






# DATA MANIPULATION AND VISUALIZATION




SWATHI P|920821106083|ECE

# COVID-19 Data Analysis with Python and Power BI

## Objective

To analyze global COVID-19 data by performing data manipulation using Python and visualizing the insights using Power BI.   

## Dataset

- **Source:** [COVID-19 Dataset by Our World in Data](#)
- **Direct Download Link:** [owid-covid-data.csv](#)   

## About the Dataset

The dataset is a comprehensive collection of COVID-19 statistics aggregated from official public health sources. It includes global data on cases, deaths, testing, vaccinations, and more. It is maintained by "Our World in Data" to support researchers, policymakers, and the general public in understanding the pandemic's progression.

## Data Features


The dataset contains the following key features:

- **Location:** The name of the country or region.
- **Date:** The specific date for each data entry.
- **Total Cases:** Cumulative confirmed COVID-19 cases.
- **New Cases:** Daily new confirmed cases.
- **Total Deaths:** Cumulative confirmed deaths due to COVID-19.
- **New Deaths:** Daily new deaths reported.
- **People Vaccinated:** Total number of individuals vaccinated.
- **Continent:** The continent to which the country belongs.
- Additional metrics like testing data, hospitalizations, and vaccination rates are also included where available.

## Steps to Complete the Project

### Step 1: Data Manipulation Using Python

#### 1.1 Load the Dataset

Use Pandas to load and preview the dataset.   




```
import pandas as pd

# Load the dataset
url = "https://covid.ourworldindata.org/data/owid-covid-data.csv"

data = pd.read_csv(url)

# Preview the dataset
print(data.head())
```

#### 1.2 Data Cleaning

Focus on essential columns and handle missing values.   

```
# Keep only relevant columns
columns_to_keep = [
    'location', 'date', 'total_cases', 'new_cases', 'total_deaths',
    'new_deaths', 'people_vaccinated', 'continent'
]

data = data[columns_to_keep]

# Drop rows with missing continent data
data = data.dropna(subset=['continent'])

# Fill missing numerical values with 0
data.fillna({'total_cases': 0, 'new_cases': 0, 'total_deaths': 0, 'new_deaths': 0,
            'people_vaccinated': 0}, inplace=True)
```

#### 1.3 Add Calculated Columns

Create meaningful metrics like **Case Fatality Rate** and **Vaccination Rate**.   







```
# Add Case Fatality Rate (%)
data['case_fatality_rate'] = (data['total_deaths'] / data['total_cases']) * 100

# Add Vaccination Rate
data['vaccination_rate'] = (data['people_vaccinated'] / data['total_cases']) * 100
```

```
# Convert date to datetime for better processing
data['date'] = pd.to_datetime(data['date'])




# Save the cleaned dataset as CSV to the specified path for power bi
file_path = r"C:\Users\swathi\Downloads\cleaned_covid_data.csv"
data.to_csv(file_path, index=False)
print(f"Cleaned dataset saved at {file_path}")
```

## Step 2: Import the Cleaned Data into Power BI

1. **Open Power BI Desktop.**   
2. **Import Data:**
  - Navigate to **Home > Get Data > Text/CSV**.
  - Select the cleaned\_covid\_data.csv file.
3. **Load Data:**
  - Review the imported data and load it into Power BI.   

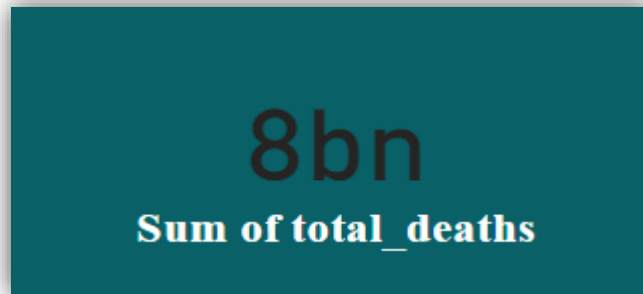
## Step 3: Data Visualization Using Power BI

### Visualizations to Include

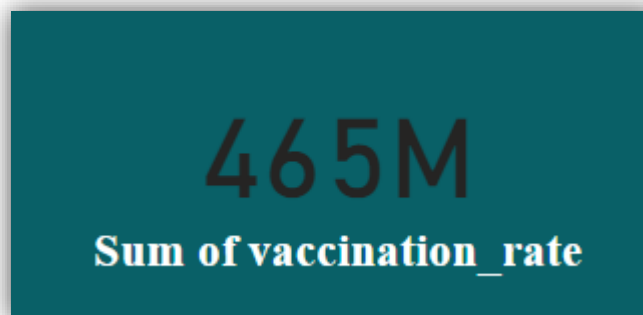
1. **Global Overview Dashboard**   
  - **KPI Cards** for:
    - Total Cases



- Total Deaths



- Total Vaccinations



- **Map Visualization:**

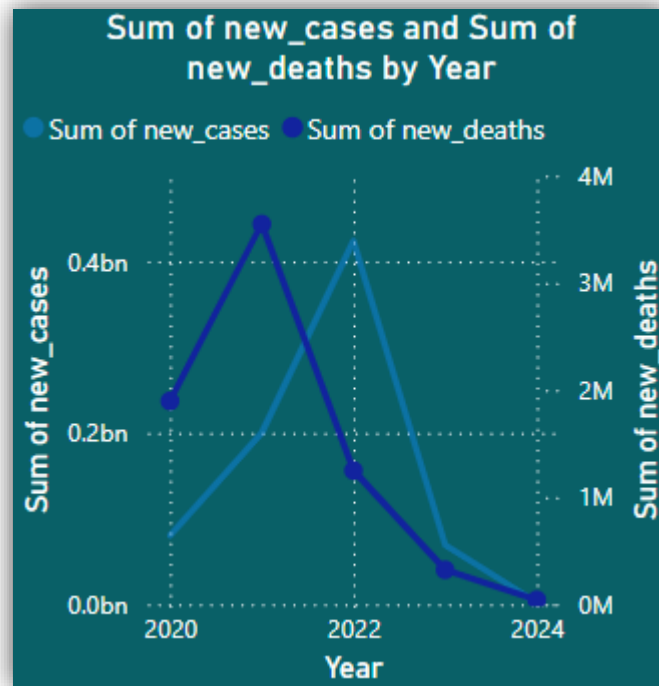
- Show total cases by country using a filled map.



## 2. Daily Trends 📅📈🔍

### ○ Line Chart:

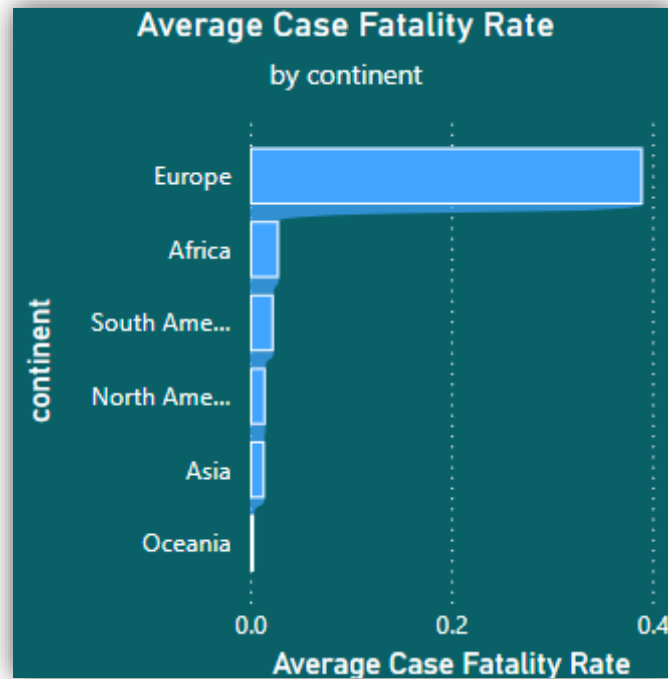
- X-axis: Date.
- Y-axis: New Cases and New Deaths (use dual axes if needed).
- Legend: Location.



## 3. Case Fatality Rate by Continent 🌐📊📈

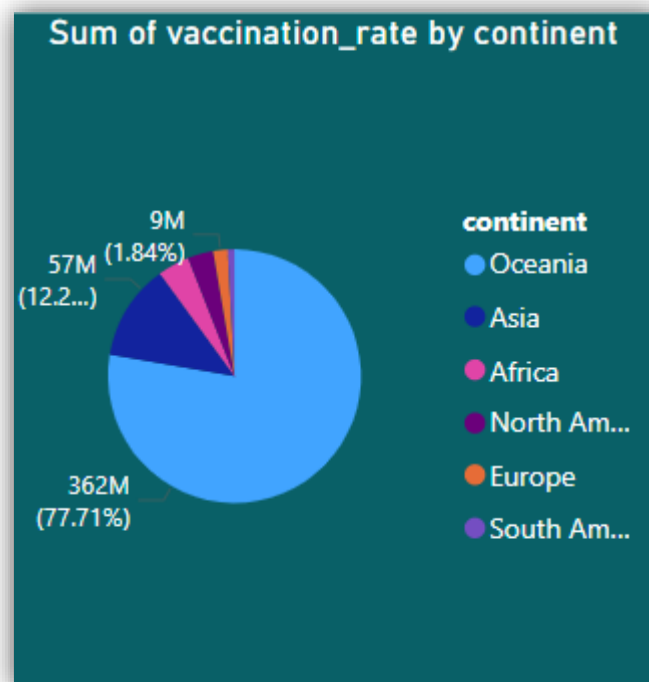
### ○ Bar Chart:

- X-axis: Continent.
- Y-axis: Average Case Fatality Rate.



#### 4. Vaccination Analysis

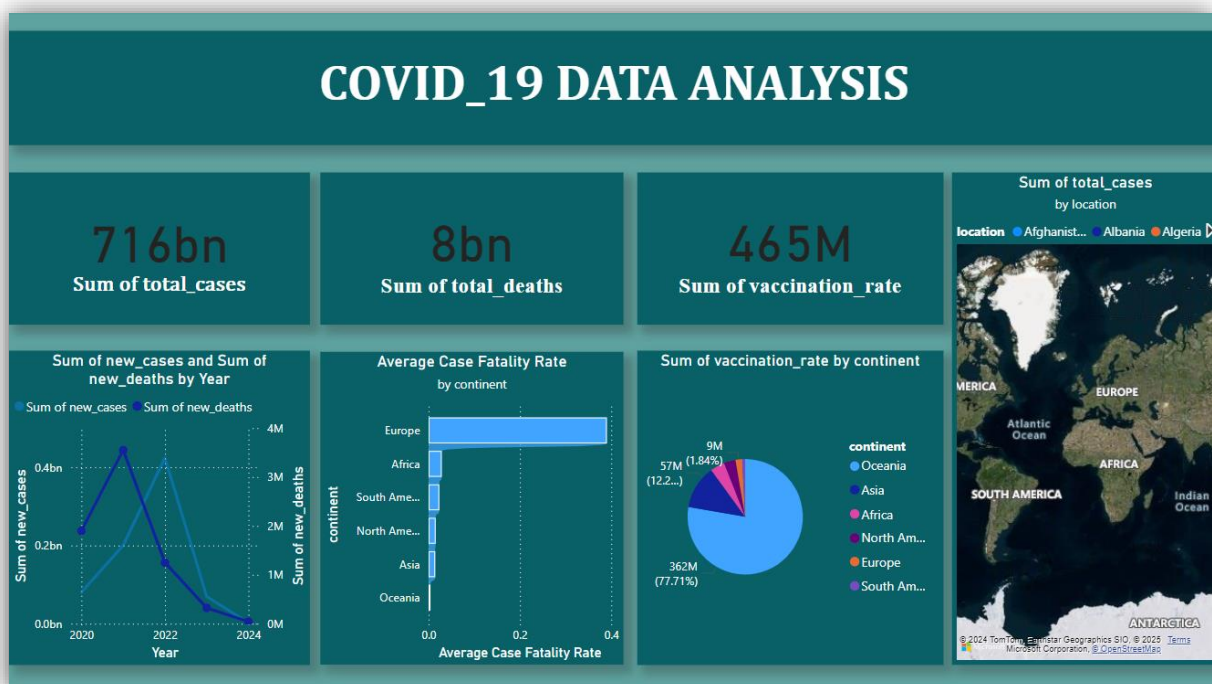
- **Pie Chart:**
  - Values: Total Vaccinations.
  - Legend: Continent.















## 5. Dynamic Filters

- Add slicers to filter data by:
  - Continent.
  - Country.
  - Date Range







## Step 4: Insights to Highlight in the Dashboard









1. **Top Countries by Total Cases:** Which countries have the most cases?   
2. **Vaccination Progress:** Which regions are leading in vaccinations?   
3. **Fatality Rate by Continent:** Which continents have the highest fatality rates?   
4. **Trends:** How are cases and deaths changing over time globally?   

## Expected Deliverables



1. A cleaned CSV file ([cleaned\\_covid\\_data.csv](#)) prepared in Python.   
2. A Power BI report/dashboard showcasing visual insights with slicers for interaction.   

#### Tools Required

- **Python:** For data manipulation.   
  - Libraries: pandas, numpy (optional).
- **Power BI:** For data visualization and dashboard creation.   

---

### Key Insights from Data Manipulation and Visualizations

#### Data Manipulation Insights

##### 1. Efficient Dataset Cleaning:

- Focused on relevant columns like cases, deaths, vaccinations, and continent to streamline the analysis.
- Handled missing values effectively by removing incomplete rows and filling numerical gaps with zero.

##### 2. Calculated Metrics for Better Insights:

- **Case Fatality Rate (CFR):** Provided a clearer perspective on the severity of the pandemic by region.
- **Vaccination Rate:** Highlighted progress in vaccination campaigns relative to reported cases.

##### 3. Enhanced Data Accessibility:

- Converted dates into datetime format for seamless temporal analysis.
- Saved the cleaned dataset for integration into Power BI, facilitating smooth visualization.

#### Data Visualization Insights

##### 1. KPI Cards:

- **Global Scale:** Total cases and deaths conveyed the pandemic's widespread impact.
- **Vaccination Progress:** Showed significant global strides in combating COVID-19 through vaccinations.

##### 2. Filled Map (Total Cases by Country):

- Identified countries with the highest burden of cases, enabling targeted interventions.
- Provided a visual comparison of regional disparities in case distribution.

### 3. Line Chart (Daily Trends):

- Highlighted spikes and drops in new cases and deaths, helping to identify outbreak waves.
- Illustrated how certain periods corresponded to improved or worsening pandemic control.

### 4. Bar Chart (Case Fatality Rate by Continent):

- Revealed variations in pandemic management effectiveness, with some continents showing notably higher fatality rates.
- Enabled insights into resource allocation and healthcare challenges by region.

### 5. Pie Chart (Vaccination Analysis by Continent):

- Highlighted continents leading in vaccinations, showcasing efforts to curb the pandemic.
- Identified areas lagging behind, providing insights for policymakers to prioritize resource deployment.

### 6. Dynamic Filters (Slicers):

- Enabled exploration of specific regions, countries, and time periods.
- Facilitated a tailored analysis for different audiences, from researchers to decision-makers.

## Overall Insights

**1. Global Trends:** The data manipulation and visualization provide an in-depth understanding of COVID-19's progression, mortality, and vaccination efforts across regions and over time.

**2. Regional Analysis:** Clear identification of regions requiring attention, either due to high fatality rates or low vaccination coverage.

**3. Temporal Patterns:** Insights into the timeline of pandemic waves, guiding future outbreak preparedness.

This analysis supports data-driven decision-making, enhances public awareness, and aids researchers in understanding pandemic dynamics.