



IS-635

COVID-19 World Data Tableau Analysis Report

By Aastha Kshatriya- 032179860

Guided by-Mohamed Abdelhamid

California State University, Long Beach

Course: IS 635 - Spring 2025



Table of Contents

1. Dataset Description

2. Data Processing with SSMS

- The dataset was processed using SQL Server Management Studio (SSMS), where SQL queries were employed to aggregate and clean the data for analysis.
- SQL queries were used to calculate metrics like total cases, total deaths, infection percentages, and vaccination progress by location and time.

3. Research Questions and Objectives

- Key Questions to Investigate
- Expected Insights from the Data

4. Data Visualization and Analysis

- Overview of the Dashboards
- Summary of Key Visualizations
 - **Global Impact and Death Count by Continent**
 - **Infection Rate by Country and Region**
 - **Vaccination Progress and Its Impact**
 - **Total Cases and Deaths by Location**

5. Findings and Insights

- Key Observations from the Data
- Trends and Patterns Identified in Visualizations

6. Business Implications and Recommendations

- Key Implications for Health Policy and Vaccination Strategies
- Recommendations for Governments and International Organizations

7. Conclusion

- Summary of the Analysis
- Final Thoughts on the COVID-19 Impact and Vaccination Efforts

8. Appendices

- Additional Insights from the Analysis



1. Dataset Description and Variables

The dataset used for this analysis comes from the **COVID-19** tracking database, which includes information on daily COVID-19 cases, deaths, and vaccination data across various countries and regions. The data spans different continents, countries, and geographic regions. The key variables in the dataset include:

Bifurcated Dataset into 2 excels

1. Covid Death Excel Dataset

- **Location:** Represents the country or region where COVID-19 cases and deaths are reported (e.g., United States, India, Europe, etc.).
- **Date:** The date when the data was recorded, typically daily.
- **New Cases:** The number of new COVID-19 cases reported on a given day.
- **New Deaths:** The number of new COVID-19 related deaths reported on a given day.
- **Total Cases:** Sum of the addition of Per Day new cases of Covid 19.
- **Population:** The population of the corresponding location.
- **New Vaccinations:** The number of new vaccinations administered on a given day.
- **Continent:** The continent to which the location belongs (e.g., Asia, Europe, North America).

2. Covid Vaccination Excel Dataset

- **Continent:** The continent to which the location belongs (e.g., Asia, Europe, North America).
- **Location:** Represents the country or region where COVID-19 cases and deaths are reported (e.g., United States, India, Europe, etc.).
- **Population:** The population of the corresponding location.
- **Date:** The date when the data was recorded, typically daily.
- **People Vaccinations:** The number of people vaccinated administered on a given day.
- **New Vaccinations:** The number of new vaccinations administered on a given day.



2. Data Processing with SSMS

The dataset was processed using SQL Server Management Studio (SSMS), a comprehensive database management system for querying, cleaning, and analyzing data. SQL queries were employed to aggregate and transform raw data into meaningful insights. The following steps summarize how SSMS was used for data processing:

- **Data Aggregation:** Using SQL queries, we aggregated data to calculate total COVID-19 cases, total deaths, and vaccination numbers by country and region. This helped in summarizing the data at a continental, national, and regional level.
- **Calculation of Infection and Death Rates:** Using the aggregated data, the percentage of the population infected and the death rate were calculated. These rates are crucial for understanding the severity of the pandemic in various regions.
- **Handling Missing Data:** SSMS queries were designed to handle missing or incomplete data. For example, locations with incomplete or unavailable data were excluded from analysis, ensuring the integrity of the results.
- **Creation of Time-Series Data:** SQL queries were used to generate time-series data for COVID-19 cases, deaths, and vaccinations. This allows us to analyze the progress of the pandemic and vaccination efforts over time, identifying trends and shifts in the data.
- **Joining Data from Multiple Tables:** Data related to COVID-19 cases and vaccinations were stored in separate tables. SQL JOIN operations were used to combine this data, enabling a holistic view of both infection and vaccination trends across countries and regions.
- **Data Cleaning:** Data cleaning operations in SSMS ensured that any discrepancies, such as mismatched location names or incorrect entries, were resolved, making the dataset ready for analysis and visualization.

The processed data from SSMS was then exported for use in Tableau, where it was visualized to explore the pandemic's trends and impacts.



3. Research Questions and Objectives

- **Key Questions to Investigate**

- **What is the overall death toll across different continents?**
- **How has the percentage of infection varied by country?**
- **How does the vaccination rollout correlate with infection rates?**
- **What are the total cases and deaths by location?**
- **How do COVID-19 cases and deaths compare globally and regionally?**

Expected Insights from the Data

- Impact of vaccination on controlling infections and deaths.
- Regional trends in cases, deaths, and vaccination rates.
- Identification of high-risk areas needing focused intervention.



4. Data Visualization and Analysis

Story Structure in Tableau

- **Slide 1: Infection Rate by Country**
 - A **map** visualization displays the **percentage population infected per country**. This map allows for geographic comparison, showcasing how different countries have been impacted by the virus. It highlights the countries with the highest infection rates, including **the United States, Mexico, and India**.
- **Slide 2: Total Cases and Deaths by Location**
 - A **table** summarizes the **total cases and deaths by location**. This slide is particularly useful for detailed comparisons between countries, with a focus on major affected regions like **North America, Europe, and South America**.
- **Slide 3: Death Count by Continent**
 - A **bar chart** visualizes the **total death count by continent**, showing which regions have been most affected by COVID-19 in terms of fatalities. This visualization highlights that **Europe and North America** have experienced the highest total deaths.
- **Slide 4: Vaccination Progress and Infection Rates**
 - This slide features a **line graph** that tracks the **average percentage of the population infected** over time. It provides insights into how vaccination efforts (represented by different colors for each country) correlate with infection rates, revealing the impact of vaccines in controlling the spread of COVID-19.

Dashboards

At the end of the story, the final slides bring all the previous visualizations together in two **comprehensive dashboards**. The first dashboard combines global summary data, death counts by continent, and infection rate maps, while the second aggregates data on total cases and deaths by location, providing an in-depth view of the regional impact of the pandemic.

The **interactive dashboards** allow users to explore the data in greater detail, with options to filter by location and view trends over time, enhancing the user experience and providing actionable insights.

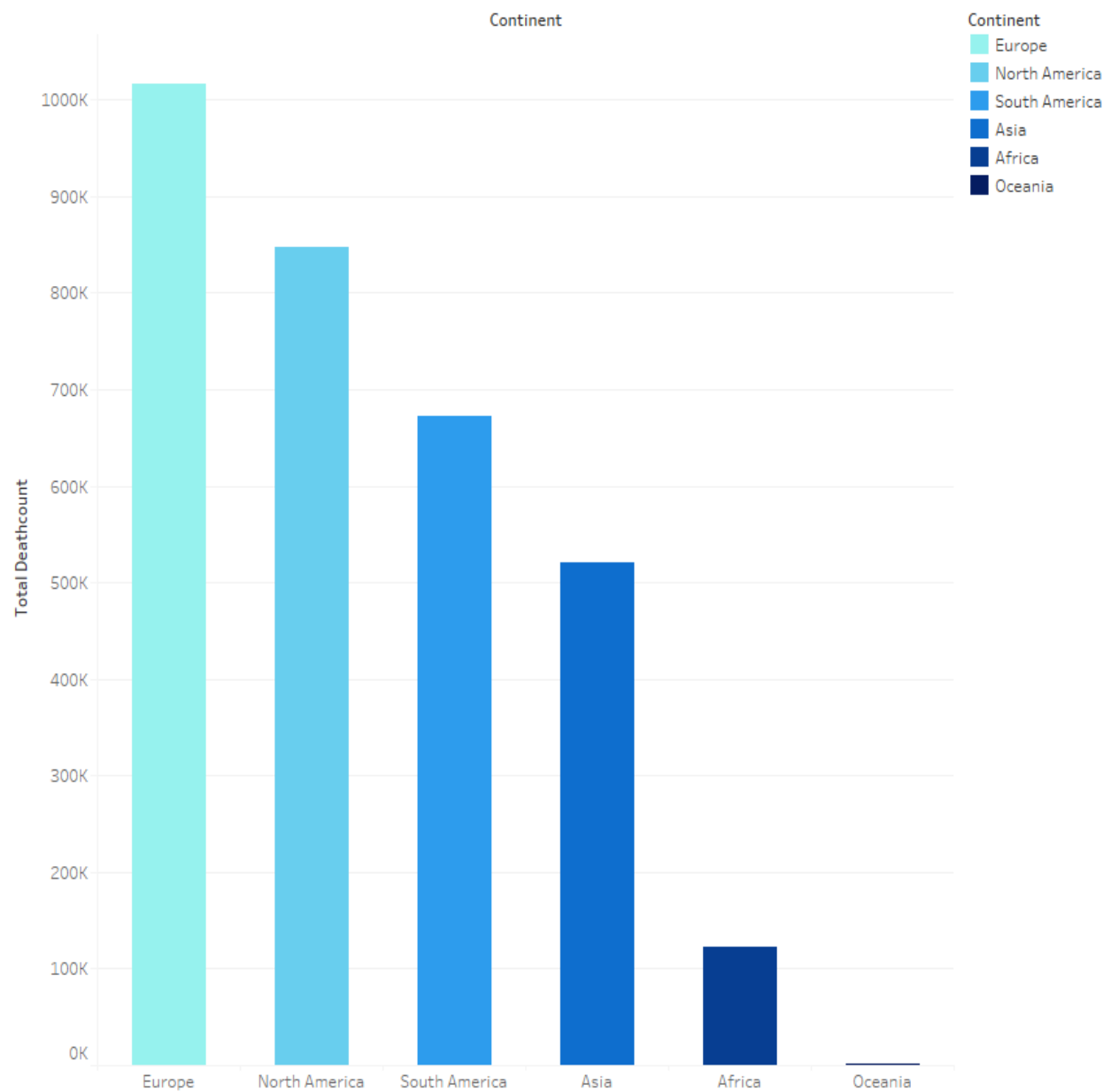


- **Summary of Key Visualizations**

- **Global Impact and Death Count by Continent**

- A bar chart visualizing the total death toll across continents.

Deathcount by Continent





- **Infection Rate by Country and Region**

- A map displaying the percentage of the population infected per country.

Percent Population Infected Per Country

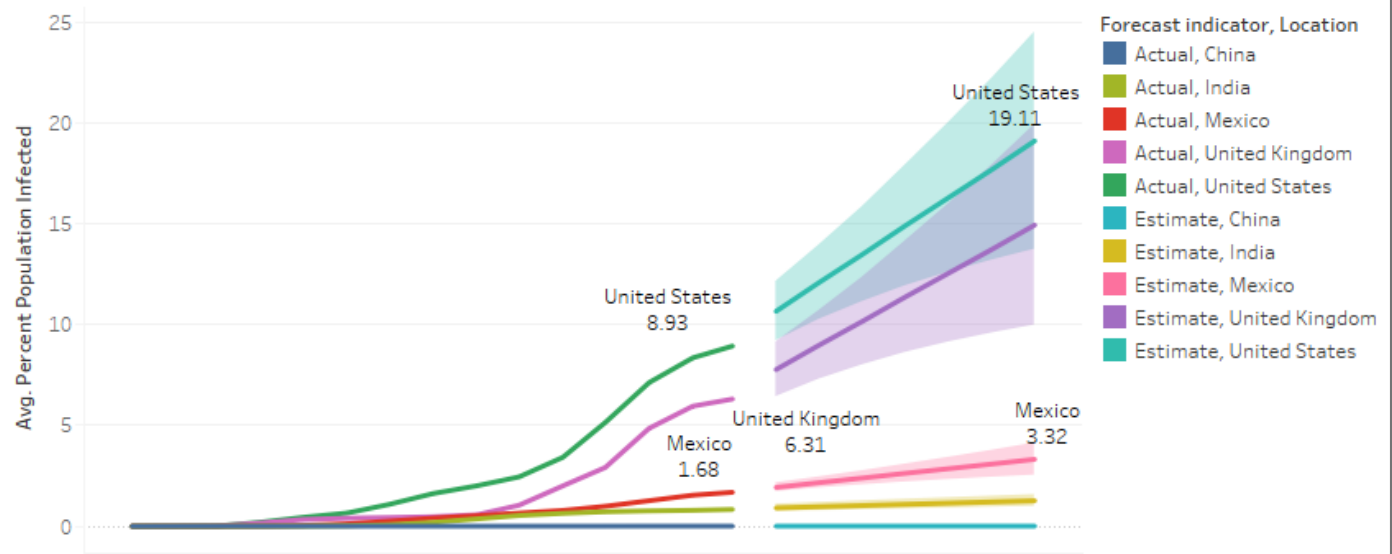




- **Vaccination Progress and Its Impact**

- Line graphs showing how vaccination rates impact infection trends over time.

Percent Population Infected





- **Total Cases and Deaths by Location**

- A table summarizing total cases and deaths for major locations.

Total Cases and Deaths by Location

Location	Total Cases	Total Deaths
Africa	4,557,699	121,784
Asia	38,702,527	520,269
Brazil	14,659,011	403,781
Europe	44,863,478	1,016,750
France	5,677,835	104,675
India	19,164,969	211,853
North America	37,529,487	847,942
Russia	4,750,755	108,290
South America	24,878,216	672,415
United States	32,346,970	576,232

- **Total Covid Cases, Total Deaths and Death Percentage all over the world**

Toal Covid cases all over the World

Covid Cases	Total Deaths	Death Percentage
150,574,977	3,180,206	2.11

Covid Cases, Death Percentage and Total Deaths.



5. Findings and Insights

Key Observations from the Data

- Europe and North America exhibit the highest total deaths.
- The United States, Mexico, and India are among the countries with the highest infection rates.
- Countries with higher vaccination rates, like the United Kingdom and the United States, show slower increases in infections.

Trends and Patterns Identified in Visualizations

- Regions with higher vaccination rates tend to have lower infection rates, supporting the effectiveness of vaccination in controlling the pandemic.
- The data highlights disparities in vaccination progress, with some regions lagging behind, particularly in Africa and parts of Asia.



6. Business Implications and Recommendations

- **Key Implications for Health Policy and Vaccination Strategies**
 - Accelerating vaccination efforts in regions with high infection rates to reduce the burden on healthcare systems.
 - Strong focus on vulnerable regions like Africa and parts of Asia where vaccination rates are low.

- **Recommendations for Governments and International Organizations**
 - Support global vaccination programs to ensure equitable distribution of vaccines, especially in developing countries.
 - Prioritize high-risk populations and regions with high infection rates for vaccination campaigns.
 - Increase investment in healthcare infrastructure to manage future pandemics.



7. Conclusion

- **Summary of the Analysis**

- The dashboards provide a clear overview of the global spread of COVID-19, highlighting the impact of vaccinations on controlling infections and deaths.
- There are significant regional differences, with certain areas facing more severe challenges due to slower vaccination rollouts and higher infection rates.

- **Final Thoughts on the COVID-19 Impact and Vaccination Efforts**

- Vaccination remains the most effective tool in controlling the spread of COVID-19. Governments must continue to prioritize vaccination efforts, especially in high-risk regions.
- Ongoing monitoring and adjustments in health policies will be essential to combat the pandemic in the long run.



8. Appendices

- **Detailed Tables and Graphs**

- Tables summarizing the total cases and deaths by location, and graphs depicting the trends in infection rates and vaccination progress.

- **Additional Insights from the Analysis**

- Exploration of specific regional challenges, such as healthcare capacity and vaccine distribution logistics.