

```

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("covid_19_clean_complete.csv")

print(df.columns)
print(df.head())
print(df.isnull().sum())

countries = ['Kenya', 'US', 'India', 'Brazil', 'Italy', 'South Africa', 'China']
df = df[df['Country/Region'].isin(countries)]
df = df.dropna(subset=['Date', 'Confirmed'])
df['Date'] = pd.to_datetime(df['Date'])
df = df.fillna(0)

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

```

	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

```

Province/State    34404
Country/Region    0
Lat               0
Long              0
Date              0
Confirmed         0
Deaths           0
Recovered        0
Active           0
WHO Region       0
dtype: int64

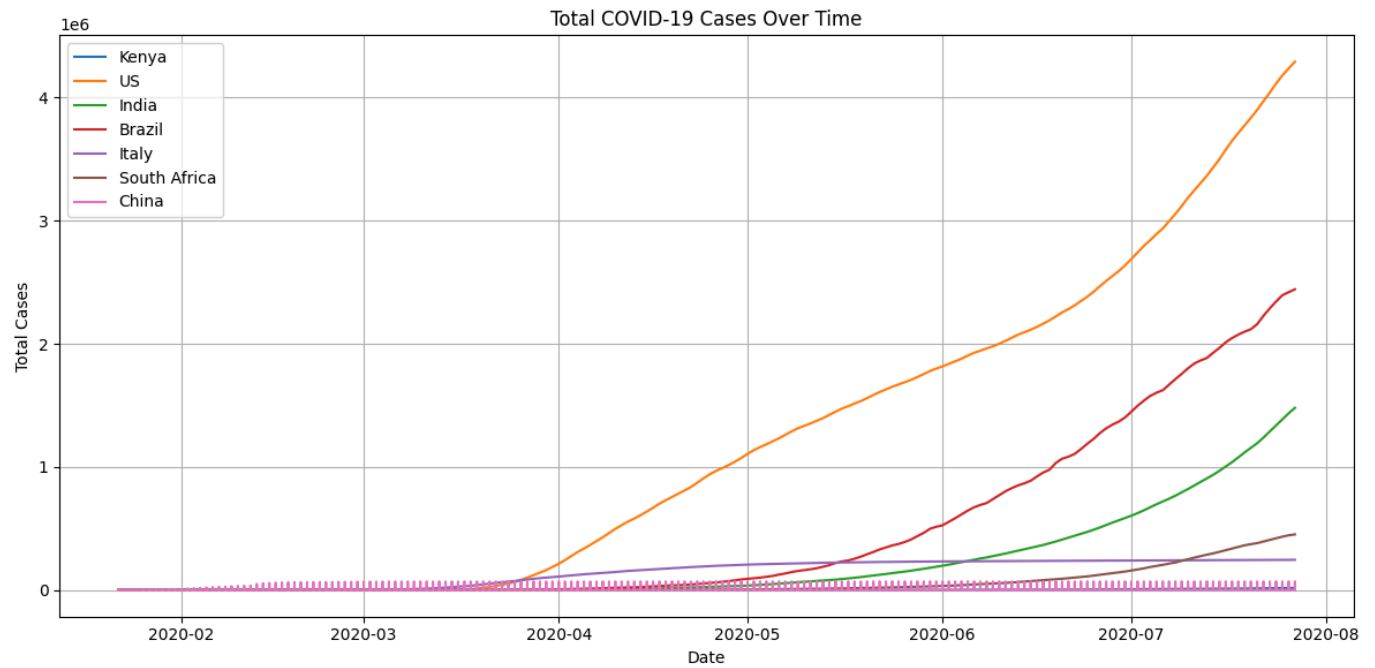
```

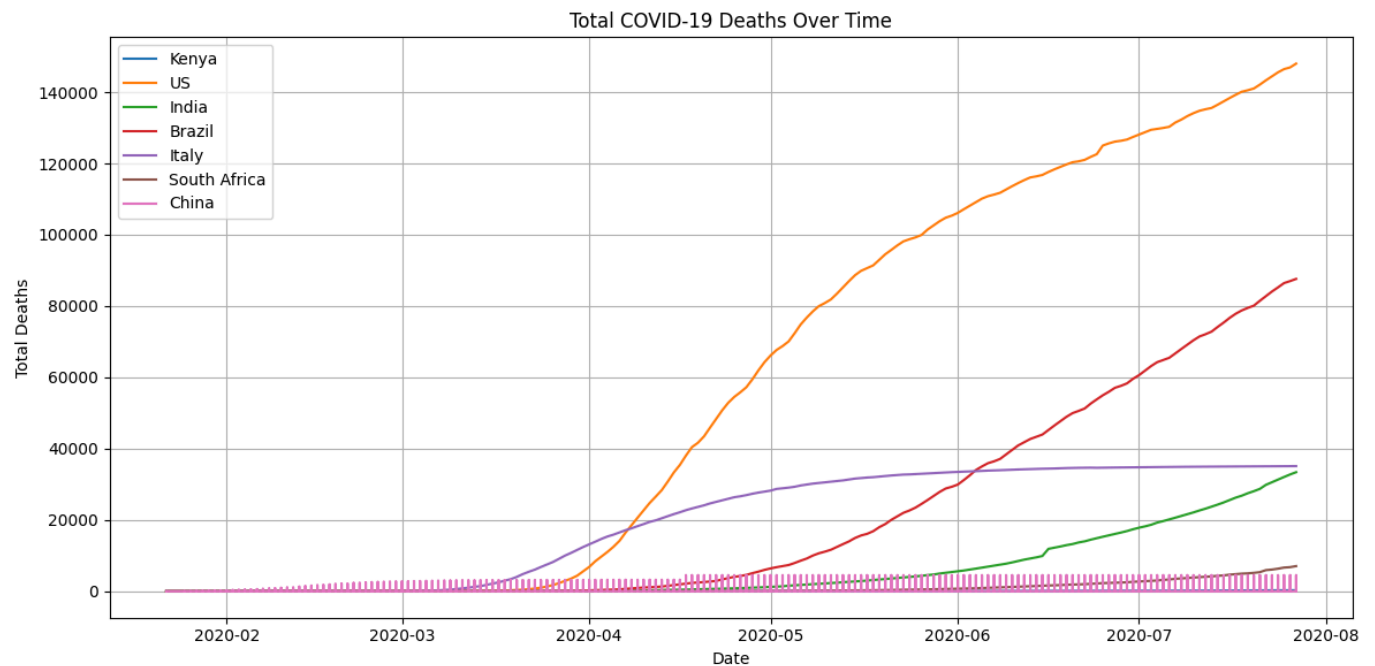
```

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

```



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



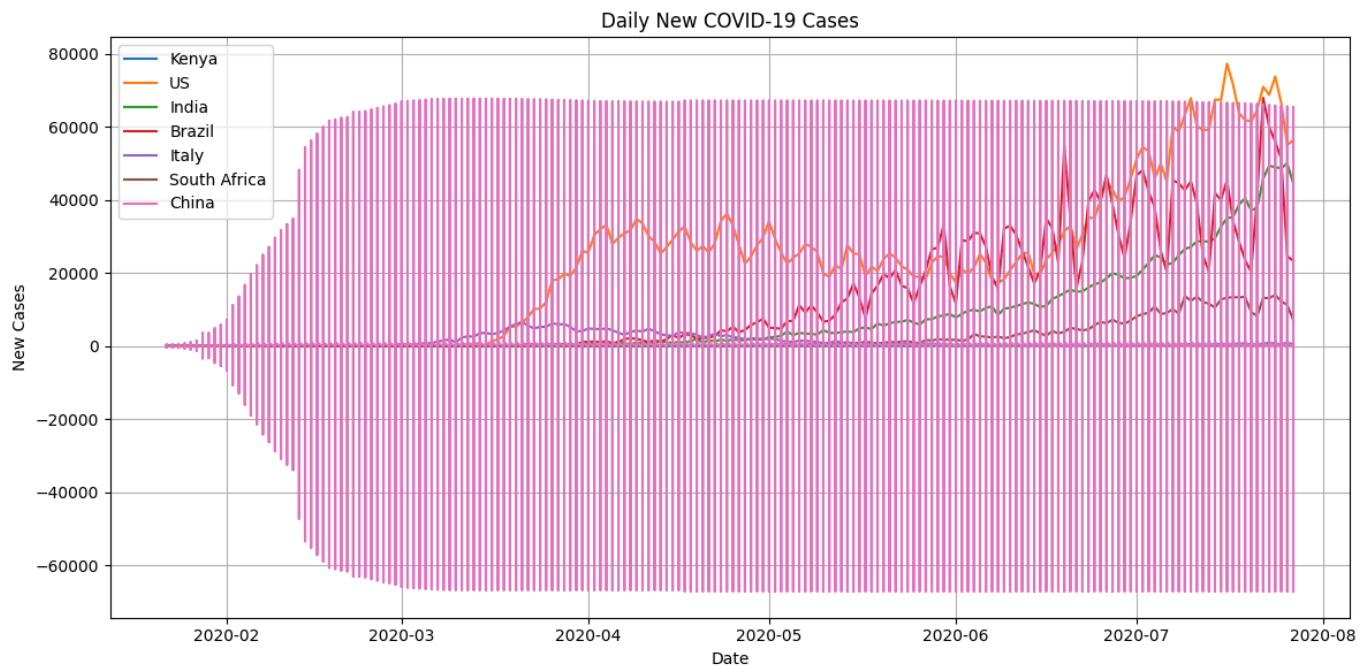
```

df['NewCases'] = df.groupby('Country/Region')['Confirmed'].diff().fillna(0)
plt.figure(figsize=(12,6))

for country in countries:
    subset = df[df['Country/Region'] == country]
    plt.plot(subset['Date'], subset['NewCases'], label=country)

plt.title('Daily New COVID-19 Cases')
plt.xlabel('Date')
plt.ylabel('New Cases')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

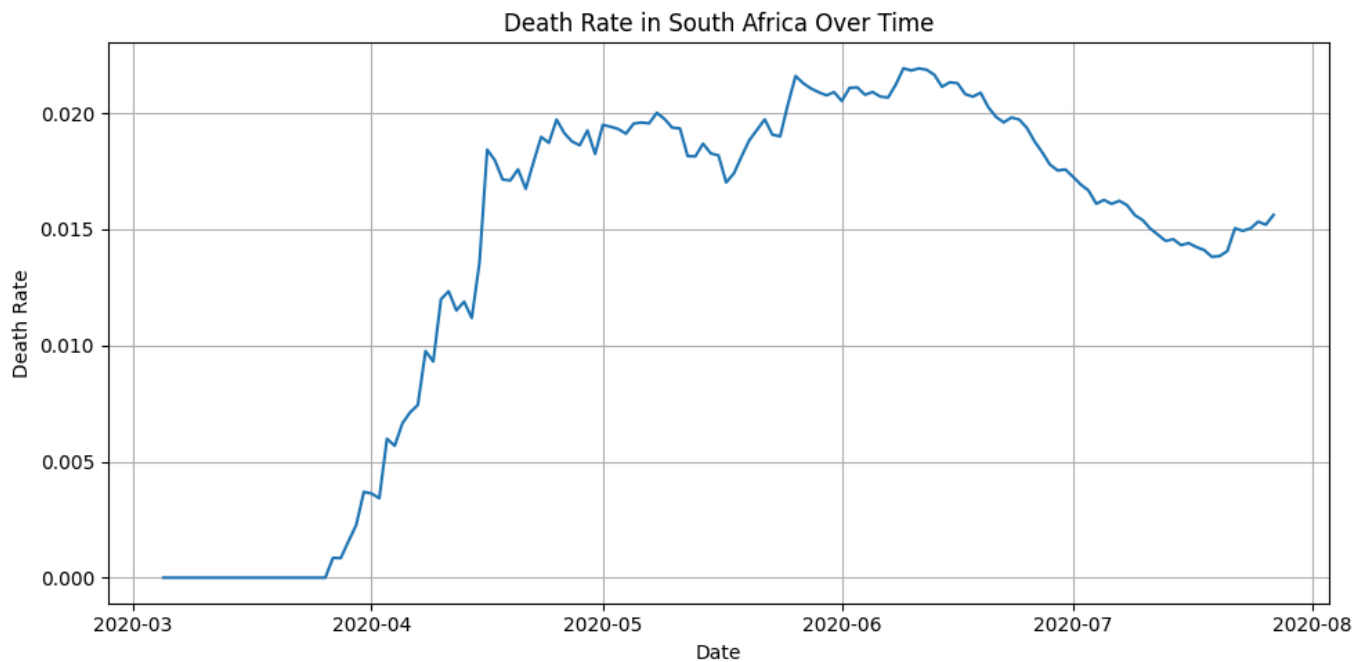
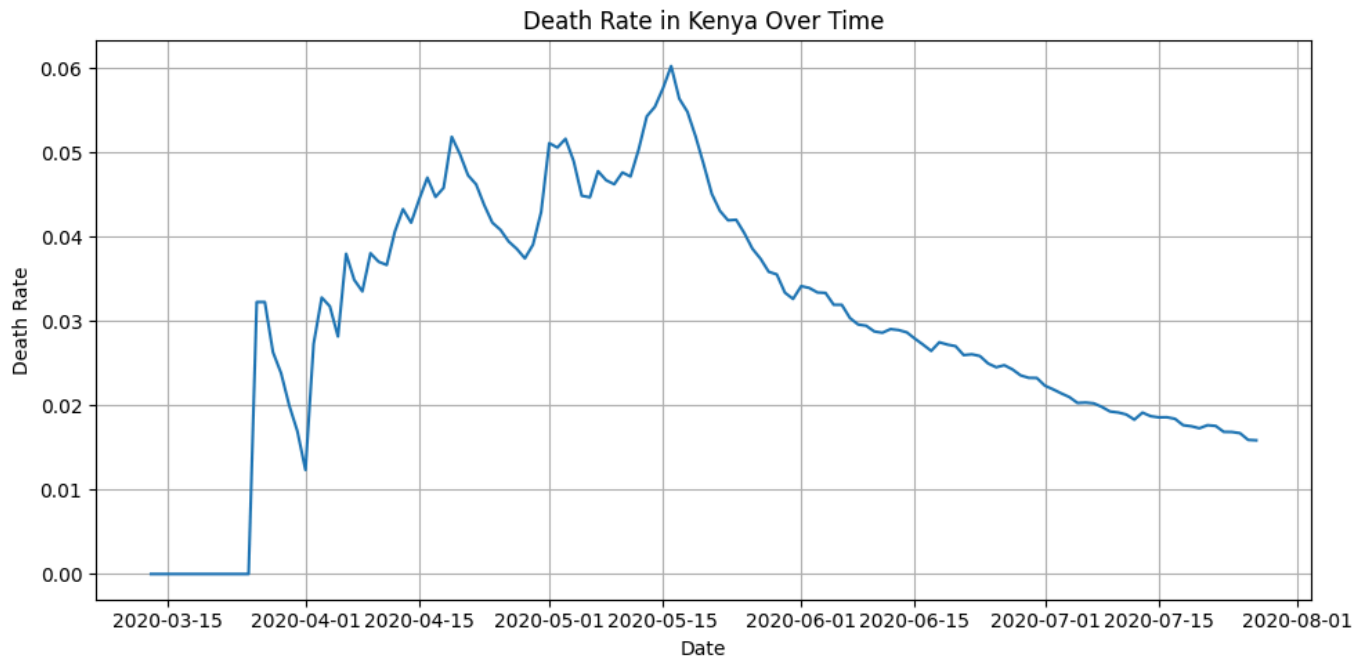
```



```

df['DeathRate'] = df['Deaths'] / df['Confirmed']
for country in ['Kenya', 'South Africa']:
    subset = df[df['Country/Region'] == country]
    plt.figure(figsize=(10,5))
    plt.plot(subset['Date'], subset['DeathRate'])
    plt.title(f'Death Rate in {country} Over Time')
    plt.xlabel('Date')
    plt.ylabel('Death Rate')
    plt.grid(True)
    plt.tight_layout()
    plt.show()

```



#### # Insights

```
print("1. Brazil and the US had the most rapid growth in total cases.")
print("2. India eventually surpassed many countries in total confirmed cases.")
print("3. Kenya and South Africa had more gradual increases compared to the Americas and Europe.")
print("4. Italy showed early spikes due to being among the first severely affected countries in Europe.")
print("5. China's early curve flattens quickly due to strict initial containment.")
```



1. Brazil and the US had the most rapid growth in total cases.
2. India eventually surpassed many countries in total confirmed cases.
3. Kenya and South Africa had more gradual increases compared to the Americas and Europe.
4. Italy showed early spikes due to being among the first severely affected countries in Europe.
5. China's early curve flattens quickly due to strict initial containment.

