

Assignment 3

Advanced Programming in Python(DSE 309)

COVID 19 Data Analysis and Visualization

Name: Ajay Choudhury

Roll no.: 18018

Date: 30th Oct 2021

Introduction

COVID 19 is an ongoing pandemic that is a matter of concern all over the world. Here COVID 19 data from the Worldometer is used to analyse and visualise the impact and change in COVID 19 cases and recovery all over the world in different countries through graphs.

These visualizations and plots were made using the matplotlib module in Python. The data set contains data pertaining to the COVID 19 pandemic such as cases, deaths and tests information for each country and continent in the world.

Data Cleaning

Data cleaning is an important process in any data analysis or machine learning project as removing incorrect, corrupted or irrelevant data helps improve the correctness or accuracy of a trained model or analysis.

After converting the fetched data to a dataframe in pandas, the continent data is removed as we wish to visualize only data from various countries. Now, for better visualization of the top 10 affected countries based on confirmed COVID 19 cases, the data is sorted in the dataframe itself.

Visualization

A. Line graph of Top 10 countries affected by COVID 19

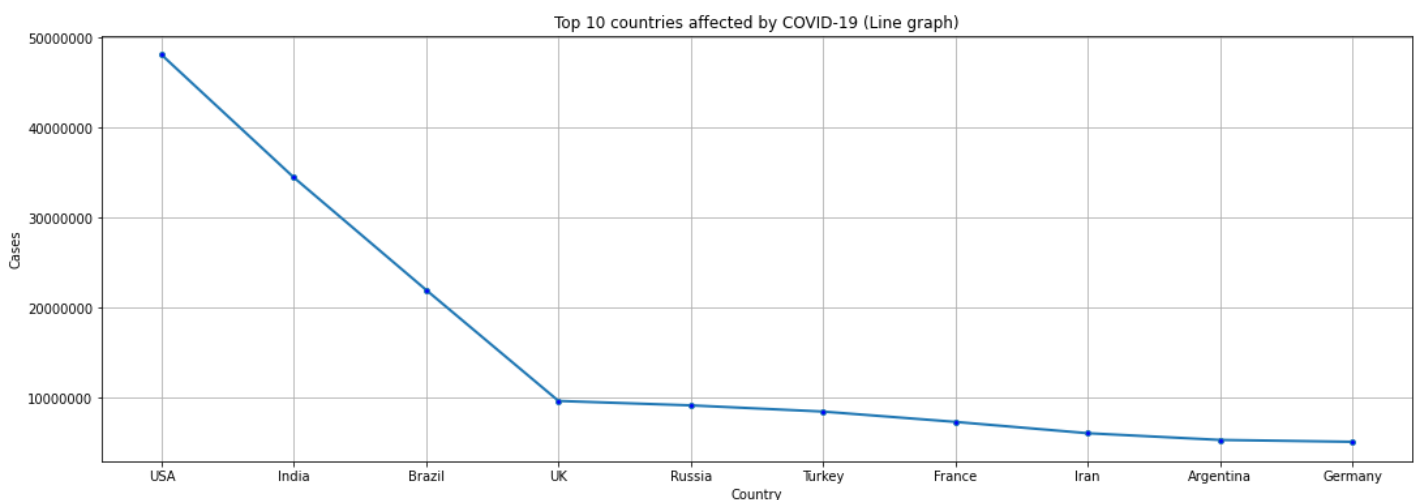


Fig. 1. Top 10 countries affected by Covid 19

This plot shows USA, India and Brazil are the seriously affected countries with the highest cases found. The other 7 countries in the list of top 10 affected countries are very close to each other and depict almost a straight horizontal line with cases below 10 million.

B. Bar graph of Top 10 countries affected by COVID 19

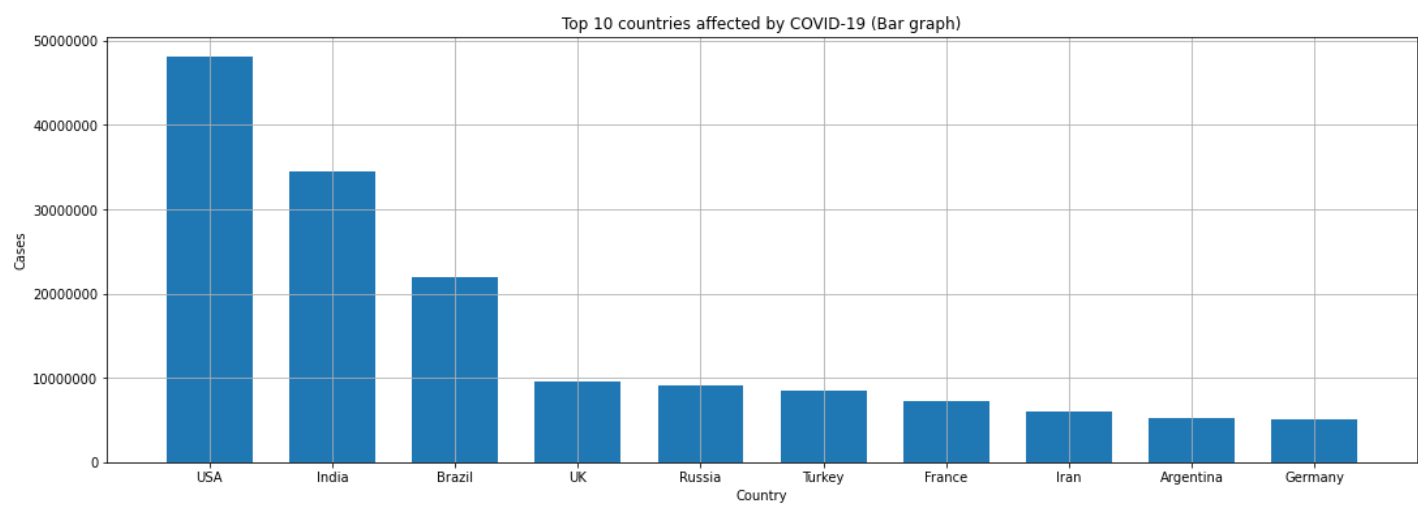


Fig. 2. Bar graph for data of Top 10 countries affected by Covid 19.

In this plot, statistics of the top 10 countries are depicted in a bar graph, similar to the line graph above.

C. Multiline graph of Top 10 countries affected by COVID 19

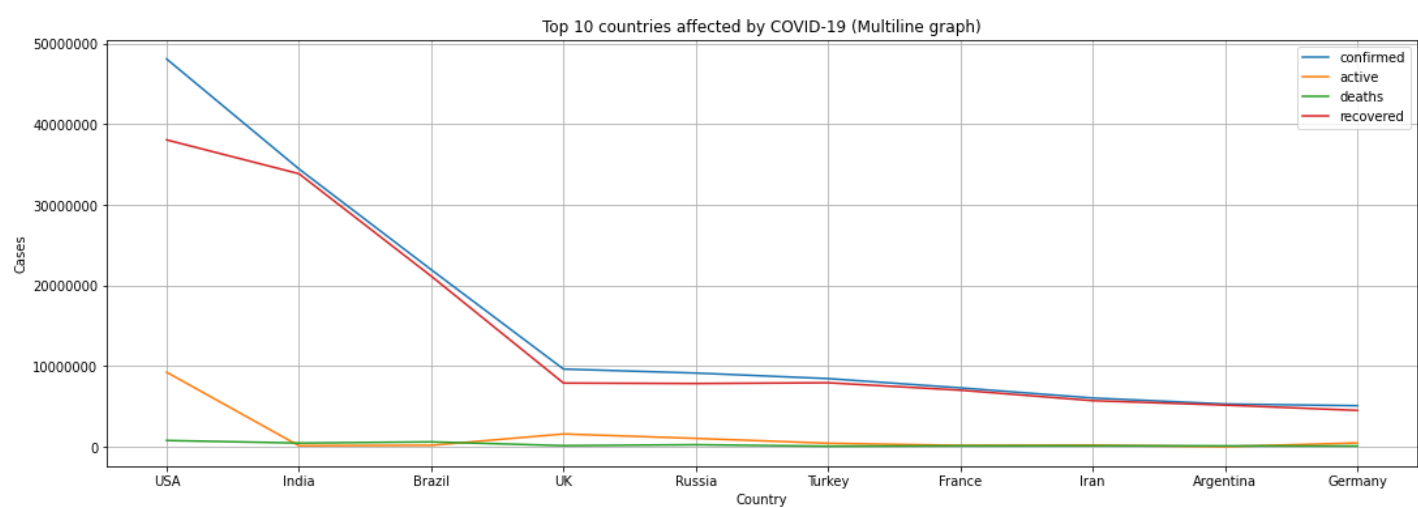


Fig. 3. Multiline graph for data of Top 10 countries affected by Covid 19.

The multiline plot shows the confirmed, active, recovered and death cases for each country in different colours. We can see how recovery and confirmed cases are high in countries with higher populations.

D. COVID 19 statistics of Top 10 countries (Multiple bar graph)

This plot shows multiple bar graphs for each country against confirmed, active, recovered and death cases. As depicted by the plot confirmed and recovered cases are in spikes whereas deaths and active cases are diminishing which shows our healthcare system has taken control of the situation recently.

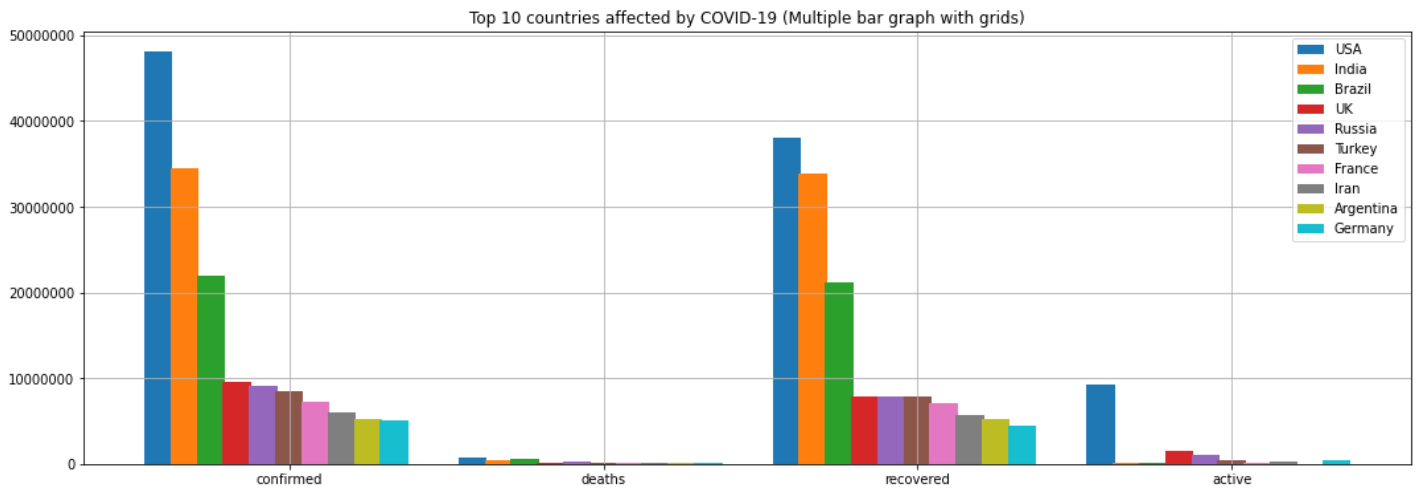


Fig. 4. Multiple bar graphs for statistics of Top 10 countries affected by Covid 19.

E. Scatter graph for statistics of Top 10 countries affected by COVID 19

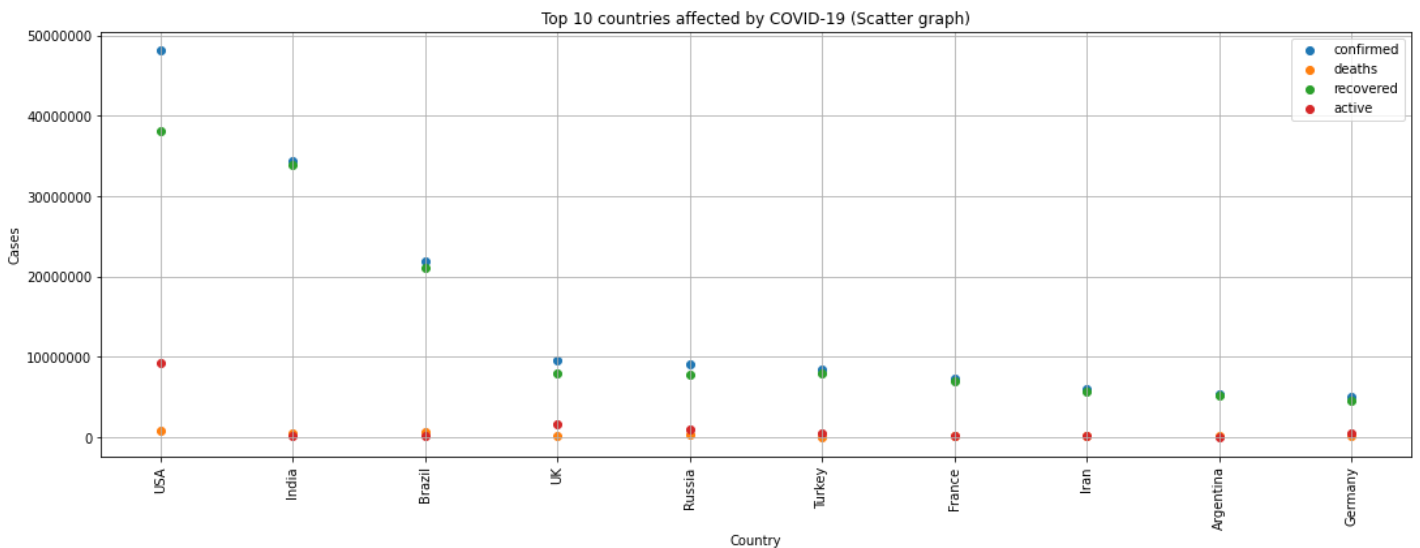


Fig.5. Scatter plot for statistics of Top 10 countries affected by Covid 19.

This plot is similar to the above multiline graph, it shows a dot for the number of confirmed, active, recovered and death cases for each country depicted with different colours (shown in legend).

F. Stacked-area graph for COVID 19 statistics of Top 10 countries

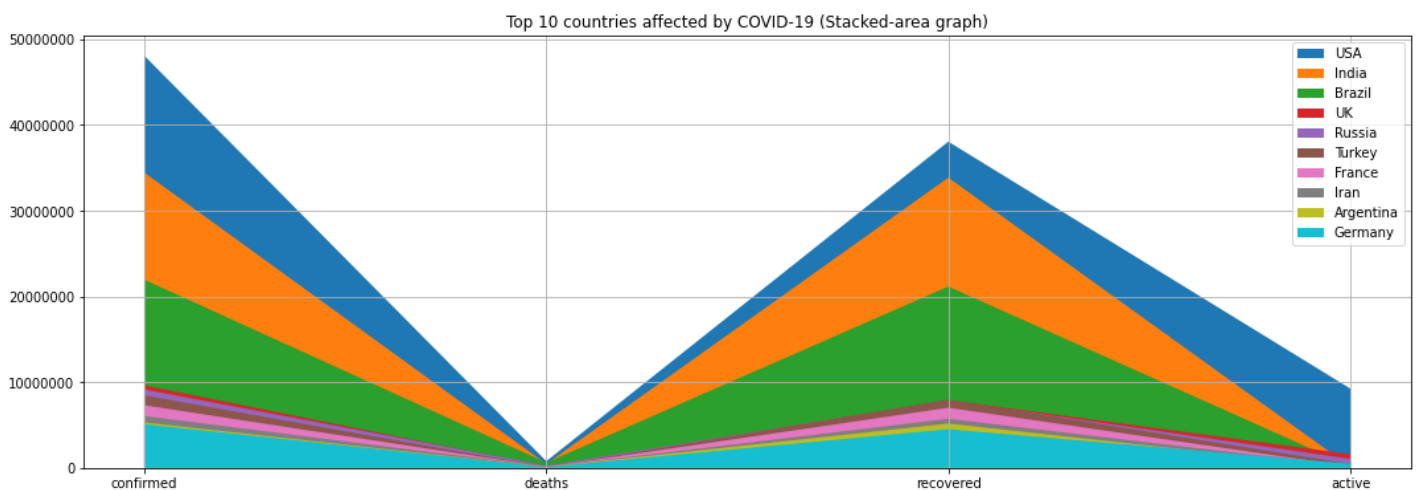


Fig.6. Stacked-area graph for statistics of Top 10 countries affected by Covid 19.

A Stacked-area graph better visualises the difference between the number of cases for each country. We can see the difference in the number of confirmed cases between the USA and India clearly. We can see how the number of recoveries is slightly greater than the number of confirmed cases in India, which is a positive sign of progress, whereas, the situation is opposite in the USA which is a matter of concern.

References

- [1] PyPI: <https://pypi.org/project/covid/>
- [2] Worldometer: <https://www.worldometers.info/coronavirus/>