Covid-19 Data Analysis and Analyzing regions of Delhi using Foursquare API

~ By Aneeka Afaq

This project will contain two parts:

1- The first part will be "**Covid-19 Data analysis as on 20th July 2020**" and I will be using the data from the World Health Organization website to analyze the world data about the spread of Covid-19. I will be using different libraries in python to read, clean, analyze, and visualize the data.

2- This part will be focused on "**Using Foursquare API to analyze the regions of Delhi**". Different features such as districts, latitudes and longitudes will be used in this section.

Data Collection

1- For the first section of this project, the data was collected from the WHO website. The data was available under the Covid-19 Dashboard and was easily retrieve as a csv file. Following is the URL from where I collected the data: https://covid19.who.int/info

As for the Folium Choropleth, I downloaded the json file 'custom.geo.json' which contained all the data of the world countries such as their names, postal codes, etc.

2- For the second Foursquare section, I first created my credentials for the Foursquare developer portal and used those credentials in my notebook for further usage.

I used the data from the following website: http://pincode.india-server.com/states/delhi/ . This website allowed me access to the data of all the districts in the state of Delhi. As for the latitudes and the longitudes of all these districts, I used Google to search for the coordinates and added them to the given dataframe manually.

Methodology

1. For the "**Covid-19 Data analysis as on 20th July 2020**" section:

   - The data was collected from the WHO – Covid -19 dashboard. The WHO provides access to open source data and that is from where I retrieved this detailed data of all the countries and their covid-19 cases.

   - The data was present for as many as 5 months for each country, i.e., from February 2020 to 20th July 2020(the day on which I used the data), so I had to do some data wrangling so that I could use the data for the latest date available (not to mention that I had to change the date time format as well because only after that, any data cleaning was possible).

   - Also, I had to remove the columns that contained –'new cases', 'new deaths', 'cumulative deaths' and 'WHO region'- as all this data was redundant and useless for what I was about to do.

   - This lead to a new dataframe which contained 'date reported'(but only the latest date available), 'Country', ' Country code' and 'Cumulative cases'.

   - Now this dataframe was ready to be used for further data visualization. I started with installing Folium which I used for the purpose of making the choropleth map indicating the density of covid-19 spread in different countries.

- For the histogram, I imported numpy, matplotlib, and matplotlib.pyplot. The relevant data was fed and a histogram was visualized.

2. For the "**<u>Using Foursquare API to analyze the regions of Delhi</u>**" section:
   - I started with the most basic step, yes - importing all necessary libraries. In this case I imported 'requests', 'random', ('pandas' and 'numpy' were already imported), 'Nominatim' from geopy.geocoders, 'json_normalize' from pandas.io.json, and a couple of more of them whose description is given in the code itself.
   - No the next step included searching for a specific venue category (in this case I chose 'medical' which implies that I was actually searching for all medical stores and clinics in the corresponding region) and defining the corresponding URL.
   - The results were examined by sending the GET request.
   - Getting the relevant part of the json file, it was then converted into pandas dataframe for further use.
   - Next, we defined the information of interest and filtered the dataframe accordingly. Here, out of all the districts in the state of Delhi, we chose the district corresponding to index=2, i.e., New Delhi.
   - The data was visualized using the folium library.

Results and Discussion

1.   For the "**Covid-19 Data analysis as on 20ᵗʰ July 2020**" section:

Our first attempt of visualizing the data was done through a Choropleth map. This map was made using the Folium library. I used the custom.geo.json file for building this choropleth map that displayed the 'Covid-19 spread as on 20ᵗʰ July 2020' with respect to all the different countries across the world. I chose the color scheme of blue and purple and thus, following was the result:
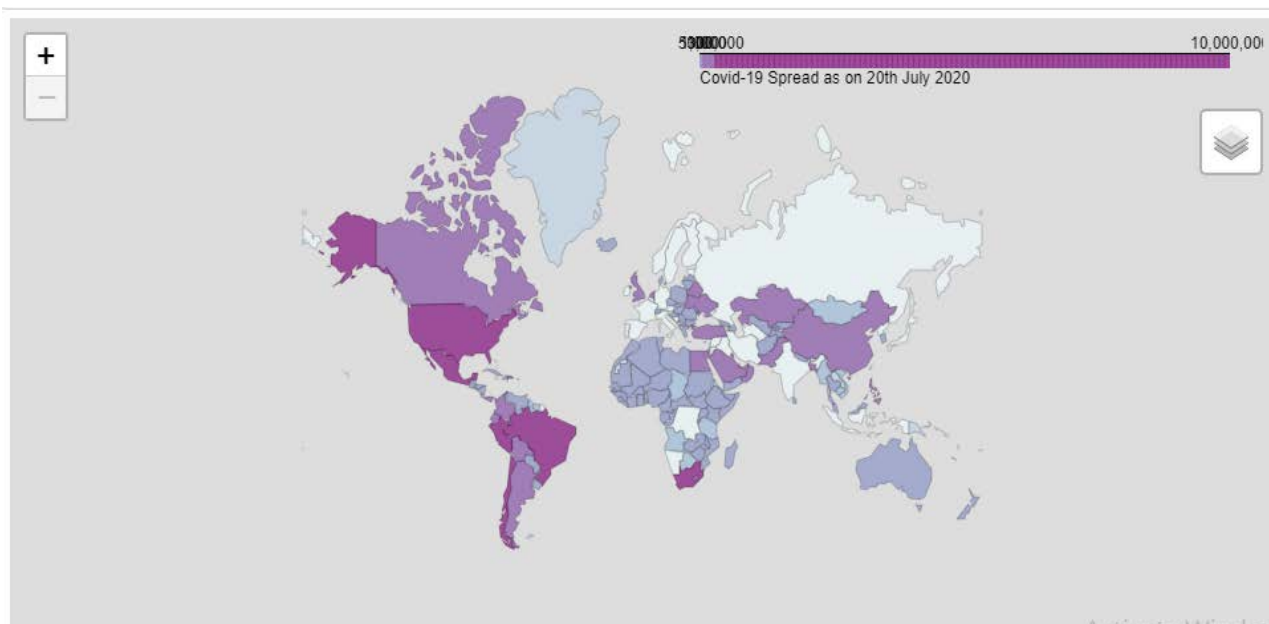

From the above map we ca see that certain regions such as Northern and Southern America are worst hit by the corona virus as compared to other countries. Canada is less affected than the USA, the Middle east being even less than Canada.

This choropleth map makes it visually easier for us to track the virus spread and its intensity in various regions. The WHO websites too uses a version of choropleth map on its Covid-19 dashboard for better data visualization and so it felt really good when as an amateur in data analytics, I was able to create something similar and useful.

Next comes a histogram that displays the number of countries vs number of cases plot.

Following id the histogram that I was able to make using the world covid-19 data:
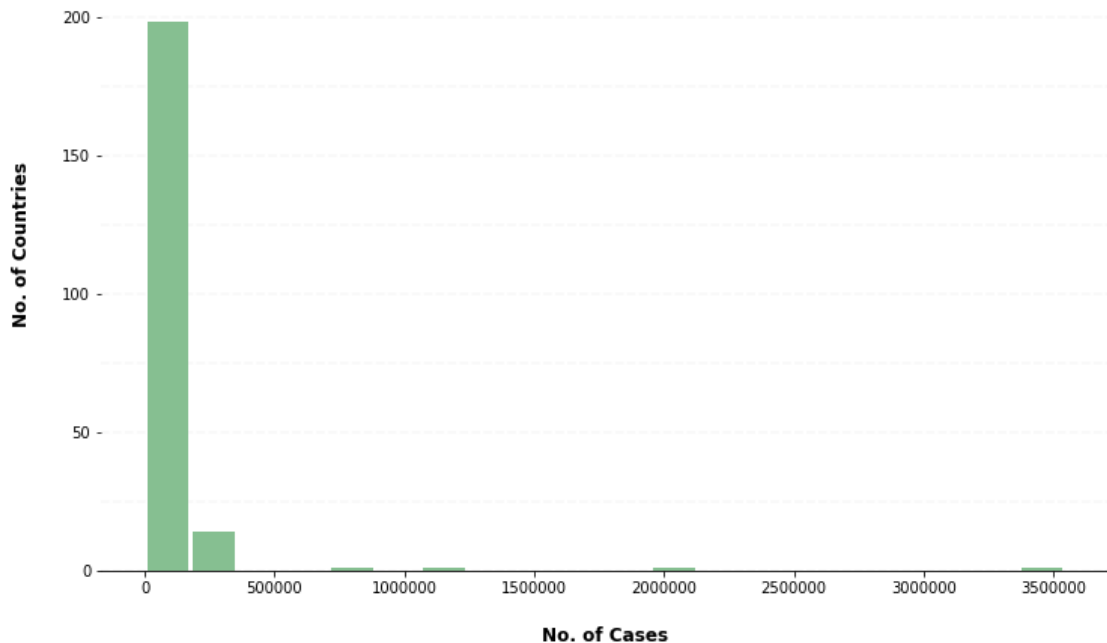

It is clearly visible from the above histogram that a majority of number of countries, i.e., around 200 countries have less cumulative cases (approximately 20,000 cases) as shown by the longest green bar, while there are certain (and luckily only a handful) countries show dangerously high number of cases (in a couple of cases it's as high as 35,00,000).

This data can be used in many ways to keep track of the countries and the number of cases of covid-19 in them. Similar plots such as line graphs and boxplots can be used to visualize different aspects of the same data. I used only histogram because I had the other part of the project waiting for me as well but I believe it would've been real fun to play with all this data and bring out different features and inferences out of it.

2. For the "**Using Foursquare API to analyze the regions of Delhi**" section:

This part was focused on using the Foursquare API. The steps of going through the entire process are given in detain in the notebook and methodology section so I would just skip to the result section. The Foursquare was used to visualize the the 'New Delhi' district of the state of Delhi. Below is the final foursquare map that I was able to bring out as a part of this project:
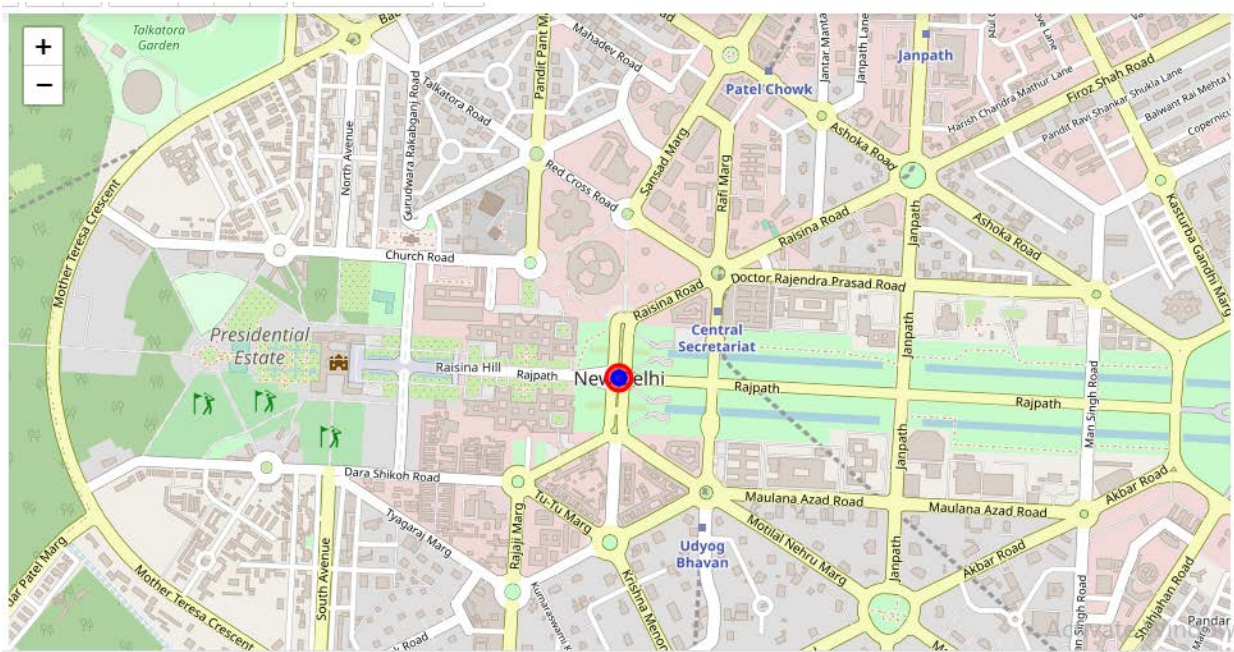
We can se how detailed and well displayed this map is. Foursquare is definitely a great platform to perform geospatial data analysis and there is a lot more that we can do and observe using this platform.

Conclusion

Since this project was not a client-based project, the conclusion is not what I can offer at this point. Still I would like to point out that the data visualization of Covid-19 data using various libraries was something worth doing. Though in this case you all already have the WHO website to go and see the data insights, but I believe what I learned through this project is going to be with me in magical ways. Who knows if someday I would come up with something absolutely original and that project would be useful to the world in its own way. And the Foursquare API was again something I would love to work with in future. The details and the features, and all the thousands of different things that we can possibly do with this platform are amazing.