# Covid19 - Country wise Analysis

# Introduction:

The analysis of COVID-19 data from a global perspective. We'll dive into the key aspects of data collection, exploration, visualization, and insights to better understand the pandemic's impact across different countries. Country-wise analysis datasets typically include a range of information such as:

* **Country/Region**: Name of the country or region.
* **Confirmed**: Total number of confirmed COVID-19 cases.
* **Deaths**: Total number of deaths due to COVID-19.
* **Recovered**: Total number of recovered COVID-19 cases.
* **Active**: Active cases (Confirmed - Deaths - Recovered).
* **New cases**: Number of new confirmed cases reported.
* **New deaths**: Number of new deaths reported.
* **New recovered**: Number of new recovered cases reported.
* **Deaths / 100 Cases**: Percentage of deaths among confirmed cases.
* **Recovered / 100 Cases**: Percentage of recovered cases among confirmed cases.
* **Deaths / 100 Recovered**: Percentage of deaths among recovered cases.
* **Confirmed last week**: Total confirmed cases reported in the previous week.
* **1 week change**: Change in confirmed cases compared to the previous week.
* **1 week % increase**: Percentage increase in confirmed cases compared to the previous week.
* **WHO Region**: World Health Organization (WHO) region to which the country belongs.

All data represents the date of reporting as opposed to date of symptom onset. All data is subject to continuous verification and may change based on retrospective updates to accurately reflect trends, changes in country case definitions and/or reporting practices. Significant data errors detected or reported to WHO may be corrected at more frequent intervals with some countries performing retrospective bulk corrections may lead to the appearance of significant spikes or negative values which are verified and validated by WHO.

**PROBLEM DEFINITION**

Since the pandemic has started, some countries are facing problem of ever-increasing cases. Through the data analysis of cases one can analyze how countries all over the world are doing in terms of controlling the pandemic. Analyzing data leads to accommodate the prevention model of the courtries that are doing great in terms of lowering the graph. Predictions are made with the dataset which available to the individual/country/associations, therefore helping them to decide how far they are able to control the pandemic or up to how much extent they should guide preventive measures.

Through the project, a step towards helping people to understand the spread and predict the cases in their country has been done. The project also gives a vision of how a country is doing in terms of limiting the spread.

Data Manipulation:

* We are building our own covid prediction
* System using jupyter notebook.

**Cleaning the dataset:**

The very first step in our project is to get a reliable and authentic dataset for the prediction and analysis. Our search for dataset ended on which is kaggle which has provided Dataset for free use and is absolutely authentic.

**Data Visualization:**

Here, we use the dataset and check the consistency of the dataset by checking the values out of the dataset randomly.

Then we do data visualization for better understanding of data by the use of various plots, graph and heat maps. All this graphs and plots gets us an insight into huge datasets easily.

**Insights:**

* US and Brazil are most number of confirmed cases.
* Qatar has 109597 confirmed cases, out of that 106328 people are recovered.
* HOLY SEE, Grenada, Dominica countries have 100% Recovery rate.
* The number of Deaths correlates strongly with Confirmed cases (0.93), indicating that regions with more confirmed cases tend to have higher numbers of deaths.
* Deaths also correlate strongly with Active cases (0.87), suggesting regions with higher active cases often have more deaths.
* There is a strong correlation between Recovered and Confirmed cases (0.91), indicating that regions with more confirmed cases tend to also have more recoveries.
* **Active cases correlate moderately with Recovered cases (0.68), suggesting that regions with more active cases also tend to have more recoveries, though not as strongly correlated as with deaths.**

**Statistical analysis:**

### ONE-WAY ANOVA USING SCIPY.STATS

* **Null hypothesis: means of the groups are equal**
* **Alternative hypothesis**: means of the groups are not equal

The p-value (0.0812) is greater than 0.05, suggesting that **we do not have sufficient evidence to reject the null hypothesis** at the 0.05 significance level. Therefore, based on this ANOVA test, we cannot conclude with statistical significance that there are differences in the mean number of 'Confirmed' cases across the WHO regions listed.

The p-value (0.0571) is greater than 0.05, suggesting that **we do not have sufficient evidence to reject the null hypothesis at the 0.05** significance level. Therefore, based on this ANOVA test, we cannot conclude with statistical significance that there are differences in the mean number of 'Confirmed' cases across the WHO regions listed.

The p-value (0.0678) is greater than 0.05, suggesting that **we do not have sufficient evidence to reject the null hypothesis at the 0.05** significance level. Therefore, based on this ANOVA test, we cannot conclude with statistical significance that there are differences in the mean number of 'Confirmed' cases across the WHO regions listed.

The p-value (0.228) is greater than 0.05, suggesting that **we do not have sufficient evidence to reject the null hypothesis at the 0.05 significance level**. Therefore, based on this ANOVA test, we cannot conclude with statistical significance that there are differences in the mean number of 'Confirmed' cases across the WHO regions listed.

**Conclusion:**

I learned how to use Python and Pandas to analyze COVID-19 data. We loaded the data, checked for missing values, and performed various analyses to understand the pandemic’s impact. We calculated the total number of confirmed, death, and recovered cases mainly in WHO region, identified regions with the highest and lowest cases. Additionally, we extracted and sorted the data based on confirmed and recovered cases, respectively. Analyzing COVID-19 data helps us gain valuable insights and make informed decisions during these challenging times.