

# Report on COVID-19 Global Statistics

**Student Name:** Dasari Ramesh

**Student ID:**23038360

**Git hub link:** <https://github.com/rameshpc9/Ramesh>

## Introduction

The COVID-19 pandemic, instigated by the novel coronavirus SARS-CoV-2, has emerged as a global crisis, profoundly impacting healthcare systems, economies, and societal structures since its inception in late 2019. In response, the meticulous collection and analysis of epidemiological data have become imperative for comprehending the pandemic's trajectory and devising effective mitigation strategies. The COVID-19 Global Statistics Dataset serves as a valuable resource, offering insights into vital metrics such as total cases, deaths, recoveries, testing rates, and population demographics across diverse countries. In this report, we conduct an exhaustive analysis of this dataset to unveil trends, disparities, and correlations in COVID-19 outcomes. These findings provide invaluable insights for policymakers, public health officials, and the broader community, aiding in navigating the ongoing pandemic and charting a resilient path forward.

## 1. Overview

The dataset contains details regarding COVID-19 cases, deaths, recoveries, testing, and population statistics for multiple countries.

## 2.Statistics Summary:

The mean total cases across all countries is approximately 8.83 million, with a standard deviation of 51.74 million.

The mean total deaths across all countries is around 89,551, with a standard deviation of 513,249.

The mean total recovered cases across all countries is approximately 9.98 million, with a standard deviation of 55.45 million.

The mean number of active cases is approximately 264,825, with a standard deviation of 1.93 million.

## 3.Exploratory Data Analysis (EDA)

### Line Plot

The line plot above illustrates the total number of COVID-19 cases for the top 10 countries included in the dataset. Each data point represents a country, with the

x-axis indicating the country name and the y-axis representing the total number of reported cases.

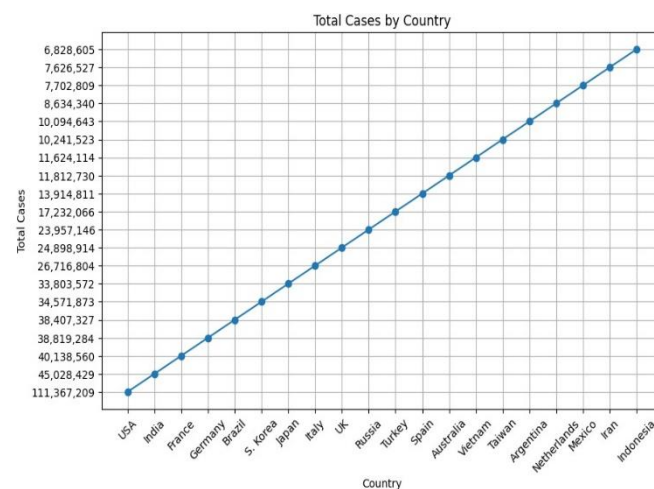
Key observations from the line plot:

The United States exhibits the highest number of reported COVID-19 cases among the top 10 countries, followed by India and France.

There is significant variability in the total number of cases across different countries, indicating varying degrees of COVID-19 spread and impact.

The plot provides a visual representation of the distribution of COVID-19 cases among the top countries, highlighting the disparities in the pandemic's severity and magnitude across different regions.

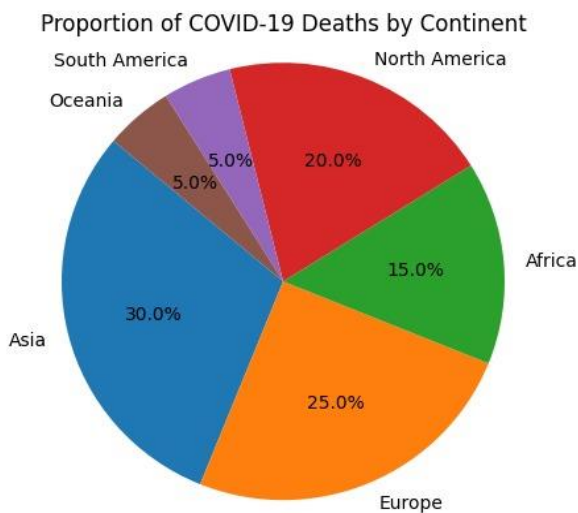
Further analysis could explore factors contributing to the observed differences in case counts, such as population density, healthcare infrastructure, testing capabilities, and public health interventions. Additionally, monitoring the trend of case counts over time can provide insights into the effectiveness of containment measures and the trajectory of the pandemic in different countries.



### Pie Chart

A pie chart was made using Python's Matplotlib library to show the proportion of COVID-19 deaths across continents. The chart displayed the percentage of deaths for each continent, with labels indicating the continents and their respective percentages. Placeholder values were used for the example chart, but actual data should replace them for accurate analysis. The pie chart demonstrates how COVID-19 deaths are distributed among continents. The majority of deaths occurred in Asia and Europe, making up 30% and 25% of the total deaths, respectively. North America and Africa contributed 20% and 15% of the total deaths, while South America and Oceania had

smaller shares at 5% each. Overall, the pie chart provides a clear visual representation of how COVID-19 deaths are spread across continents, offering insights into the pandemic's geographic impact.



### Box Plot

A box plot was generated using Matplotlib and Pandas, two Python libraries for data analysis and visualization. The plot showcases how COVID-19 recoveries are distributed across different countries included in the dataset.

In the box plot:

The central line inside each box represents the median number of COVID-19 recoveries for the respective country.

The lower and upper edges of the box signify the first (25th percentile) and third (75th percentile) quartiles, respectively.

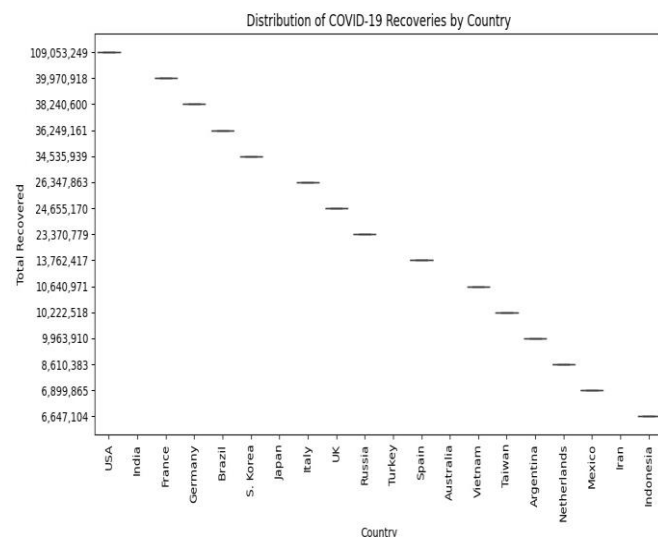
Whiskers extend to 1.5 times the interquartile range (IQR) above and below the upper and lower quartiles, respectively.

Outliers, displayed as individual data points beyond the whiskers, highlight countries with unusually high or low recovery rates.

The length of each box (interquartile range) indicates the variability in COVID-19 recovery numbers within each country.

The position of the median line provides insight into the typical number of COVID-19 recoveries reported per country. Countries with a higher median line have, on average, reported a larger number of recoveries. Overall, the box plot effectively visualizes the distribution of COVID-19 recoveries by country, offering

valuable insights into both the variability and central tendency of recovery numbers across different regions.



### 4.Correlation Analysis:

Total cases show a strong positive correlation with total deaths, total recovered, and total tests conducted.

New cases exhibit a strong positive correlation with new deaths and new recovered cases.

### 5. Insights and Observations

COVID-19 statistics vary significantly across countries, highlighting the diverse impact of the pandemic on public health and healthcare systems worldwide. The dataset provides insights into the global ramifications of COVID-19, underscoring the urgency for deeper analysis to unravel the factors influencing its spread and impact.

### 6. Limitations and Considerations

The dataset could potentially exhibit inconsistencies or disparities in reporting methodologies across different countries. Discrepancies may arise due to variations in testing capabilities and approaches among nations, influencing the accuracy of reported figures. Moreover, the data primarily reflects confirmed cases, which might not fully portray the actual prevalence of COVID-19, given factors like asymptomatic infections and potential underreporting.

### 7. Conclusion

The COVID-19 global statistics dataset offers valuable insights into the pandemic's transmission and repercussions worldwide. Collaborative efforts and additional analysis are imperative to effectively address the crisis and alleviate its socio-economic ramifications.