





COVID-19 Data Analysis Report

1 Introduction

This report analyzes COVID-19 data, focusing on case trends, deaths, vaccinations, and continental distributions. The dataset used was sourced from **Our World in Data**, and insights were derived using Python's **pandas, numpy, matplotlib, and seaborn** libraries.

2 Data Cleaning & Preparation

Before analyzing trends, the dataset was cleaned using the following steps:  Converted **date** column to datetime format (pd.to_datetime()).  Filtered data for **Kenya, USA, and India** for focused analysis.  Dropped rows with missing **critical values** (total_cases, total_deaths).  Handled missing numeric values using **forward-fill & back-fill interpolation** to ensure smooth trends.

3 Exploratory Data Analysis (EDA)

Total Cases & Deaths Trends

- **COVID-19 Cases Growth:** Exponential increases observed in all selected countries.
- **Mortality Rates:** Death rates vary significantly across countries, highlighting differences in healthcare systems and pandemic response strategies.

Continental Distribution

- A **bar chart** reveals the number of affected countries per continent.
- Disparities observed, with **Asia & Europe** having the highest dataset representation.

Pie Chart Analysis (Total Deaths)

- The **top 5 countries** with the highest COVID-19 deaths were plotted using a pie chart.
- This visualization highlights variations in mortality rates, possibly due to differences in medical infrastructure and pandemic mitigation efforts.

4 Vaccination Progress & Insights

Cumulative Vaccinations Over Time

- A **line chart** tracks vaccination rollout progress across Kenya, USA, and India.
- Clear upward trends indicate the success of mass vaccination drives.

Comparison of % Vaccinated Population

- A **bar chart** presents vaccination rates by country, highlighting nations that have reached higher immunity thresholds.

Pie Chart: Vaccinated vs. Unvaccinated

- A pie chart was used to visualize the proportion of vaccinated vs. unvaccinated populations per country.
- Data highlights gaps in vaccine accessibility and distribution.

5 Conclusion & Key Takeaways

📌 **Data Cleaning ensures accurate trend visualization**—interpolating missing values prevents abrupt gaps. 📌 **COVID-19 trends differ vastly among countries**, reflecting healthcare and containment measures. 📌 **Continental disparities exist in affected countries representation**—Europe and Asia dominate in dataset records. 📌 **Vaccination campaigns have made significant progress**, but gaps remain in immunity coverage worldwide.