ANALYSING COVID-19 PATIENTS DISTRIBUTION FOR VACCINATION

DIVYA NAUTIYAL

25/JAN/2020

1. INTRODUCTION

On February 11, 2020, the World Health Organization announced an official name for the disease that is causing the 2019 novel coronavirus outbreak. The new name of this disease is coronavirus disease 2019, abbreviated as COVID-19. In COVID-19, 'CO' stands for 'corona,' 'VI' for 'virus,' and 'D' for disease. Formerly, this disease was referred to as "2019 novel coronavirus" or "2019-nCoV."

Coronaviruses, named for the crown-like spikes on their surfaces, are a large family of viruses that are common in people and many different species of animals, including camels, cattle, cats, and bats. There are many types of human coronaviruses, including some that commonly cause mild upper-respiratory tract illnesses. COVID-19 is a new disease, caused by a novel (or new) coronavirus that has not previously been seen in humans.

Animal coronaviruses rarely infect people and then spread between people. This occurred with two earlier coronaviruses, MERS-CoV and SARS-CoV.
SARS-CoV-2 virus is a beta coronavirus, like MERS-CoV and SARS-CoV. All three of these viruses have their origins in bats. The sequences from U.S. patients are similar to the one that China initially posted, suggesting a likely single, recent emergence of this virus from an animal reservoir. However, the exact source of this virus has not been identified.

1.2 Problem

The latest news on the covid-19 hit the world with the release of its vaccine.In November, companies reported that the vaccine was more than 90 percent effective in late-stage trials and were granted emergency use authorization on Dec. 11.

In midst of all this situation , the major challenge lies with the distribution of the vaccine among the people. Especially to those who have been affected with the virus and in need of urgent care. A quick visualization of all those affected are needed so the nearest hospital can provide them with the vaccination.

'The supply of COVID-19 vaccines in the United States is currently limited and the County of Sonoma is following State and Federal guidelines for distribution. Priority administration of the vaccine has been determined by a number of factors including risk of exposure from work or living environments, as well as vulnerabilities due to age and medical conditions.'

Hence it becomes necessary to sort out these factors and keeping the factors in mind we can reduce the amount of time for distribution of vaccines.

2.Data acquisition and cleaning

For the sample visualization, the dataset of Ontario, Canada was taken. Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values like the exact locations. The data was taken from the official govt. site. The file used were the one with the *COVID-19 cases* in Ontario and the second one were the *geospatial data* to get the coordinates of the places.

DATA CLEANING:

The files contained a lot of irrelevant data like assigned id, ids etc. They had to be arranged in a way that the elderly age group could be of priority . The irrelevant columns were dropped and rest were created as a new data frame to be merged with the other file of geospatial data.

```
df=pd.read_csv("COVID19 cases.csv")
df.head()
```

	_id	Assigned_ID	Outbreak Associated	Age Group	Neighbourhood Name	FSA	Source of Infection	Classification	Episode Date	Reported Date	Client Gender	Outcome	Currently Hospitalized	Currently in ICU	Currently Intubated	Ever Hospitalized	Ever in ICU
0	873349	1	Sporadic	50 to 59 Years	Willowdale East	M2N	Travel	CONFIRMED	2020-01-22	2020-01-23	FEMALE	RESOLVED	No	No	No	No	No
1	873350	2	Sporadic	50 to 59 Years	Willowdale East	M2N	Travel	CONFIRMED	2020-01-21	2020-01-23	MALE	RESOLVED	No	No	No	Yes	No
2	873351	3	Sporadic	20 to 29 Years	Parkwoods-Donalda	M3A	Travel	CONFIRMED	2020-02-05	2020-02-21	FEMALE	RESOLVED	No	No	No	No	No
3	873352	4	Sporadic	60 to 69 Years	Church-Yonge Corridor	M4W	Travel	CONFIRMED	2020-02-16	2020-02-25	FEMALE	RESOLVED	No	No	No	No	No
4	873353	5	Sporadic	60 to 69 Years	Church-Yonge Corridor	M4W	Travel	CONFIRMED	2020-02-20	2020-02-26	MALE	RESOLVED	No	No	No	No	No

```
[12]: data=pd.merge(covid,data,on='Postal Code')
      data.head(20)
```

[12]:

	Neighbourhood Name	Postal Code	Classification	Client Gender	Latitude	Longitude
0	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
1	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
2	Willowdale West	M2N	CONFIRMED	MALE	43.77012	-79.408493
3	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
4	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
5	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
6	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
7	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
8	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
9	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
10	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
11	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
12	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
13	Willowdale East	M2N	PROBABLE	MALE	43.77012	-79.408493
14	Willowdale West	M2N	CONFIRMED	MALE	43.77012	-79.408493
15	Lansing-Westgate	M2N	CONFIRMED	MALE	43.77012	-79.408493
16	Bay Street Corridor	M2N	CONFIRMED	FEMALE	43.77012	-79.408493

3. Feature Selection

Now after data cleaning, there were 77872 samples and 4 features in the data. Upon examining the meaning of each feature, it was clear that there was some redundancy in the features. For example, there was a feature in which the number neighbourhood names were way too much to make a clear visual map . Hence just for the starters we have to clean it up a little bit and select a particular neighborhood like *Willow East* and visualize the hospitals around it.

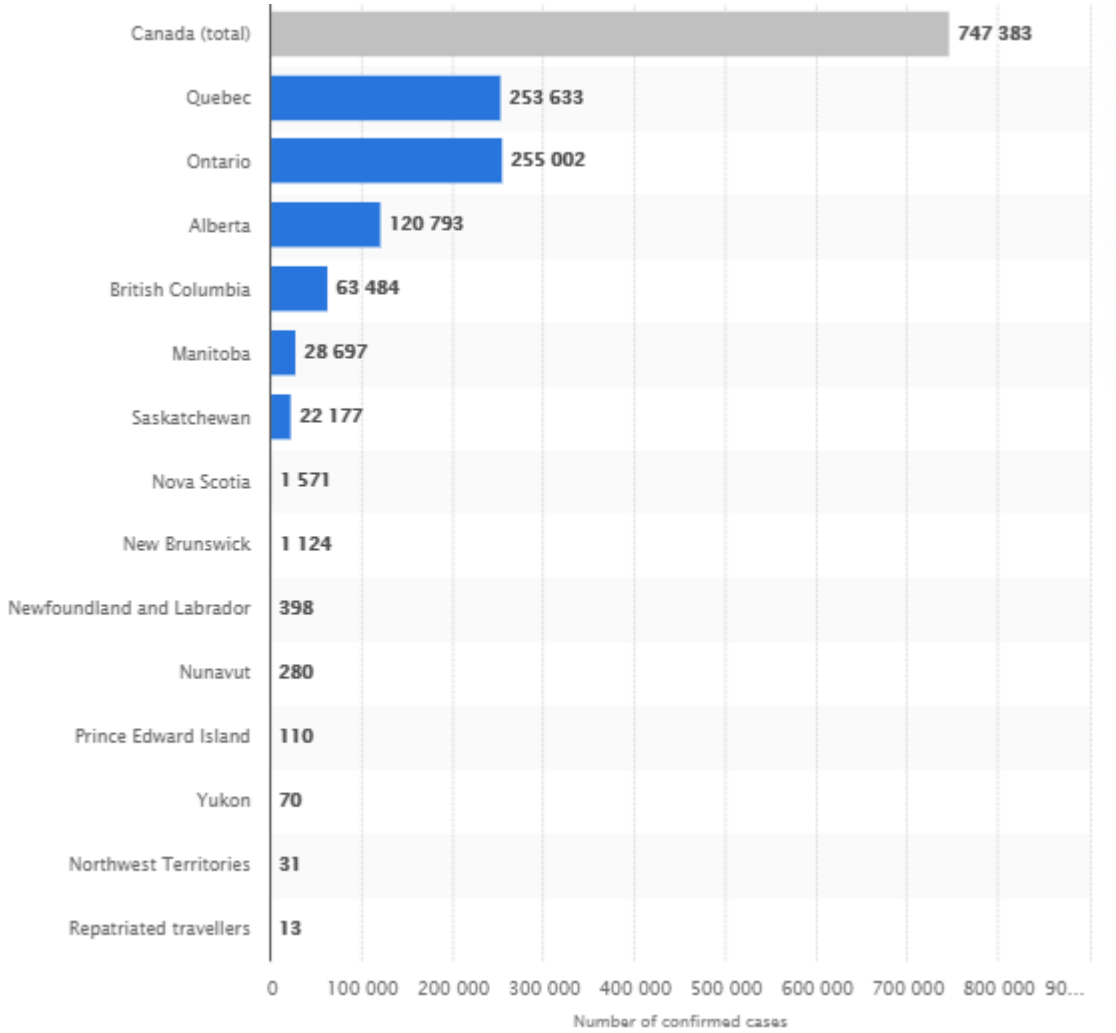
Primary features selected were Client Gender and age range, for this we visualize the dependency among these features.

3.Exploratory Data Analysis

The following histogram shows the dependency of the age group in case of the covid- 19 cases.

CURRENT POPULATION AFFECTED IN CANADA

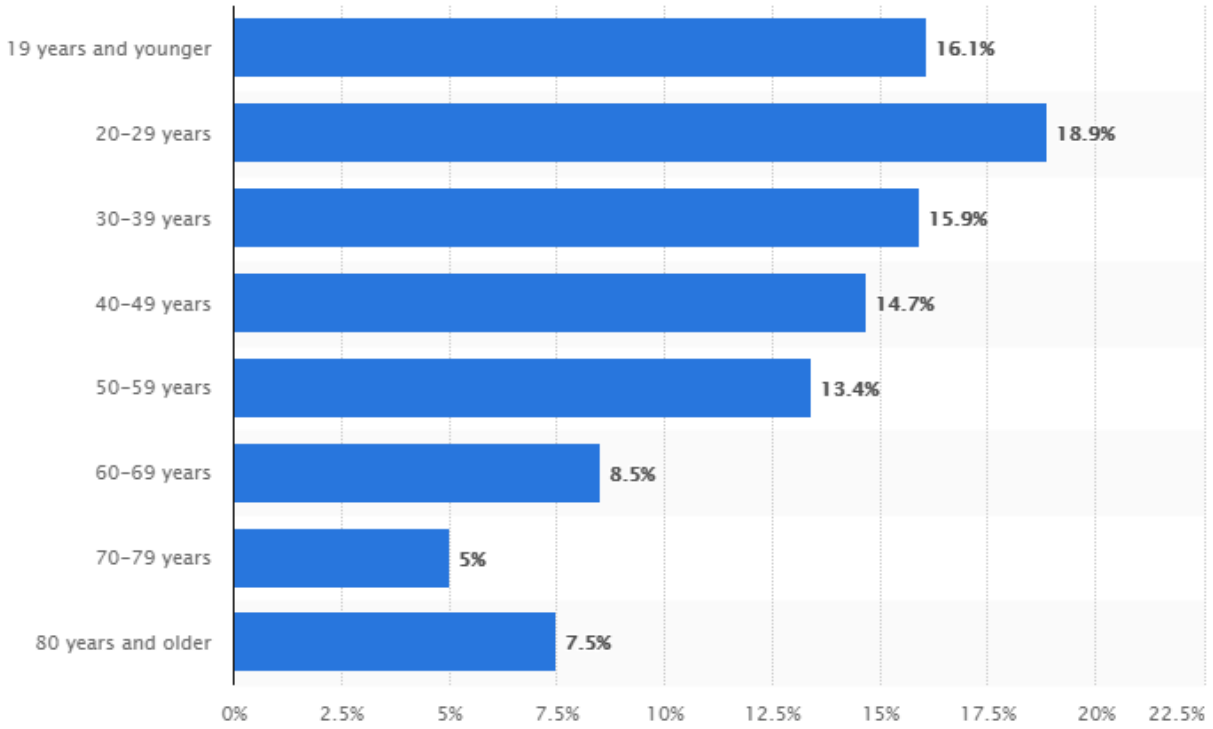
As of January 24, 2021, there had been 747,383 confirmed cases of coronavirus (COVID-19) in Canada, including 13 cases among repatriated travelers. As of this date, the coronavirus had been confirmed in every province and territory, with the province of Quebec having the highest number of confirmed cases.



TRENDS TO BE NOTICED

From the above graph we can notice that the top two most affected areas are Quebec and Ontario. We will be focusing on Ontario data set for now and dividing it further for an easy compilation.

We have to visualize the trend of affected people according to the age group and see which group of people are more likely to be affected by the virus and are in-need of urgent care.



People of which Age group are at higher risk?

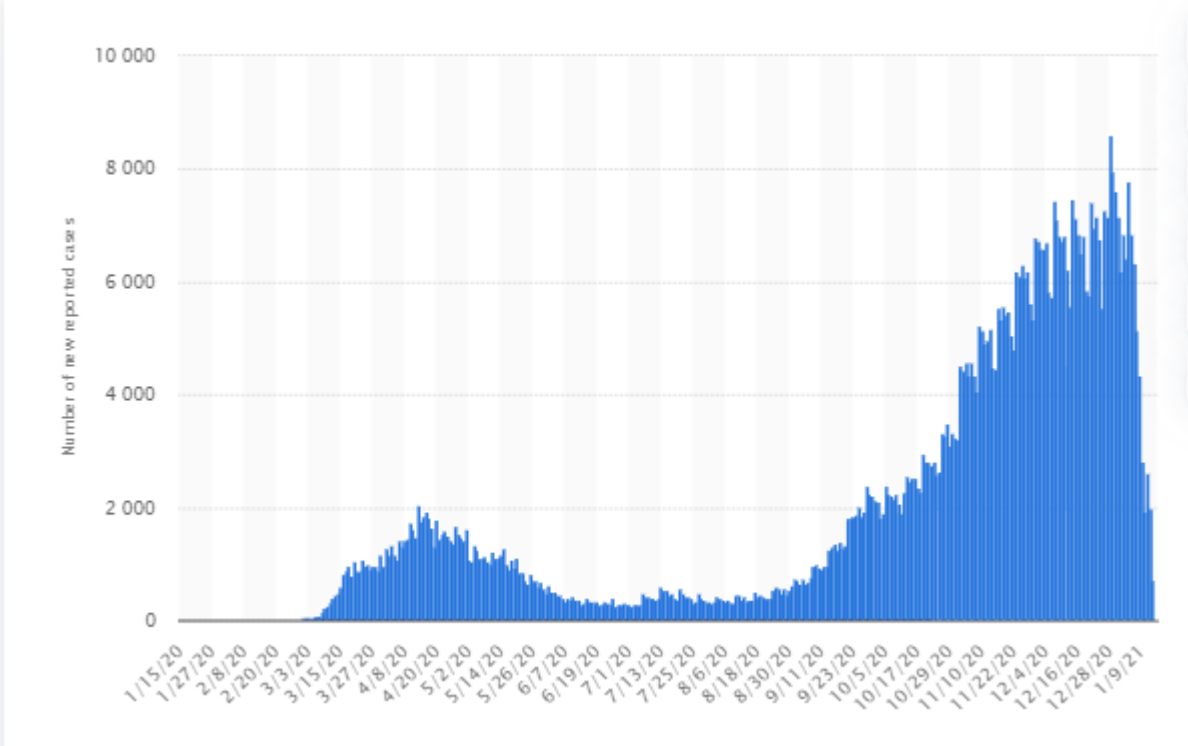
Canada appeared to have its COVID-19 outbreak under control, but the risk of infection remains high as the number of new daily cases increased as winter began. Given the risk to older people and those with preexisting conditions, it is not surprising that adults aged 55 years and older are more concerned about contracting the disease than younger age groups. During these challenging times, it is important to protect older people living alone and those in care facilities. Groups should not be discriminated against because of age, and all communities need to be supported.

Protecting yourself and others from the coronavirus

Many steps can be taken to prevent the spread of COVID-19. Standard recommendations include washing your hands frequently, regularly cleaning surfaces, and staying away from large gatherings. Governments and health institutes have advised people to avoid nonessential travel

and to limit contact with people at higher risk, such as older adults. However, increased social isolation can lead to further problems. Therefore, try to move appointments with friends and relatives online. High-risk groups should also consider stockpiling critical prescription medicines.

4.VISUALIZING NUMBER OF ACTIVE CASES

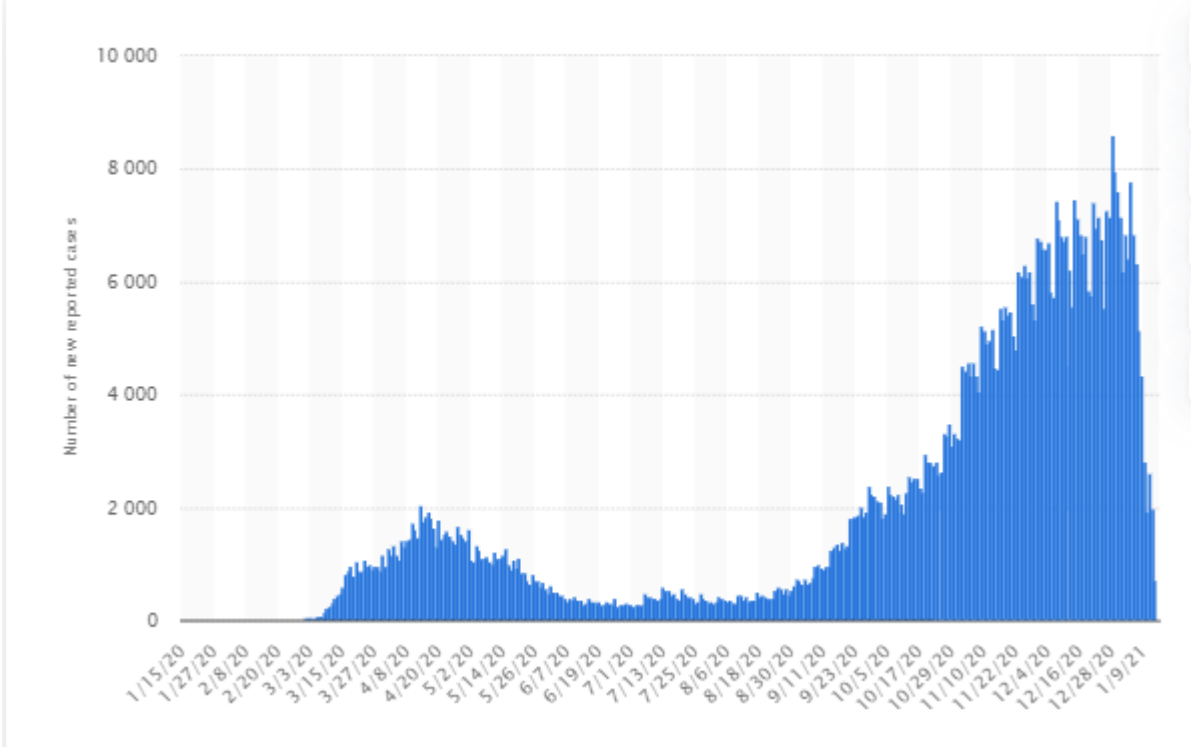


The above graph depicts the number of active cases against the dates plotted as well. We can see that the graph sees a sudden increase in the cases from May-2020.
We will try to examine these cases against the selected location in the current date schema.

CURRENT RATE OF COVID TESTS

	Neighbourhood Name	Postal Code	Classification	Client Gender	Latitude	Longitude
0	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
1	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
3	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
4	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
5	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
6	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
7	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
8	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
9	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
10	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
11	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
12	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
13	Willowdale East	M2N	PROBABLE	MALE	43.77012	-79.408493
17	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
19	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493
20	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
22	Willowdale East	M2N	CONFIRMED	MALE	43.77012	-79.408493
23	Willowdale East	M2N	CONFIRMED	FEMALE	43.77012	-79.408493

As of the 31st of March, 2020, there had been a cumulative total of 236,851 tests for COVID-19 that took place in Canada, rising to 3,946,565 by the 30th of July, 2020. The statistic illustrates the cumulative number of COVID-19 tests performed in Canada from March 12 to July 30, 2020.



5. EXPLORING AND VISUALIZING TORONTO DATA

We will be using the FOURSQUARE API to explore the neighborhood and to get all the data required for the report.

Here is a snippet of the final dataframe

```
address= 'Toronto, CA'
geolocator=Nominatim(user_agent='cn_explorer')
location=geolocator.geocode(address)
latitude=location.latitude
longitude=location.longitude

print('The geograpical coordinate of Ontario are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Ontario are 43.6534817, -79.3839347.
```

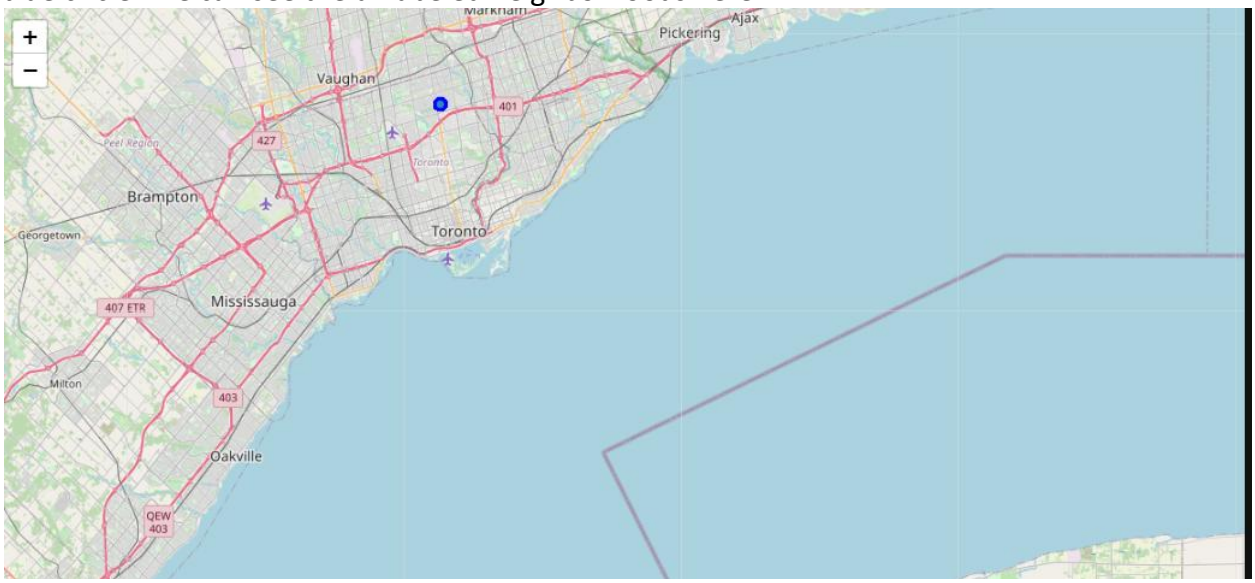
```
# create map of New York using latitude and longitude values

map_ontario=folium.Map(location=[latitude,longitude], zoom_start=10)

#add markers too map
for lat,lng,gen,neighborhood,classi in zip(new_covid['Latitude'], new_covid['Longitude'], new_covid['Client Gender'], new_covid['Neighbourhood Name'],new_cov
    label='{},{}'.format(neighborhood,classi)
    label=folium.Popup(label,parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_ontario)

map_ontario
```

The result of this is the map of our current location in Toronto, Ontario, depicted by a blue circle. We can see the unlabeled neighborhoods here.



We can't yet see the neighboring hospitals here. We will have to run the FOURSQUARE API again to get the list of hospitals in the area by passing 'HOSPITAL' key in the search_query.

Searching for hospitals near the current location

```
search_query = 'Hospital'
radius = 500
print(search_query + ' .... OK!')
```

```
Hospital .... OK!
```

```
url_h = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID,
url_h
```

```
'https://api.foursquare.com/v2/venues/search?client_id=LDMJ2QUV1CHFILPCH021GF1VITI1QIIANWDQ4X5X5ANNENHE&client_secret=PZ20JILHNNIUPM0MD324AP0B4FE2B0B3WQVDSYV
GPG1ES4&ll=43.6534817,-79.3839347&oauth_token=FL0CQ00TX0YNDLEPYOWDNBPPHCTIC55W42TVBC0ELY5V4IV0&v=20180604&query=Hospital&radius=500&limit=100'
```

```
# Send GET request for result
results = requests.get(url_h).json()
results
```

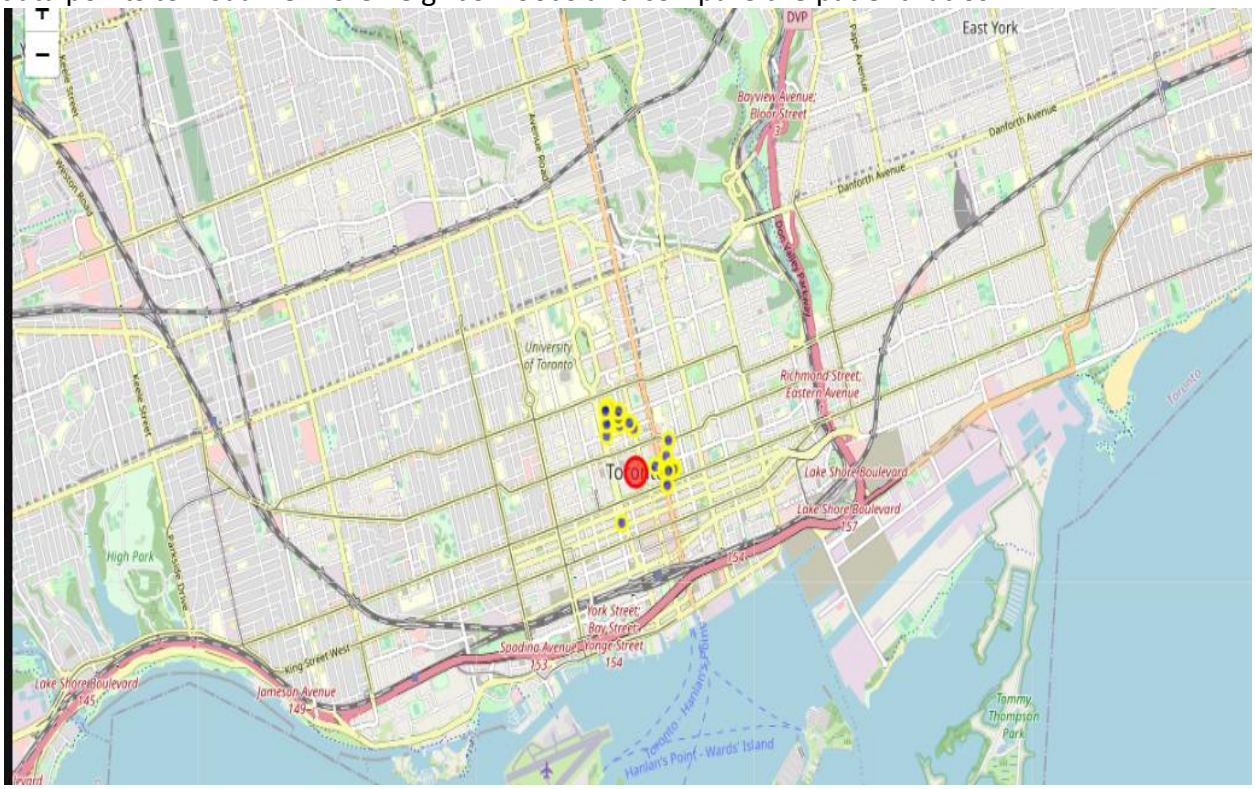

```
{'meta': {'code': 200, 'requestId': '600e5f33520d1f65397b87e7'},
 'notifications': [{'type': 'notificationTray', 'item': {'unreadCount': 0}}],
 'response': {'venues': [{'id': '4ad4c064f964a5206ef820e3',
    'name': 'The Hospital for Sick Children (SickKids)',
    'location': {'address': '555 University Ave.',
     'crossStreet': 'at Gerrard St.',
     'lat': 43.657498668962646,
     'lng': -79.3865121609307,
     'labeledLatLngs': [{'label': 'display',
       'lat': 43.657498668962646,
       'lng': -79.3865121609307}],
     'distance': 492,
```


Here we can see the results , listing all the data about the neighboring Hospitals in the areas. Now we can convert this into a dataframe with all the needed columns .

CONCLUSION

Visualize the nearby hospitals

Here the hospitals near to the current location are displayed. We can further add more data points to visualize more neighborhoods and compare the patient ratios.



We can clearly see that the hospitals which are near our location. This makes it a lot easier to spot the red zones as well as the hot spots for the people. It can be used to spot all the areas where there can be the distributions of vaccines in order to make it reach to a larger group of people much easier.

This project on a large scale can immensely help in solving this problem and within a much shorter amount of time.