

COVID-19 PROJECT

May 5, 2025

```
[1]: import pandas as pd

# Load data directly from OWID URL
url = "https://covid.ourworldindata.org/data/owid-covid-data.csv"
df = pd.read_csv(url)

# Preview the data
df.head()
```

```
[1]: iso_code continent    location    date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan  2020-01-05         0.0         0.0
1      AFG      Asia  Afghanistan  2020-01-06         0.0         0.0
2      AFG      Asia  Afghanistan  2020-01-07         0.0         0.0
3      AFG      Asia  Afghanistan  2020-01-08         0.0         0.0
4      AFG      Asia  Afghanistan  2020-01-09         0.0         0.0

      new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  ...  \
0                NaN           0.0         0.0                NaN  ...
1                NaN           0.0         0.0                NaN  ...
2                NaN           0.0         0.0                NaN  ...
3                NaN           0.0         0.0                NaN  ...
4                NaN           0.0         0.0                NaN  ...

      male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0             NaN                37.746                0.5
1             NaN                37.746                0.5
2             NaN                37.746                0.5
3             NaN                37.746                0.5
4             NaN                37.746                0.5

      life_expectancy  human_development_index  population  \
0             64.83                0.511    41128772
1             64.83                0.511    41128772
2             64.83                0.511    41128772
3             64.83                0.511    41128772
4             64.83                0.511    41128772
```

	excess_mortality_cumulative_absolute	excess_mortality_cumulative	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	

	excess_mortality	excess_mortality_cumulative_per_million
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

[5 rows x 67 columns]

```
[39]: # Check for missing values
df.isnull().sum()
```

```
[39]: iso_code          0
continent          26525
location           0
date               0
total_cases       17631
...
population         0
excess_mortality_cumulative_absolute  416024
excess_mortality_cumulative          416024
excess_mortality                     416024
excess_mortality_cumulative_per_million  416024
Length: 67, dtype: int64
```

```
[40]: # Check the column names
df.columns
```

```
[40]: Index(['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'],
dtype='object')

```

```

[52]: # Convert date to datetime format
df['date'] = pd.to_datetime(df['date'])

# Filter for selected countries
countries = ['Africa', 'United States', 'India']
df_filtered = df[df['location'].isin(countries)]

```

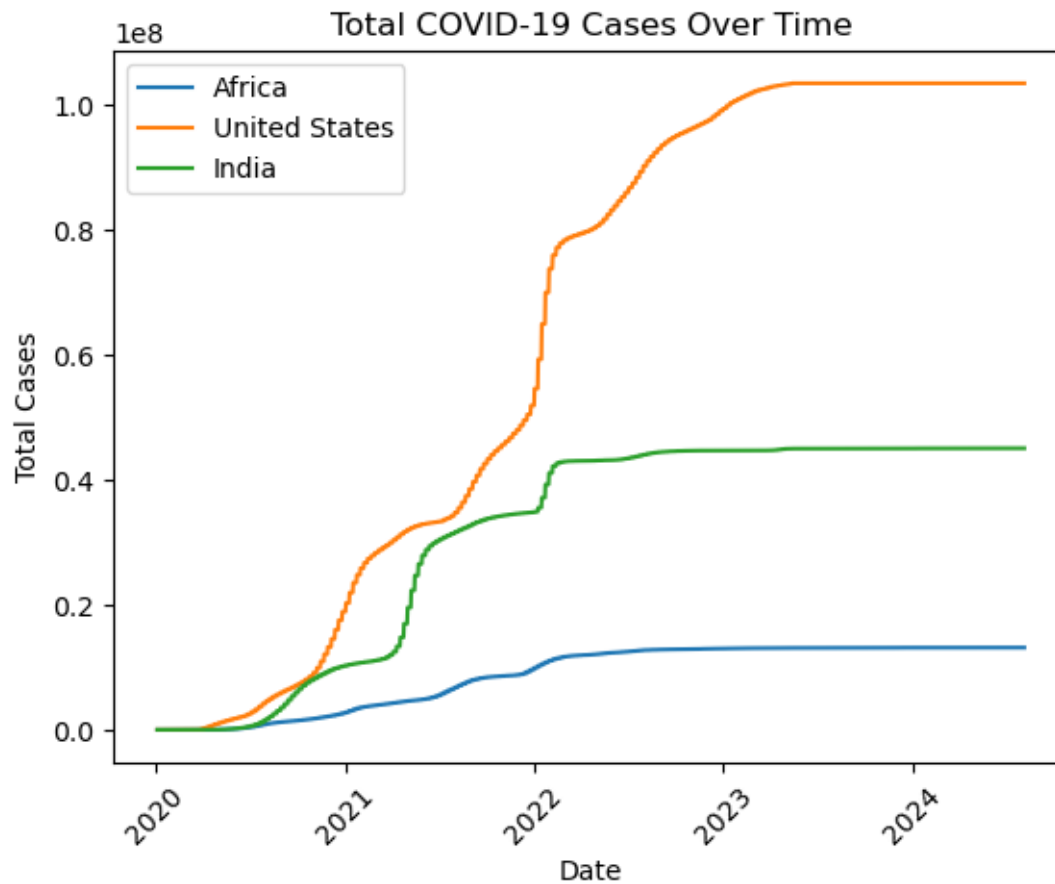
```

[56]: import matplotlib.pyplot as plt

# Plot total cases over time for each country
for country in countries:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_cases'], label=country)

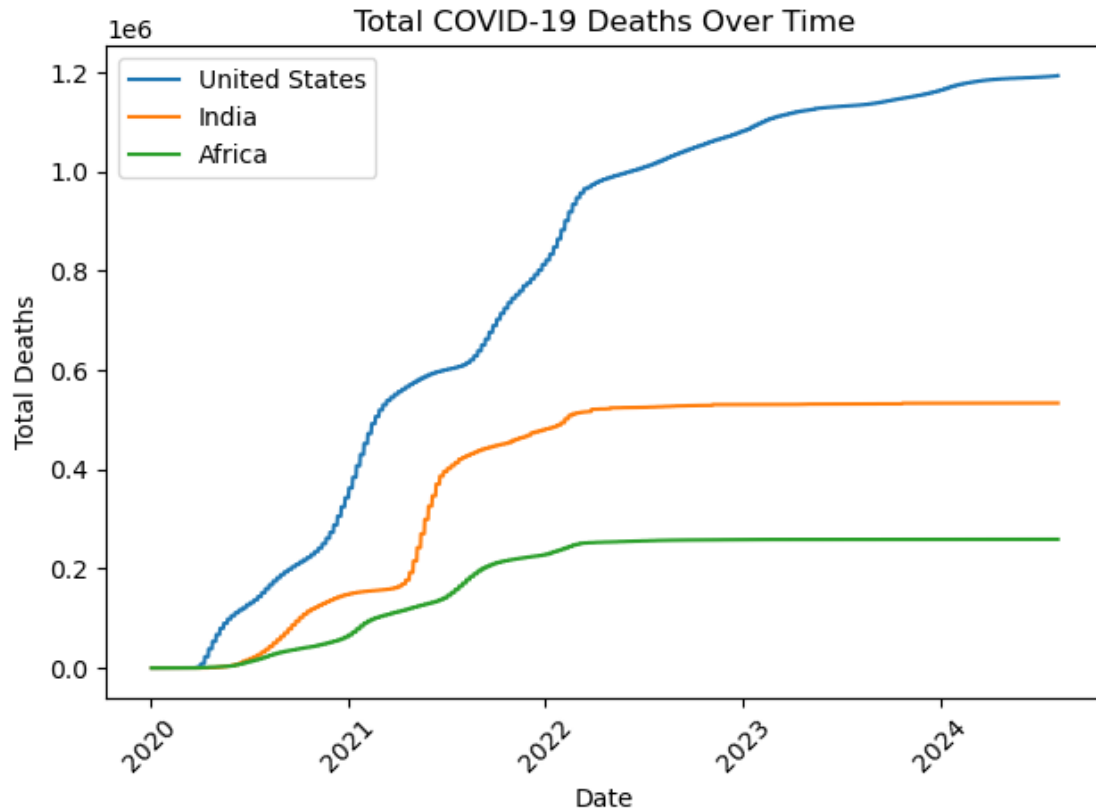
plt.title("Total COVID-19 Cases Over Time")
plt.xlabel("Date")
plt.ylabel("Total Cases")
plt.legend()
plt.xticks(rotation=45)
plt.show()

```



```
[36]: # Plot total deaths over time
for country in countries:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_deaths'], label=country)

plt.title("Total COVID-19 Deaths Over Time")
plt.xlabel("Date")
plt.ylabel("Total Deaths")
plt.legend()
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

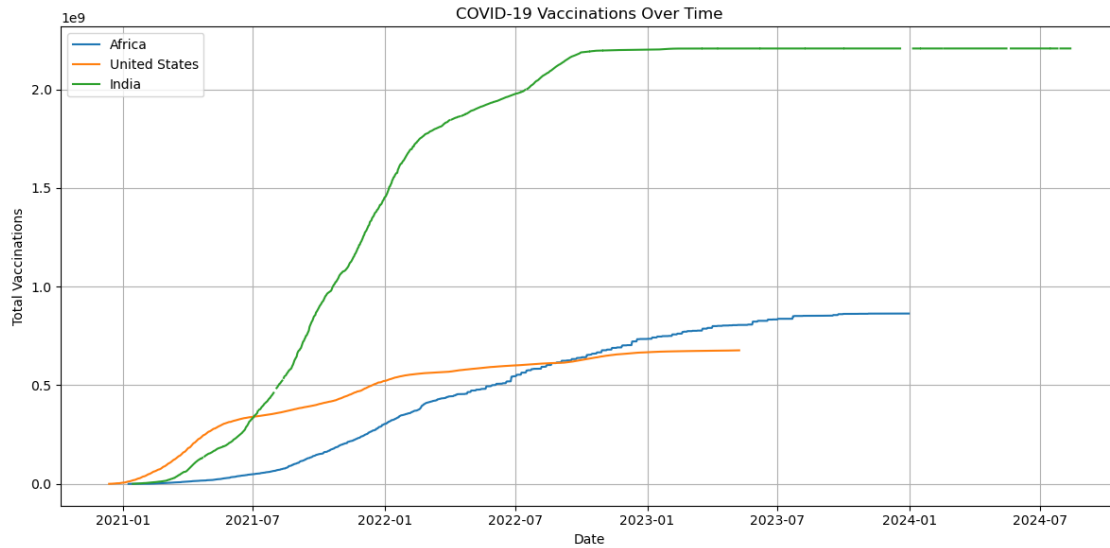


```
[57]: # Compare % of fully vaccinated population
import matplotlib.pyplot as plt

plt.figure(figsize=(12, 6))

for country in countries:
    country_data = df[df['location'] == country]
    plt.plot(country_data['date'], country_data['total_vaccinations'],
             label=country)

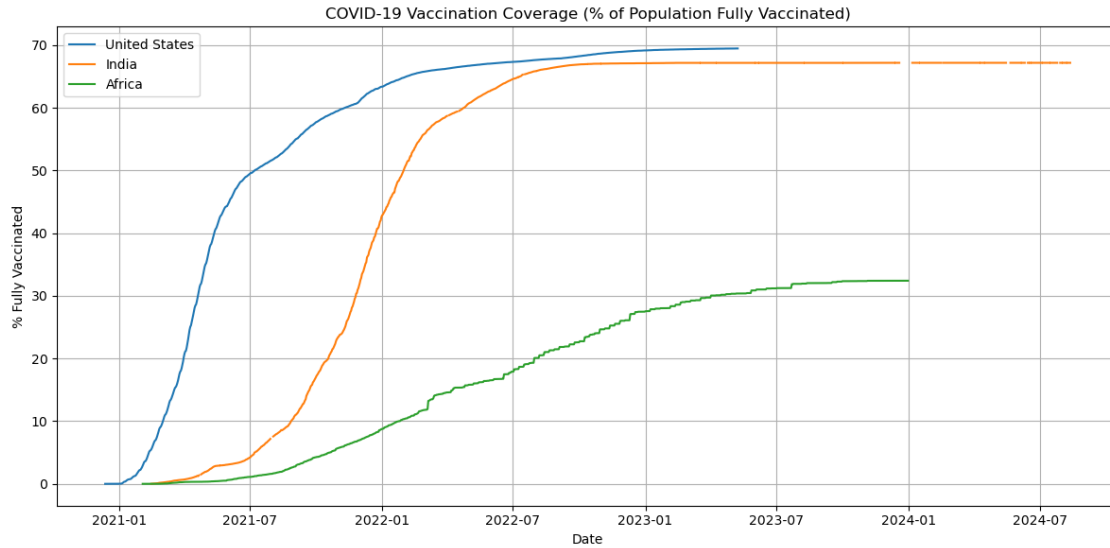
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.title('COVID-19 Vaccinations Over Time')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
[38]: plt.figure(figsize=(12, 6))

for country in countries:
    country_data = df[df['location'] == country]
    plt.plot(
        country_data['date'],
        country_data['people_fully_vaccinated_per_hundred'],
        label=country
    )

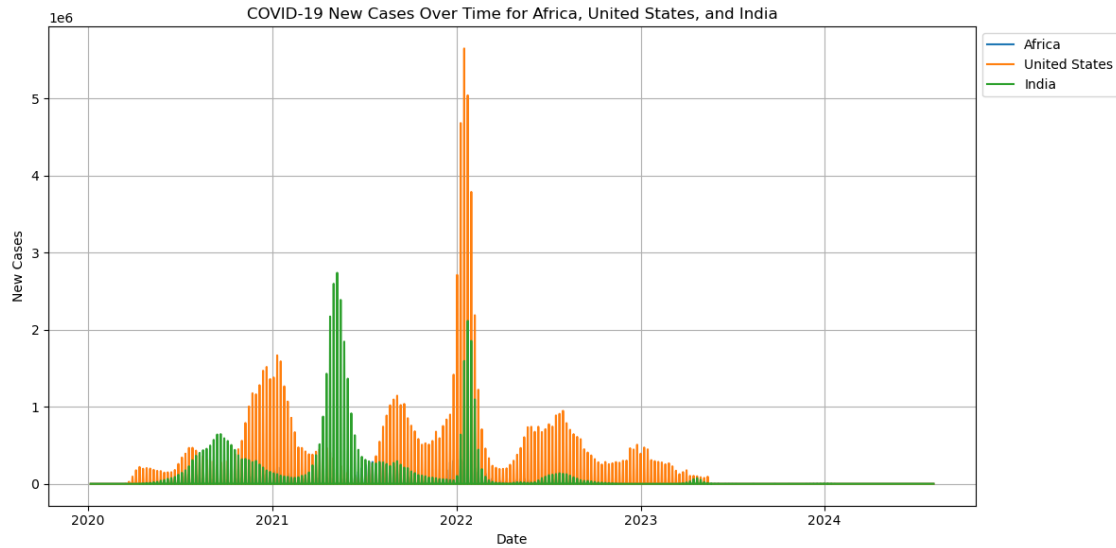
plt.xlabel('Date')
plt.ylabel('% Fully Vaccinated')
plt.title('COVID-19 Vaccination Coverage (% of Population Fully Vaccinated)')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
[58]: # Plot daily new cases for selected countries
plt.figure(figsize=(12, 6))

for country in countries:
    country_df = df_filtered[df_filtered['location'] == country].copy() # Make_
    ↪ a copy of the dataframe
    country_df['new_cases'] = country_df['total_cases'].diff() # Calculate_
    ↪ daily new cases safely
    plt.plot(country_df['date'], country_df['new_cases'], label=country)

plt.xlabel('Date')
plt.ylabel('New Cases')
plt.title('COVID-19 New Cases Over Time for Africa, United States, and India')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.grid(True)
plt.tight_layout()
plt.show()
```

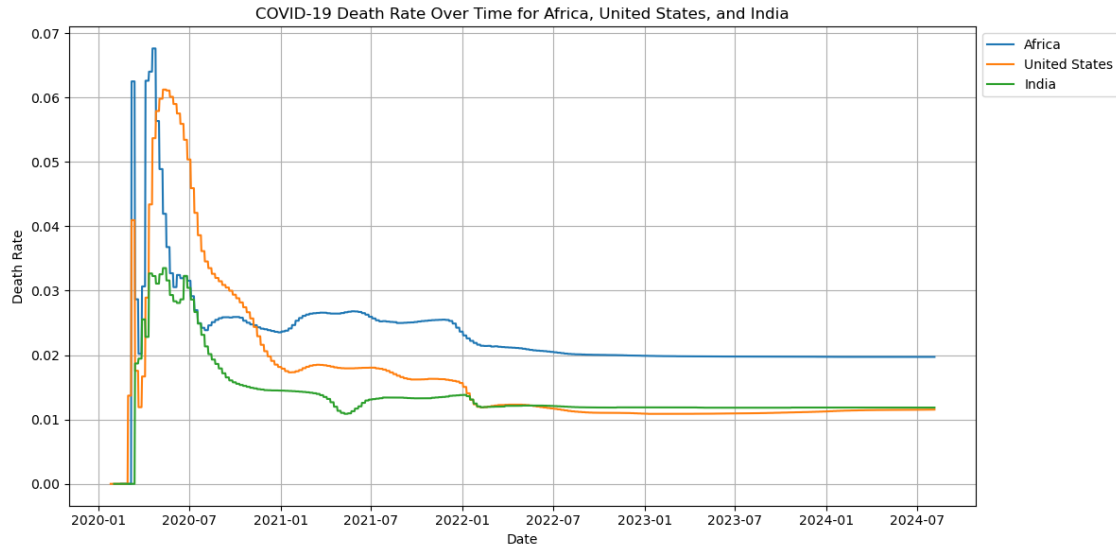


```
[81]: # Calculate the death rate: total_deaths / total_cases
df_filtered_copy['death_rate'] = df_filtered_copy['total_deaths'] / \
    df_filtered_copy['total_cases']

# Plot the death rate over time for selected countries
plt.figure(figsize=(12, 6))

for country in countries:
    country_df = df_filtered_copy[df_filtered_copy['location'] == country]
    plt.plot(country_df['date'], country_df['death_rate'], label=country)

plt.xlabel('Date')
plt.ylabel('Death Rate')
plt.title('COVID-19 Death Rate Over Time for Africa, United States, and India')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.grid(True)
plt.tight_layout()
plt.show()
```

[]: ## Key Insights and Analysis

1. **United States had High Total Cases**

- **United States** has the highest total number of COVID-19 cases among the selected countries, surpassing 100 million cases by 2022. Despite the high case count, the country managed to implement one of the largest vaccination programs globally. By mid-2022, the U.S. had vaccinated over 60% of its population, which helped to reduce the spread of the virus, despite having a higher number of cases compared to other countries.

2. **India had the a Fast Vaccine Rollout**

- **India** demonstrated an extremely rapid vaccine rollout, reaching over 60% of its population vaccinated in a few months during 2021. This quick deployment significantly helped to curb the spread of COVID-19 and reduce the severity of the pandemic, despite India initially being one of the hardest-hit countries in terms of cases and deaths.

3. **Africa had a Lower Overall Case Count**

- Many African countries, as represented in the **Africa** category, experienced relatively lower total case counts compared to the global averages. This could be attributed to several factors such as less widespread testing, younger populations, and differences in healthcare infrastructure. However, countries like **South Africa** and **Egypt** were significant outliers, experiencing higher case counts and mortality rates.

4. **Vaccination Trends Across the Region**

- The vaccination rates in **Africa** have been slower compared to countries like the U.S. and India, mainly due to challenges in vaccine access, distribution, and infrastructure. However, African countries have gradually ramped up their vaccination efforts, with nations like **Kenya** and **South Africa** making notable progress by 2022. Despite these efforts, the overall vaccination coverage across the continent remains lower than in other regions, which could influence future COVID-19 case trends.