Covid-19 Global Tracker

May 14, 2025

[1]:
```
# Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime

# Set style for plots
plt.style.use('ggplot')
%matplotlib inline

# Load the dataset
data = pd.read_csv('owid-covid-data.csv')

# Display basic information about the dataset
print("Dataset shape:", data.shape)
print("\nFirst 5 rows:")
data.head()


# Display column information
print("Columns in the dataset:")
print(data.columns.tolist())

# Check data types and missing values
print("\nData types and missing values:")
data_info = pd.DataFrame({
    'Data Type': data.dtypes,
    'Missing Values': data.isnull().sum(),
    'Missing %': (data.isnull().sum() / len(data)) * 100
})
data_info.sort_values(by='Missing %', ascending=False)



# Convert date column to datetime format
data['date'] = pd.to_datetime(data['date'])
```

```
# Select columns we'll focus on for analysis
key_columns = [
    'iso_code', 'continent', 'location', 'date',
    'total_cases', 'new_cases', 'total_deaths', 'new_deaths',
    'total_cases_per_million', 'new_cases_per_million',
    'total_deaths_per_million', 'new_deaths_per_million',
    'total_vaccinations', 'people_vaccinated', 'people_fully_vaccinated',
    'population'
]


# Create a filtered dataframe with our key columns
df = data[key_columns].copy()

# Fill missing values for vaccination data with 0 (assuming no vaccinations yet)
vaccination_cols = ['total_vaccinations', 'people_vaccinated',␣
 ↪'people_fully_vaccinated']
df[vaccination_cols] = df[vaccination_cols].fillna(0)

# Calculate some additional metrics
df['case_fatality_rate'] = (df['total_deaths'] / df['total_cases']) * 100
df['vaccination_rate'] = (df['people_vaccinated'] / df['population']) * 100
df['full_vaccination_rate'] = (df['people_fully_vaccinated'] /␣
 ↪df['population']) * 100

# Display cleaned data info
print("Cleaned dataset info:")
df.info()




# Create a global aggregate by date
global_df = df.groupby('date').agg({
    'new_cases': 'sum',
    'new_deaths': 'sum',
    'total_cases': 'sum',
    'total_deaths': 'sum',
    'people_vaccinated': 'sum',
    'people_fully_vaccinated': 'sum'
}).reset_index()

# Calculate 7-day rolling averages for smoothing
global_df['new_cases_7day_avg'] = global_df['new_cases'].rolling(7).mean()
global_df['new_deaths_7day_avg'] = global_df['new_deaths'].rolling(7).mean()
```

```
# Plot global new cases and deaths over time
plt.figure(figsize=(16, 10))

# New cases plot
plt.subplot(2, 1, 1)
plt.plot(global_df['date'], global_df['new_cases_7day_avg'], color='blue')
plt.title('Global Daily New COVID-19 Cases (7-day Average)')
plt.ylabel('Cases')
plt.grid(True)

# New deaths plot
plt.subplot(2, 1, 2)
plt.plot(global_df['date'], global_df['new_deaths_7day_avg'], color='red')
plt.title('Global Daily New COVID-19 Deaths (7-day Average)')
plt.ylabel('Deaths')
plt.grid(True)

plt.tight_layout()
plt.show()



# Get the most recent data for each country
latest_data = df.sort_values('date').groupby('location').last().reset_index()

# Select top 20 countries by total cases
top_countries = latest_data.sort_values('total_cases', ascending=False).head(20)

# Plot total cases by country
plt.figure(figsize=(12, 8))
sns.barplot(x='total_cases', y='location', data=top_countries,␣
 ↪palette='viridis')
plt.title('Top 20 Countries by Total COVID-19 Cases')
plt.xlabel('Total Cases')
plt.ylabel('Country')
plt.show()

# Plot total deaths by country
plt.figure(figsize=(12, 8))
sns.barplot(x='total_deaths', y='location', data=top_countries, palette='magma')
plt.title('Top 20 Countries by Total COVID-19 Deaths')
plt.xlabel('Total Deaths')
plt.ylabel('Country')
plt.show()
```

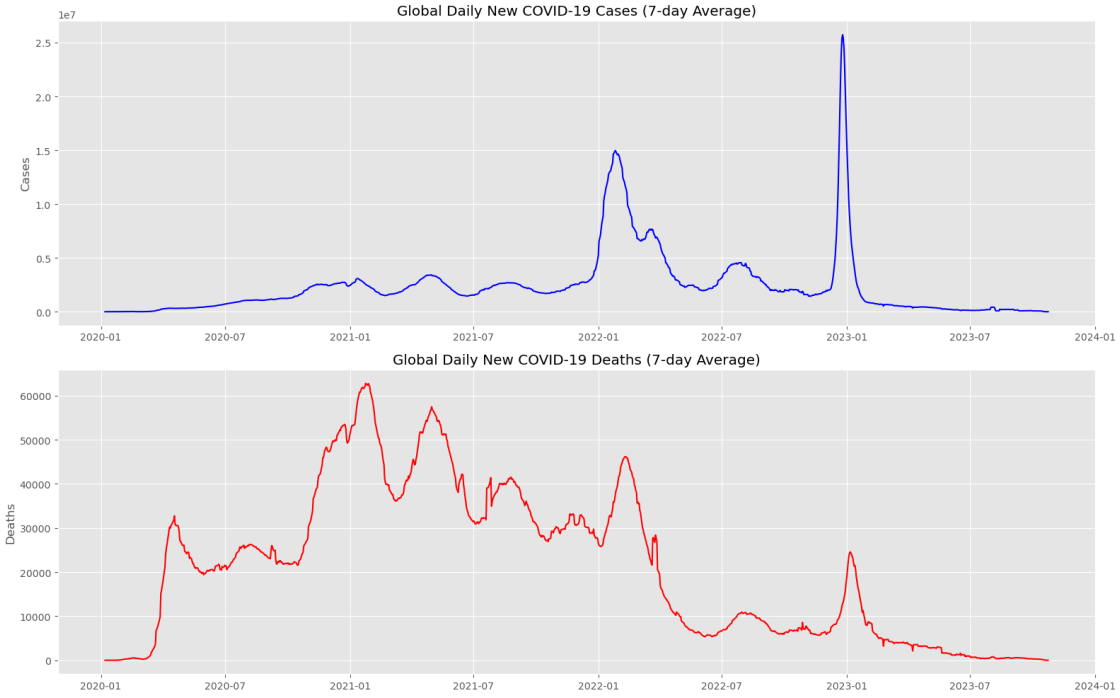
```
# Filter countries with at least 1000 cases to get meaningful CFR
cfr_df = latest_data[latest_data['total_cases'] > 1000].
 ↪sort_values('case_fatality_rate', ascending=False)

# Plot top 20 countries by CFR
plt.figure(figsize=(12, 8))
sns.barplot(x='case_fatality_rate', y='location', data=cfr_df.head(20),␣
 ↪palette='plasma')
plt.title('Top 20 Countries by COVID-19 Case Fatality Rate (Minimum 1,000␣
 ↪Cases)')
plt.xlabel('Case Fatality Rate (%)')
plt.ylabel('Country')
plt.show()


# Select countries with vaccination data
vaccination_df = latest_data[latest_data['people_vaccinated'] > 0].
 ↪sort_values('vaccination_rate', ascending=False)

# Plot top 20 countries by vaccination rate
plt.figure(figsize=(12, 8))
sns.barplot(x='vaccination_rate', y='location', data=vaccination_df.head(20),␣
 ↪palette='coolwarm')
plt.title('Top 20 Countries by COVID-19 Vaccination Rate (% of Population)')
plt.xlabel('Vaccination Rate (%)')
plt.ylabel('Country')
plt.show()

# Plot full vaccination rates
full_vaccination_df = latest_data[latest_data['people_fully_vaccinated'] > 0].
 ↪sort_values('full_vaccination_rate', ascending=False)

plt.figure(figsize=(12, 8))
sns.barplot(x='full_vaccination_rate', y='location', data=full_vaccination_df.
 ↪head(20), palette='summer')
plt.title('Top 20 Countries by Full COVID-19 Vaccination Rate (% of␣
 ↪Population)')
plt.xlabel('Fully Vaccinated Rate (%)')
plt.ylabel('Country')
plt.show()



# Select some countries for comparison
```

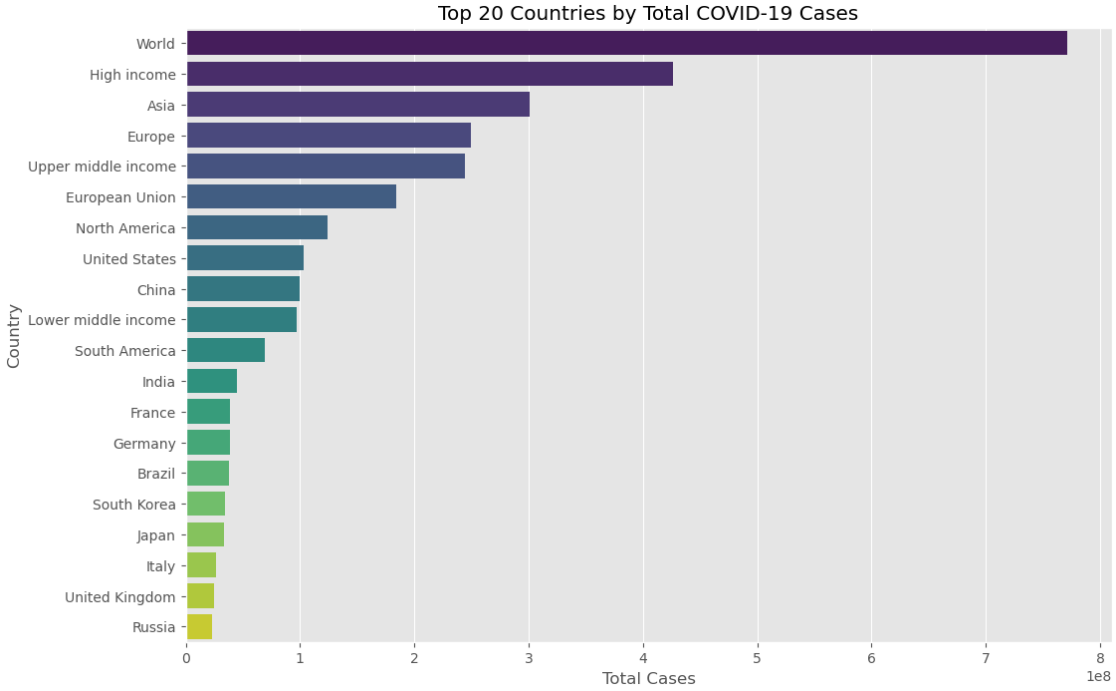
```
selected_countries = ['United States', 'India', 'Brazil', 'United Kingdom',␣
 ↪'Germany', 'Japan', 'South Africa']

# Filter data for selected countries
country_df = df[df['location'].isin(selected_countries)]

# Plot new cases over time for selected countries
plt.figure(figsize=(16, 8))
for country in selected_countries:
    country_data = country_df[country_df['location'] == country]
    plt.plot(country_data['date'], country_data['new_cases'].rolling(7).mean(),␣
 ↪label=country)

plt.title('Daily New COVID-19 Cases (7-day Average) in Selected Countries')
plt.xlabel('Date')
plt.ylabel('New Cases')
plt.legend()
plt.grid(True)
plt.show()

# Plot vaccination progress over time for selected countries
plt.figure(figsize=(16, 8))
for country in selected_countries:
    country_data = country_df[country_df['location'] == country]
    plt.plot(country_data['date'], country_data['vaccination_rate'],␣
 ↪label=f'{country} - Vaccinated')
    plt.plot(country_data['date'], country_data['full_vaccination_rate'], '--',␣
 ↪label=f'{country} - Fully Vaccinated')

plt.title('COVID-19 Vaccination Progress in Selected Countries')
plt.xlabel('Date')
plt.ylabel('Percentage of Population')
plt.legend()
plt.grid(True)
plt.show()




# Select numeric columns for correlation analysis
numeric_cols = [
    'total_cases', 'total_deaths', 'total_cases_per_million',
    'total_deaths_per_million', 'case_fatality_rate',
    'vaccination_rate', 'full_vaccination_rate', 'population'
]

# Calculate correlation matrix
```

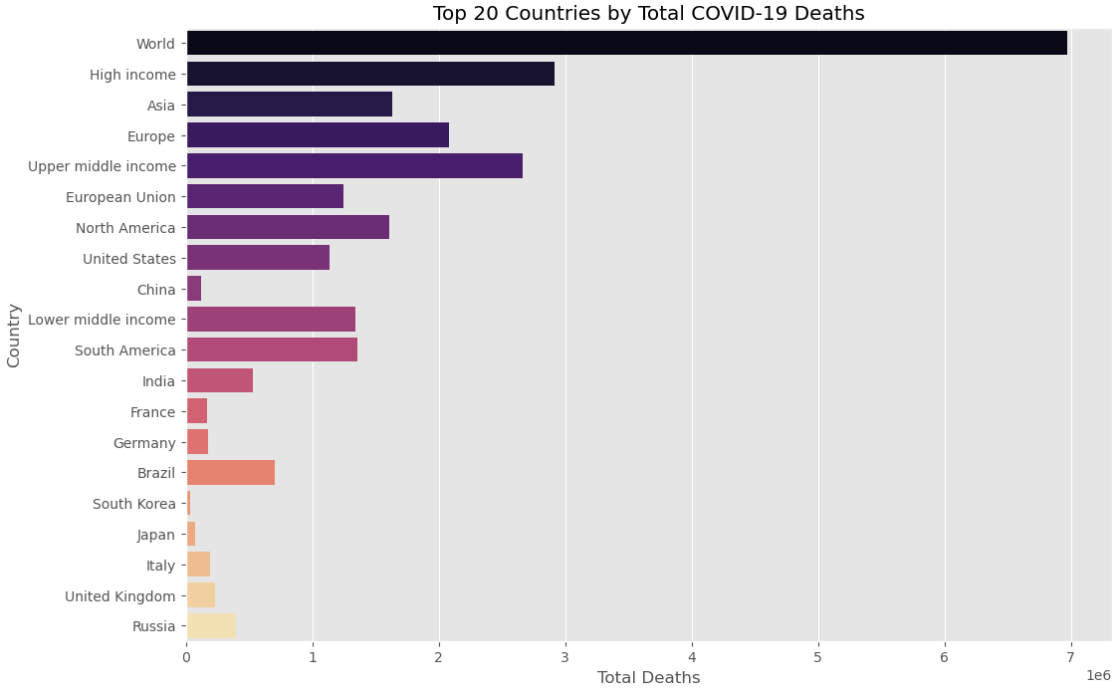
```
corr_matrix = latest_data[numeric_cols].corr()

# Plot correlation heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', center=0)
plt.title('Correlation Matrix of COVID-19 Metrics')
plt.show()




df.to_csv('cleaned_covid_data.csv', index=False)
print("Analysis complete. Cleaned data saved to 'cleaned_covid_data.csv'")
```

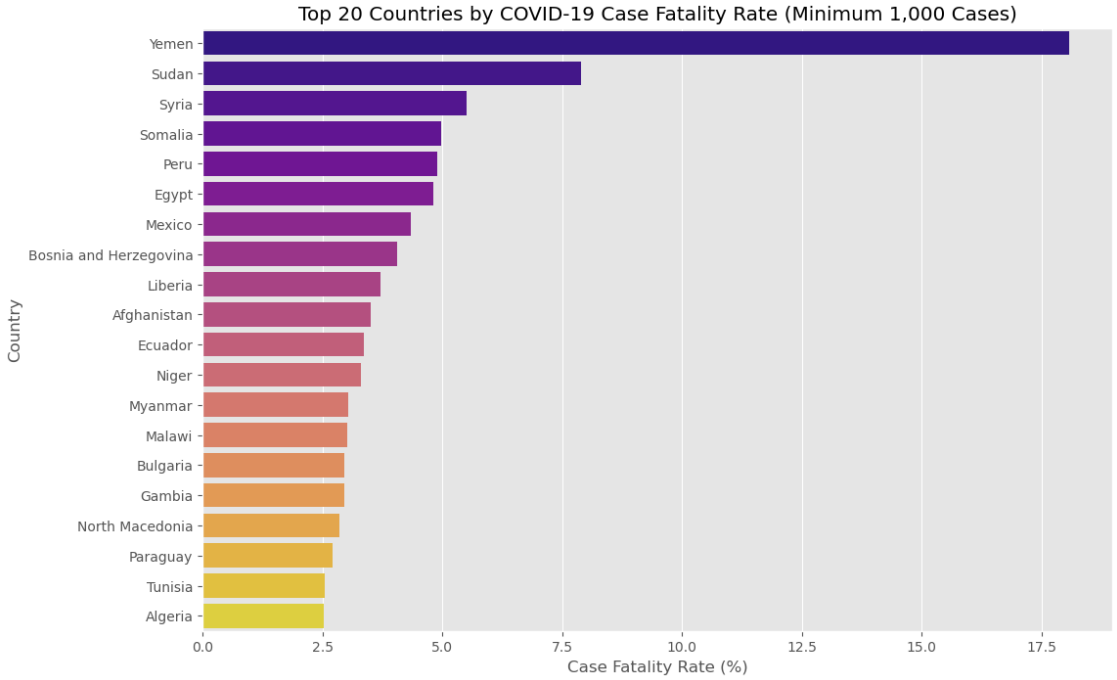
```
Dataset shape: (350085, 67)

First 5 rows:
Columns in the dataset:
['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases',
'new_cases_smoothed', 'total_deaths', 'new_deaths', 'new_deaths_smoothed',
'total_cases_per_million', 'new_cases_per_million',
'new_cases_smoothed_per_million', 'total_deaths_per_million',
'new_deaths_per_million', 'new_deaths_smoothed_per_million',
'reproduction_rate', 'icu_patients', 'icu_patients_per_million',
'hosp_patients', 'hosp_patients_per_million', 'weekly_icu_admissions',
'weekly_icu_admissions_per_million', 'weekly_hosp_admissions',
'weekly_hosp_admissions_per_million', 'total_tests', 'new_tests',
'total_tests_per_thousand', 'new_tests_per_thousand', 'new_tests_smoothed',
'new_tests_smoothed_per_thousand', 'positive_rate', 'tests_per_case',
'tests_units', 'total_vaccinations', 'people_vaccinated',
'people_fully_vaccinated', 'total_boosters', 'new_vaccinations',
'new_vaccinations_smoothed', 'total_vaccinations_per_hundred',
'people_vaccinated_per_hundred', 'people_fully_vaccinated_per_hundred',
'total_boosters_per_hundred', 'new_vaccinations_smoothed_per_million',
'new_people_vaccinated_smoothed', 'new_people_vaccinated_smoothed_per_hundred',
'stringency_index', 'population_density', 'median_age', 'aged_65_older',
'aged_70_older', 'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate',
'diabetes_prevalence', 'female_smokers', 'male_smokers',
'handwashing_facilities', 'hospital_beds_per_thousand', 'life_expectancy',
'human_development_index', 'population', 'excess_mortality_cumulative_absolute',
'excess_mortality_cumulative', 'excess_mortality',
'excess_mortality_cumulative_per_million']

Data types and missing values:
Cleaned dataset info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 350085 entries, 0 to 350084
```

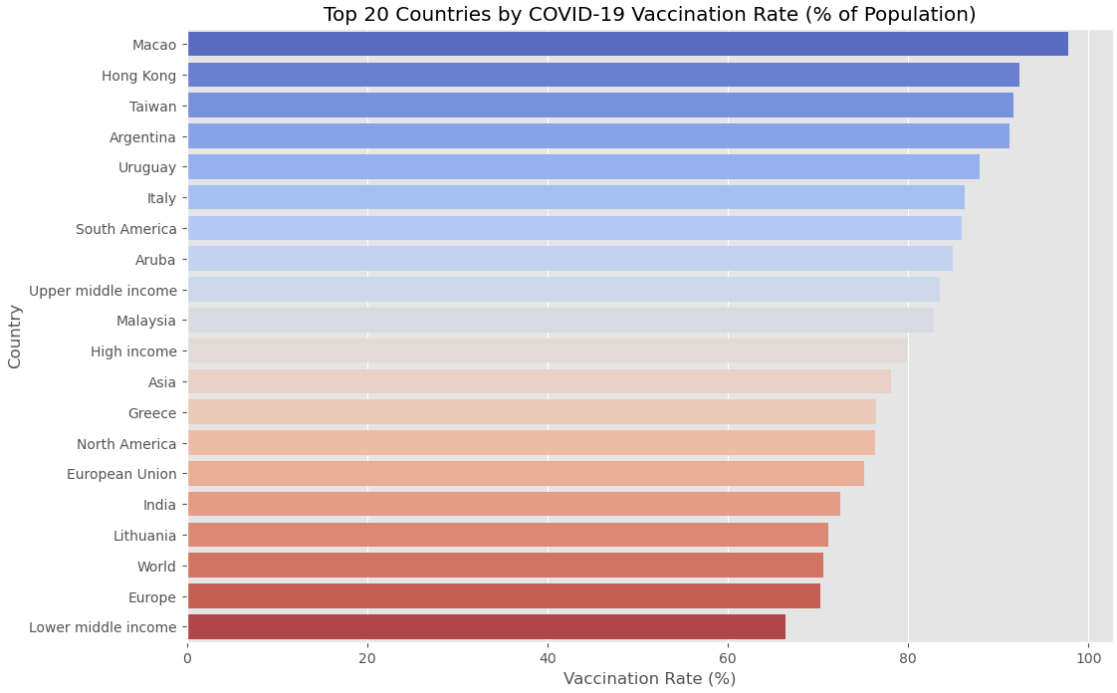
```
Data columns (total 19 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   iso_code                  350085 non-null  object
 1   continent                 333420 non-null  object
 2   location                  350085 non-null  object
 3   date                      350085 non-null  datetime64[ns]
 4   total_cases               312088 non-null  float64
 5   new_cases                 340457 non-null  float64
 6   total_deaths              290501 non-null  float64
 7   new_deaths                340511 non-null  float64
 8   total_cases_per_million   312088 non-null  float64
 9   new_cases_per_million     340457 non-null  float64
 10  total_deaths_per_million  290501 non-null  float64
 11  new_deaths_per_million    340511 non-null  float64
 12  total_vaccinations        350085 non-null  float64
 13  people_vaccinated         350085 non-null  float64
 14  people_fully_vaccinated   350085 non-null  float64
 15  population                350085 non-null  float64
 16  case_fatality_rate        290479 non-null  float64
 17  vaccination_rate          350085 non-null  float64
 18  full_vaccination_rate     350085 non-null  float64
dtypes: datetime64[ns](1), float64(15), object(3)
memory usage: 50.7+ MB
```

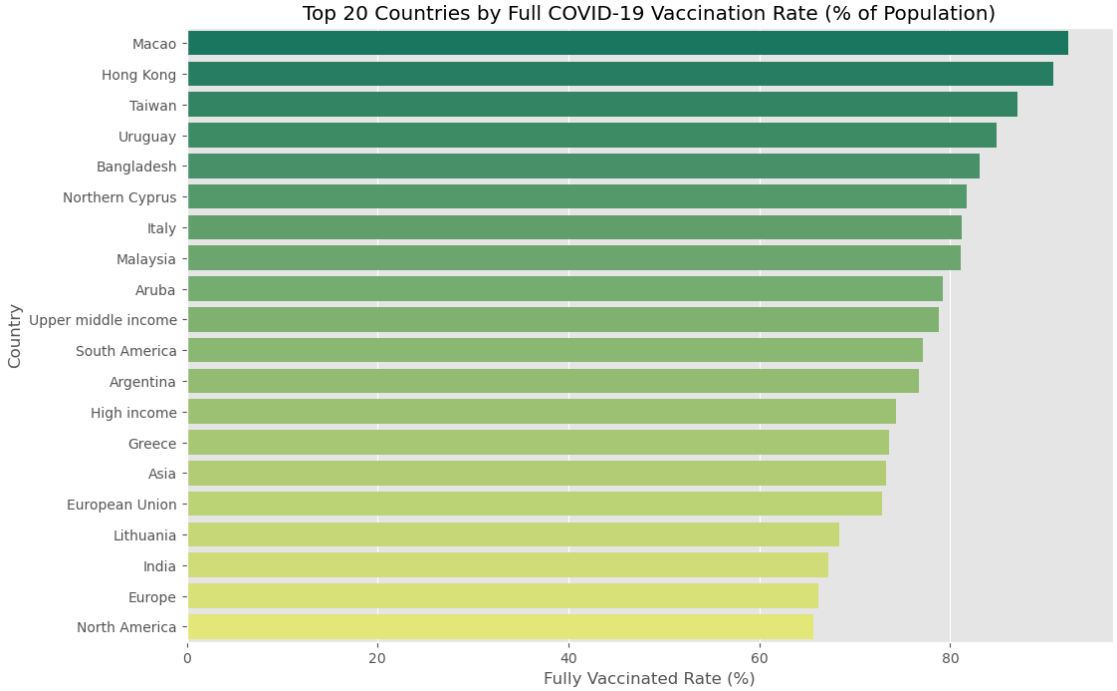


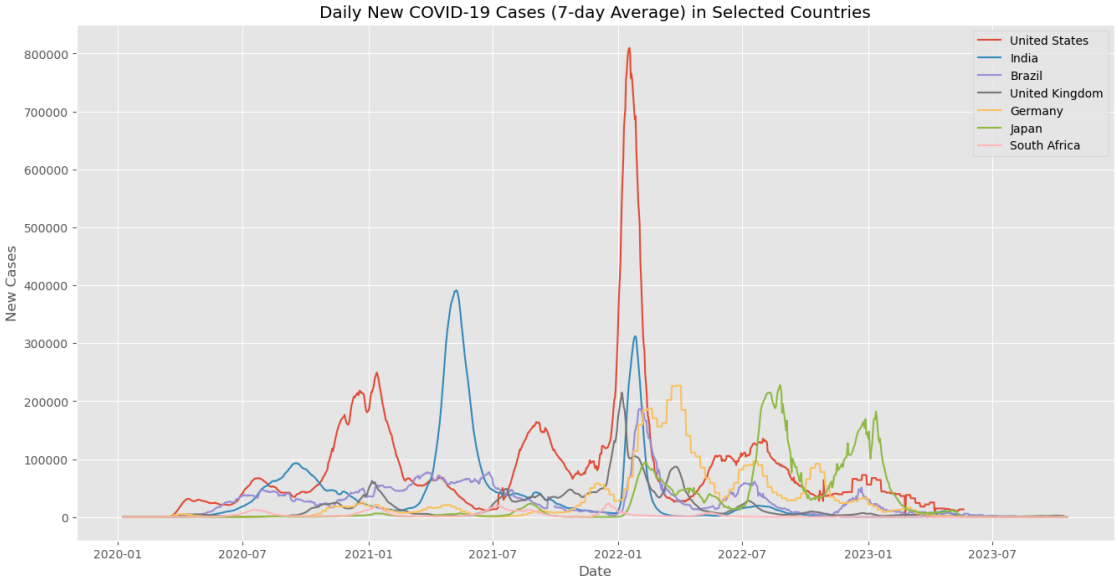

Top 20 Countries by Total COVID-19 Cases
World
High income
Asia
Europe
Upper middle income
European Union
North America
United States
China
Lower middle income
South America
India
France
Germany
Brazil
South Korea
Japan
Italy
United Kingdom
Russia
Country
0
1
2
3
4
5
6
7
8
Total Cases
1e8

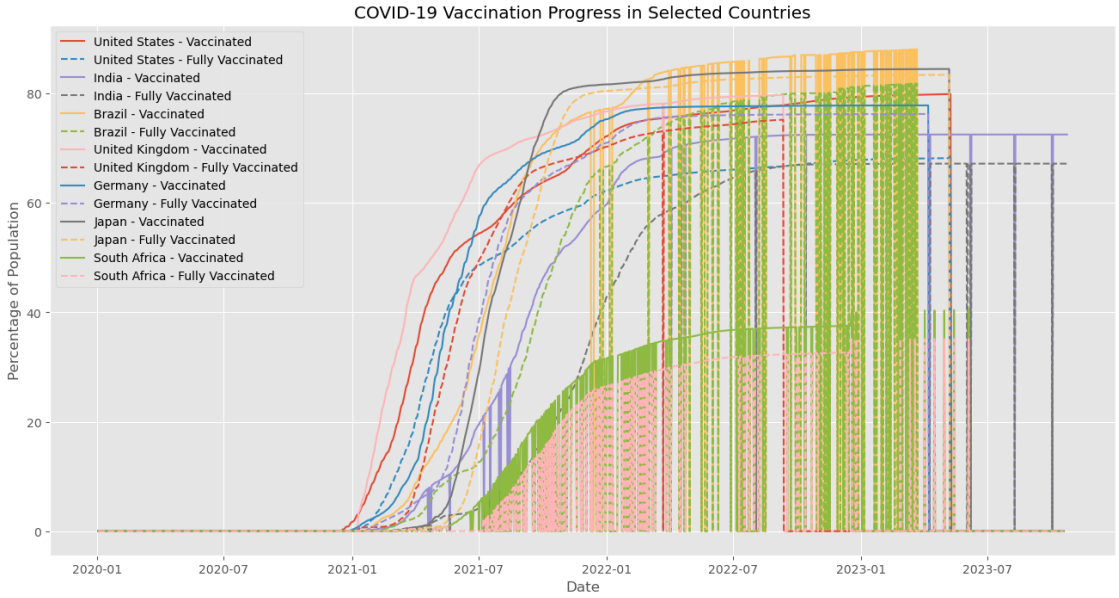

Top 20 Countries by Total COVID-19 Deaths
World
High income
Asia
Europe
Upper middle income
European Union
North America
United States
China
Lower middle income
South America
India
France
Germany
Brazil
South Korea
Japan
Italy
United Kingdom
Russia
Country
0
1
2
3
4
5
6
7
Total Deaths
1e6

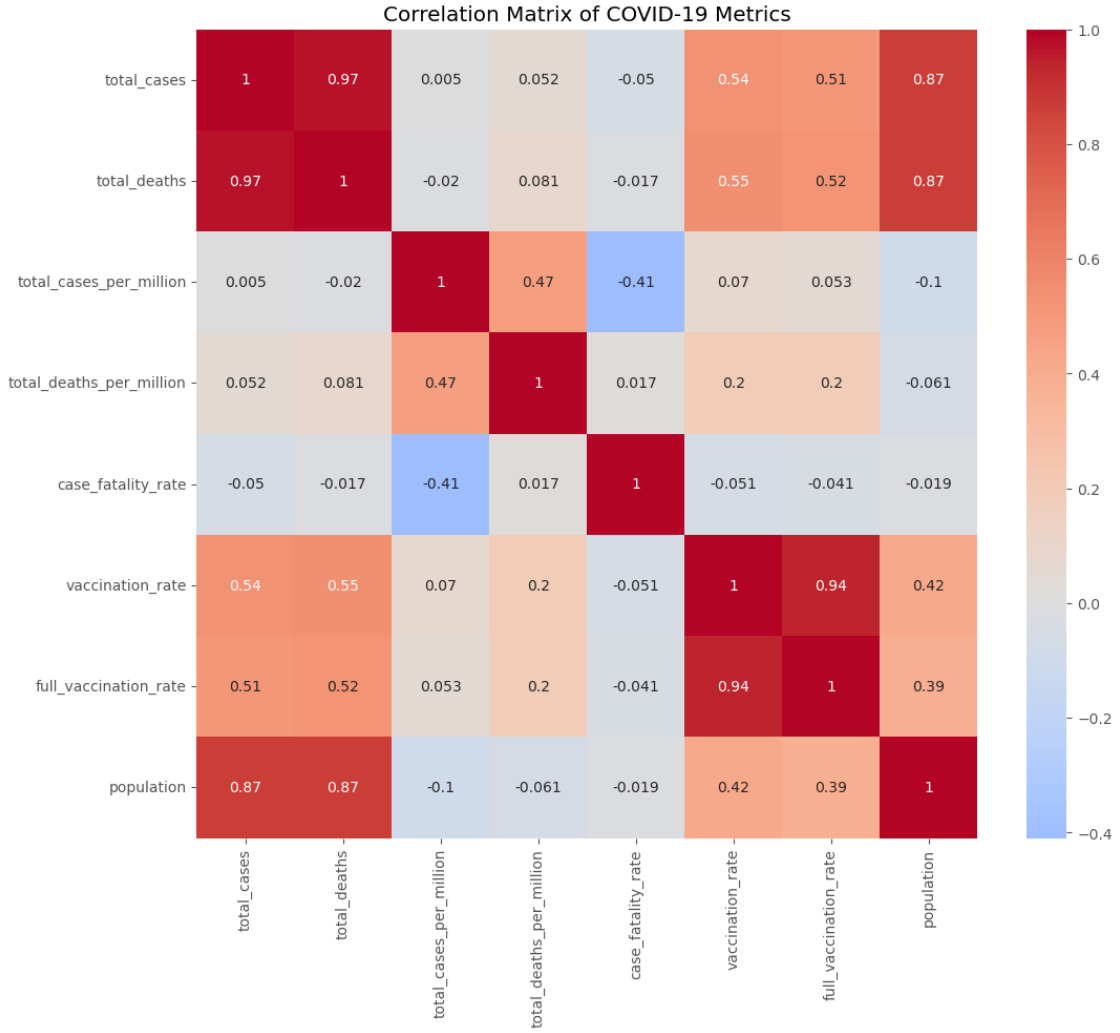

Top 20 Countries by COVID-19 Case Fatality Rate (Minimum 1,000 Cases)
Yemen
Sudan
Syria
Somalia
Peru
Egypt
Mexico
Bosnia and Herzegovina
Liberia
Afghanistan
Ecuador
Niger
Myanmar
Malawi
Bulgaria
Gambia
North Macedonia
Paraguay
Tunisia
Algeria
Country
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Case Fatality Rate (%)


Top 20 Countries by COVID-19 Vaccination Rate (% of Population)
Macao
Hong Kong
Taiwan
Argentina
Uruguay
Italy
South America
Aruba
Upper middle income
Malaysia
High income
Asia
Greece
North America
European Union
India
Lithuania
World
Europe
Lower middle income
Country
0
20
40
60
80
100
Vaccination Rate (%)


Top 20 Countries by Full COVID-19 Vaccination Rate (% of Population)
Macao
Hong Kong
Taiwan
Uruguay
Bangladesh
Northern Cyprus
Italy
Malaysia
Aruba
Upper middle income
South America
Argentina
High income
Greece
Asia
European Union
Lithuania
India
Europe
North America
Country
0
20
40
60
80
Fully Vaccinated Rate (%)


Daily New COVID-19 Cases (7-day Average) in Selected Countries
United States
India
Brazil
United Kingdom
Germany
Japan
South Africa
800000
700000
600000
500000
400000
300000
200000
100000
0
New Cases
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
Date


COVID-19 Vaccination Progress in Selected Countries
United States - Vaccinated
United States - Fully Vaccinated
India - Vaccinated
India - Fully Vaccinated
Brazil - Vaccinated
Brazil - Fully Vaccinated
United Kingdom - Vaccinated
United Kingdom - Fully Vaccinated
Germany - Vaccinated
Germany - Fully Vaccinated
Japan - Vaccinated
Japan - Fully Vaccinated
South Africa - Vaccinated
South Africa - Fully Vaccinated
Percentage of Population
0
20
40
60
80
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
Date




```
Analysis complete. Cleaned data saved to 'cleaned_covid_data.csv'
```

[ ]: