**Abstract**

The COVID-19 pandemic (2020–2024) reshaped global health, society, and economics. Using the Kaggle dataset, “Daily COVID-19 Data (2020-2024),” this exploratory data analysis (EDA) highlights critical pandemic trends. By leveraging granular data metrics like cumulative cases, deaths, testing, and vaccinations. This report identifies temporal, geographical, and policy-driven factors shaping pandemic outcomes.

**Link:** [**Daily COVID-19 Data (2020-2024)**](https://www.kaggle.com/datasets/abdoomoh/daily-covid-19-data-2020-2024)

**Introduction**

COVID-19, caused by SARS-CoV-2, emerged in late 2019, rapidly evolving into a global pandemic. It has led to over 750 million confirmed cases and more than 6.8 million deaths globally by 2024. The accompanying dataset, sourced from Kaggle, provides daily records of cases, deaths, and vaccinations across countries and regions, facilitating comprehensive pandemic analysis. The goal of this EDA is to uncover key trends, understand disparities, and evaluate the impact of public health interventions.

**Overview of the Dataset**

The dataset encompasses daily COVID-19 statistics across continents, providing metrics such as cumulative confirmed cases, deaths, testing rates, and vaccinations. The data is segmented by WHO regions, enabling regional and cross-country comparisons. Specific data points include:

* **Global Numbers:**
  + **Cases - Cumulative Total:** 751 million.
  + **Deaths - Cumulative Total:** 6.8 million.
* **Regional Distribution (WHO Regions):**
  + **Europe:** 279.5M cases, 2.27M deaths.
  + **Americas:** 192.7M cases, 3.02M deaths.
  + **Africa:** 9M cases, 174K deaths.
  + **South-East Asia:** 61.3M cases, 808K deaths.
  + **Western Pacific:** 208.5M cases, 421K deaths.

This segmentation highlights significant disparities in pandemic outcomes between regions.

**Analysis**

**a. Temporal Trends**

* Case and death trajectories exhibit distinct waves, with peaks in late 2020 and 2021, attributed to new variants and fluctuating public health measures.
* By 2024, cases and deaths show a marked decline, paralleling global vaccination campaigns.

**b. Regional Distribution**

* Europe and the Americas account for ~60% of global cases and ~77% of deaths, making them the hardest-hit regions.
* Africa and South-East Asia reported lower per capita cases, though underreporting and limited testing may contribute.

**c. Testing and Positivity Rates**

* Testing volume increased significantly over time, reducing positivity rates. Countries with robust testing strategies reported better containment and fewer deaths.

**d. Vaccination Impact**

* Europe and the Americas achieved high vaccination rates early, resulting in reduced severe outcomes.
* Correlation analyses reveal a direct link between higher vaccination rates and lower death rates in 2023-2024.

**Findings**

1. **High-Impact Regions:** Europe and the Americas experienced the greatest burden, reflecting urban density and travel hubs' impact.
2. **Temporal Trends:** Peaks in 2020-2021 align with variant surges, while 2023-2024 highlights the role of vaccinations in reducing case severity.
3. **Testing Variance:** Countries with advanced testing protocols identified and mitigated outbreaks more effectively.
4. **Vaccination Outcomes:** Regions with high vaccine uptake observed a faster recovery and fewer deaths.

**Conclusion**

This EDA reveals the pandemic's profound global and regional disparities. Europe and the Americas suffered disproportionately, while Africa and South-East Asia faced unique challenges like underreporting. Vaccinations and robust testing emerged as critical factors in mitigating severe outcomes. These findings underscore the importance of equitable access to healthcare resources and coordinated public health responses for future pandemics.