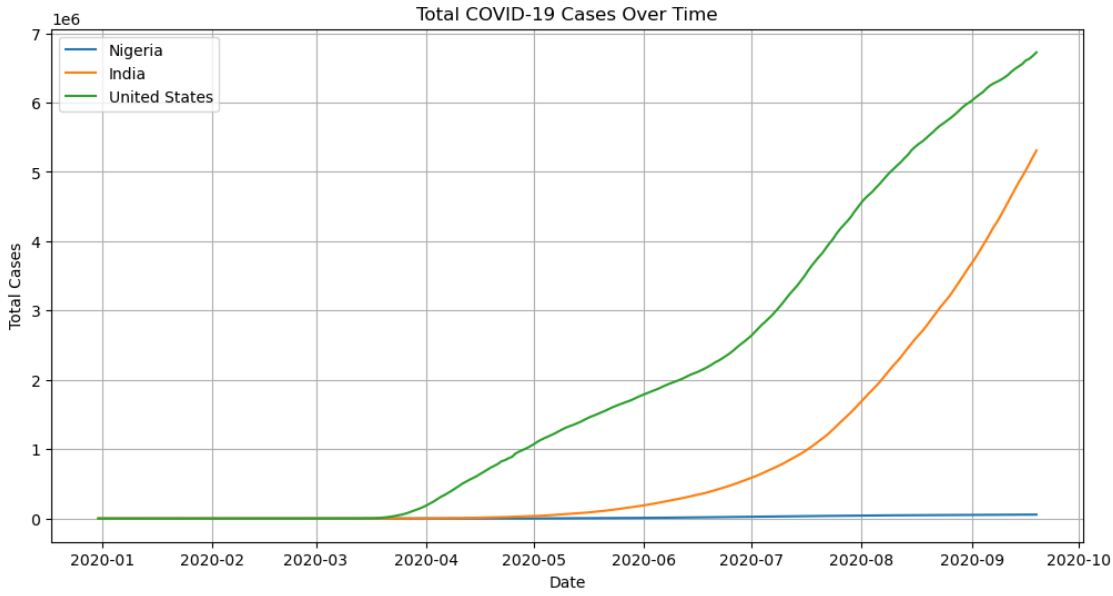
covid19 dataset

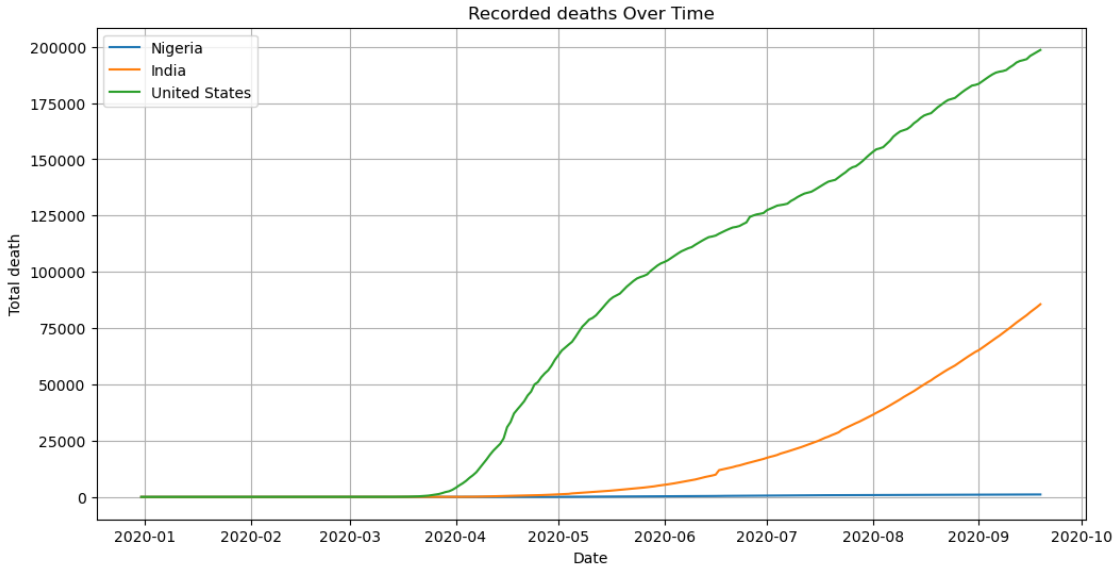
May 16, 2025

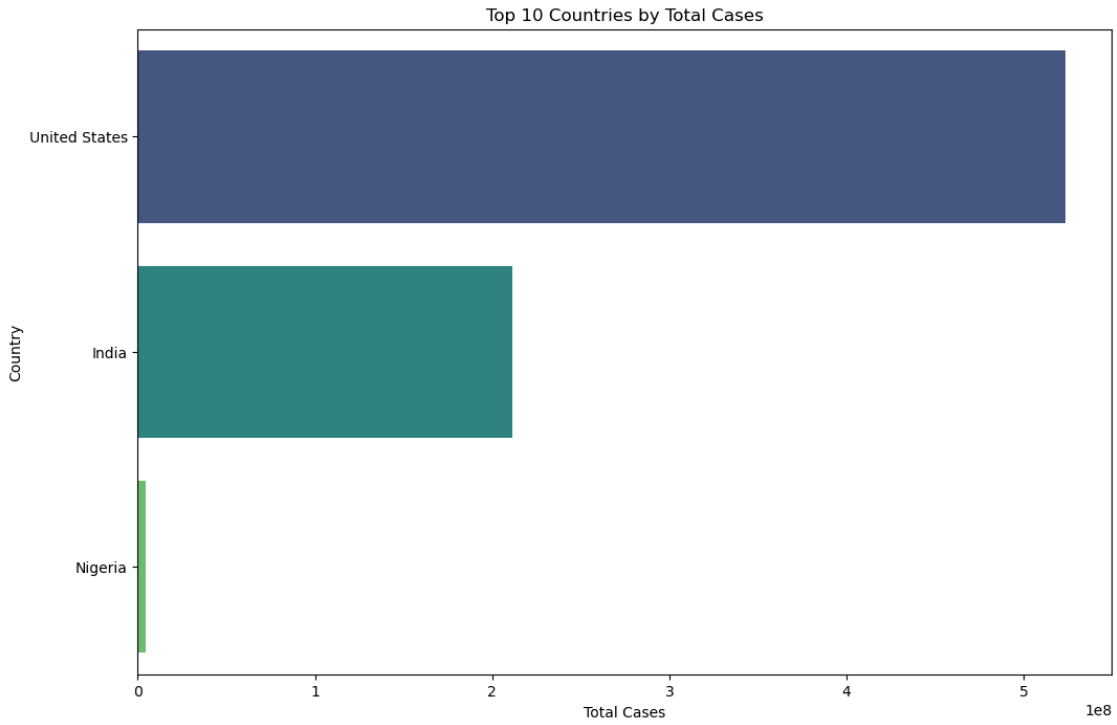
[35]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px


df = pd.read_csv('owid-covid-data.csv')
print(df.columns)
df.head()


df['date'] = pd.to_datetime(df['date'])

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

# Drop rows with missing critical values
df = df.dropna(subset=['total_cases', 'total_deaths'])

# Fill missing numeric values
df = df.fillna(method='ffill')

# Plot total cases over time
plt.figure(figsize=(12,6))
for country in countries:
    subset = df[df['location'] == country]
    plt.plot(subset['date'], subset['total_cases'], label=country)
plt.title('Total COVID-19 Cases Over Time')
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.legend()
plt.grid(True)
plt.show()
```

```
# Recorded death over time
plt.figure(figsize=(12,6))
for country in countries:
    subset = df[df['location'] == country]
    plt.plot(subset['date'], subset['total_deaths'], label=country)
plt.title('Recorded deaths Over Time')
plt.xlabel('Date')
plt.ylabel('Total death')
plt.legend()
plt.grid(True)
plt.show()

country_cases = df.groupby('location')['total_cases'].sum().reset_index()

# Sort countries by total cases in descending order
top_countries = country_cases.sort_values(by='total_cases', ascending=False).
 ↪head(10)

# Plotting a bar chart
plt.figure(figsize=(12, 8))
sns.barplot(x='total_cases', y='location', data=top_countries,␣
 ↪palette='viridis')
plt.title('Top 10 Countries by Total Cases')
plt.xlabel('Total Cases')
plt.ylabel('Country')
plt.show()

# Create choropleth map (latest date)
latest_date = df['date'].max()
latest_df = df[df['date'] == latest_date]



import pandas as pd
import plotly.express as px

fig = px.choropleth(
    df,
    locations='location',             # Column with country codes
    locationmode='ISO-3',             # Specify 'ISO-3' for country codes
    color='total_cases',              # Data to color by
    color_continuous_scale='Reds',
    title='Total COVID-19 Cases by Country',
    labels={'total_cases': 'Total Cases'}
)
print(df['location'].unique())
fig.show()
```

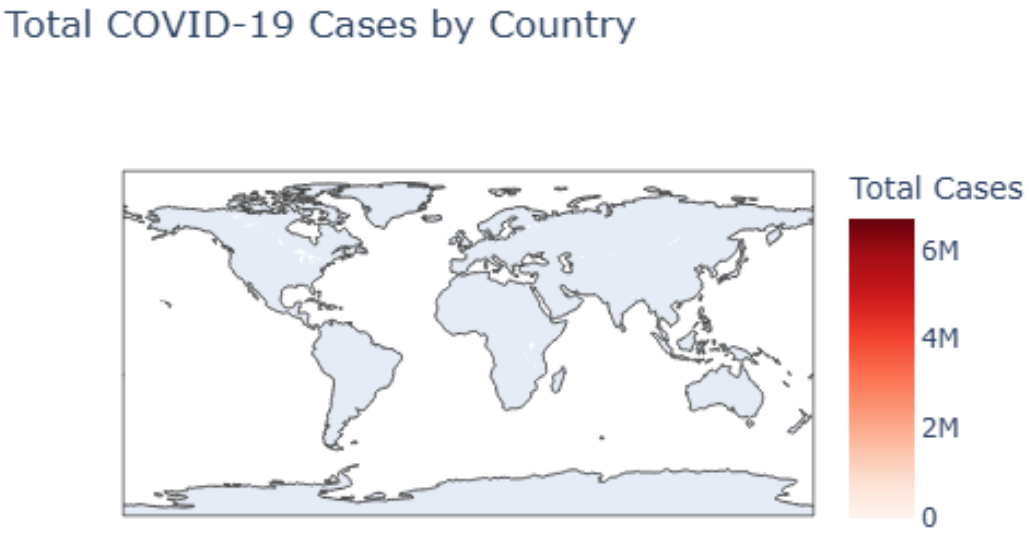
```
Index(['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases',
       'new_cases_smoothed', 'total_deaths', 'new_deaths',
       'new_deaths_smoothed', 'total_cases_per_million',
       'new_cases_per_million', 'new_cases_smoothed_per_million',
       'total_deaths_per_million', 'new_deaths_per_million',
       'new_deaths_smoothed_per_million', 'new_tests', 'total_tests',
       'total_tests_per_thousand', 'new_tests_per_thousand',
       'new_tests_smoothed', 'new_tests_smoothed_per_thousand',
       'tests_per_case', 'positive_rate', 'tests_units', 'stringency_index',
       'population', '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'],
      dtype='object')
```








```
['India' 'Nigeria' 'United States']
```


Total COVID-19 Cases by Country
Total Cases
6M
4M
2M
0

[ ]: