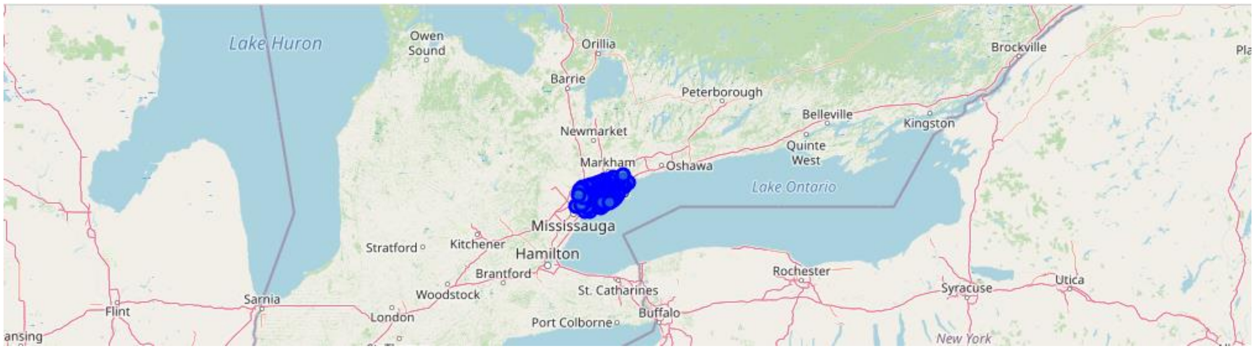
IBM Data Science



Battle of the Neighborhoods

PART-II Week 5

Introduction

The Data Science Capstone project Battle of the Neighborhood – Part II allows course participants to create their own project that build on top of Part-I where neighborhoods were reviewed and mapped for the Toronto region.

My project will take this neighborhood project one level further by addressing a very important need of the day which is to balance incidences of covid-19 in specific neighborhoods and quide first responders to the appropriate hospital that can best serve the needs of patients.

Problem Statement

Hospital overcrowding and lack of beds and personal protective equipment become commonplace in the event of pandemics as well as natural disasters and large scale emergencies. When this happens, an uneven utilization of available assets across area hospitals leads to the need for temporary facilities so that the delivery of assets and logistics can be better managed. The current practice of creating makeshift hospitals in stadiums and conference centers while aimed at reducing logistics planning overheads flies in the face of social separation. A large stadium full of COVID-19 patients cannot be a healthy environment for care givers. More than the pathogenic environment it fosters, the psychological impact on patience being cast shoulder to shoulder in tents separated by semi-transparent plastic sheeting is not emblematic of a first world country!

Solution

The problem of uneven utilization of available assets or locating assets in one concentrated area leads to high mobility rates in terms of movement of the assets or the physical movement of people needing those assets. This problem is not very different from mobile networks where equally distributed assets cause uneven loading and lead to addition of more assets rather than using proven “load balancing” measures. The solution envisioned for this problem is to borrow the load balancing methods from mobile networks to the problem of predicting where assets need to be deployed in order to cater to surge conditions.

Target Users:

First responders, City planners, Hospital planners,

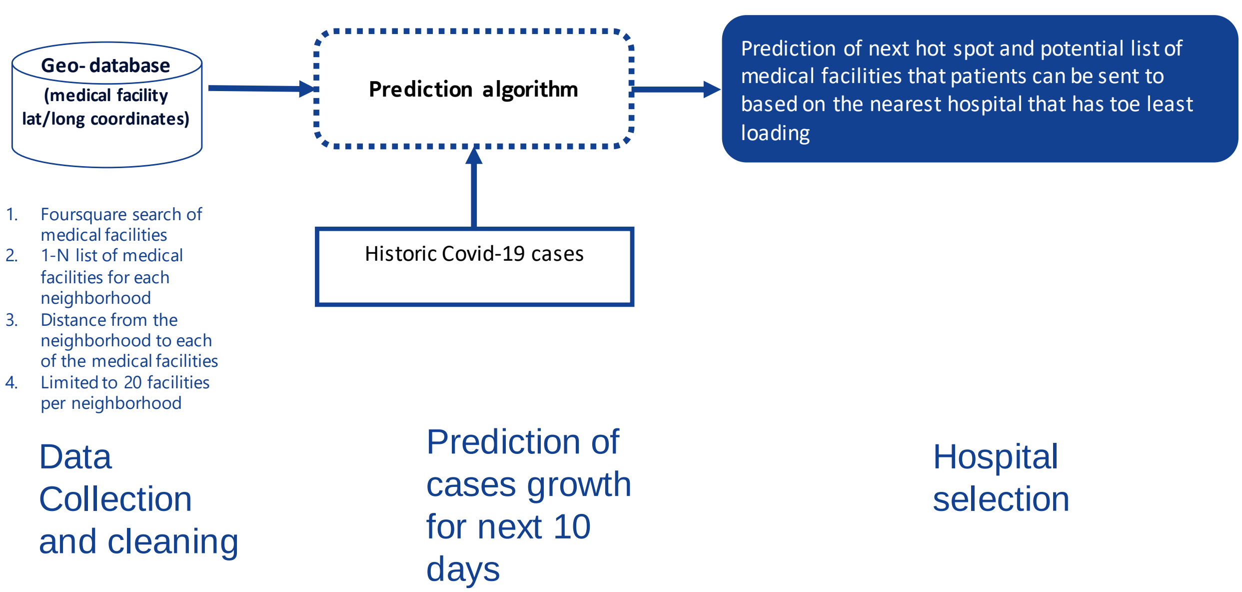
Methodology

The methodology for implementing the solution is based on the following steps:

1. Read the hospital list from foursquare for each neighborhood
2. Find the Postal code of each hospital and add the hospital to the neighborhood(s) that are in that postal code
3. (*) Count the number of hospitals in each neighborhood
4. Create an assumed capacity for each hospital
5. (*) Calculate the total capacity for the neighborhood
6. Create a predicted demand that indicates the number of COVID-19 patients that falls in the capacity range for a given neighborhood.
7. (*) If the predicted demand exceeds the available neighborhood capacity, find the closest neighborhood that has excess capacity: we call this "Load Balancing”.
8. Show on a map, the number of hospitals in each Toronto neighborhood
9. (*) Using k-means clustering or other methods, to validate the load balancing methodology

(*)- steps that are not implemented in this study due to lack of granular data

The graphic below summarizes the steps taken in the methodology followed:



Implementation Steps

We first find the category ID for hospitals and create a http query to FourSquare to get 10-20 hospitals centred around each neighborhood. The range of 10-20 will be compared to gauge the number of unique hospitals returned in the search. The distance from the search point to each hospital is also of immense interest to the solution. The foursquare database returns only postal codes for each hospital so we will find all hospitals around each postal code. Get rid of data that we will not need. Since we are iterating for each neighborhood and finding the 10-20 closest hospitals.

Foursquare URL API query:

foursquare_url = 'https://api.foursquare.com/v2/venues/search?ll={},{}&categoryId={}&client_id={}&client_secret={}&limit={}&v={}'.format(lat, lng, categoryID, CLIENT_ID, CLIENT_SECRET, LIMIT, VERSION)

This results in (After proper cleaning):

```
hospital_names= df_hosp['name'].value_counts()
hospital_names
```

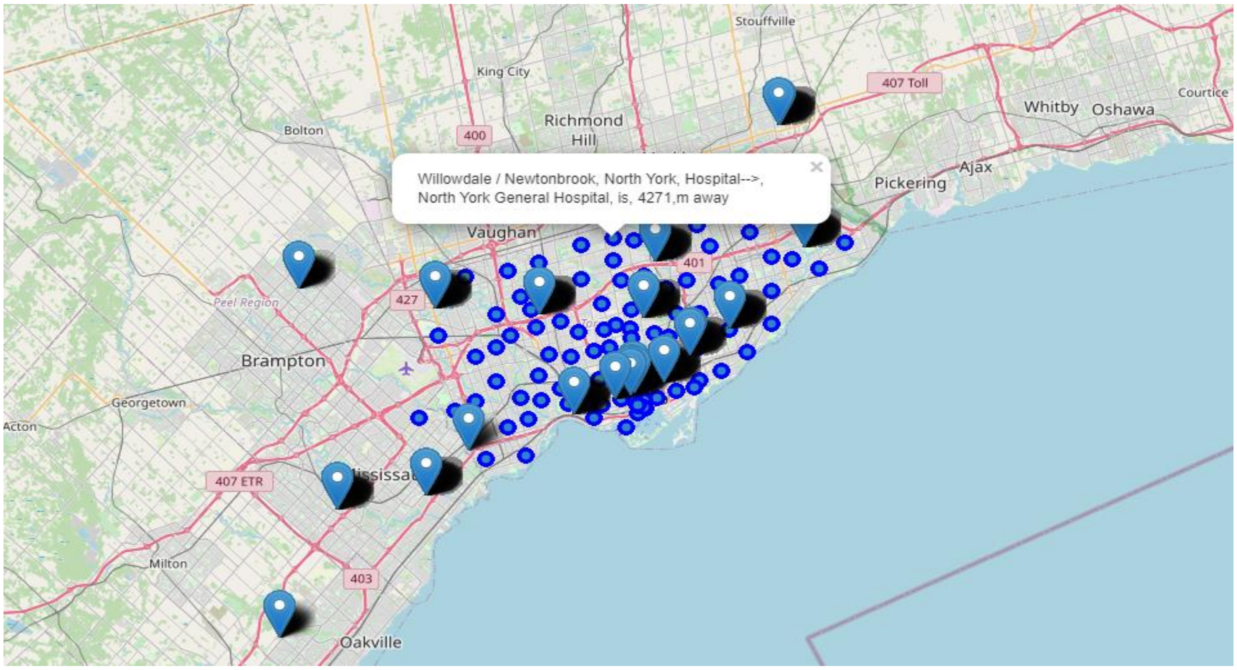
```
Out[18]: St. Joseph's Health Centre                                       103
         Humber River Hospital                                            103
         Bridgepoint Health                                               103
         Toronto Rehabilitation Institute                                 103
         Michael Garron Hospital                                          103
         Mount Sinai Hospital                                             103
         North York General Hospital                                      103
         The Hospital for Sick Children (SickKids)                        103
         Princess Margaret Cancer Centre                                  103
         Toronto Western Hospital                                         103
         Providence HealthCare                                            103
         Rouge Valley Centenary Hospital                                  103
         Sunnybrook Health Sciences Centre                                103
         Women's College Hospital                                         103
         Markham Stouffville Hospital                                     102
         Etobicoke General Hospital                                        99
         Trillium Health Centre                                            98
         Credit Valley Hospital                                            81
         Toronto General Hospital                                          58
         Southlake Regional Health Centre EMERGENCY                        46
         Southlake Regional Health Centre                                  46
         Brampton Civic Hospital                                           34
         Sunnybrook Hospital                                               32
         Oakville Trafalgar Memorial Hospital                              13
         Childbirth & Children's Centre Wing at Markham Stouffville Hospital  5
         Emergency Toronto East General Hopital                             2
         M-wing: Sunnybrook                                                 2
         Name: name, dtype: int64
```

At the next step, the distance and postal codes along with the hospital names are the data of interest to map to the neighborhood data that we already have mapped

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Hospital Name	Hospital lat	Hospital long	Hospital Postal	Hospital Distance	Nearest Hospital	Nearest Distance
0	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Women's College Hospital	43.661491	-79.387602	M5S 1B2	22424	Women's College Hospital	22424
1	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	The Hospital for Sick Children (SickKids)	43.657499	-79.386512	M5G 1X8	22687	Women's College Hospital	22424
2	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Humber River Hospital	43.724337	-79.488066	M3M 0B2	25329	Women's College Hospital	22424
3	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Michael Garron Hospital	43.689573	-79.326173	M4C 3E7	16802	Michael Garron Hospital	16802
4	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Southlake Regional Health Centre	44.061136	-79.452311	L3Y 2P9	35070	Michael Garron Hospital	16802
5	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Toronto General Hospital	43.658762	-79.388292	M5G 2C4	22682	Michael Garron Hospital	16802
6	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Toronto Western Hospital	43.653434	-79.406074	M5T 2S7	24105	Michael Garron Hospital	16802
7	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Southlake Regional Health Centre EMERGENCY	44.060452	-79.452570	L3Y 2P9	35021	Michael Garron Hospital	16802
8	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Toronto Rehabilitation Institute	43.656307	-79.389910	M5G 2A2	22970	Michael Garron Hospital	16802
9	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353	Markham Stouffville Hospital	43.883569	-79.232452	L3P 7P3	9088	Markham Stouffville Hospital	9088

And we can map the neighborhoods along with the hospitals that are closest to that neighborhood.

NOTE: the callout shows the distance to the closest hospital



The final implementation step is to collect historic covid-19 incidence rates. Given the absence of neighborhood level data for covid-19 spread, we will use this website with the assumption that the growth rates for he US at a country level matches the rate at our worst affected area: NY City.

The website used for the covid-19 case history is:

https://raw.githubusercontent.com/datasets/covid-19/master/data/key-countries-pivoted.csv

This site has daily data from early February for the top 5-6 affected cities. The format of this data is:

```
Date,China,US,United_Kingdom,Italy,France,Germany,Spain,Iran
2020-01-22,548,1,0,0,0,0,0,0
2020-01-23,643,1,0,0,0,0,0,0
2020-01-24,920,2,0,0,2,0,0,0
2020-01-25,1406,2,0,0,3,0,0,0
2020-02-04,23707,11,2,2,6,12,1,0
2020-02-05,27440,11,2,2,6,12,1,0
```

First six days of outbreak

(USA is highlighted)

```
2020-04-18,83787,732197,115314,175925,149149,143342,191726,8086
2020-04-19,83805,758809,121172,178972,154097,145184,198674,8221
2020-04-20,83817,784326,125856,181228,156480,147065,200210,8350
2020-04-21,83853,811865,130172,183957,159297,148291,204178,8480
2020-04-22,83868,840351,134638,187327,157125,150648,208389,8599
2020-04-23,83884,869170,139246,189973,159460,153129,213024,8702
```

Last six days of outbreak

(USA is highlighted)

Conclusion

The project allows us to map city neighborhoods to medical facilities and identifies a set up to 20 hospitals with distances from each neighborhood to the hospitals listed by the distances between them.

With this 10-day prediction horizon, we can distribute the predicted loading across the granularity of our regions (neighborhoods). As a next step, a similar capability to monitor current hospital load factors will allow us to direct COVID-19 patients to the other hospitals. Absent hospital load factors, the solution will only predict demand in terms of resources in specific regions without the capability to direct new cases to unloaded hospitals.