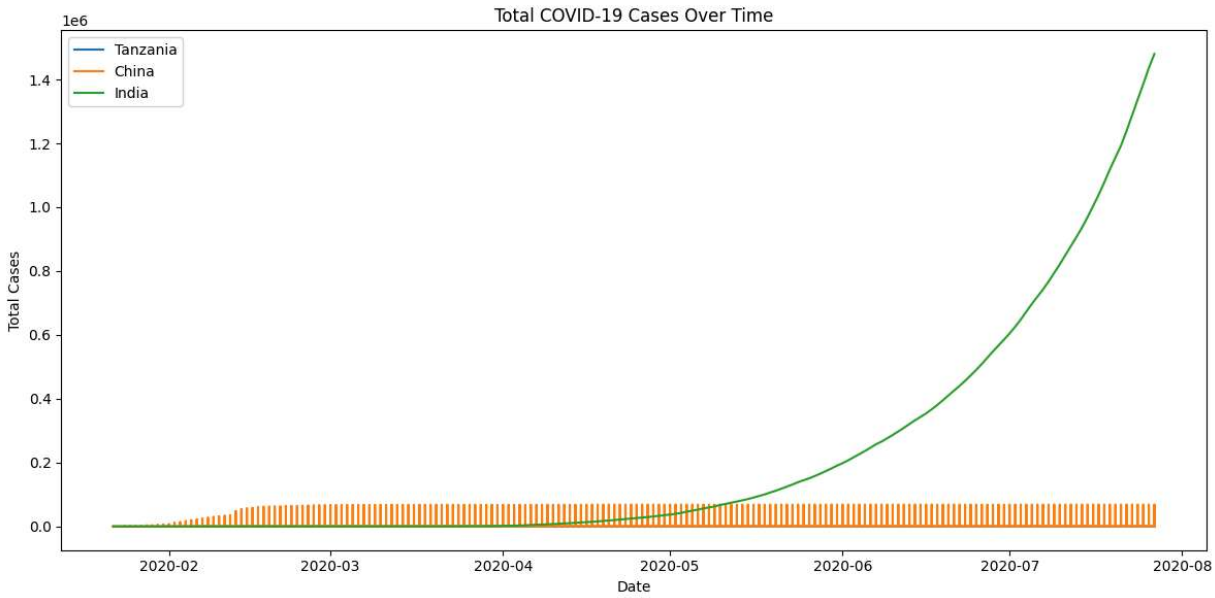
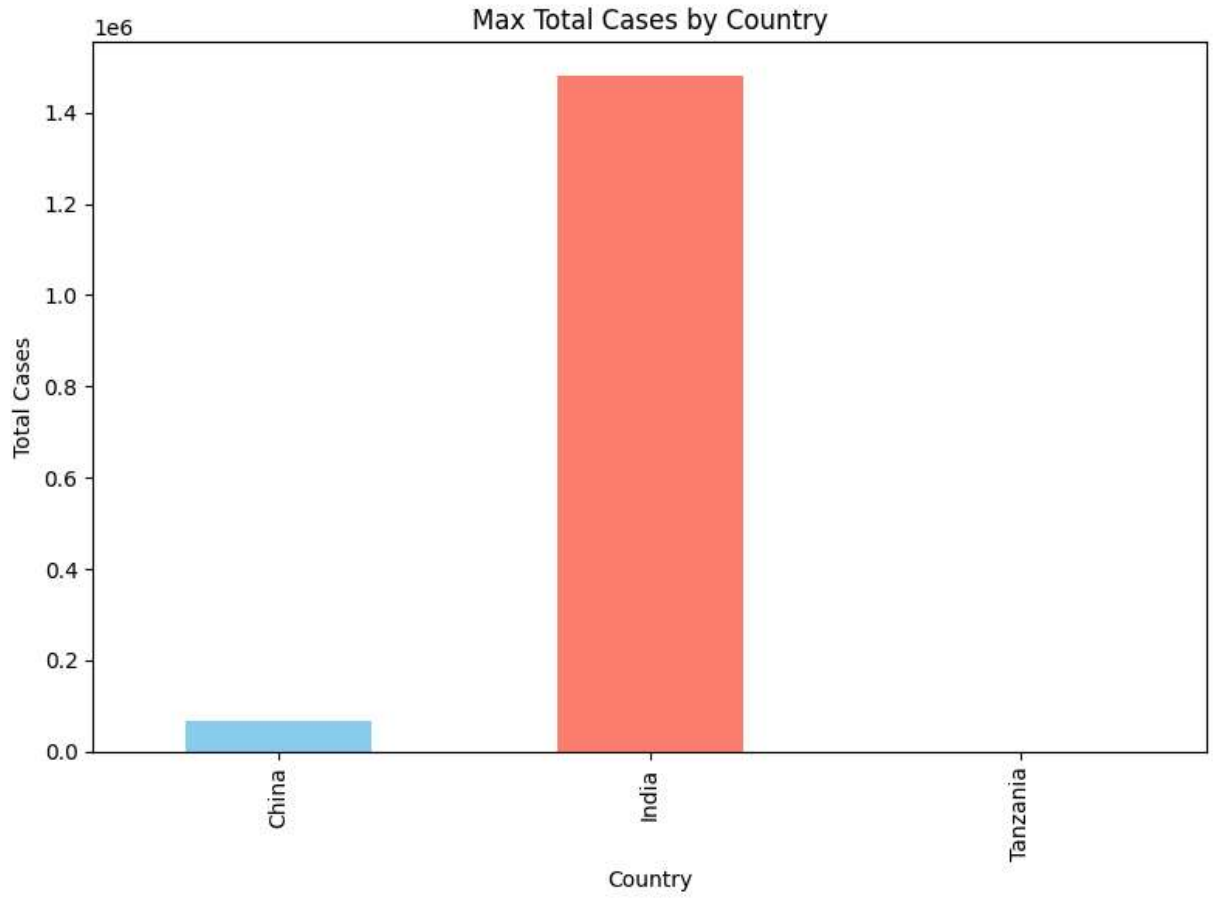
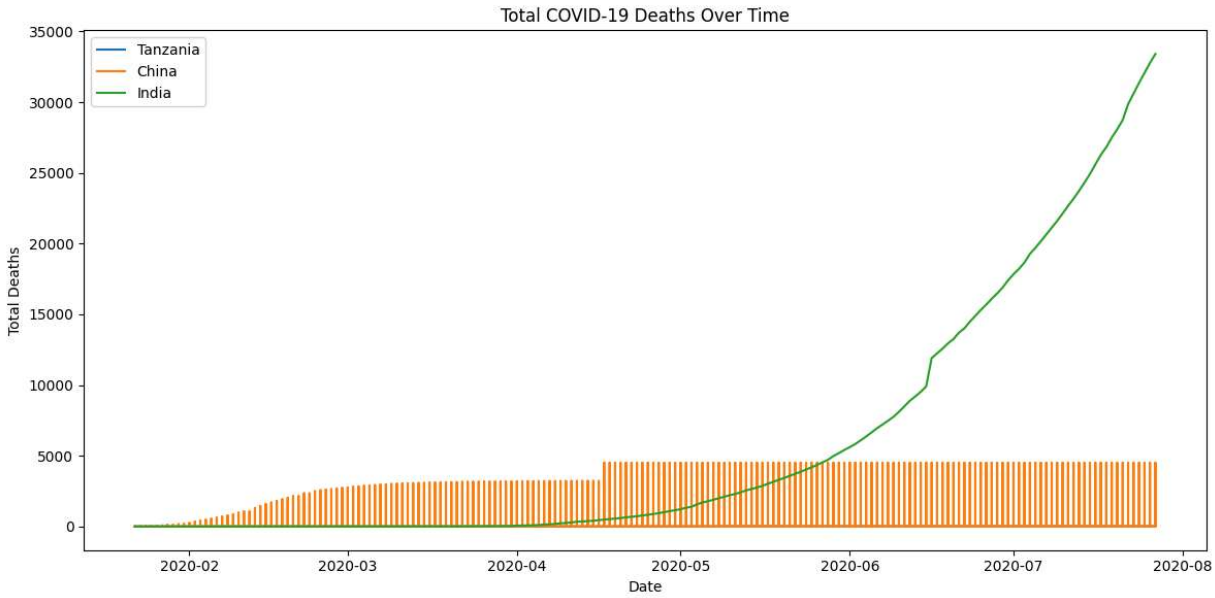
In [18]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

try:
    df = pd.read_csv("owid-covid-data.csv")
    print("Data loaded successfully.\n")
except FileNotFoundError:
    print("File not found. Please check the file path.")
    df = None

if df is not None:
    print(" Dataset columns:")
    print(df.columns)
    print("\n Preview of data:")
    print(df.head())

    countries = ['Tanzania', 'China', 'India']
    df = df[df['Country/Region'].isin(countries)]

    df['date'] = pd.to_datetime(df['Date'])
    df.fillna(0, inplace=True)

    df['death_rate'] = df['Deaths'] / df['Confirmed']
    df['death_rate'] = df['death_rate'].replace([np.inf, -np.inf], 0)

    print("\n Summary Statistics:")
    print(df.describe())

    print("\n Mean values grouped by country:")
    print(df.groupby('Country/Region')[['Confirmed', 'Deaths', 'Recovered']].mean()

    plt.figure(figsize=(12, 6))
    for country in countries:
        country_data = df[df['Country/Region'] == country]
        plt.plot(country_data['date'], country_data['Confirmed'], label=country)
    plt.title('Total COVID-19 Cases Over Time')
    plt.xlabel('Date')
    plt.ylabel('Total Cases')
    plt.legend()
    plt.tight_layout()
    plt.show()

    plt.figure(figsize=(8, 6))
    total_cases_by_country = df.groupby('Country/Region')['Confirmed'].max()
    total_cases_by_country.plot(kind='bar', color=['skyblue', 'salmon', 'limegreen'
    plt.title('Max Total Cases by Country')
    plt.xlabel('Country')
    plt.ylabel('Total Cases')
    plt.tight_layout()
    plt.show()
```

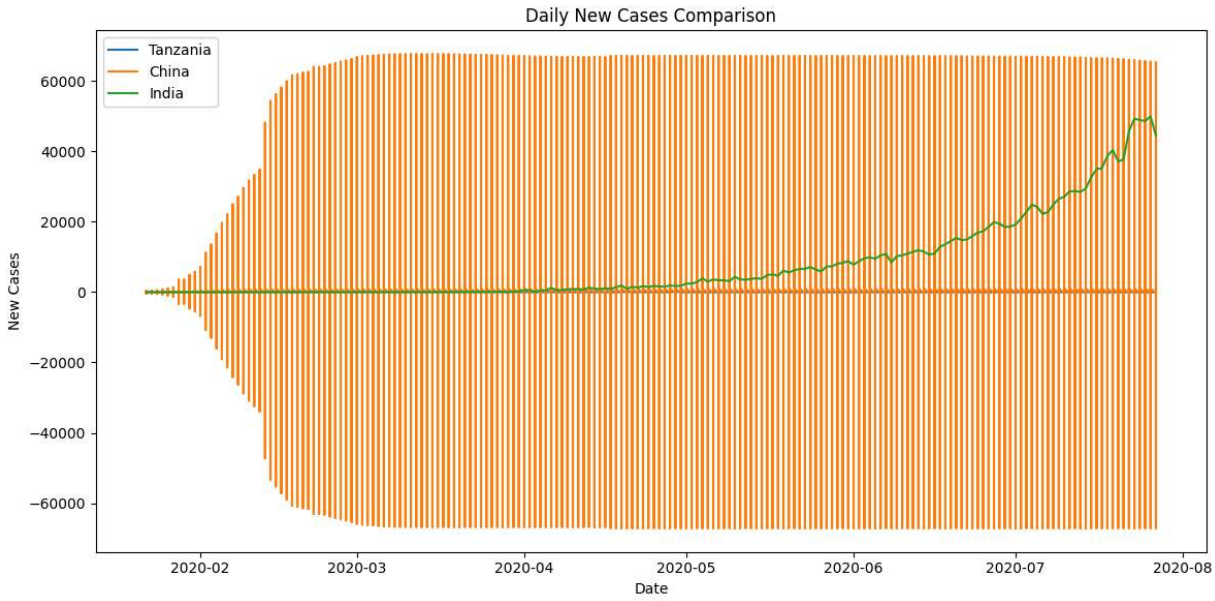
```
    plt.figure(figsize=(12, 6))
    for country in countries:
        country_data = df[df['Country/Region'] == country]
        plt.plot(country_data['date'], country_data['Deaths'], label=country)
    plt.title('Total COVID-19 Deaths Over Time')
    plt.xlabel('Date')
    plt.ylabel('Total Deaths')
    plt.legend()
    plt.tight_layout()
    plt.show()

    plt.figure(figsize=(12, 6))
    for country in countries:
        country_data = df[df['Country/Region'] == country]
        new_cases = country_data['Confirmed'].diff().fillna(0)
        plt.plot(country_data['date'], new_cases, label=country)
    plt.title('Daily New Cases Comparison')
    plt.xlabel('Date')
    plt.ylabel('New Cases')
    plt.legend()
    plt.tight_layout()
    plt.show()

    if 'total_vaccinations' in df.columns and 'Population' in df.columns:
        plt.figure(figsize=(12, 6))
        for country in countries:
            country_data = df[df['Country/Region'] == country]
            plt.plot(country_data['date'], country_data['total_vaccinations'], labe
        plt.title('Total Vaccinations Over Time')
        plt.xlabel('Date')
        plt.ylabel('Total Vaccinations')
        plt.legend()
        plt.tight_layout()
        plt.show()

        latest = df.groupby('Country/Region').last()
        latest['vaccination']()

    print("\n Summary Insights:")
    print("- India has the highest number of total cases and deaths among the three
    print("- China shows moderate case numbers with a relatively high number of rec
    print("- Tanzania has the lowest reported cases and deaths but also fewer recov
    print("- The death rate varies across countries, likely influenced by testing a
    print("- Vaccination progress and daily new cases trends should be analyzed to
```

```
Data loaded successfully.

 Dataset columns:
Index(['Province/State', 'Country/Region', 'Lat', 'Long', 'Date', 'Confirmed',
       'Deaths', 'Recovered', 'Active', 'WHO Region'],
      dtype='object')

 Preview of data:
  Province/State Country/Region       Lat       Long        Date  Confirmed  \
0            NaN    Afghanistan  33.93911  67.709953  2020-01-22          0
1            NaN        Albania  41.15330  20.168300  2020-01-22          0
2            NaN        Algeria  28.03390   1.659600  2020-01-22          0
3            NaN        Andorra  42.50630   1.521800  2020-01-22          0
4            NaN         Angola -11.20270  17.873900  2020-01-22          0

   Deaths  Recovered  Active             WHO Region
0       0          0       0  Eastern Mediterranean
1       0          0       0                 Europe
2       0          0       0                 Africa
3       0          0       0                 Europe
4       0          0       0                 Africa

 Summary Statistics:
                Lat         Long     Confirmed        Deaths      Recovered  \
count  6580.000000  6580.000000  6.580000e+03   6580.000000    6580.000000
mean     31.414770   108.651126  8.368510e+03    271.463374    5359.774316
min      -6.369028    34.888822  0.000000e+00      0.000000       0.000000
25%      26.078900   106.165500  1.250000e+02      0.000000      43.000000
50%      31.692700   113.424400  2.560000e+02      2.000000     173.000000
75%      37.895700   117.323000  7.920000e+02      6.000000     619.250000
max      47.862000   127.761500  1.480073e+06  33408.000000  951166.000000
std       9.687884    16.495345  7.049068e+04   1904.332182   44037.300905

              Active                 date   death_rate
count    6580.000000                 6580  6499.000000
mean     2737.271884  2020-04-24 12:00:00     0.011337
min        -6.000000  2020-01-22 00:00:00     0.000000
25%         0.000000  2020-03-08 18:00:00     0.000789
50%         5.000000  2020-04-24 12:00:00     0.007905
75%        67.000000  2020-06-10 06:00:00     0.014377
max    495499.000000  2020-07-27 00:00:00     1.000000
std     25209.917774                  NaN     0.019313

 Mean values grouped by country:
                   Confirmed       Deaths      Recovered
Country/Region
China             2277.885558   108.383785    1848.302063
India           217465.234043  5913.994681  126509.148936
Tanzania           262.377660    10.558511      88.984043
```


Total COVID-19 Cases Over Time
1e6
Tanzania
China
India
Total Cases
Date
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08


Max Total Cases by Country
1e6
Total Cases
China
India
Tanzania
Country





Summary Insights:
- India has the highest number of total cases and deaths among the three countries.
- China shows moderate case numbers with a relatively high number of recoveries.
- Tanzania has the lowest reported cases and deaths but also fewer recoveries compared to the others.
- The death rate varies across countries, likely influenced by testing and healthcare differences.
- Vaccination progress and daily new cases trends should be analyzed to understand the current situation better.

In [ ]: