward St.	43.655445	-79.386335	Nan	ON	

Table 2 Screenshot of pharmacy in Toronto (partial list)

The data extracted from Foursquare is surprisingly small. We assume that health facilities including hospital, emergency unity, and pharmacy, are not populated in Foursquare site. Not many users post a review on those facility.

Figure 1 visualization the data from Table 1 and 2 on folium function. We can found that most of health facilities are centred near two university zones. This imbalanced distribution of health related facilities make the problem worsen. That might link why Toronto has suffered from overwhelmed patients and a shortage of hospitals, despite that Toronto is the largest city in Canada.

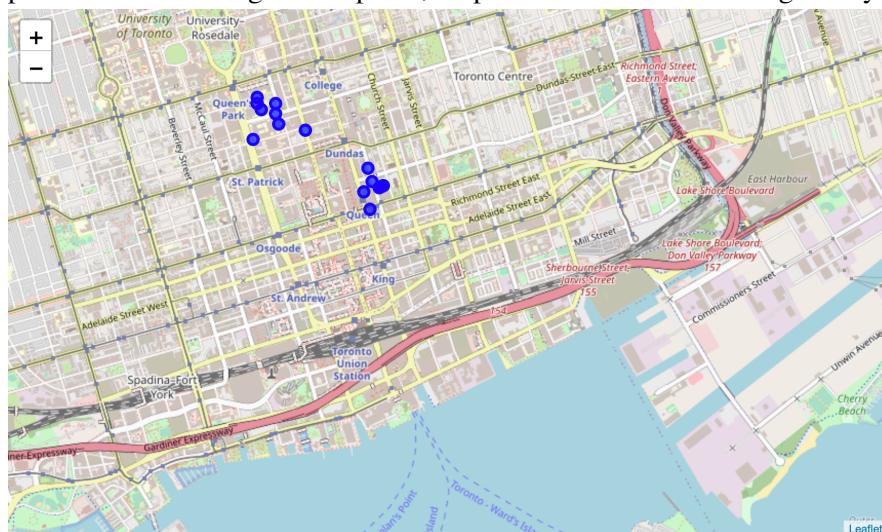


Figure 1 Distribution of Hospitals and Pharmacies in Toronto (folium map)

Further, we run a k -means to cluster the neighborhood into 6 clusters. The purpose was to examine each cluster and determine the discriminating venue categories that distinguish each cluster. Here, we have not limited search query. We intended to examine general tendency and distribution of neighborhood of Toronto. Neighbourhood in main area of Toronto is likely to show restricted popular venue (Cluster 1 in red). The cluster 1 is also populated with many restaurants and also gym& yoga studio. Table below and Figure demonstrate two things: first most of venue are labeled about entertainment, food & drinks. The data include health-related venue, however, as the data is relatively small, we cannot extract any dominant results from this.

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th N Com Ven
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Coffee Shop	Health Food Store	Trail	Asian Restaurant	Pub
1	M4K	East Toronto	The Danforth West / Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Furniture / Home Store	Book
2	M4L	East Toronto	India Bazaar / The Beaches West	43.668999	-79.315572	0	Fast Food Restaurant	Gym	Pub	Sandwich Place	Burrit Place
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Brewery	Gastropub	Bakei
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Park	Bus Line	Swim School	Dessert Shop	Ethio Resta

Table 3 Screenshot of Cluster analysis (partial list)

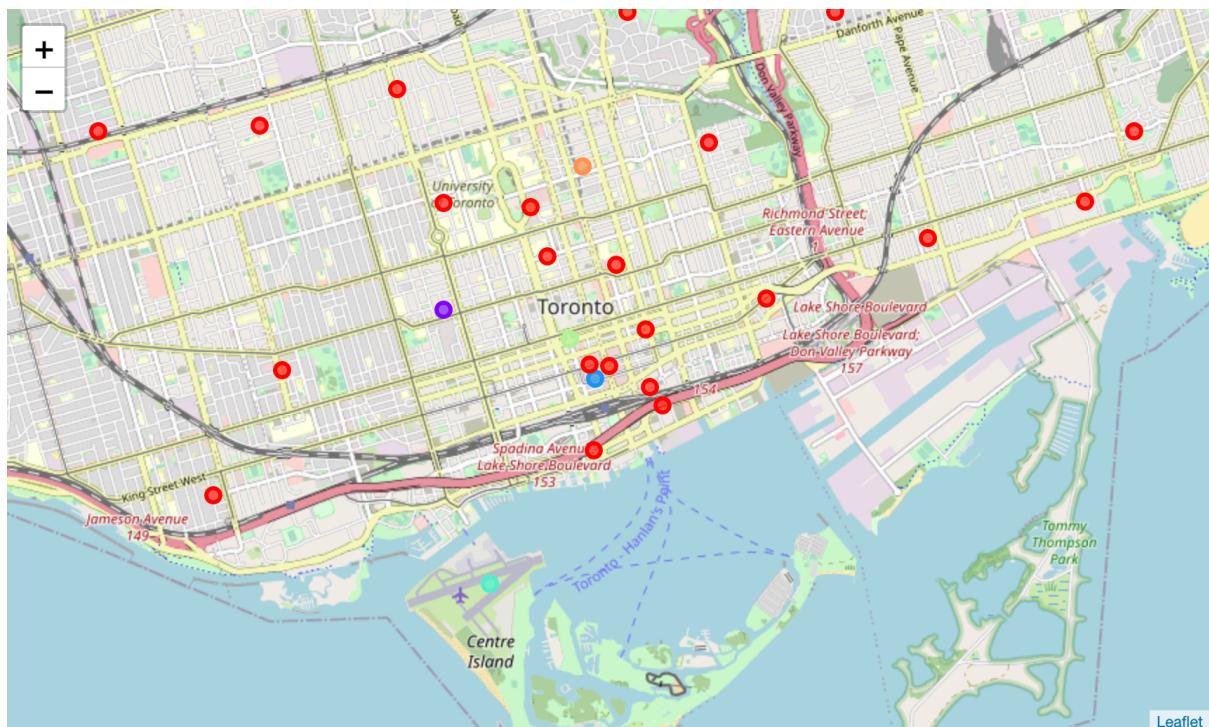


Figure 2 K-mean cluster analysis (**folium map**)

4. Discussions and conclusions

As per analysis done in this project, it is recommended to disperse hospital facility in neighbourhood of Toronto Cluster 1 includes cities such as CN Tower / King and Spadina and University area, they are the one's with high density of hospital facility

Note that the study has many limitations. For instance, the results might be biased as long as the current capstone rely on Foursquare data only for venue marked which is highly popular for "attractions" and "restaurants",

With the help of Foursquare API and various machine learning techniques we can perform analysis on various other venues and can answer many business problem. Basis on the Data visualization we can have a good understanding why a particular neighbourhood is overwhelmed with hospitals, and others are not.