

# COVID-19 Data Analysis

---

## Tools & Technologies:

- **Programming Language:** Python
  - **Data Manipulation:** Pandas, NumPy
  - **Data Visualization:** Matplotlib, Seaborn
  - **Machine Learning:** Scikit-learn
  - **Development Environment:** Jupyter Notebook
- 

## 1. Introduction

This project analyzes global COVID-19 data to uncover trends in cases, deaths, and recoveries. By leveraging Python and data analysis libraries, the study aims to provide insights into the pandemic's progression and its impact across different regions and countries.

---

## 2. Objectives

1. Examine global trends in COVID-19 cases and deaths.
  2. Visualize and compare data by continent and country.
  3. Explore temporal patterns in cases and deaths.
  4. Model the relationship between cases and deaths using regression.
  5. Conduct a focused analysis of COVID-19 data specific to India.
- 

## 3. Dataset Description

The dataset `owid-covid-data.csv` includes the following key attributes:

- **Date:** The date of the record.
- **Location:** Country or region name.
- **Continent:** Continent of the country.
- **Total Cases:** Cumulative confirmed cases.
- **Total Deaths:** Cumulative deaths.
- **Total Vaccinations:** Total vaccine doses administered.

## Data Cleaning Steps:

- Removed irrelevant columns such as excess mortality-related data.
- Handled missing values in key columns (e.g., `total_cases`, `total_deaths`, `continent`).

- Extracted the year from the `date` column for temporal analysis.
- 

## 4. Exploratory Data Analysis (EDA)

### 4.1 Analysis by Continent

- Aggregated total cases by continent and visualized them using a bar chart.
- **Insight:** Identified the continent with the highest cases.

### 4.2 Country-Level Analysis

- Identified and visualized the top 10 countries with the most cases and deaths.
- **Insight:** Highlighted heavily impacted countries.

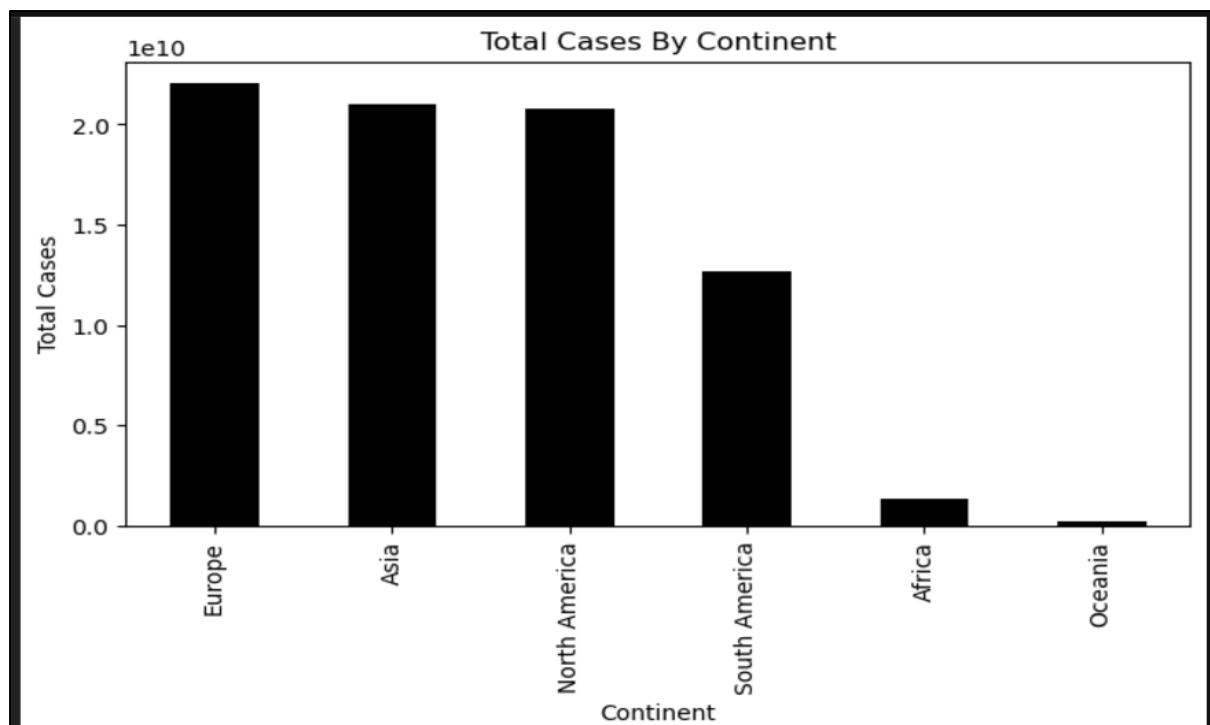
### 4.3 Temporal Trends

- Grouped cases and deaths by year to analyze trends over time.
- 

## 5. Key Visualizations

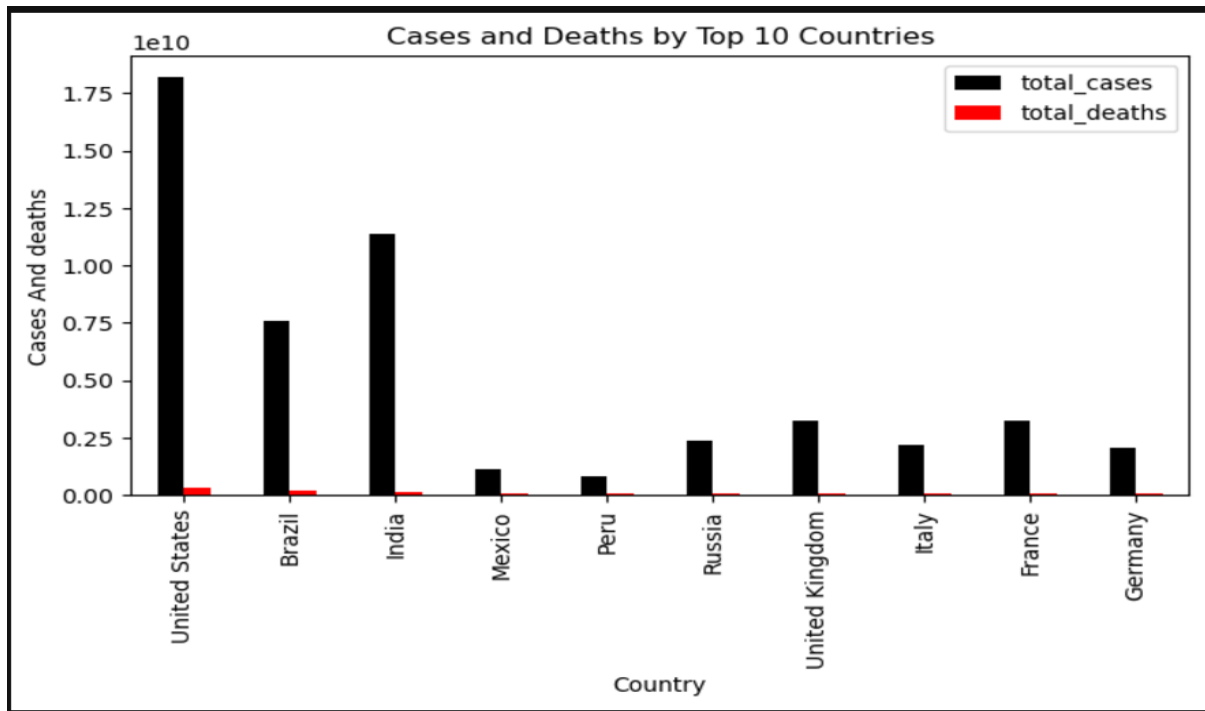
### 5.1 Total Cases by Continent

A bar plot illustrating the total number of COVID-19 cases across continents.



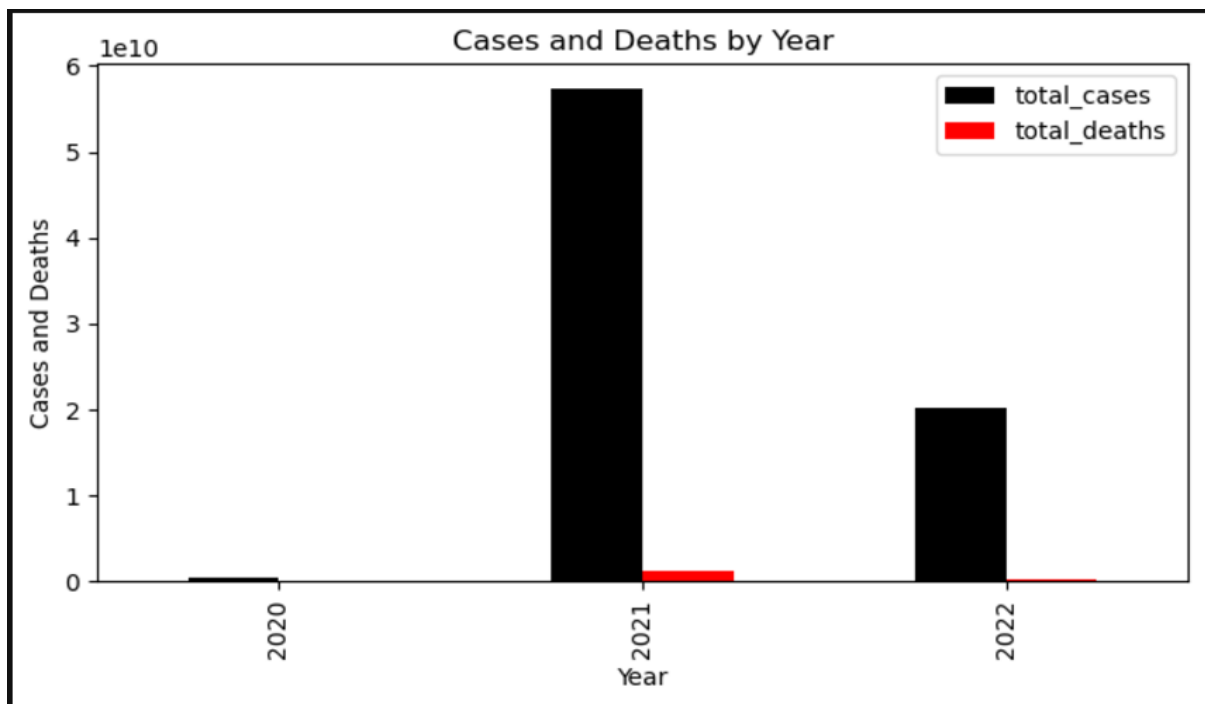
## 5.2 Cases and Deaths by Top 10 Countries

A comparative bar chart showcasing cases (black) and deaths (red) for the top 10 affected countries.



## 5.3 Yearly Trends in Cases and Deaths

Bar plots showing yearly totals of cases and deaths globally.



## 6. Linear Regression Analysis

### Objective

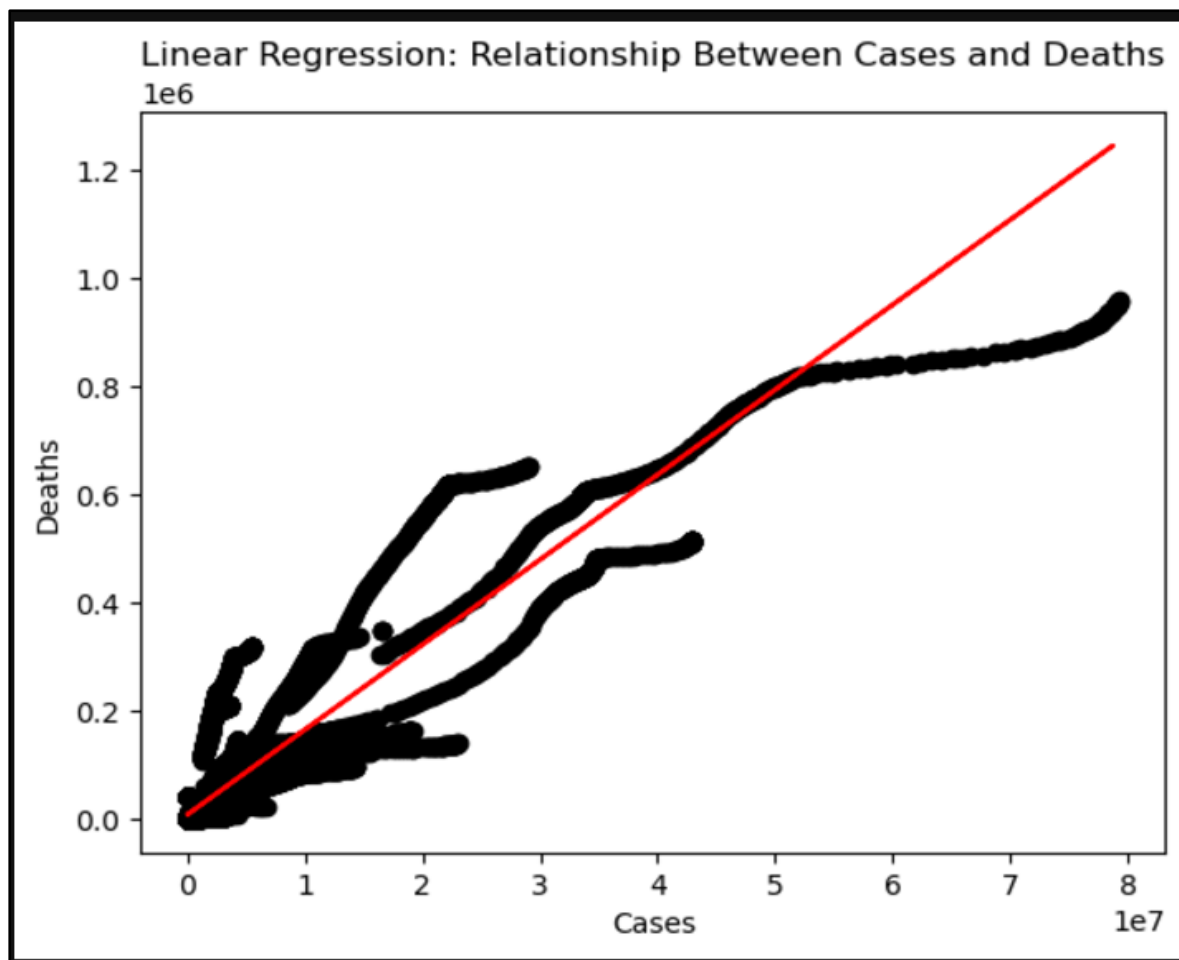
To explore the relationship between total cases and total deaths.

### Methodology

- **Feature:** `total_cases` (independent variable).
- **Target:** `total_deaths` (dependent variable).
- **Model:** Linear regression using Scikit-learn.

### Result

A scatter plot with the regression line visualizes this relationship.



## 7. India-Specific Analysis

Focused analysis on total deaths in India:

- Filtered and grouped data specific to India.
- Aggregated total deaths for the country.

**Insight:** India's total deaths were [147638153.0].

---

## 8. Conclusion and Future Work

### Conclusion

The analysis provides a comprehensive overview of the pandemic's progression. It highlights the impact of COVID-19 on continents and countries, uncovers temporal patterns, and establishes relationships between cases and deaths.

### Future Work

1. Integrate vaccination data to provide further insights into the pandemic's trajectory.
  2. Develop predictive models for case forecasting.
  3. Examine the socioeconomic impacts of the pandemic on a global scale.
-