**Table of Contents**

* [Project Overview](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Project-Overview)
* [Data Source](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Data-Source)
* [Tools Used](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Tools-Used)
* [Data Preparation and Data Cleaning](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Data-Preparation-and-Data-Cleaning-Process-in-MySQL-Workbench)
* [Data Analysis](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Questions-Answered-in-Data-Analysis-Process-in-MySQL-Workbench)
* [Visualization in Power BI](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Visualization-in-Power-BI)
* [Findings](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Findings-from-the-Analysis)
* [Limitations](https://github.com/Bhushan0403/HR_Analytics_SQL_Project/blob/main/README.md#Limitations)

# **COVID-19 Global Data Analysis Project**

## Project Overview

This project aims to analyze the global impact of COVID-19 by examining cumulative cases, deaths, and other relevant metrics across different countries and WHO regions. Utilizing a dataset provided by the World Health Organization (WHO), we derive insights on the spread and severity of the pandemic to inform public health decisions and strategies.

### About the Data

The dataset, "WHO-COVID-19-global-table-data," consists of 12 columns and 239 rows, providing comprehensive information on COVID-19 statistics across various countries and WHO regions.

* **Columns**:
  1. **Name**: The name of the country or region.
  2. **WHO Region**: The World Health Organization region to which the country belongs.
  3. **Cases - cumulative total**: The total number of COVID-19 cases reported cumulatively.
  4. **Cases - cumulative total per 100000 population**: The cumulative total of COVID-19 cases per 100,000 population.
  5. **Cases - newly reported in last 7 days**: The number of new COVID-19 cases reported in the last 7 days.
  6. **Cases - newly reported in last 24 hours**: The number of new COVID-19 cases reported in the last 24 hours.
  7. **Deaths - cumulative total**: The total number of COVID-19 deaths reported cumulatively.
  8. **Deaths - cumulative total per 100000 population**: The cumulative total of COVID-19 deaths per 100,000 population.
  9. **Deaths - newly reported in last 7 days**: The number of new COVID-19 deaths reported in the last 7 days.
  10. **Deaths - newly reported in last 24 hours**: The number of new COVID-19 deaths reported in the last 24 hours.
  11. **Transmission Classification**: The classification of the transmission type (e.g., community transmission, clusters of cases).
  12. **Population**: The total population of the country or region.
* **Rows**:
  1. The dataset contains 239 rows, each representing a specific country or region and providing detailed COVID-19 statistics.

### **Data Source**

The primary dataset utilized for this analysis is sourced from the "WHO-COVID-19-global-table-data Project.csv" file, comprising comprehensive COVID-19 case and death statistics essential for this analysis project. ====The dataset includes detailed information on cumulative total cases, cumulative total deaths, newly reported cases and deaths in the last 7 days, and cases and deaths per 100,000 population for various countries and WHO regions. This data is crucial for understanding the global impact of the COVID-19 pandemic and informing public health strategies.====

### Tools Used

* **SQL Server Management Studio (SSMS)**: For querying and analyzing the data.
* **Power BI**: For transforming and visualizing the outcomes of the analysis.
  + **Data Transformation**: Changed the data type to decimal number to ensure accurate calculations.

### Data Cleaning Process in Excel

### The data cleaning process involved several steps to ensure the dataset was ready for analysis:

1. **Removed the Global Column**: The global column was removed to focus on country-specific and regional data.
2. **Added Commas to Columns**: Commas were added to the columns (except for the Name and WHO\_REGION columns) to enhance readability.
3. **Filled Empty Spaces with Zero**: All empty spaces were filled with zero to prevent calculation errors.
4. **Changed Data Type**: In Power BI, the data type of relevant columns was changed to decimal number using Power Query to ensure accurate calculations.

#### Database Creation and Query Execution

In SQL Server Management Studio (SSMS), a database named WHO-COVID-19 was created to organize and manage the COVID-19 data analysis. Within this database, various queries were executed to answer specific research questions related to the global impact of COVID-19. The process involved inserting the questions into the database and writing the corresponding SQL queries to derive the necessary insights.

### Questions Answered in the Data Analysis Process

1. Total cumulative cases and deaths globally
2. Total cumulative cases and deaths by WHO region
3. Cumulative Number of Cases and Deaths by Country
4. Top 10 Countries with the Highest Cumulative Total of COVID-19 Cases
5. Top 10 Countries with the Highest Cumulative Total of COVID-19 Deaths
6. Top 5 Countries with the Lowest Cumulative Total of COVID-19 Cases per 100,000 Population
7. Top 5 Countries with the Lowest Cumulative Total of COVID-19 Deaths per 100,000 Population
8. Total Cases Reported in the Last 24 Hours Globally
9. Total Deaths Reported in the Last 24 Hours Globally
10. Top 5 Countries with the Lowest Cases Reported in the Last 24 Hours
11. Top 5 Countries with the Lowest Deaths Reported in the Last 24 Hours
12. Top 10 countries with the highest cases in the last 7 days
13. Top 10 countries with the highest deaths in the last 7 days
14. Average Number of Cases and Deaths
15. Top 5 Countries with the Highest Average Number of Cases
16. Top 5 Countries with the Highest Average Number of Deaths

*Note: Data analysis SQL files are attached for reference.*

Visualization in Power BI

The excel file was exported to Power BI to create a visually appealing dashboard.

The primary purpose of this Power BI dashboard was to enhance understanding of the outcomes derived from the SQL data analysis. It’s important to note that the Power BI dashboard is interactive; It will enhance user engagement by enabling them to explore and analyze data from multiple perspectives and dimensions.

### Findings from the Analysis

1. Total Cumulative Cases and Deaths Globally:

- The dataset reveals the total number of cumulative COVID-19 cases and deaths reported globally. This provides an overall picture of the pandemic's impact worldwide.

2. Total Cumulative Cases and Deaths by WHO Region:

- The data is segmented by WHO regions, allowing for an understanding of the regional distribution of COVID-19 cases and deaths. This helps identify which regions are most affected.

3. Cumulative Number of Cases and Deaths by Country:

- The dataset provides country-specific data on the cumulative number of cases and deaths. This helps in identifying the countries with the highest and lowest numbers, giving a clear view of the global distribution.

4. Top 10 Countries with the Highest Cumulative Total of COVID-19 Cases:

- The analysis identifies the top 10 countries with the highest cumulative total of COVID-19 cases. This highlights the countries that have experienced the most significant outbreaks.

5. Top 10 Countries with the Highest Cumulative Total of COVID-19 Deaths:

- Similar to the previous point, this analysis identifies the top 10 countries with the highest cumulative total of COVID-19 deaths. This provides insight into the countries with the highest mortality rates.

6. Top 5 Countries with the Lowest Cumulative Total of COVID-19 Cases per 100,000 Population:

- This metric normalizes the cumulative total of COVID-19 cases by the population size, identifying the countries with the lowest case rates per 100,000 people. This helps in understanding the relative impact of the pandemic on smaller populations.

7. Top 5 Countries with the Lowest Cumulative Total of COVID-19 Deaths per 100,000 Population:

- Similar to the previous metric, this analysis identifies the countries with the lowest death rates per 100,000 people, providing a relative measure of mortality.

8. Total Cases Reported in the Last 24 Hours Globally:

- The dataset includes data on the total number of new COVID-19 cases reported globally in the last 24 hours, highlighting the current rate of new infections.

9. Total Deaths Reported in the Last 24 Hours Globally:

- This metric provides data on the total number of new COVID-19 deaths reported globally in the last 24 hours, offering insight into the current mortality rate.

10. Top 5 Countries with the Lowest Cases Reported in the Last 24 Hours:

- The analysis identifies the top 5 countries with the lowest number of new cases reported in the last 24 hours, helping to understand which countries have the most controlled outbreaks.

11. Top 5 Countries with the Lowest Deaths Reported in the Last 24 Hours:

- This metric identifies the top 5 countries with the lowest number of new deaths reported in the last 24 hours, highlighting those with the most effective measures in reducing mortality.

12. Top 10 Countries with the Highest Cases in the Last 7 Days:

- The analysis highlights the top 10 countries with the highest number of new cases reported in the last 7 days, providing a view of recent surges.

13. Top 10 Countries with the Highest Deaths in the Last 7 Days:

- This metric identifies the top 10 countries with the highest number of new deaths reported in the last 7 days, offering insight into recent increases in mortality.

14. Average Number of Cases and Deaths:

- The dataset provides the average number of cases and deaths across all countries, offering a global perspective on the pandemic's impact.

15. Top 5 Countries with the Highest Average Number of Cases:

- The analysis identifies the top 5 countries with the highest average number of COVID-19 cases, highlighting those with consistently high infection rates.

16. Top 5 Countries with the Highest Average Number of Deaths:

- This metric identifies the top 5 countries with the highest average number of COVID-19 deaths, providing insight into countries with consistently high mortality rates.