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

1. Methodology

* Data collection
* Data analysis process

1. Results
2. Discussions and conclusions
3. Introduction
   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.

A close up of a map

Description automatically generated

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

Toronto is also marked as a venerable 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. 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)*

1. Methodology
   1. Data collection

|  |  |  |
| --- | --- | --- |
| Location & Neighbourhood | Wikipedia Canada Ontario-Toronto postcode data | https://en.wikipedia.org/wiki/ List\_of\_postal\_codes\_of\_Canada:\_[M](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/ |

* 1. 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.
6. Results
   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.

A screenshot of a cell phone

Description automatically generated

Table 1 Screenshot of hospitals in Toronto (partial list)

A screenshot of a cell phone

Description automatically generated

Table 2 Screenshot of pharmacy in Toronto (partial list)

The data extracted from Foursquare is surprisingly small. We assume that health facilities including hoptial, 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 overhelmed patients and a shortage of hospitals, despite that Toronto is the largest city in Canada.

A close up of a map

Description automatically generated

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. Neighboorhood 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 venu, however, as the data is relatively small, we cannot etract any dominant results from this.

A screenshot of a cell phone

Description automatically generated

Table 3 Screenshot of Cluster analysis (partial list)

A close up of a map

Description automatically generated

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

1. 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 Spading 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.