Corona

Data Analysis

Probabilistic Modeling and Reasoning

Objective

Objective of the Study:

- To explore and analyze COVID-19 data using Exploratory Data Analysis (EDA) techniques.

- To identify trends, patterns, and key factors impacting the survival rates of COVID-19 patients.

- To estimate the probability of survival based on various demographic and clinical factors (age, pre-existing conditions, geographic region, etc.).

- To provide actionable insights that could help in resource allocation and policy-making during a pandemic response.

Library used

- Pandas - Data Manipulation and Analysis
- Matplotlib - Data Visualization
- Seaborn - Statistical Data Visualization

COVID-19 Data Analysis

Importing the required libraries

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv("corona Data/worldometer_data.csv")
```

Shape of the dataframe

```python
df.shape
```

(209, 16)

```
countries=df["Country/Region"]
population=df["Population"]
max_index = population.idxmax()
largest_country = countries[max_index]
largest_population = population[max_index]

print(f"The country with the largest population is {largest_country} with a population of {largest_population}.")
```

```
The country with the largest population is India with a population of 1381344997.0.
```

Population

1. **Per Capita Analysis**

Relative Comparisons: Simply looking at the raw number of cases or deaths doesn't give a true picture of how severely a country is affected. A country with a large population will naturally have more cases than a small one. To make meaningful comparisons, metrics like "cases per million" or "deaths per 100,000 people" are used.

Understanding Severity: Per capita metrics allow for a clearer understanding of how widespread the virus is within a population, regardless of the country's size. A small country with a relatively high number of cases per capita could be struggling more than a large country with higher total numbers.

Test Performed

```
Highest test performed

Total_tests=df["TotalTests"]
highest_test_index=Total_tests.idxmax()
country_with_highest_test= countries[highest_test_index]
highest_test_number = Total_tests[highest_test_index]

print(f"The country with the Highest test is {country_with_highest_test} with a count of {highest_test_number}.")

The country with the Highest test is USA with a count of 63139605.0.
```

- **Understanding Spread:** The number of tests performed helps in understanding the extent of virus spread. If a country performs widespread testing, it can detect more cases, including asymptomatic and mild cases that otherwise might go unreported.

- **Undercounting Risks:** In countries with limited testing, the true number of COVID-19 cases may be significantly undercounted, which can give a false sense of security. Fewer tests might only detect the most severe cases, hiding the actual scale of the outbreak.
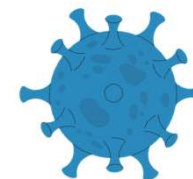
Total Deadth
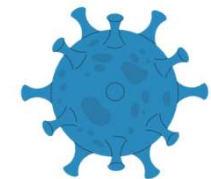
```
Highest deadth in a country

    deadth_column=df["TotalDeaths"]
    maximum_deadth_index=deadth_column.idxmax()
    highest_deadth_country=countries[maximum_deadth_index]
    highest_deadth_number=deadth_column[maximum_deadth_index]
    print(f"The country with the Highest deadth is {highest_deadth_country} with a count of {highest_deadth_number}.")

The country with the Highest deadth is USA with a count of 162804.0.
```

Factors Influencing High Death Rates:
- **Healthcare Capacity: Overwhelmed healthcare systems led to higher death rates.**
- **Vaccine Rollout: Slower or unequal vaccine distribution contributed to more deaths.**
- **Age Demographics: Older populations saw higher death rates.**
- **Variants: Deadly variants like Delta increased mortality.**
- **Economic and Social Factors: Poverty and limited access to healthcare exacerbated death rates in some regions.**

Survival Rate according the the Total cases observed

Calculate Survival Rate:

The survival rate for each country is calculated based on the number of deaths relative to the total number of confirmed cases.

Sort Countries by Survival Rate:

The countries are sorted in ascending order by their survival rate, meaning countries with the lowest survival rates (highest mortality rates) are listed first.

	Country/Region	SurvivalRate
131	Yemen	71.266968
19	France	84.505682
11	UK	84.937397
15	Italy	85.880243
35	Belgium	86.144917
103	Hungary	86.948010
5	Mexico	89.081891
40	Netherlands	89.201853
179	Sint Maarten	90.000000
208	Western Sahara	90.000000

Total Deadth

Top 10 Countries

Top 10 Countries by Total Deaths

Top 10 Countries by Total Tests
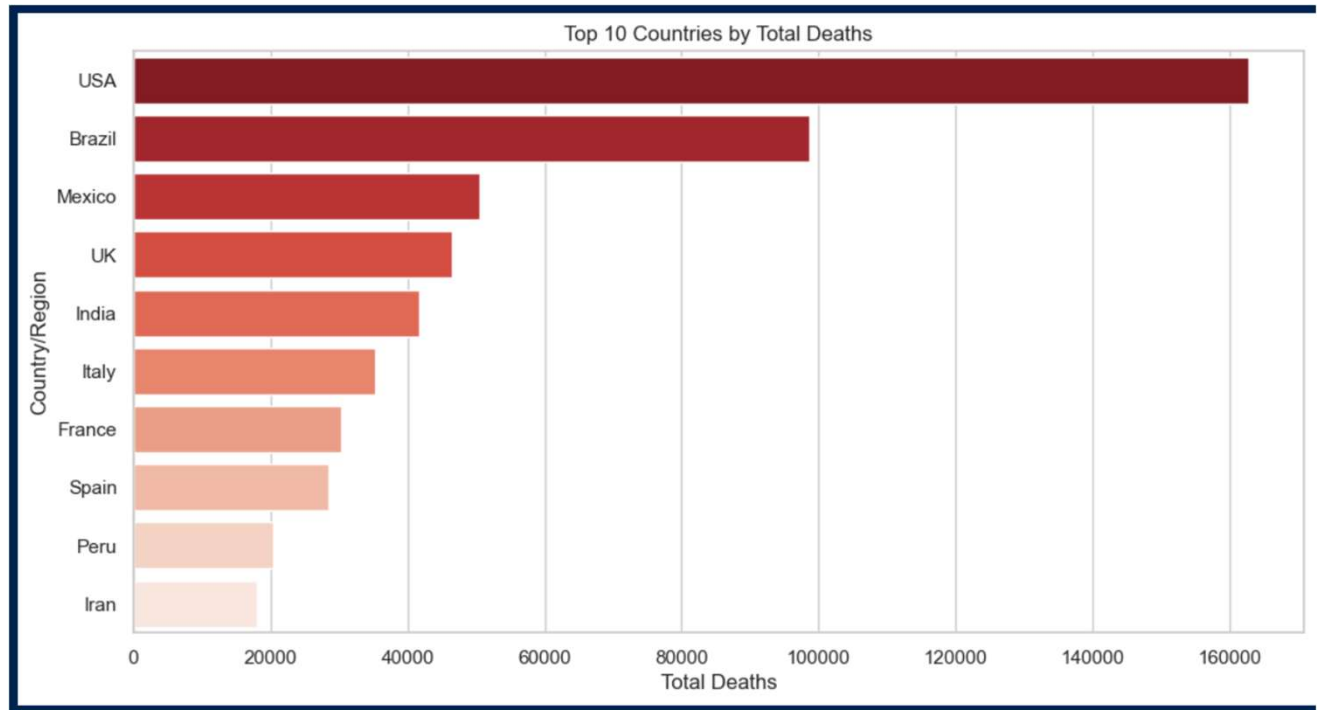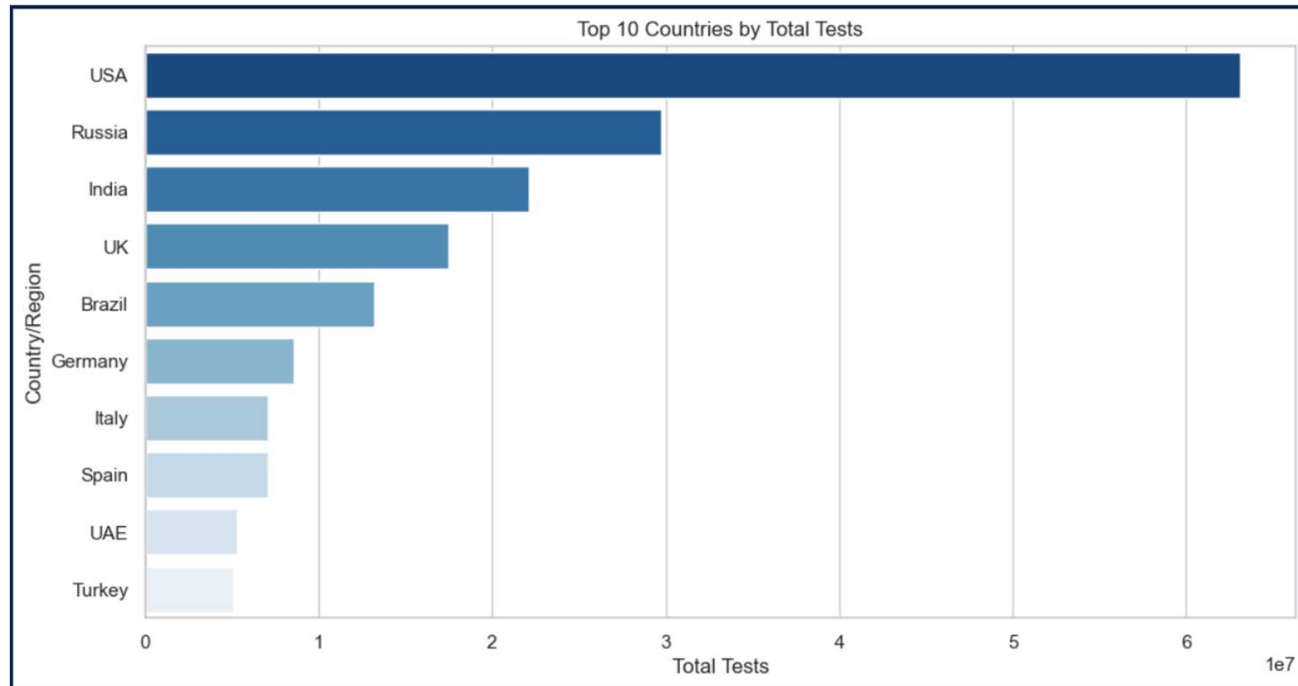
Total Tests
Top 10 Countries

Total Cases Vs Total Deadth

Continents

Total Cases vs Total Deaths
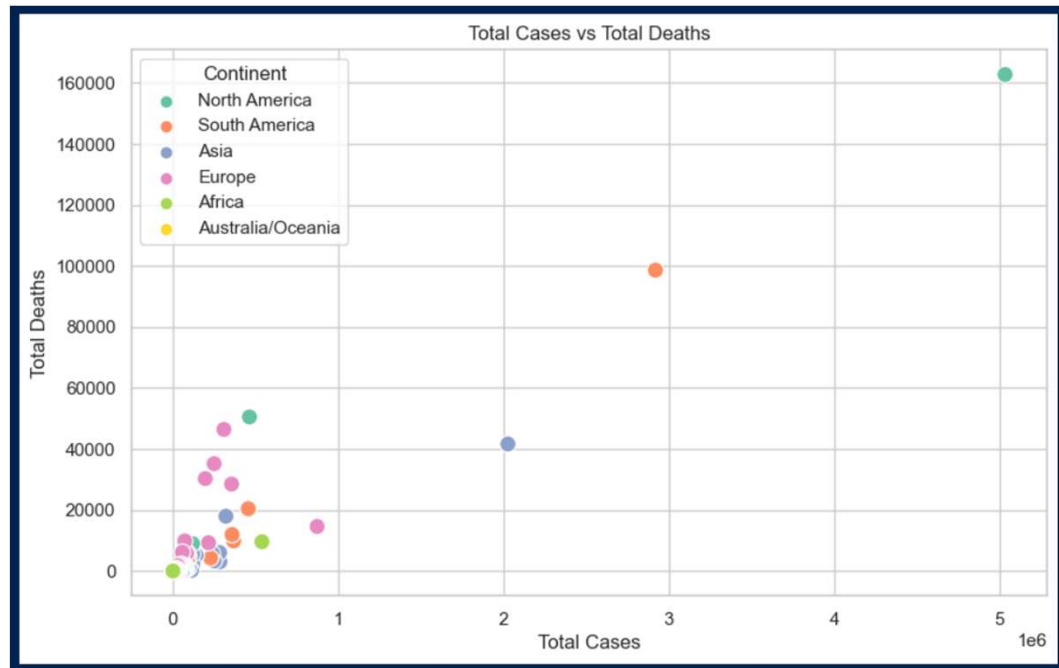
Mortality Rate

Top 10 Countries

Top 10 Countries by Mortality Rate (%)

Thank You

https://github.com/Sameershrinath/CoronaDataAnalysis.git
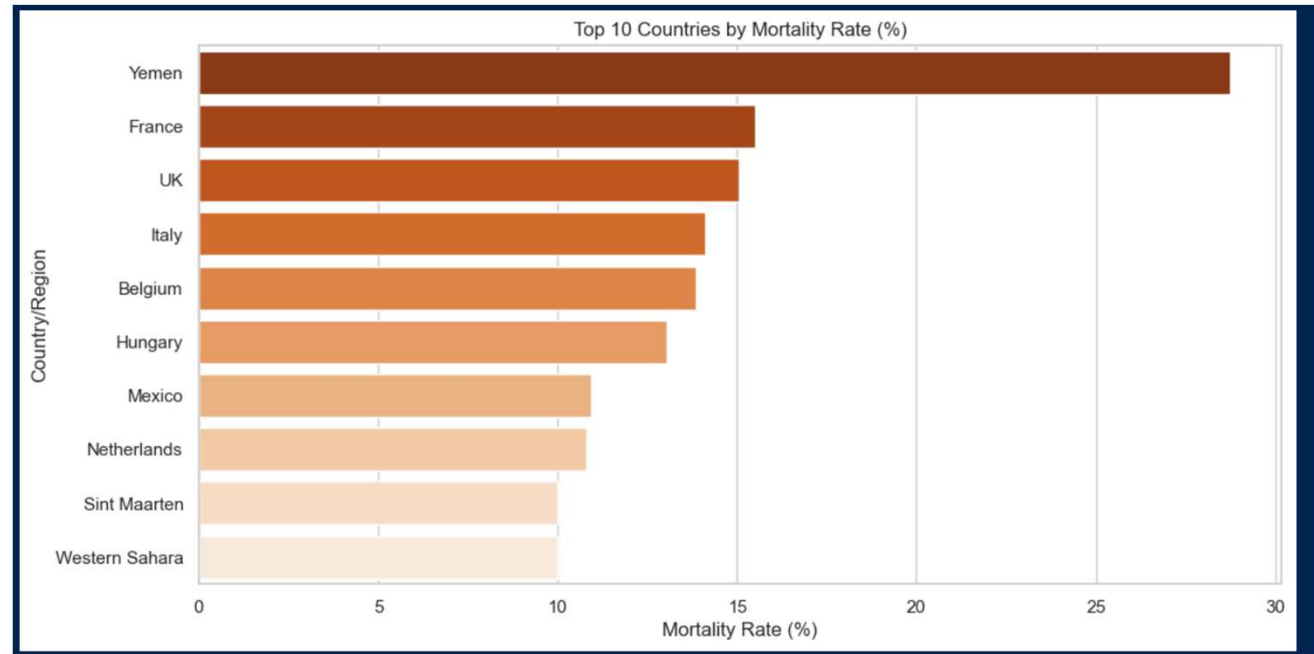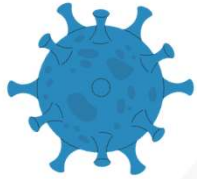
Submitted By:
Samir Srinath