# COVID-19 Analysis Report

Project Title: COVID-19 Data Analysis and Visualization

Tools: A Comprehensive Analysis Using Excel and Power BI

Submitted by: Mumtaj

Date: [Submission Date]

**🔷 Introduction**

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, had a profound global impact on public health, economies, and daily life. Governments, researchers, and the public needed timely, accurate, and insightful data to make informed decisions. Analyzing the global spread, mortality, and recovery trends through reliable datasets helps understand the pandemic’s trajectory and impact across countries and over time.

This project uses publicly available COVID-19 datasets to explore key trends, identify patterns, and create interactive dashboards that offer clear and meaningful insights into the pandemic.

**🔷 Problem Statement**

Although vast COVID-19 data is available globally, it is often underutilized due to its raw format and complexity. Policymakers, healthcare professionals, and the general public require clean, visual, and actionable insights to understand:

* How the virus spread over time
* Which countries were most affected
* How recovery and death rates vary by region
* When peaks occurred

The lack of accessible analytics limits quick response and public understanding.

**🔷 Objectives**

The main goals of this project are:

1. **Clean and preprocess** country-wise and day-wise COVID-19 data
2. **Identify and manage outliers** in the data using statistical methods (IQR)
3. **Analyze trends** in confirmed cases, deaths, recoveries, and active cases
4. **Visualize global and country-wise data** using charts, KPIs, and maps
5. Provide clear **comparisons between countries and regions**
6. Track **daily trends** and identify key turning points during the pandemic
7. Create a dashboard or summary report that can support **decision-making and awareness**

**🔷 About the Dataset**

This project utilizes two structured datasets related to the global COVID-19 pandemic. These datasets provide comprehensive information on confirmed cases, deaths, recoveries, and active cases, both at the country level and over time. They serve as the foundation for performing statistical analysis, visualizations, and trend identification.

**📁 1. country\_wise\_latest.csv**

This dataset presents the most recent cumulative data on COVID-19 cases for each country or region. It is useful for comparing how different countries are affected and understanding key ratios like recovery and death rates.

**Important columns:**

* Country/Region: Name of the country or region
* Confirmed: Total confirmed cases
* Deaths: Total deaths reported
* Recovered: Total recoveries reported
* Active: Current active cases
* New cases, New deaths, New recovered: Newly reported figures since last update
* Deaths / 100 Cases: Percentage of confirmed cases that resulted in death
* Recovered / 100 Cases: Percentage of confirmed cases that have recovered
* WHO Region: Geographical region classification by WHO

💡 **Use in project**: Helps compare country-wise severity, death/recovery rates, and regional trends.

**📁 2. day\_wise.csv**

This dataset provides a chronological, global view of the pandemic with daily cumulative totals. It supports time series analysis, helping identify trends, peaks, and changes in COVID-19 metrics over time.

**Important columns:**

* Date: Date of observation
* Confirmed, Deaths, Recovered, Active: Cumulative global counts
* New cases, New deaths, New recovered: Daily increase in numbers
* Deaths / 100 Cases, Recovered / 100 Cases: Calculated mortality and recovery rates
* No. of countries: Number of countries contributing data on a given date

💡 **Use in project**: Useful for tracking daily case trends, analyzing pandemic waves, and identifying periods of spikes or improvements.

**🔧 Data Cleaning and Preprocessing**

Before performing any analysis or visualization, it is essential to clean and preprocess the data to ensure accuracy and reliability. The following steps were taken to prepare the country\_wise\_latest.csv and day\_wise.csv datasets:

**1. Handling Missing Values**

Both datasets were checked for missing values across all columns:

* In the **country-wise dataset**, no missing values were found in any of the numeric columns such as Confirmed, Deaths, Recovered, Active, etc.
* Similarly, the **day-wise dataset** also did not contain missing values in key columns like Confirmed, Deaths, Active, and New cases.

Thus, no imputation or removal was needed for missing data.

**2. Detecting Outliers using the IQR Method**

Outliers were identified in multiple numeric columns using the Interquartile Range (IQR) method. This method calculates outliers as values falling below:

Lower Bound = Q1 - 1.5 \* IQR

Upper Bound = Q3 + 1.5 \* IQR

**📊 Country-Wise Dataset Outliers:**

Outliers were detected in several columns, such as:

| **Column** | **Outliers Found** | **Upper Bound** |
| --- | --- | --- |
| Confirmed | 24 | 99,480.25 |
| Deaths | 33 | 1,807.25 |
| Recovered | 25 | 55,575.25 |
| Active | 32 | 22,660.25 |
| New Cases | 25 | 1,042.75 |
| New Recovered | 31 | 552.5 |
| Deaths / 100 Cases | 10 | 8.27 |
| Deaths / 100 Recovered | 19 | 13.42 |
| 1 Week Change | 22 | 7,856.5 |

**Action Taken:** All detected outliers were **capped at the upper bound** to minimize their impact without completely removing them from the analysis.

**📈 Day-Wise Dataset Outliers:**

Fewer outliers were found here, mainly in:

| **Column** | **Outliers Found** | **Upper Bound** |
| --- | --- | --- |
| Recovered | 6 | 8,450,327.5 |
| New Recovered | 3 | 195,533.25 |
| Deaths / 100 Recovered | 15 | 48.88 |

All other key metrics such as Confirmed, Deaths, Active, New Cases, and New Deaths showed **no outliers**.

**Action Taken:** Outliers were capped at the calculated upper bounds to preserve data consistency while reducing skewness.

These preprocessing steps ensured that the data is clean, consistent, and ready for meaningful analysis and visualization.

**📊 Exploratory Data Analysis (EDA)**

The goal of Exploratory Data Analysis (EDA) is to understand the structure of the data, uncover patterns, spot anomalies, and generate meaningful insights that support decision-making. EDA was conducted separately for both the **country-wise** and **day-wise** datasets.

**📊 Country-wise Analysis – COVID-19 Data (from country\_wise\_latest.csv)**

**1️⃣ Top 10 Countries by Confirmed Cases**

**Chart Type**: Horizontal Bar Chart  
**What It Shows**:  
This chart displays the **top 10 countries** with the highest number of **confirmed COVID-19 cases**.

**Key Insight**:

* The **United Kingdom** has the highest confirmed cases, followed by **India**, **Brazil**, and **Russia**.
* These countries were among the most severely impacted during the global waves of COVID-19.

**2️⃣ Top 10 Countries by Deaths**

**Chart Type**: Bar Chart  
**What It Shows**:  
This chart presents the countries with the highest **COVID-19-related death tolls**.

**Key Insight**:

* The **United States** had the highest number of deaths, followed by **Brazil** and **Mexico**.
* High death counts often reflect overwhelmed healthcare systems and delayed response measures.

**3️⃣ Top 10 Countries by Recovered Cases**

**Chart Type**: Bar Chart  
**What It Shows**:  
This chart highlights the countries with the most **total recovered cases**.

**Key Insight**:

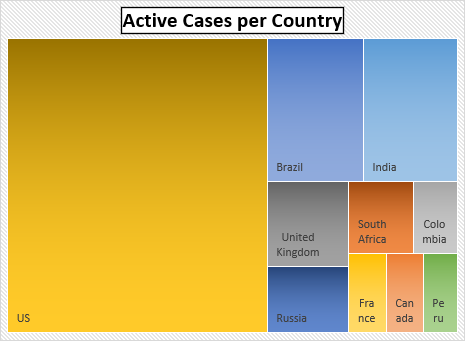
* **Brazil** and the **US** lead in the number of recoveries, followed by **India** and **Russia**.
* High recovery figures are usually proportional to high case counts and successful medical response.

**4️⃣ Active Cases per Country**

**Chart Type**: Treemap  
**What It Shows**:  
This visualization shows the countries with the **most active (currently infected) cases**,

**Key Insight**:

* The **US** dominates the active case count, indicating ongoing pressure on healthcare.
* **Brazil**, **India**, and the **UK** also carry large active burdens.



**5️⃣ Global COVID-19 Summary (Confirmed, Deaths, Recovered)**

**Chart Type**: Pie Chart  
**What It Shows**:  
A **global-level breakdown** of total confirmed, deaths, and recovered cases.

**Key Insight**:

* About **62% of cases were confirmed**, **36% recovered**, and **2% resulted in deaths**.
* This gives a high-level view of the overall pandemic outcome across the world.

**6️⃣ Death Rate vs Recovery Rate by Country**

**Chart Type**: Scatter Plot  
**What It Shows**:  
A comparison of **Death Rate** and **Recovery Rate** for each country.  
Formulas used:

* Recovery Rate = Recovered / Confirmed
* Death Rate = Deaths / Confirmed

**Key Insight**:

* Most countries have **high recovery rates** (closer to 1.0) and **low death rates**.
* A few outliers exist with relatively higher death rates, suggesting healthcare or crisis management challenges.

📆 **Day-wise COVID-19 Analysis (from day\_wise.csv)**

**1️⃣ Total Confirmed Cases Over Time**

**Chart Type:** Line Chart  
**What It Shows:**  
This graph shows the increasing trend of total confirmed COVID-19 cases from January to July 2020.

**Key Insight:**  
There was a very sharp rise in total confirmed cases after March 2020. The curve shows exponential growth, reaching over **17 million cases** by July.

**2️⃣ Confirmed vs Deaths vs Recovered Over Time**

**Chart Type:** Multi-Line Chart  
**What It Shows:**  
This chart compares the trend of **confirmed**, **deaths**, and **recovered** cases over time.

**Key Insight:**

* Confirmed cases increased the fastest.
* Recoveries also grew significantly, showing improvement in treatment and recovery.
* The number of deaths remained much lower than recoveries, showing that the **death rate was relatively low** compared to the number of infections.

**3️⃣ Daily New Confirmed Cases**

**Chart Type:** Bar Chart (Monthly Aggregation)  
**What It Shows:**  
This bar chart shows the number of new confirmed cases each month.

**Key Insight:**

* Cases were very low in January and February.
* A big rise started in March and continued growing each month.
* July had the **highest number of new cases**, indicating the pandemic was still spreading rapidly.

**4️⃣ Active Cases Trend Over Time**

**Chart Type:** Area Chart  
**What It Shows:**  
This chart displays the number of **active cases over time**, calculated as:

Active = Confirmed - Deaths - Recovered

**Key Insight:**  
Active cases rose steadily, peaking around July 2020. This shows ongoing stress on the healthcare system, as many people were still infected and under treatment.

**5️⃣ Recovery Rate and Death Rate Trend**

**Chart Type:** Line Chart  
**What It Shows:**  
This chart tracks how the **recovery rate** and **death rate** changed over time.

**Formulas Used:**

* Recovery Rate = Recovered / Confirmed
* Death Rate = Deaths / Confirmed

**Key Insight:**

* The recovery rate improved over time and reached above **55%** by July.
* The death rate remained low and mostly below **5%**, which shows that most patients survived the infection.

**📍 Conclusion**

This project provided a detailed analysis of the COVID-19 pandemic using both country-wise and day-wise data. The analysis showed that the virus spread rapidly after March 2020, with a major rise in confirmed cases globally, peaking around July. Countries like the United States, Brazil, India, and the United Kingdom were among the most affected in terms of total cases, deaths, and active cases.

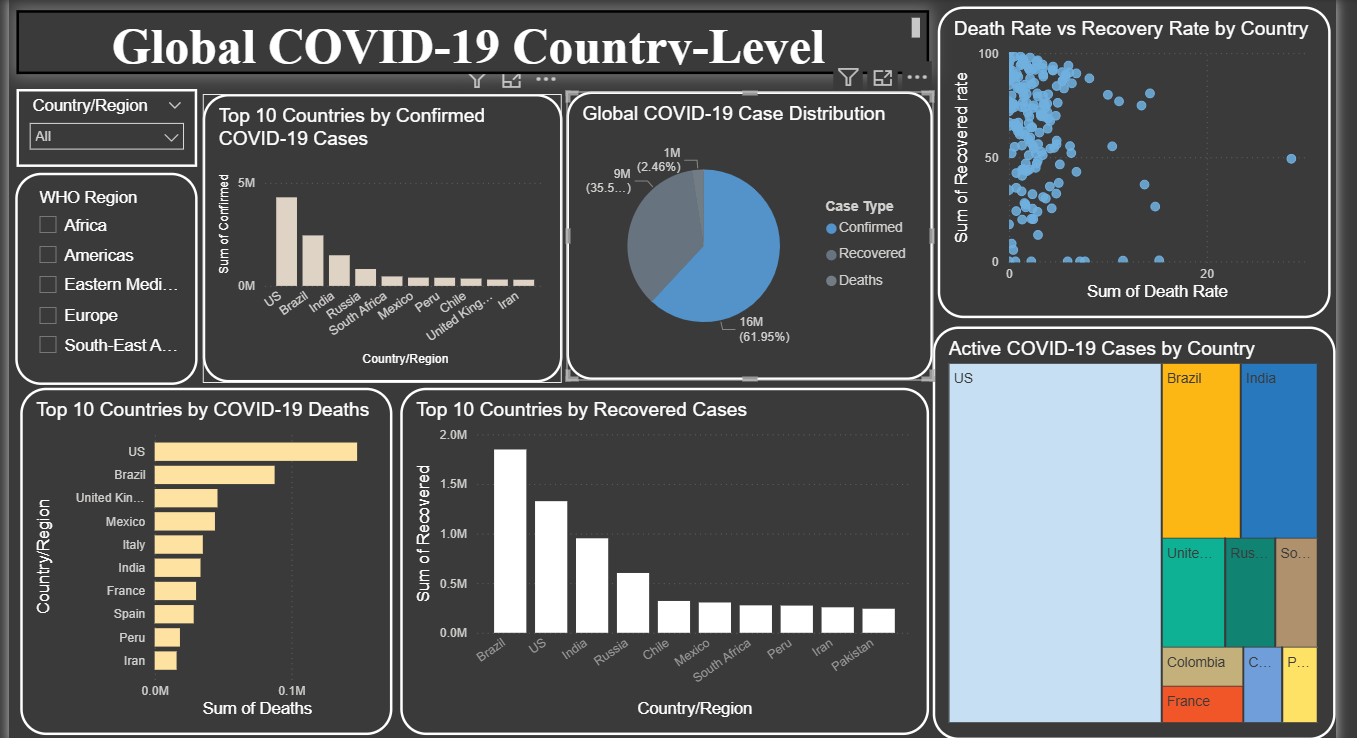
Despite the large number of confirmed cases, recovery rates improved steadily over time, and the overall global death rate remained below 5%. Countries with stronger healthcare systems managed to achieve higher recovery rates. The visualizations helped identify important patterns such as active case burdens, daily new spikes, and differences in recovery and death trends across regions.

This analysis not only helps in understanding the impact of the pandemic but also highlights the importance of timely data reporting, public health readiness, and international cooperation during global health crises.

**Power BI Dashboard**

To effectively communicate insights from the COVID-19 datasets, two interactive dashboards were developed using **Power BI**. These dashboards visually represent the key metrics, trends, and country-wise comparisons using dynamic charts, KPIs, and slicers.

**🗺️ 1. Country-Wise COVID-19 Dashboard**

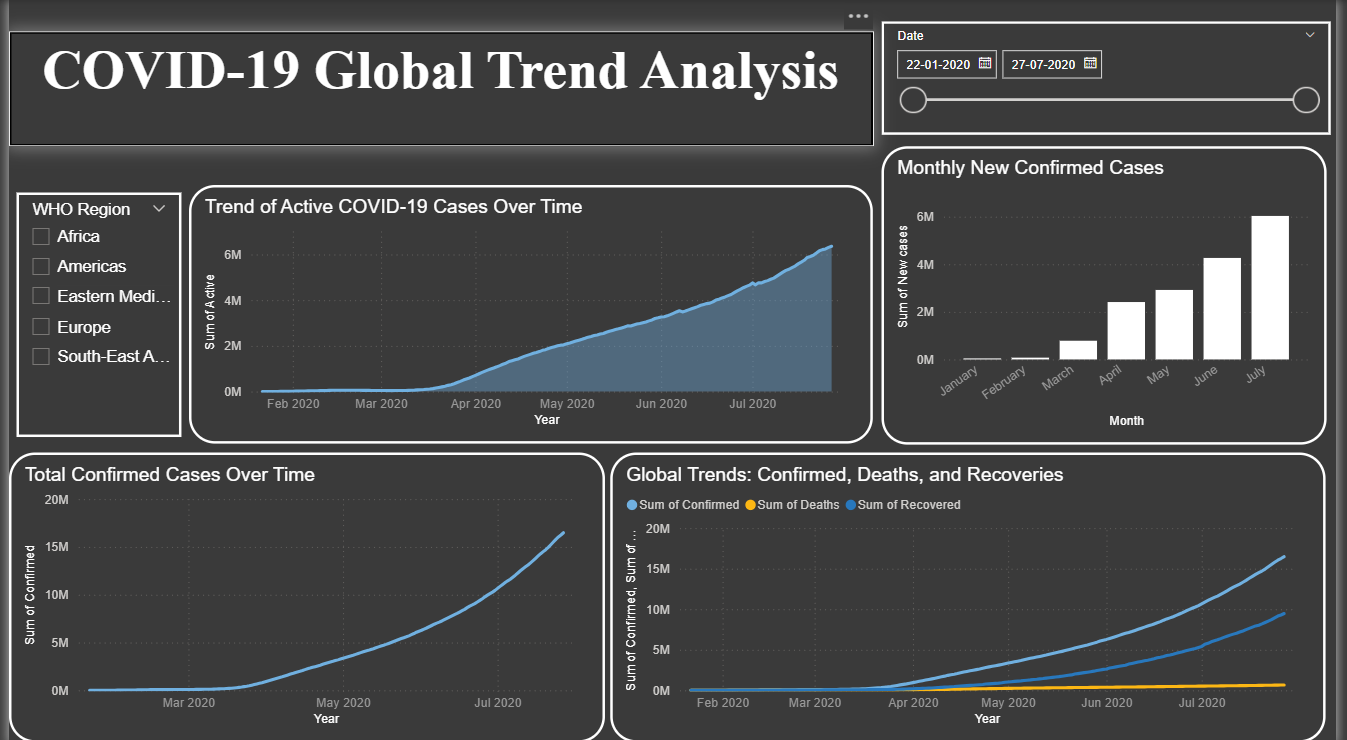
****

This dashboard focuses on the global distribution of COVID-19 cases at the country level. It includes:

* **Top 10 Countries by Confirmed Cases**: A bar chart showing the highest confirmed case counts.
* **Top 10 by Deaths and Recovered**: Charts identifying the countries with the most deaths and recoveries.
* **Active Cases Treemap**: Displays the countries with the highest number of active cases.
* **Global Case Distribution (Donut Chart)**: Summarizes confirmed, recovered, and death cases worldwide.
* **Scatter Plot (Death Rate vs Recovery Rate)**: Highlights variations in country outcomes.
* **Interactive Filters**: Slicers by Country/Region and WHO Region allow users to explore specific areas.

📌 **Insight**: The dashboard shows that countries like the US, Brazil, and India had the highest confirmed and active cases, while most countries had higher recovery rates compared to death rates.

**📈 2. Time-Series COVID-19 Dashboard**

****

This dashboard focuses on the progression of the pandemic over time using the day-wise dataset. It includes:

* **Total Confirmed Cases Over Time** (Line Chart): Shows the exponential rise in confirmed cases.
* **Confirmed vs Deaths vs Recovered** (Multi-Line Chart): Displays how each metric changed over time.
* **Monthly New Confirmed Cases** (Bar Chart): Highlights the peak months of the pandemic.
* **Active Cases Over Time** (Area Chart): Tracks the number of people actively infected over time.

📌 **Insight**: The time-series analysis revealed that cases surged significantly after March 2020 and peaked around July. Recovery rates improved steadily, and death rates remained relatively low.

**🛠 Tools Used:**

* **Power BI Desktop**
* Charts: Bar, Line, Area, Donut, Treemap, Scatter
* Filters: Slicers for Date, WHO Region, and Country

**✅ Conclusion:**

These dashboards provide a visual and interactive way to understand the pandemic’s global spread and progression. They support data-driven decision-making and make it easier to identify trends, peak periods, and regional differences.