

In [1]:

```
#Importing Libraries needed for the Project
import pandas as pd
import matplotlib.pyplot as plt
```

Create a DataFrame from the CSV file.

In [2]:

```
data=pd.read_csv('time-series-19-covid-combined.csv',parse_dates=['Date']) #reading the csv filr of covid-19 data
copydata=data
data
```

Out[2]:

| | Date | Country/Region | Province/State | Confirmed | Recovered | Deaths |
|--------|------------|----------------|----------------|-----------|-----------|--------|
| 0 | 2020-01-22 | Afghanistan | NaN | 0 | 0.0 | 0 |
| 1 | 2020-01-23 | Afghanistan | NaN | 0 | 0.0 | 0 |
| 2 | 2020-01-24 | Afghanistan | NaN | 0 | 0.0 | 0 |
| 3 | 2020-01-25 | Afghanistan | NaN | 0 | 0.0 | 0 |
| 4 | 2020-01-26 | Afghanistan | NaN | 0 | 0.0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 231739 | 2022-04-12 | Zimbabwe | NaN | 247094 | 0.0 | 5460 |
| 231740 | 2022-04-13 | Zimbabwe | NaN | 247160 | 0.0 | 5460 |
| 231741 | 2022-04-14 | Zimbabwe | NaN | 247208 | 0.0 | 5462 |
| 231742 | 2022-04-15 | Zimbabwe | NaN | 247237 | 0.0 | 5462 |
| 231743 | 2022-04-16 | Zimbabwe | NaN | 247237 | 0.0 | 5462 |

231744 rows × 6 columns

Merge the data for countries with multiple regions in order to provide a single time-series for each Country

In [3]:

```
#creating a single time-series for each country by using groupby and sum
ts=data.groupby(['Country/Region','Date']).sum()
ts
```

Out[3]:

| | | Confirmed | Recovered | Deaths |
|----------------|------------|-----------|-----------|--------|
| Country/Region | Date | | | |
| Afghanistan | 2020-01-22 | 0 | 0.0 | 0 |
| | 2020-01-23 | 0 | 0.0 | 0 |
| | 2020-01-24 | 0 | 0.0 | 0 |
| | 2020-01-25 | 0 | 0.0 | 0 |
| | 2020-01-26 | 0 | 0.0 | 0 |
| ... | ... | ... | ... | ... |
| Zimbabwe | 2022-04-12 | 247094 | 0.0 | 5460 |
| | 2022-04-13 | 247160 | 0.0 | 5460 |
| | 2022-04-14 | 247208 | 0.0 | 5462 |
| | 2022-04-15 | 247237 | 0.0 | 5462 |
| | 2022-04-16 | 247237 | 0.0 | 5462 |

161568 rows × 3 columns

Print the total number of confirmed cases and number of deaths in each country in the last reported day.

In [4]:

```
#filling None in the place of NaN Province/State values
copydata['Province/State']=copydata['Province/State'].fillna('None')
copydata
```

Out[4]:

| | Date | Country/Region | Province/State | Confirmed | Recovered | Deaths |
|--------|------------|----------------|----------------|-----------|-----------|--------|
| 0 | 2020-01-22 | Afghanistan | None | 0 | 0.0 | 0 |
| 1 | 2020-01-23 | Afghanistan | None | 0 | 0.0 | 0 |
| 2 | 2020-01-24 | Afghanistan | None | 0 | 0.0 | 0 |
| 3 | 2020-01-25 | Afghanistan | None | 0 | 0.0 | 0 |
| 4 | 2020-01-26 | Afghanistan | None | 0 | 0.0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 231739 | 2022-04-12 | Zimbabwe | None | 247094 | 0.0 | 5460 |
| 231740 | 2022-04-13 | Zimbabwe | None | 247160 | 0.0 | 5460 |
| 231741 | 2022-04-14 | Zimbabwe | None | 247208 | 0.0 | 5462 |
| 231742 | 2022-04-15 | Zimbabwe | None | 247237 | 0.0 | 5462 |
| 231743 | 2022-04-16 | Zimbabwe | None | 247237 | 0.0 | 5462 |

231744 rows × 6 columns

In [5]:

```
#Last reported day data
date_filter=copydata.loc[copydata['Date']==copydata['Date'].max()]
date_filter
```

Out[5]:

| | Date | Country/Region | Province/State | Confirmed | Recovered | Deaths |
|--------|------------|----------------------|----------------|-----------|-----------|--------|
| 815 | 2022-04-16 | Afghanistan | None | 178387 | 0.0 | 7676 |
| 1631 | 2022-04-16 | Albania | None | 274462 | 0.0 | 3496 |
| 2447 | 2022-04-16 | Algeria | None | 265739 | 0.0 | 6874 |
| 3263 | 2022-04-16 | Andorra | None | 40709 | 0.0 | 153 |
| 4079 | 2022-04-16 | Angola | None | 99194 | 0.0 | 1900 |
| ... | ... | ... | ... | ... | ... | ... |
| 228479 | 2022-04-16 | West Bank and Gaza | None | 656617 | 0.0 | 5656 |
| 229295 | 2022-04-16 | Winter Olympics 2022 | None | 535 | 0.0 | 0 |
| 230111 | 2022-04-16 | Yemen | None | 11817 | 0.0 | 2148 |
| 230927 | 2022-04-16 | Zambia | None | 318467 | 0.0 | 3973 |
| 231743 | 2022-04-16 | Zimbabwe | None | 247237 | 0.0 | 5462 |

284 rows × 6 columns

In [6]:

```
#grouping data by country and state
t=date_filter.groupby(['Country/Region', 'Province/State']).max()
t
```

Out[6]:

| | | Date | Confirmed | Recovered | Deaths |
|----------------------|----------------|------------|-----------|-----------|--------|
| Country/Region | Province/State | | | | |
| Afghanistan | None | 2022-04-16 | 178387 | 0.0 | 7676 |
| Albania | None | 2022-04-16 | 274462 | 0.0 | 3496 |
| Algeria | None | 2022-04-16 | 265739 | 0.0 | 6874 |
| Andorra | None | 2022-04-16 | 40709 | 0.0 | 153 |
| Angola | None | 2022-04-16 | 99194 | 0.0 | 1900 |
| ... | ... | ... | ... | ... | ... |
| West Bank and Gaza | None | 2022-04-16 | 656617 | 0.0 | 5656 |
| Winter Olympics 2022 | None | 2022-04-16 | 535 | 0.0 | 0 |
| Yemen | None | 2022-04-16 | 11817 | 0.0 | 2148 |
| Zambia | None | 2022-04-16 | 318467 | 0.0 | 3973 |
| Zimbabwe | None | 2022-04-16 | 247237 | 0.0 | 5462 |

284 rows × 4 columns

In [7]:

```
#sum of confirmed cases and deaths of all the states of each country
res=t.groupby('Country/Region')['Confirmed', 'Deaths'].sum()
res
```

C:\Users\Kirti\AppData\Local\Temp\ipykernel_21324\4132366157.py:2: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.

```
res=t.groupby('Country/Region')['Confirmed', 'Deaths'].sum()
```

Out[7]:

| | Confirmed | Deaths |
|----------------------|-----------|--------|
| Country/Region | | |
| Afghanistan | 178387 | 7676 |
| Albania | 274462 | 3496 |
| Algeria | 265739 | 6874 |
| Andorra | 40709 | 153 |
| Angola | 99194 | 1900 |
| ... | ... | ... |
| West Bank and Gaza | 656617 | 5656 |
| Winter Olympics 2022 | 535 | 0 |
| Yemen | 11817 | 2148 |
| Zambia | 318467 | 3973 |
| Zimbabwe | 247237 | 5462 |

198 rows × 2 columns

What are the 10 countries with the highest number of confirmed COVID-19 cases?

In [8]:

```
res.nlargest(10, 'Confirmed')['Confirmed']
```

Out[8]:

```
Country/Region
US                80625120
India             43042097
Brazil            30250077
France            27874269
Germany           23416663
United Kingdom    21916961
Russia            17801103
Korea, South      16305752
Italy             15659835
Turkey            14991669
Name: Confirmed, dtype: int64
```

What are the 10 countries with the highest number of deaths?

In [9]:

```
res.nlargest(10, 'Deaths')['Deaths']
```

Out[9]:

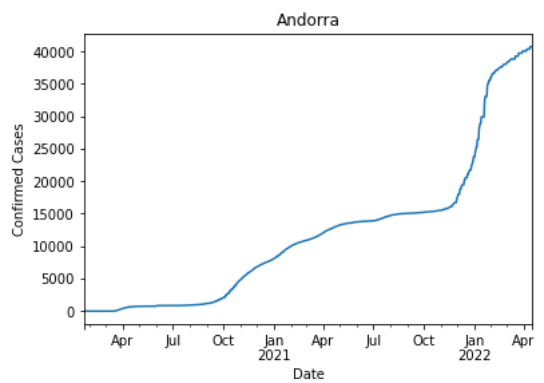
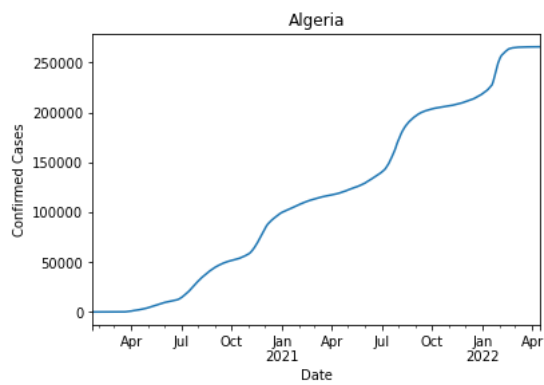
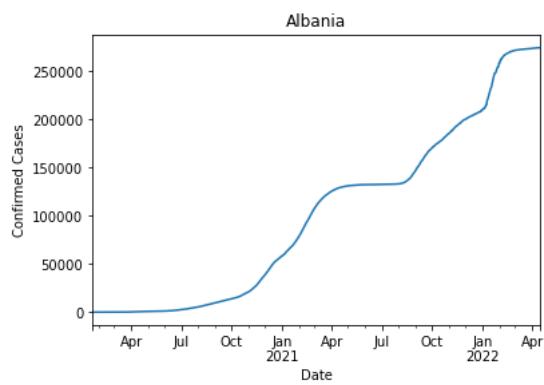
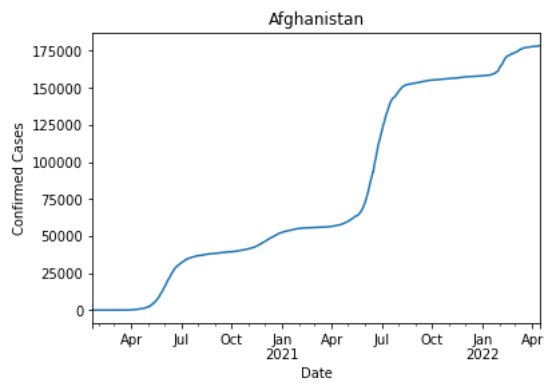
| Country/Region | |
|----------------|--------|
| US | 988609 |
| Brazil | 662185 |
| India | 521751 |
| Russia | 365774 |
| Mexico | 323938 |
| Peru | 212619 |
| United Kingdom | 172014 |
| Italy | 161602 |
| Indonesia | 155844 |
| France | 145159 |

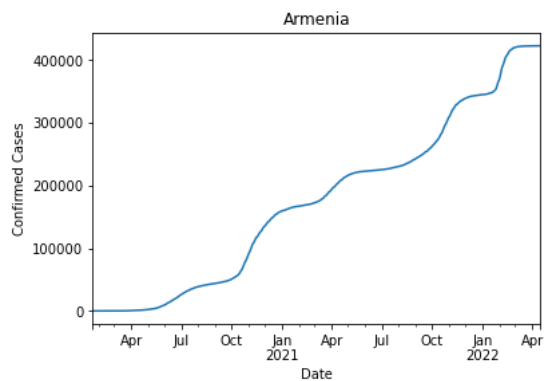
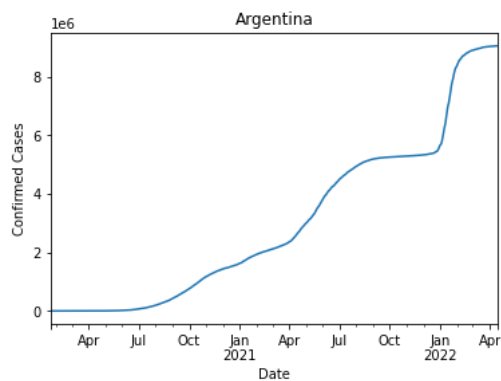
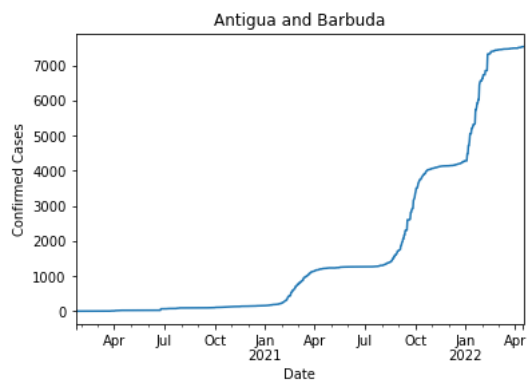
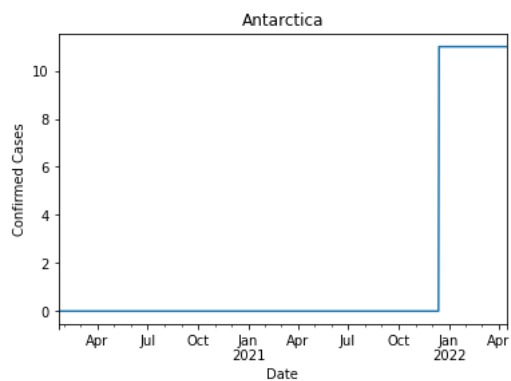
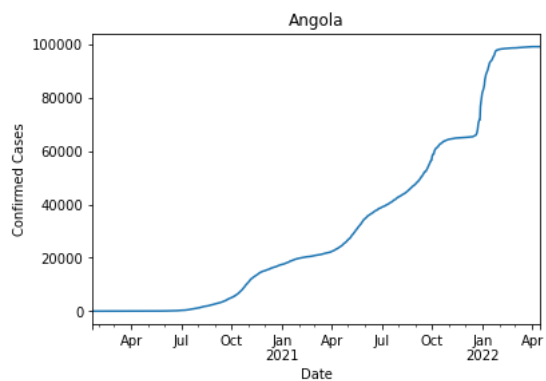
Name: Deaths, dtype: int64

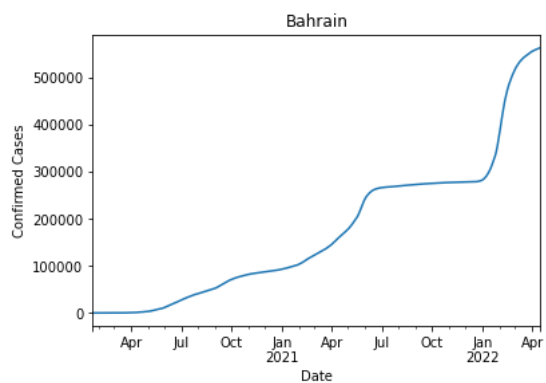
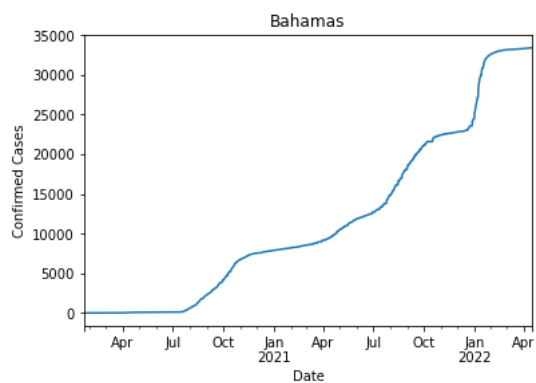
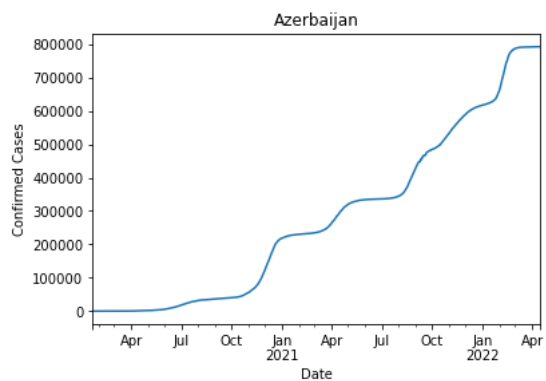
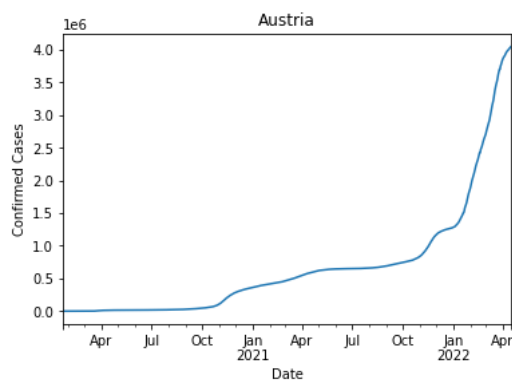
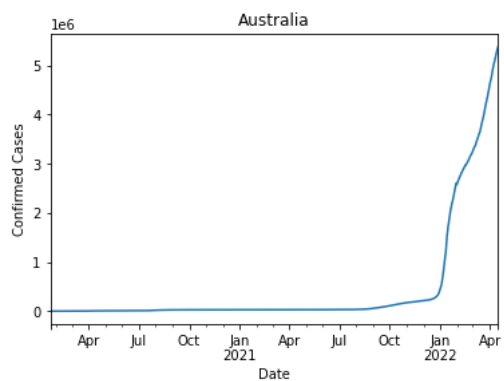
Plot a graph of the number of confirmed cases over time for each country.

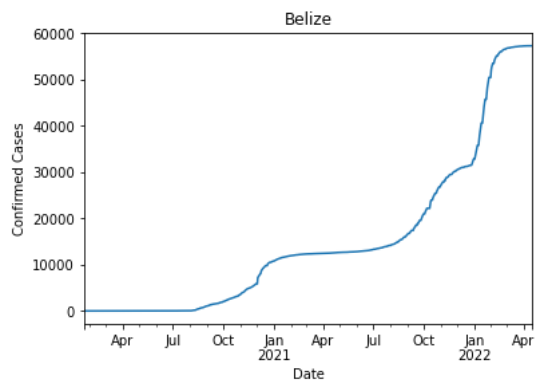
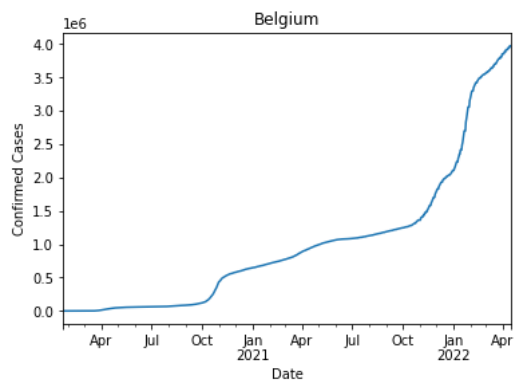
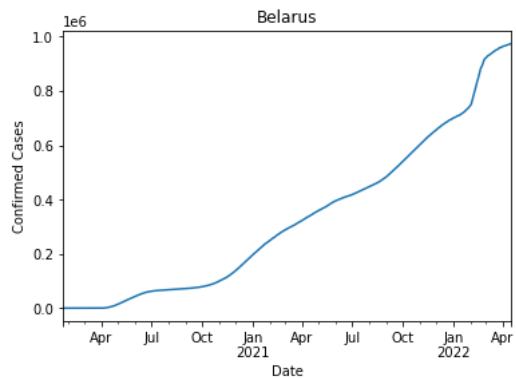
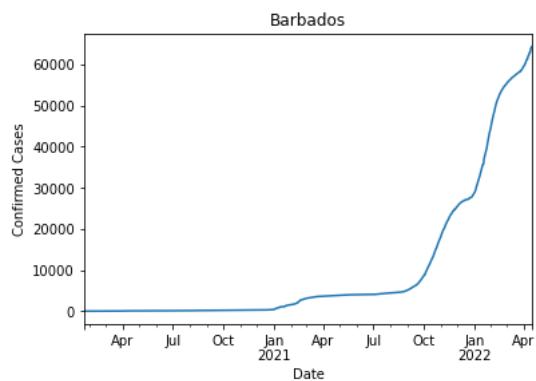
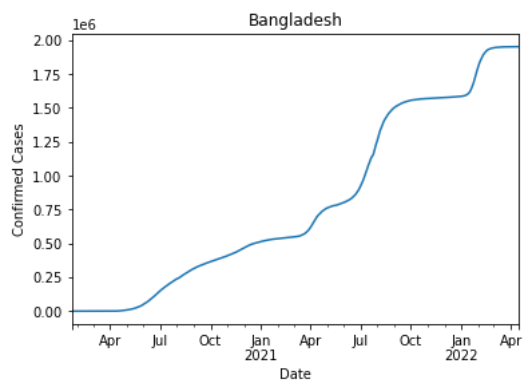
In [10]:

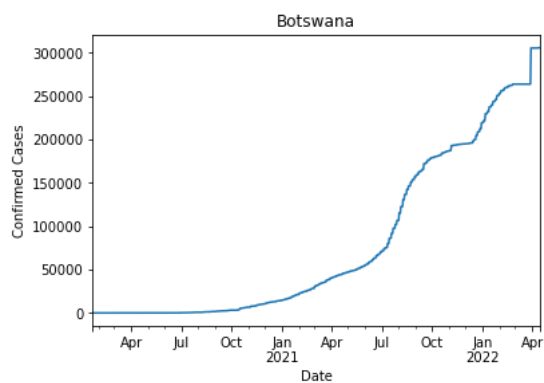
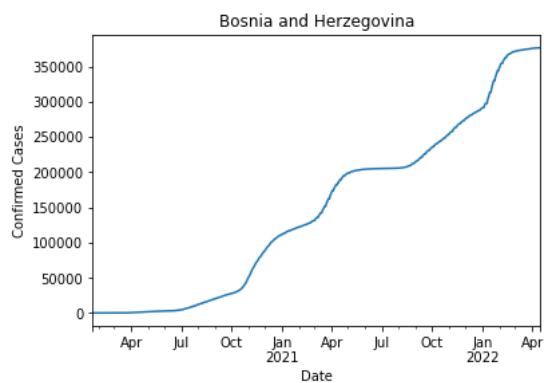
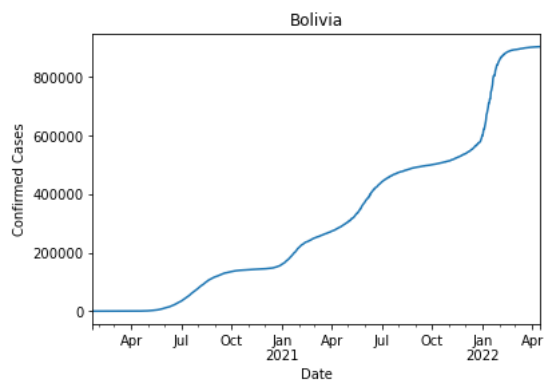
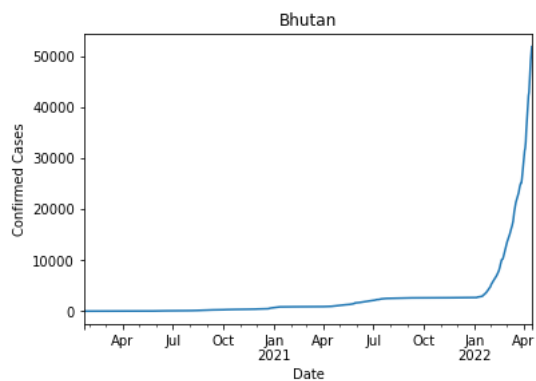
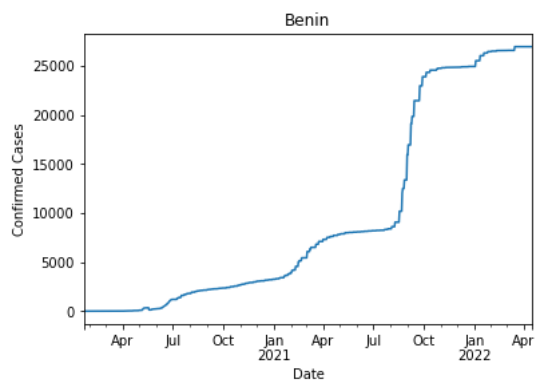
```
#pl stores the value of confirmed cases for each country per day
pl=data.groupby(['Country/Region', 'Date'])['Confirmed'].sum()
for i in data['Country/Region'].unique():
    plt.xlabel('Date')
    plt.ylabel('Confirmed Cases')
    plt.title(i)
    pl.loc[i].plot()
    plt.show()
```

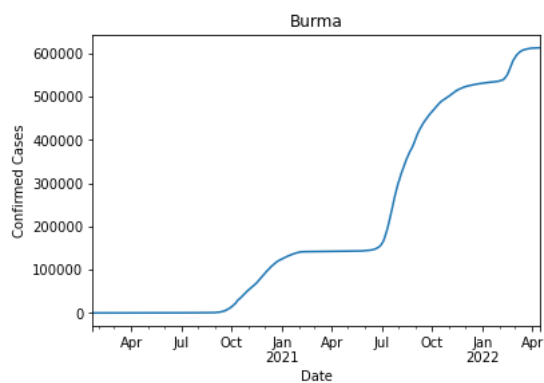
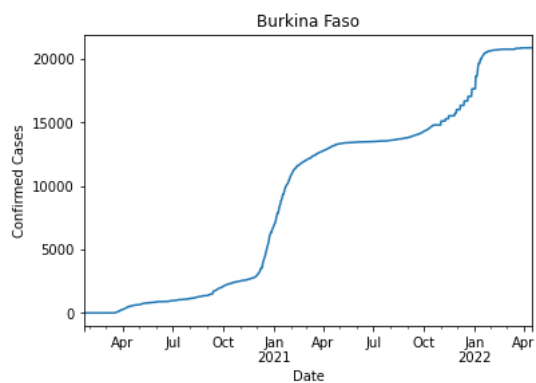
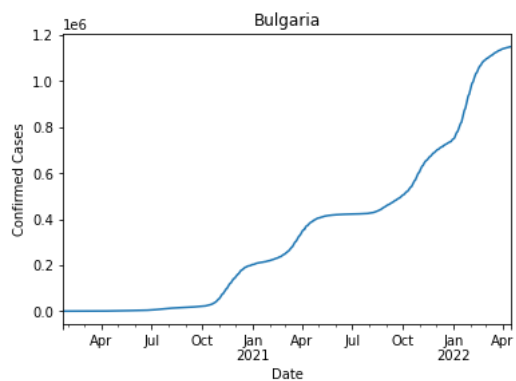
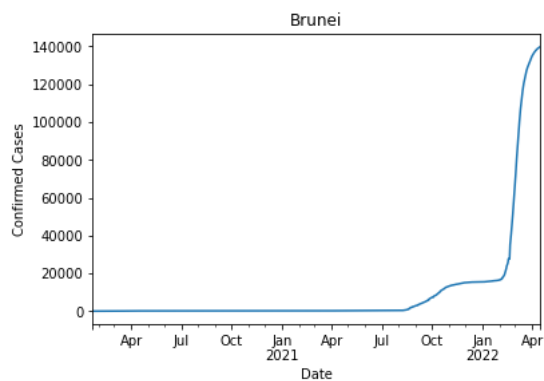
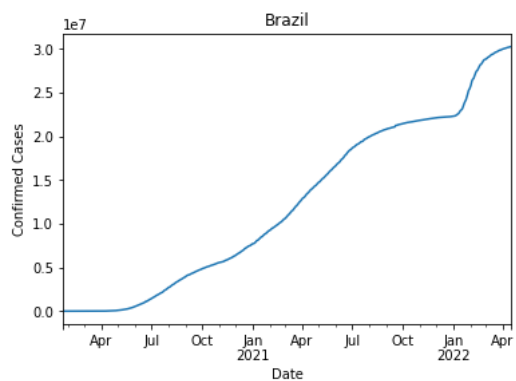


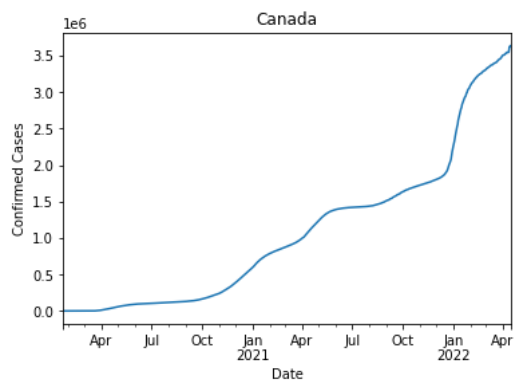
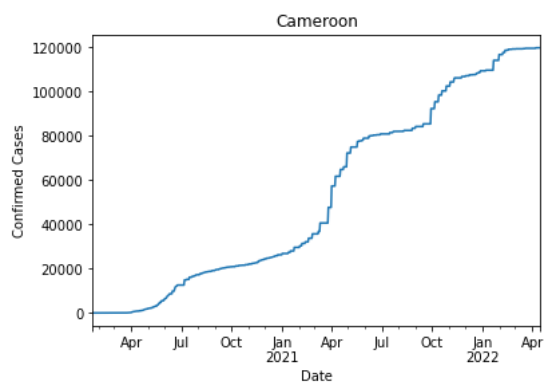
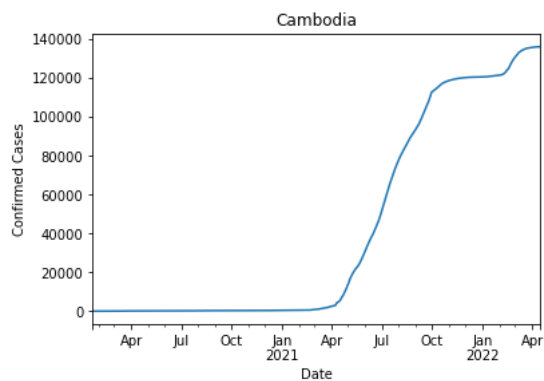
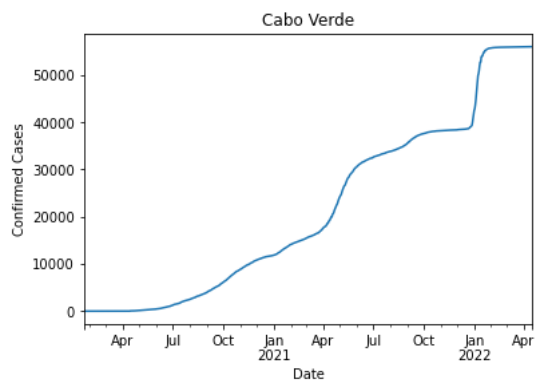
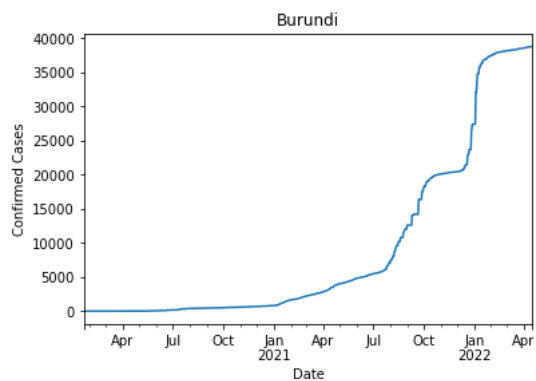


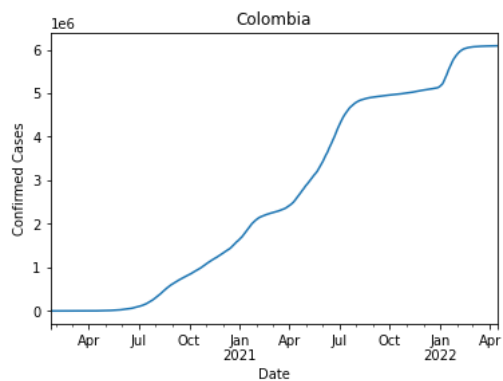
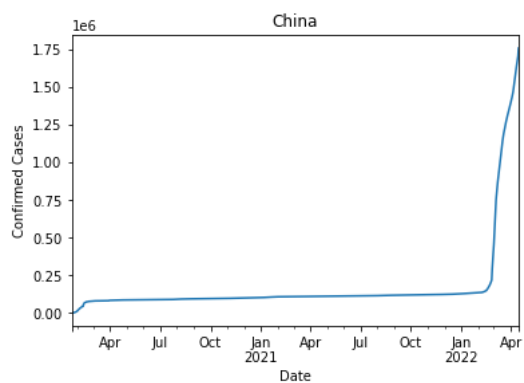
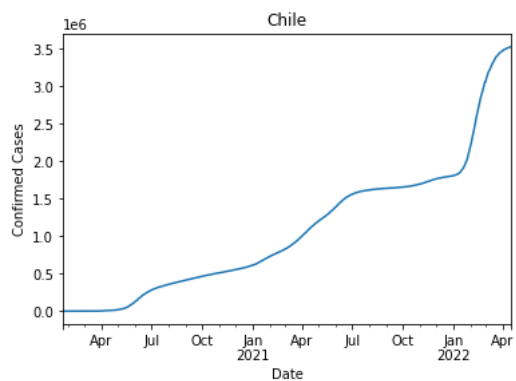
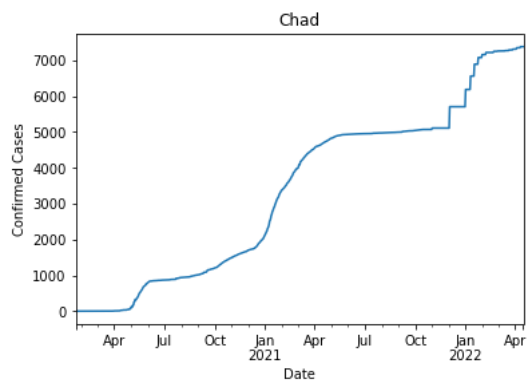
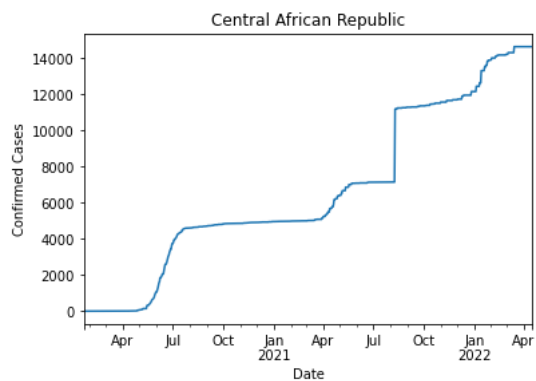


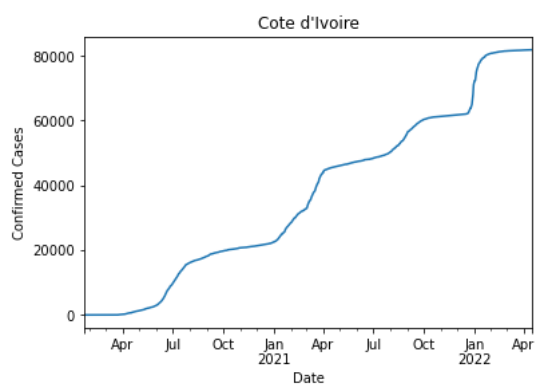
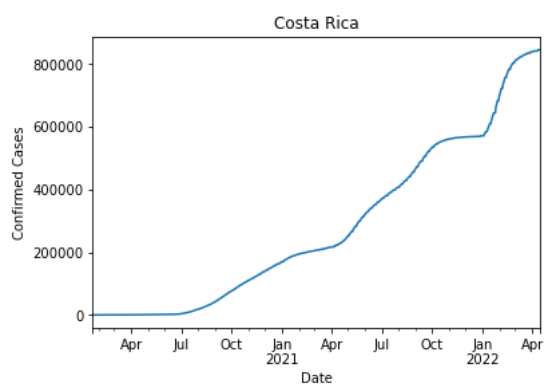
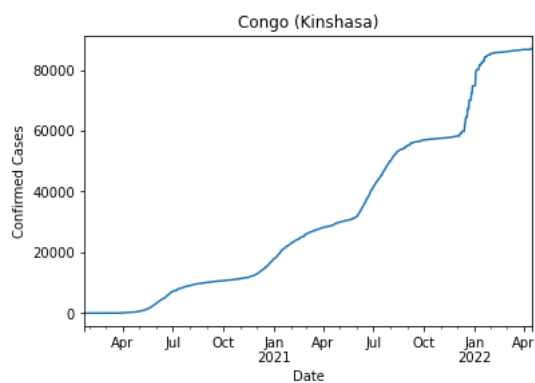
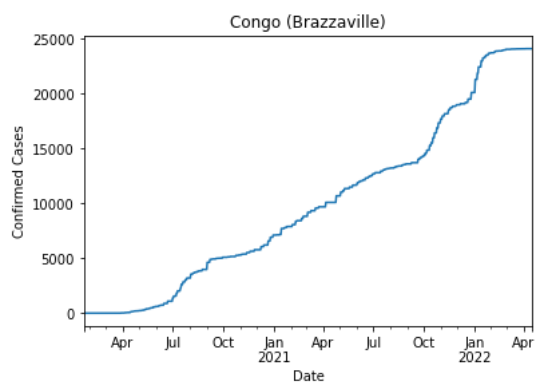
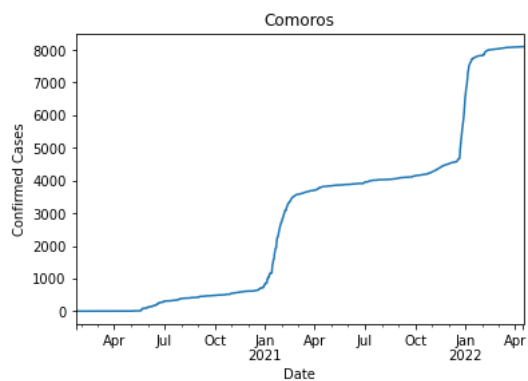


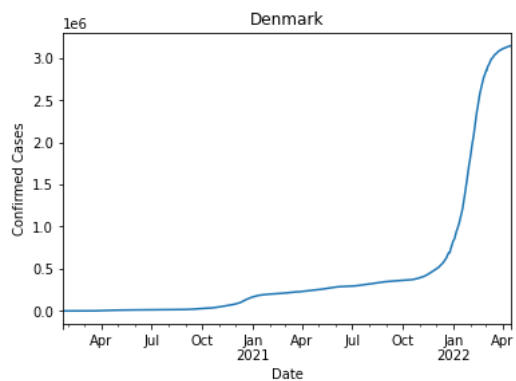
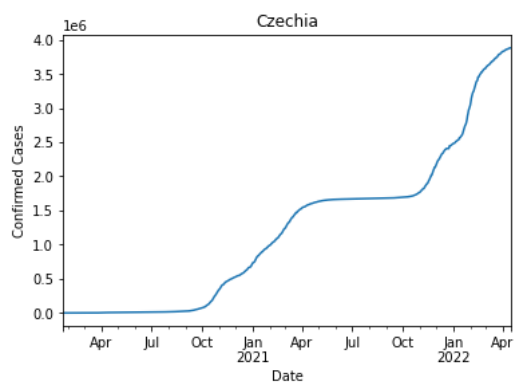
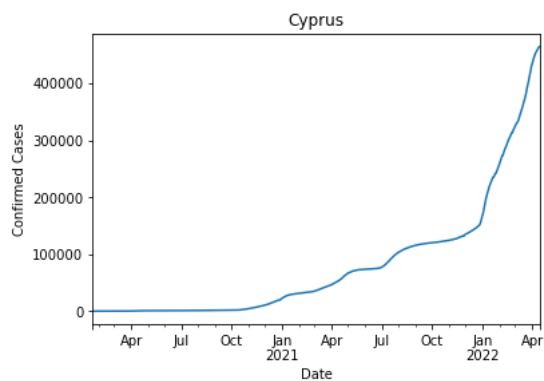
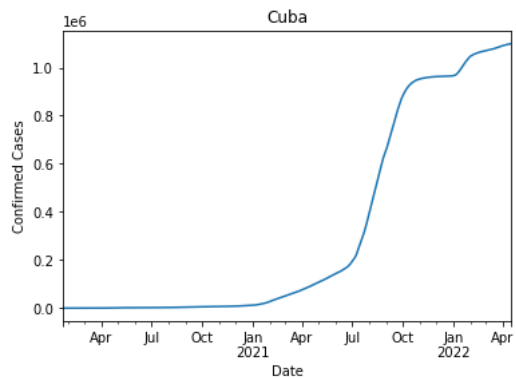
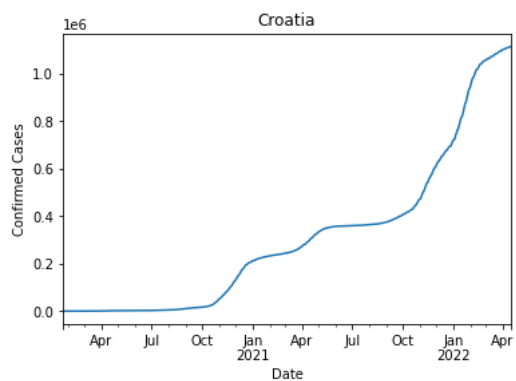


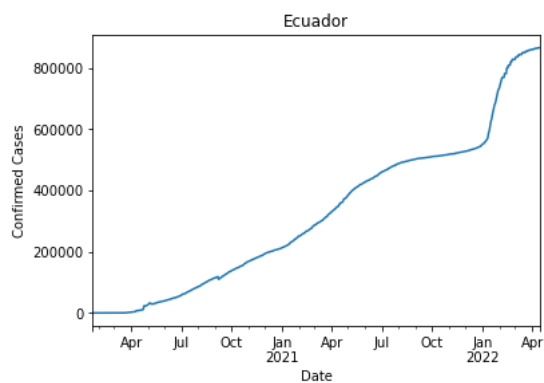
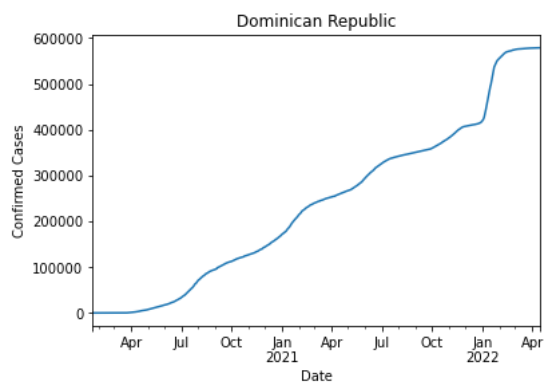
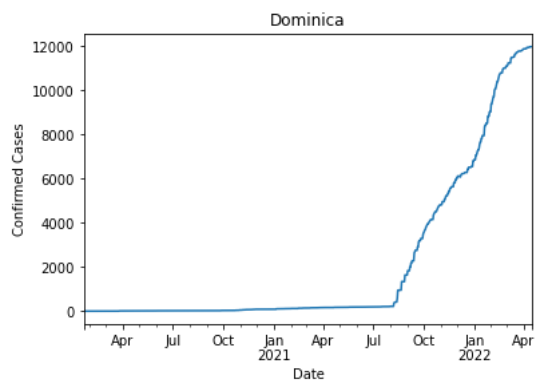
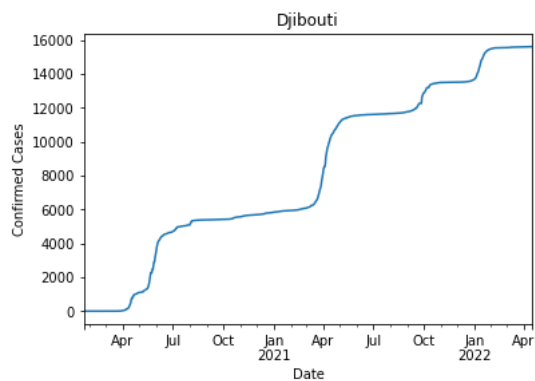
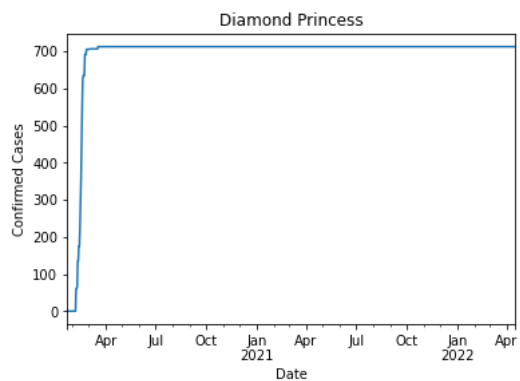


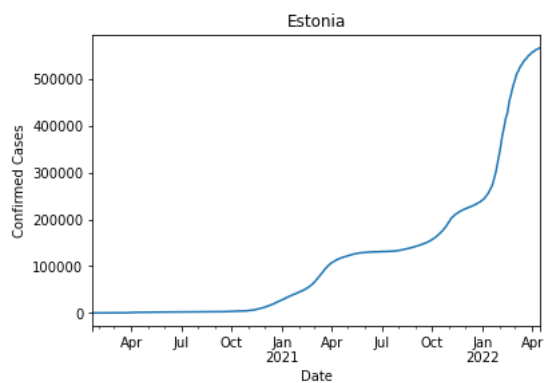
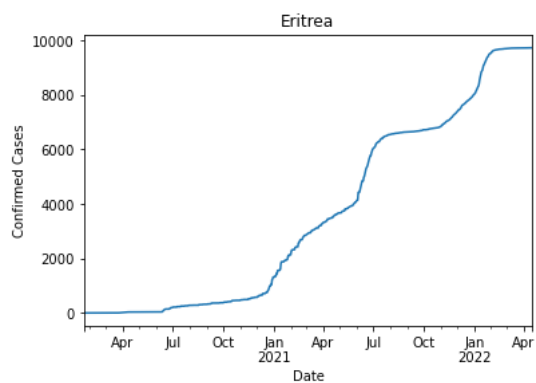
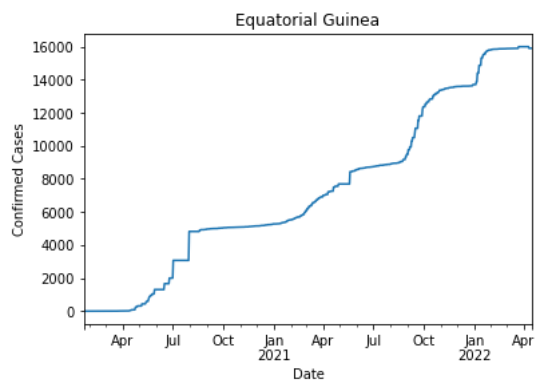
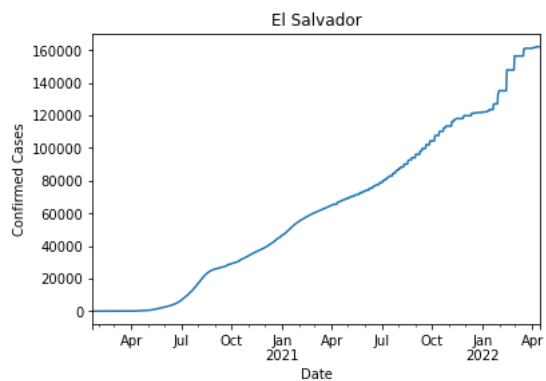
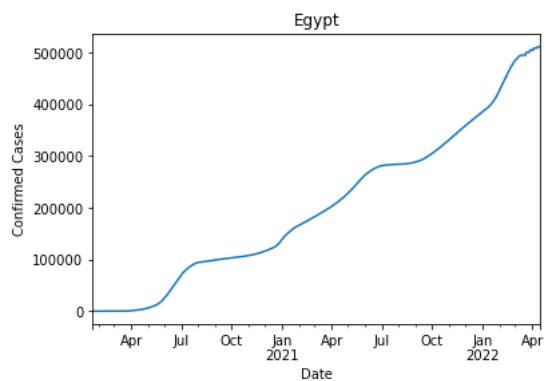


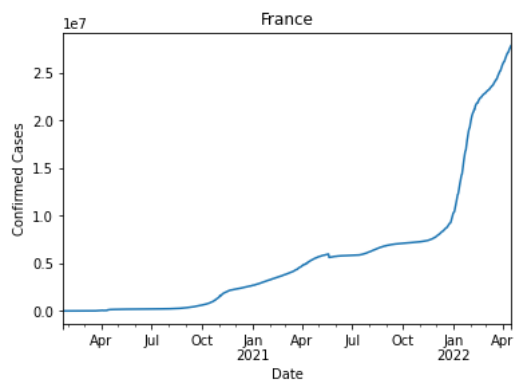
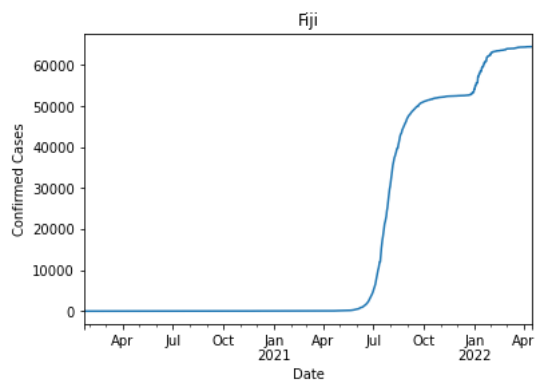
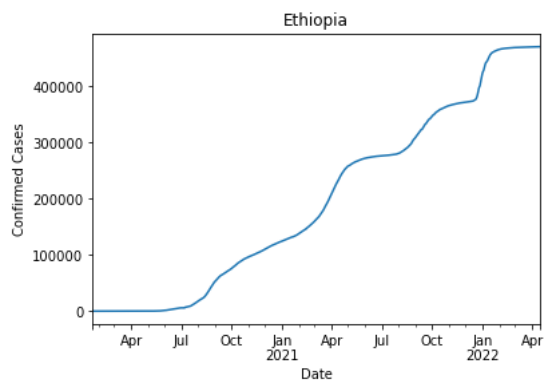
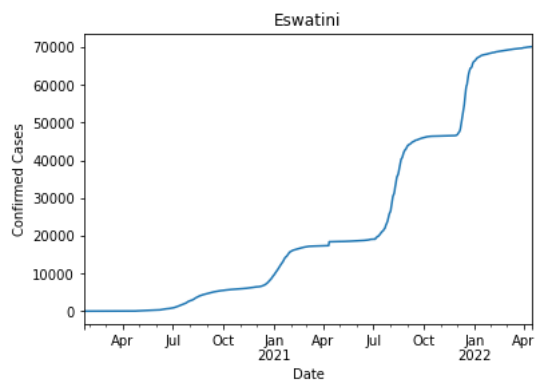


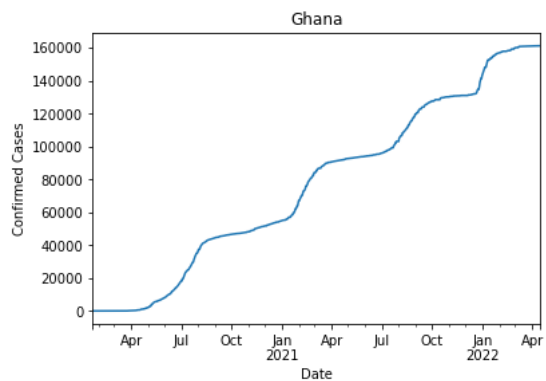
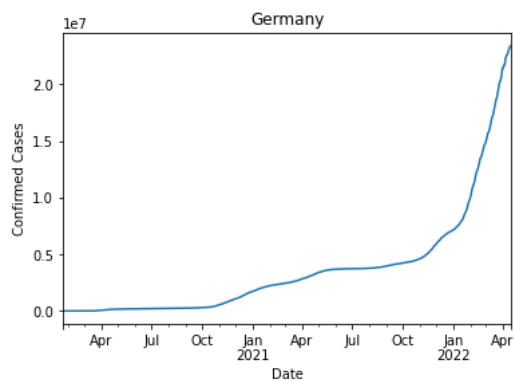
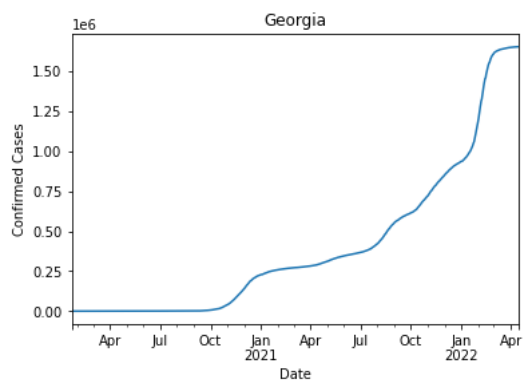
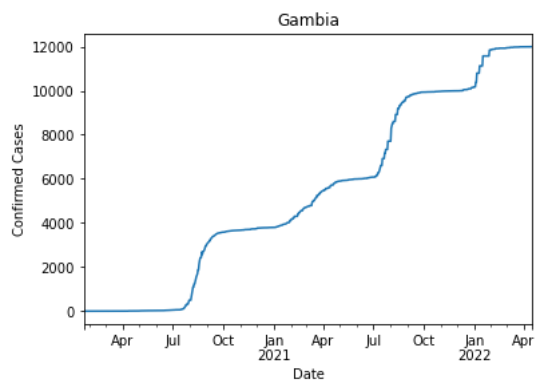
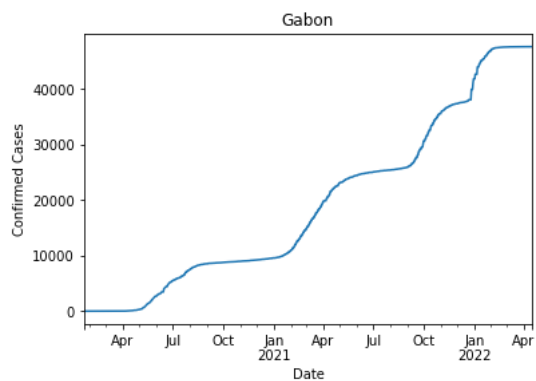


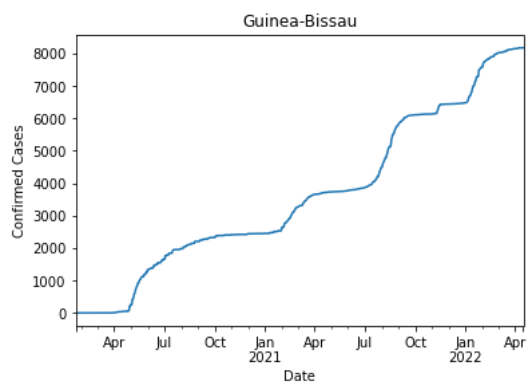
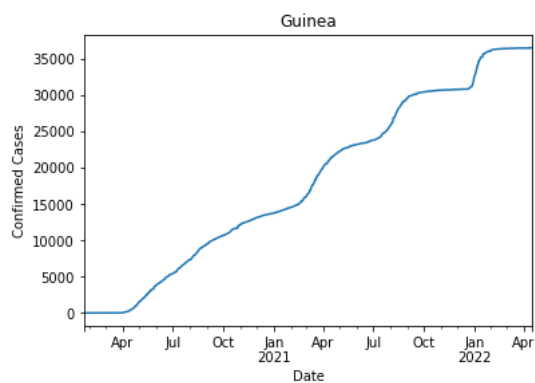
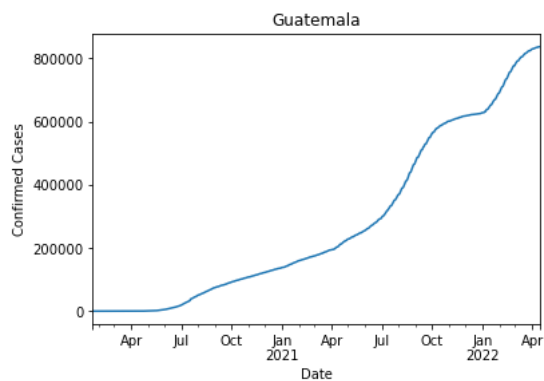
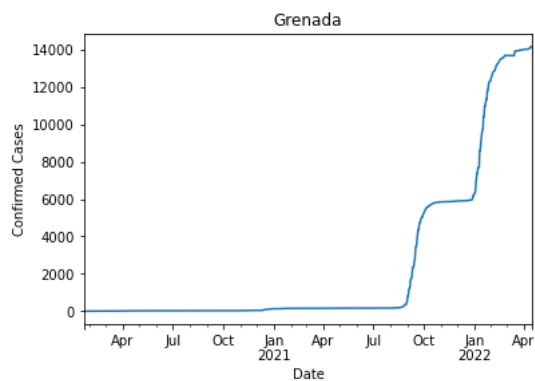
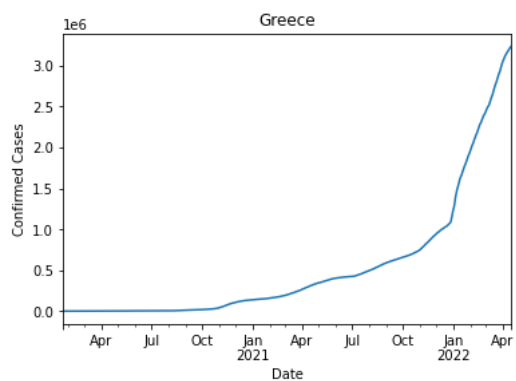


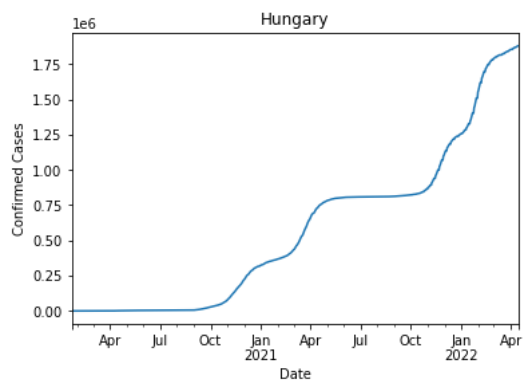
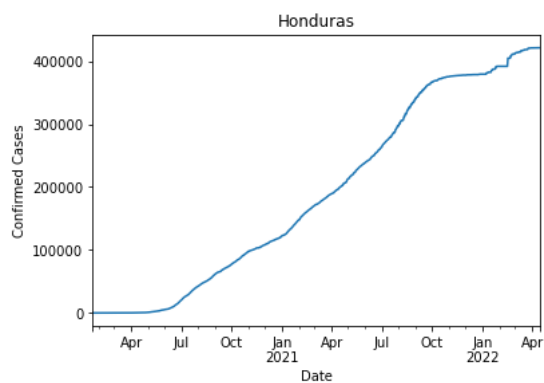
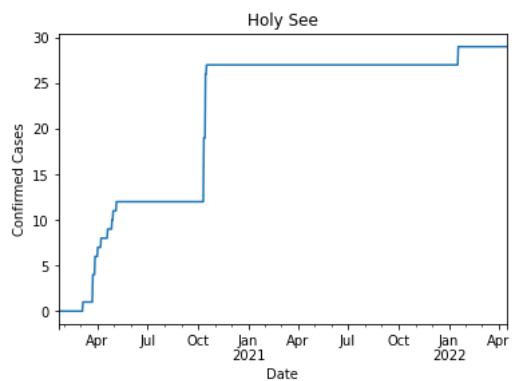
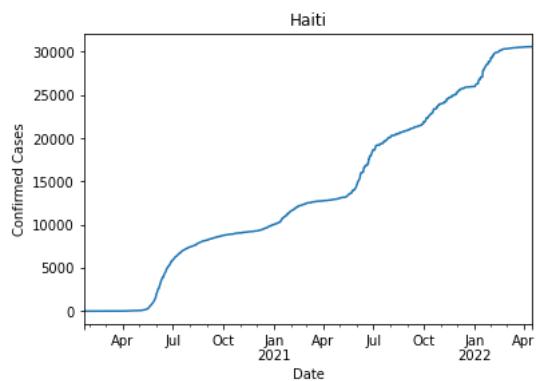
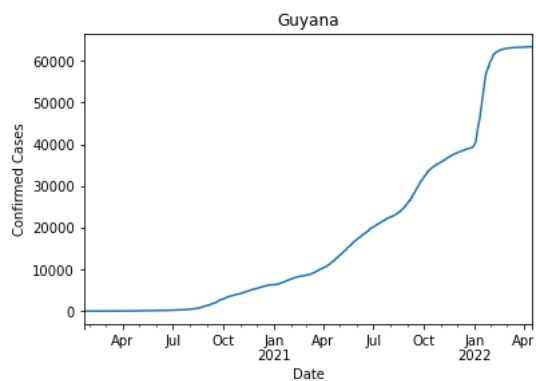


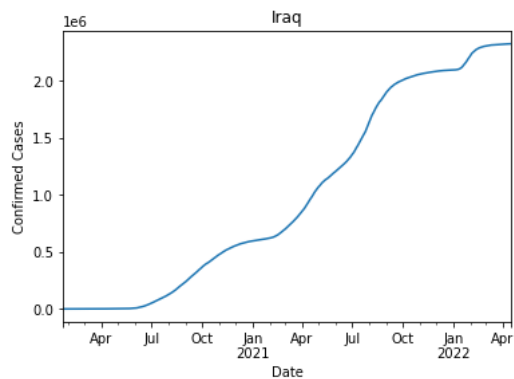
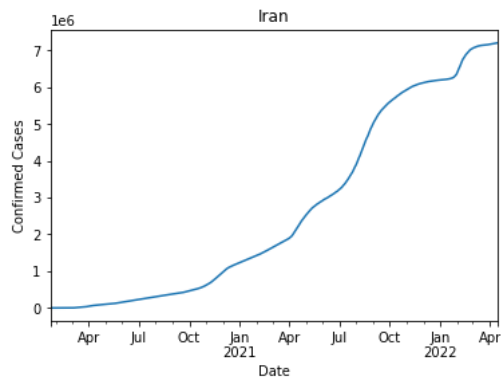
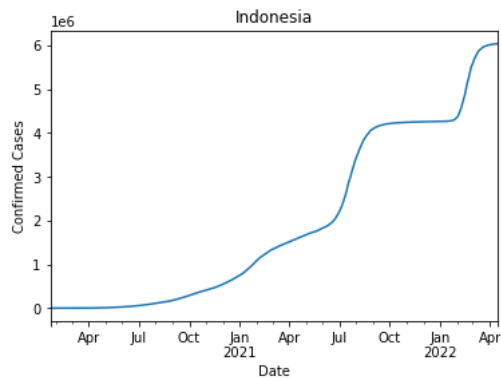
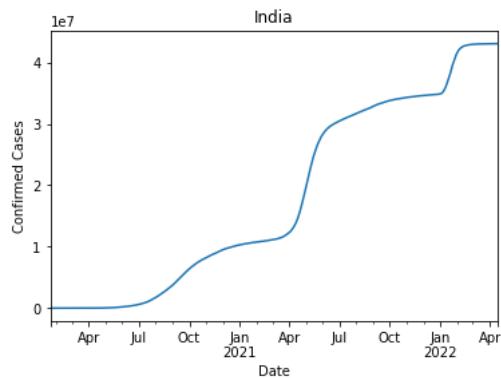
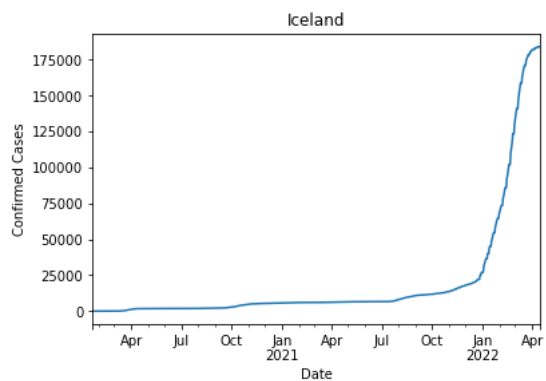


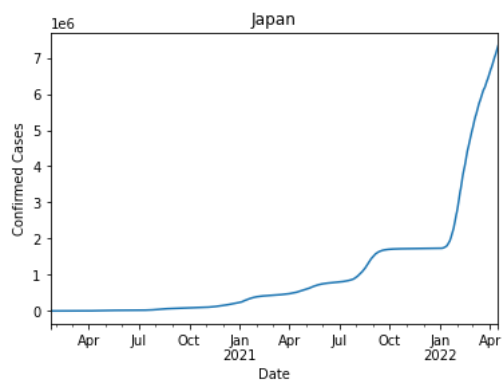
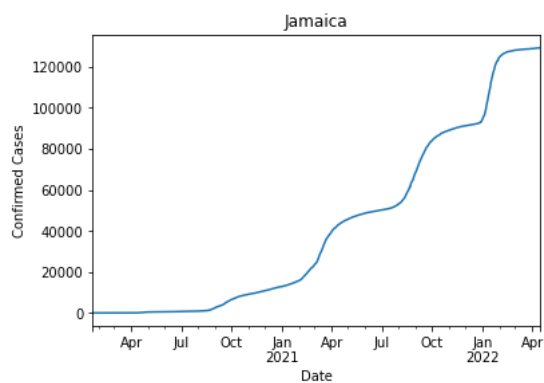
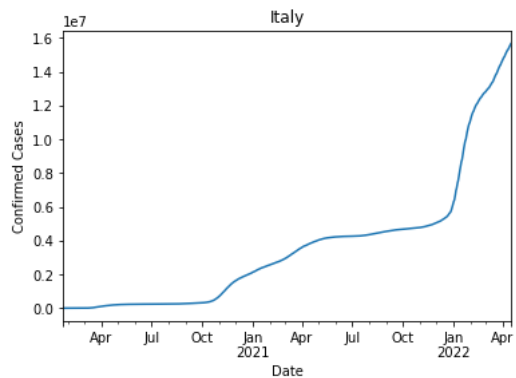
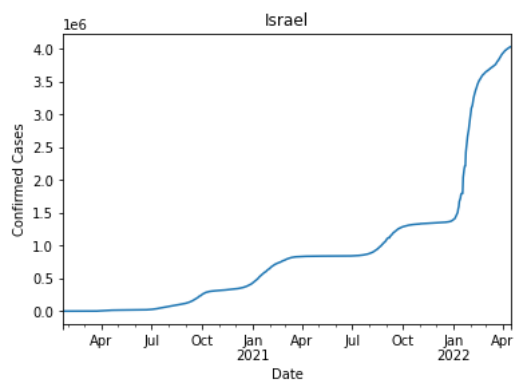
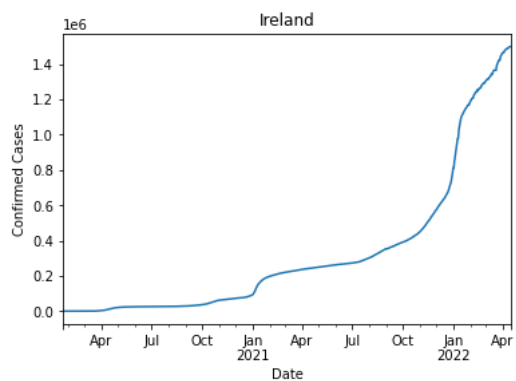


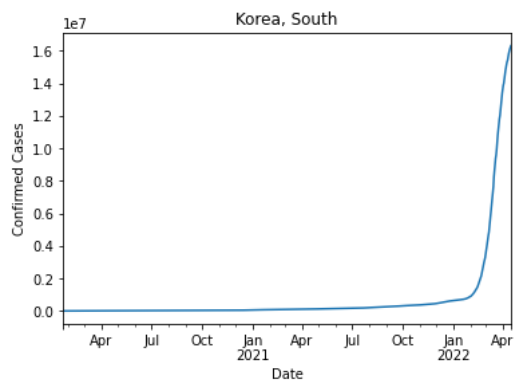
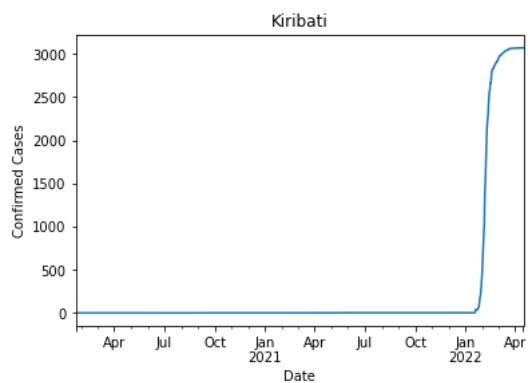
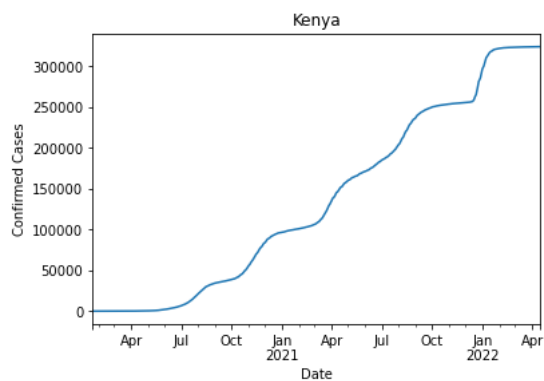
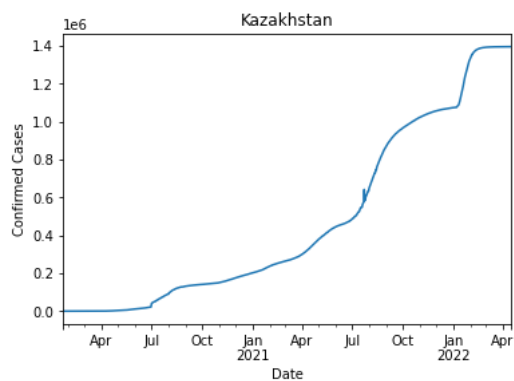
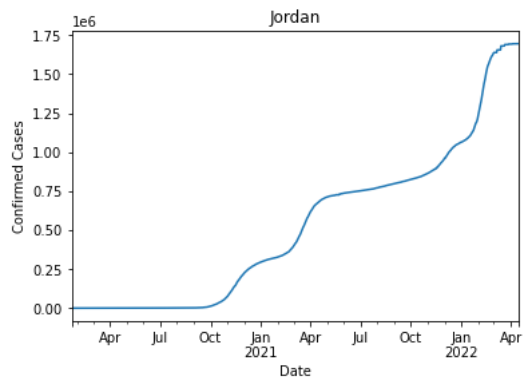


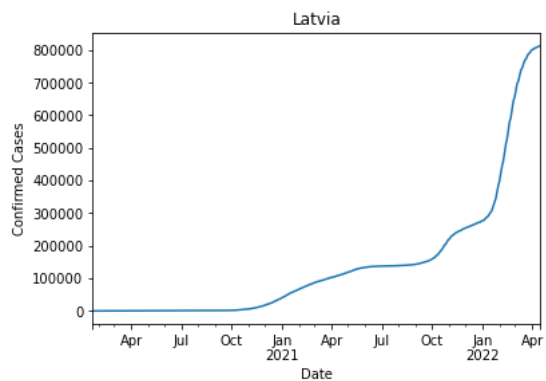
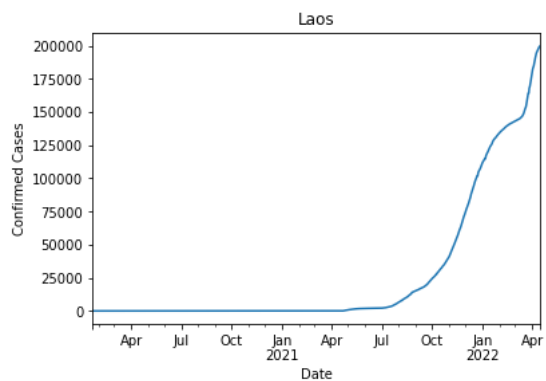
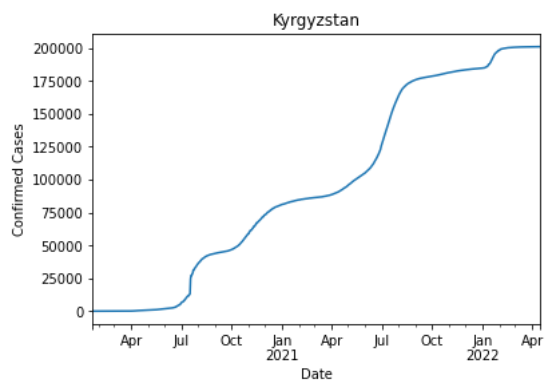
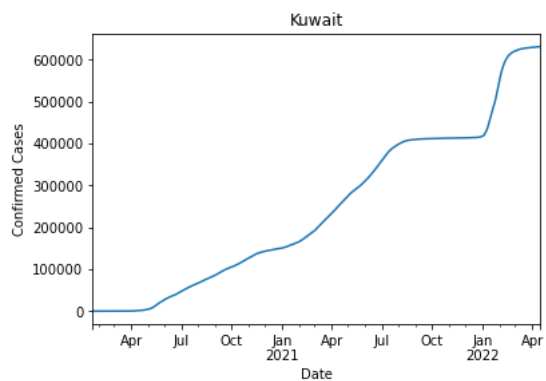
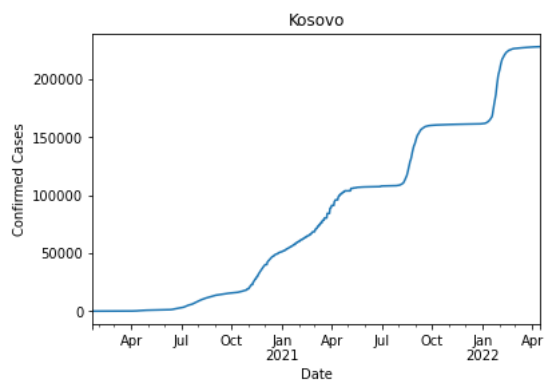


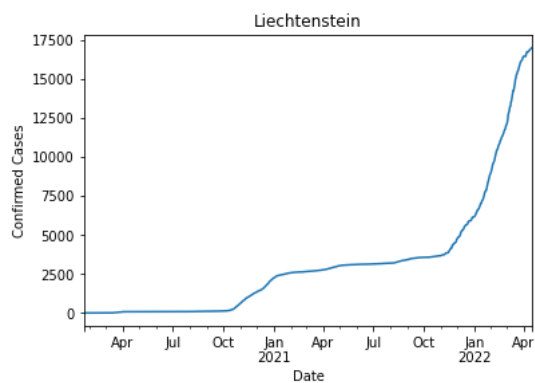
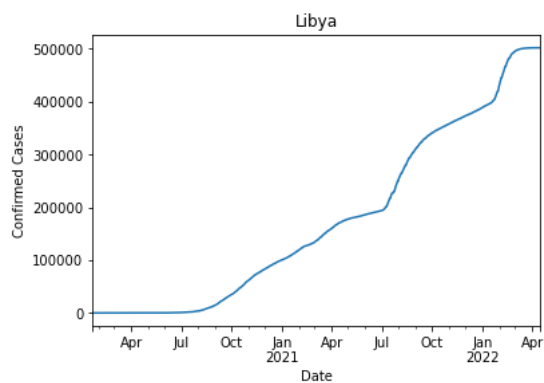
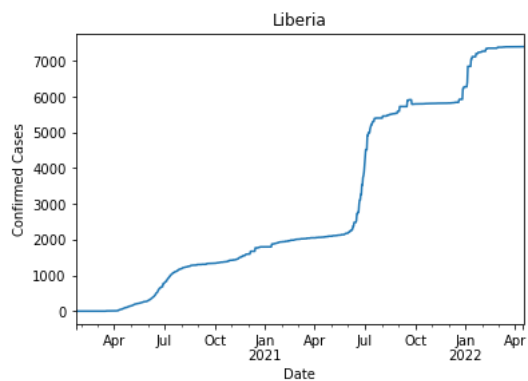
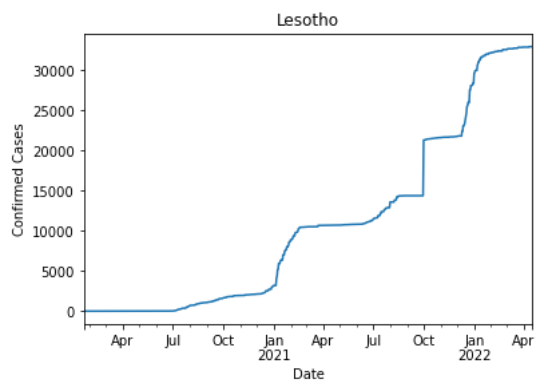
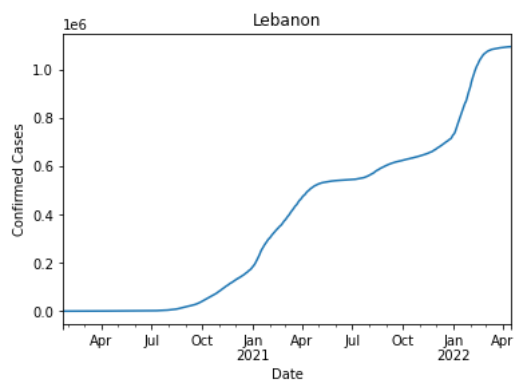


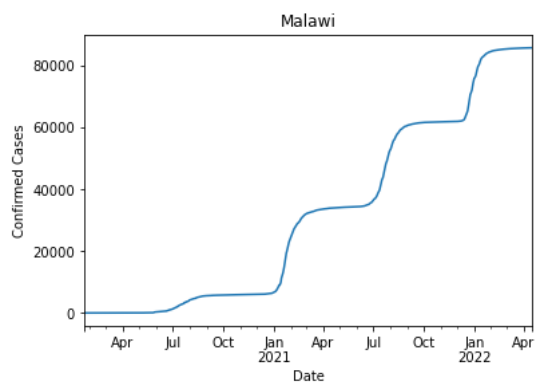
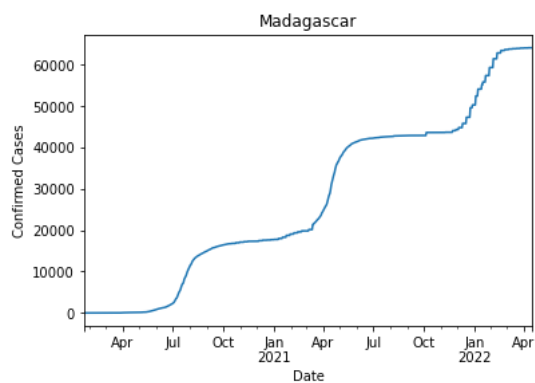
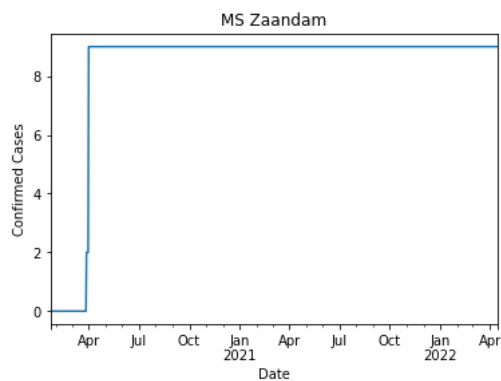
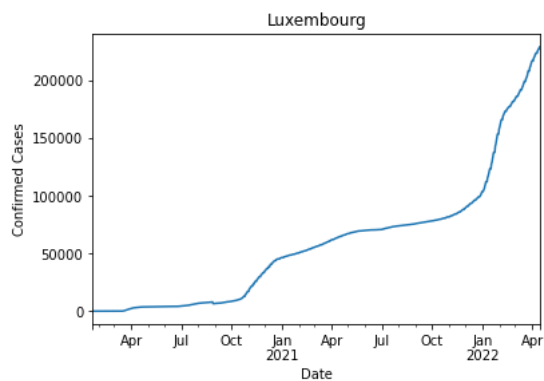
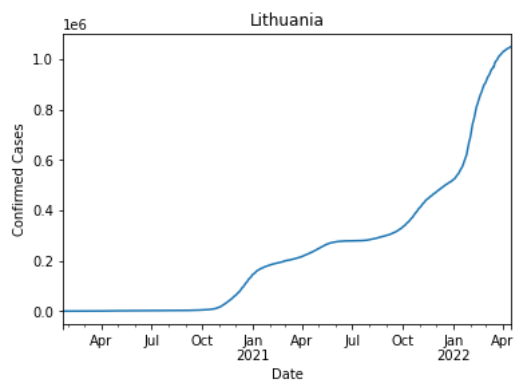


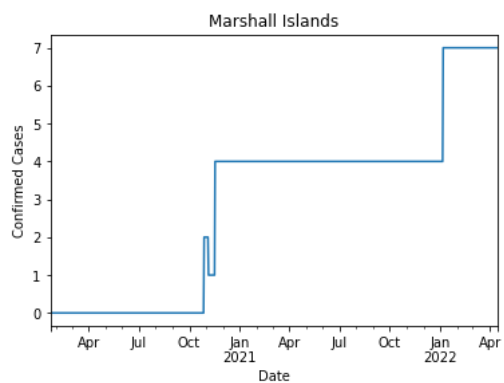
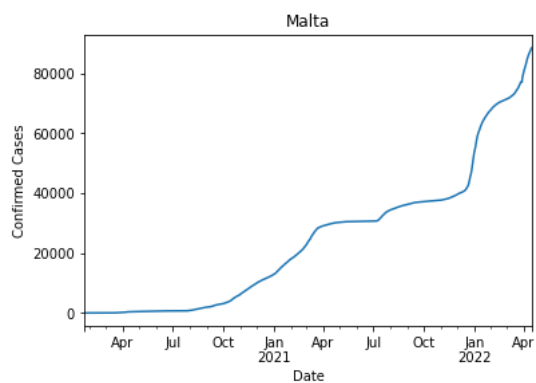
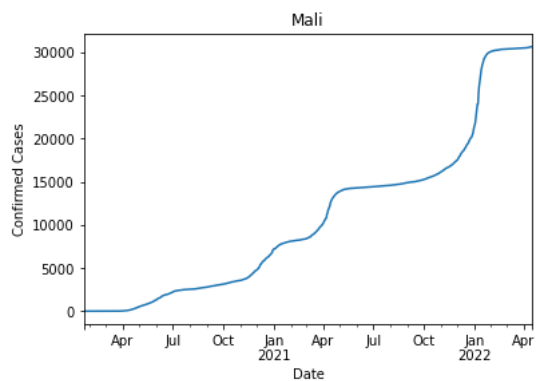
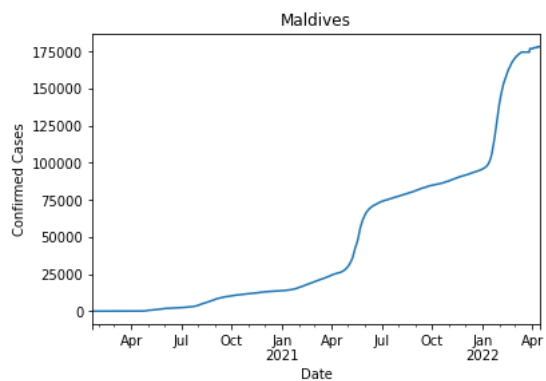
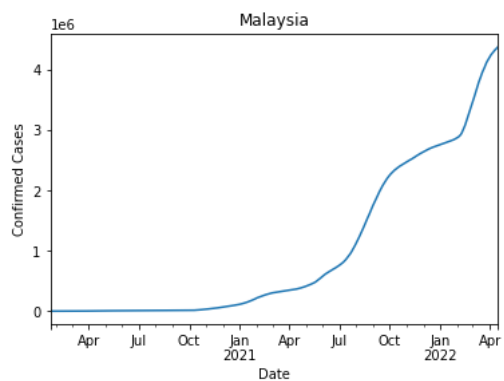


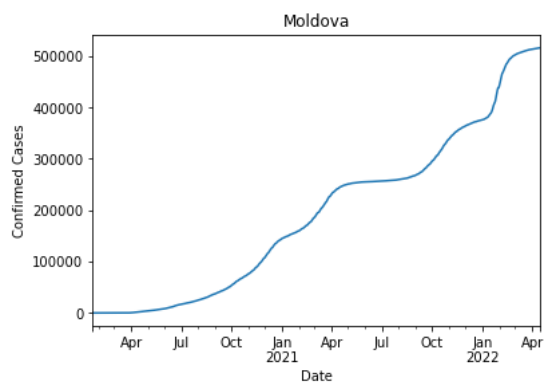
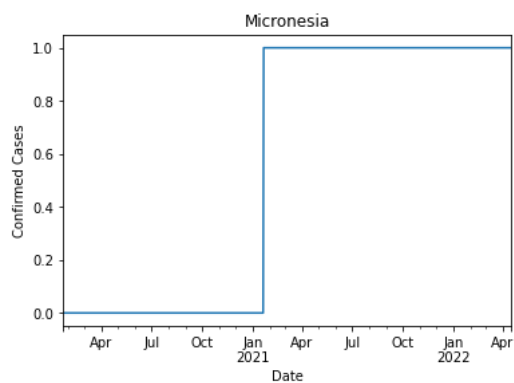
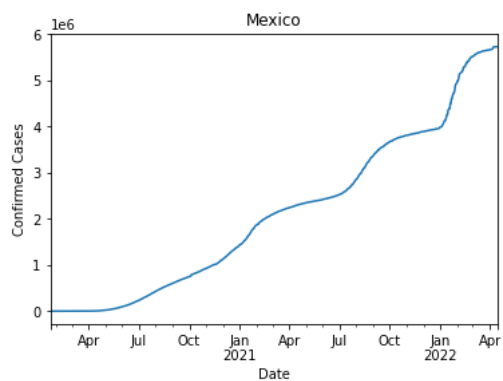
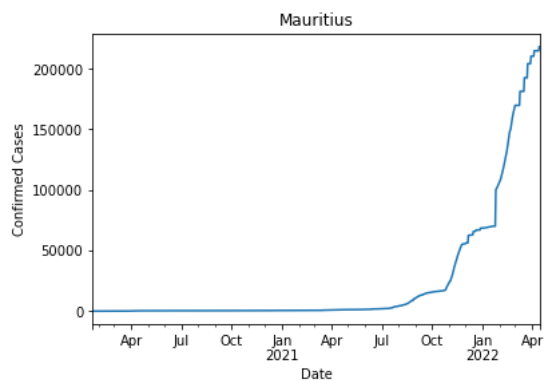
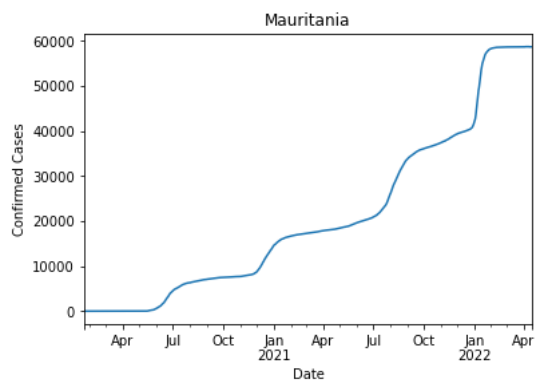


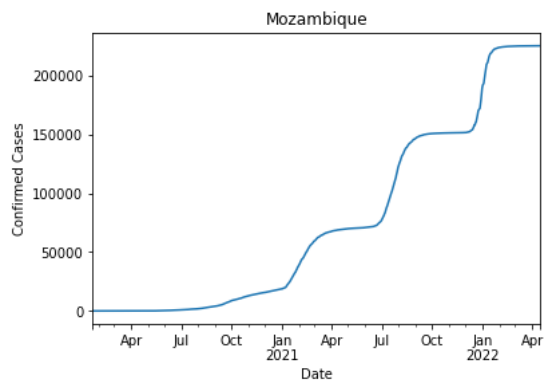
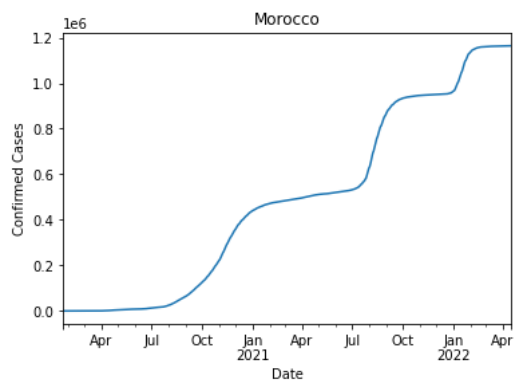
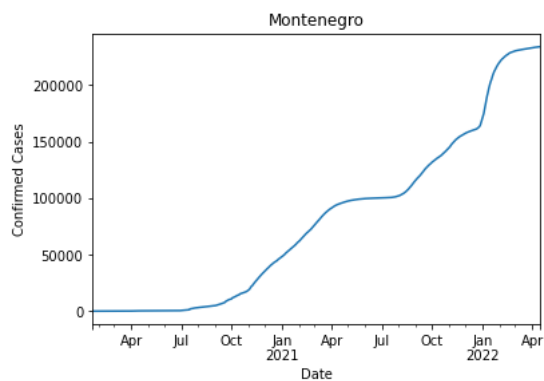
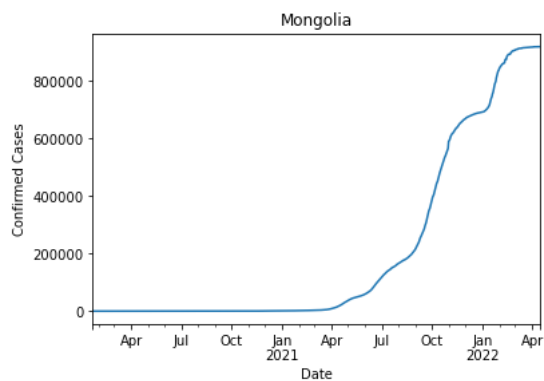
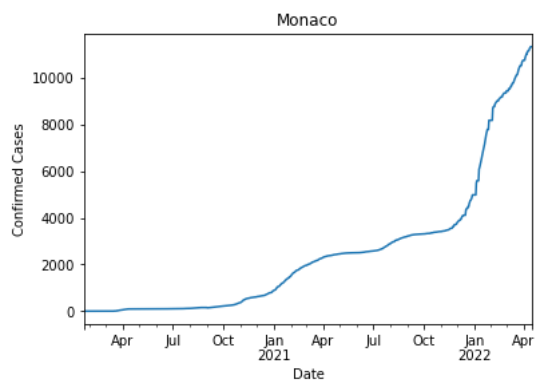


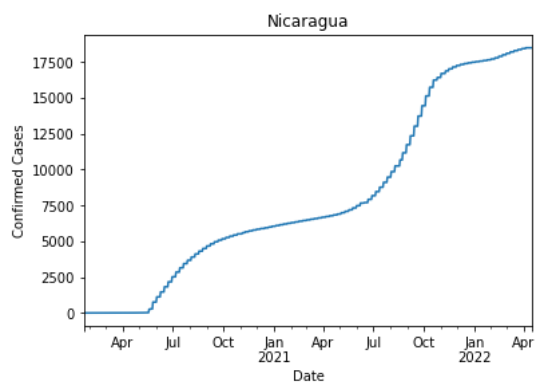
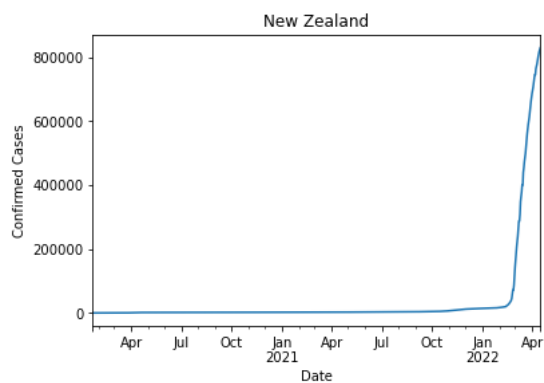
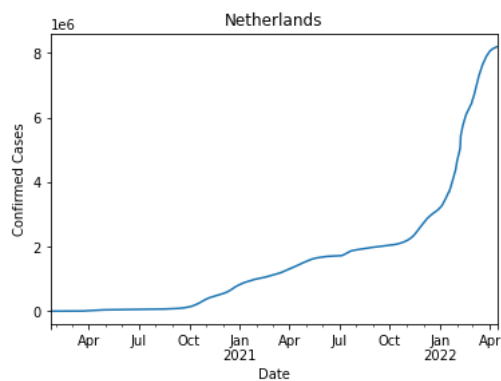
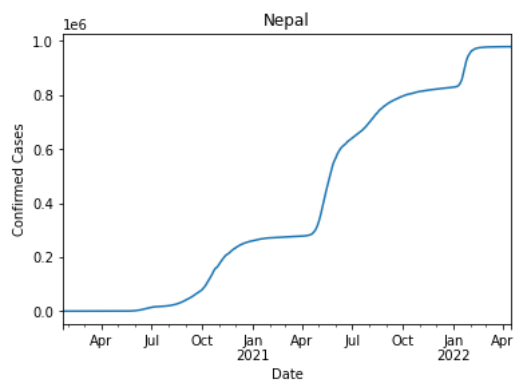
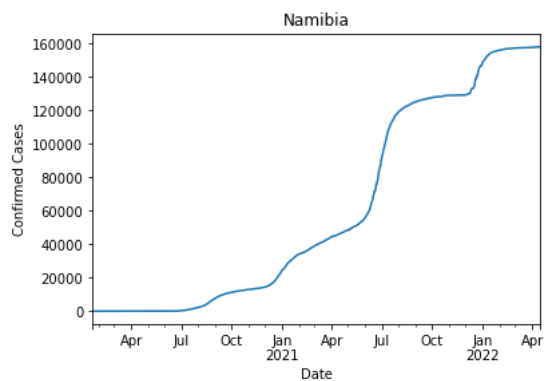


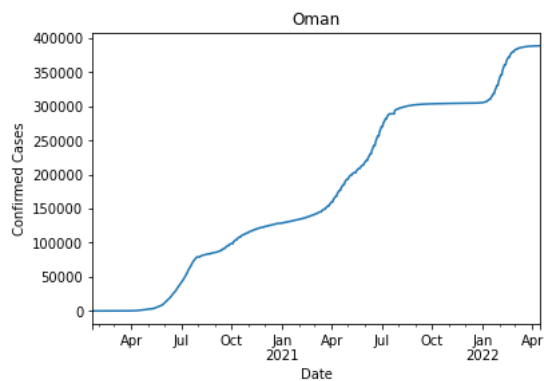
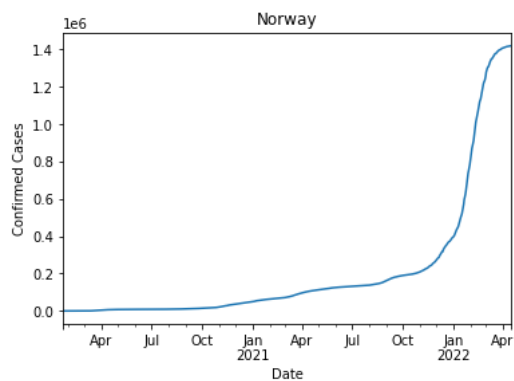
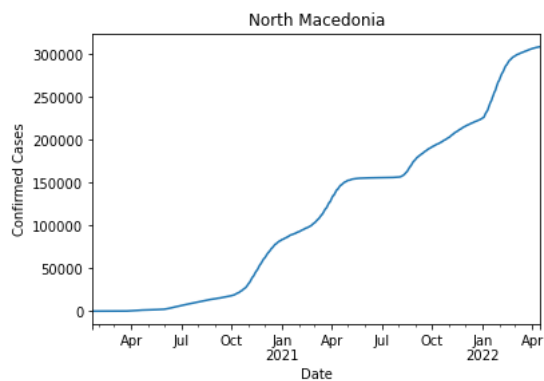
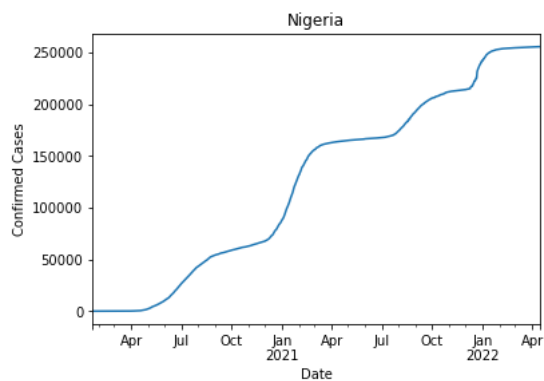
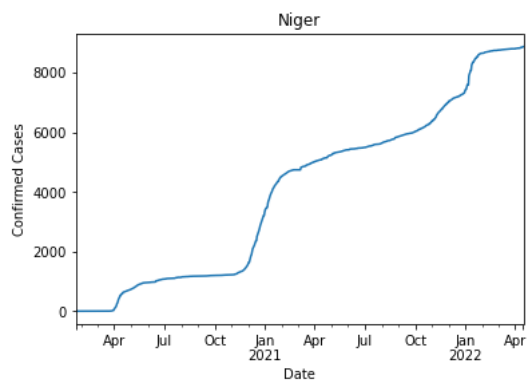


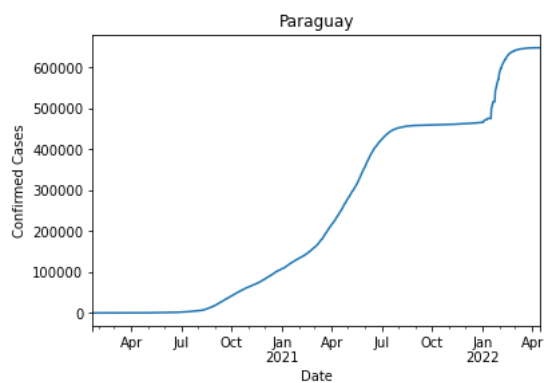
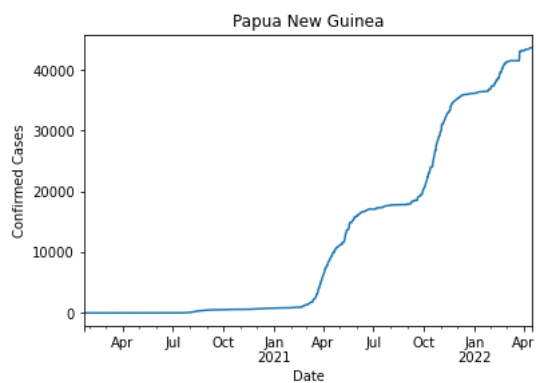
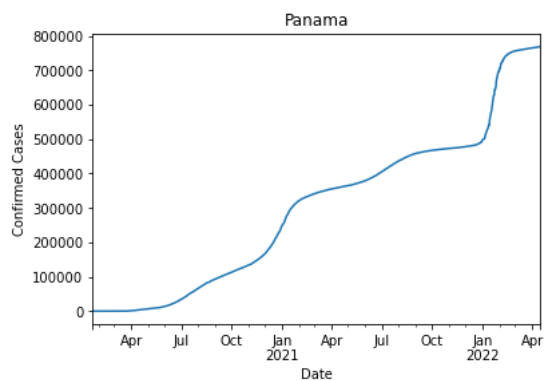
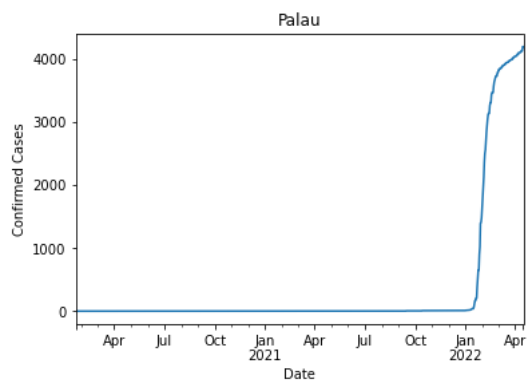
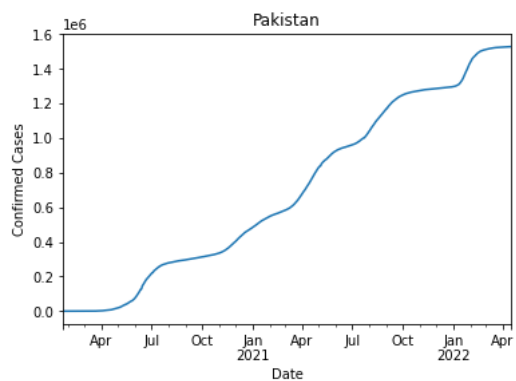


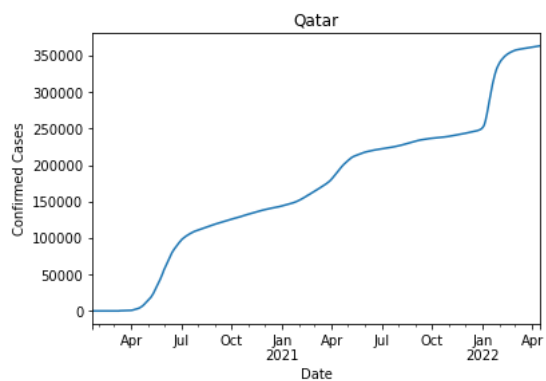
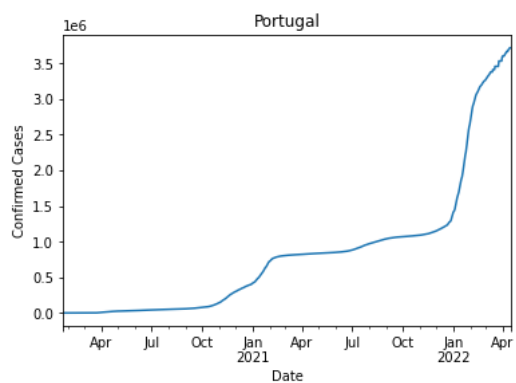
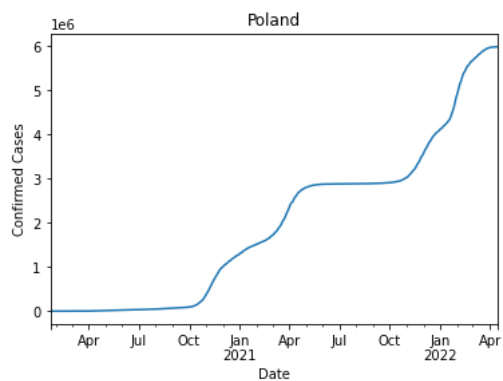
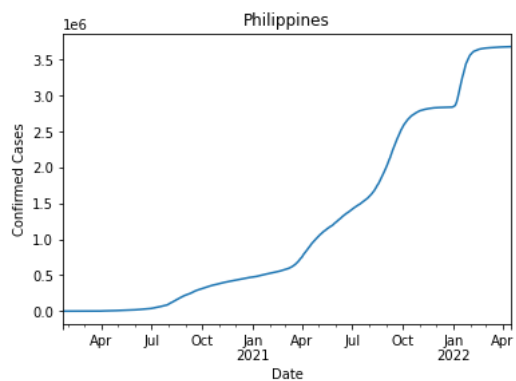
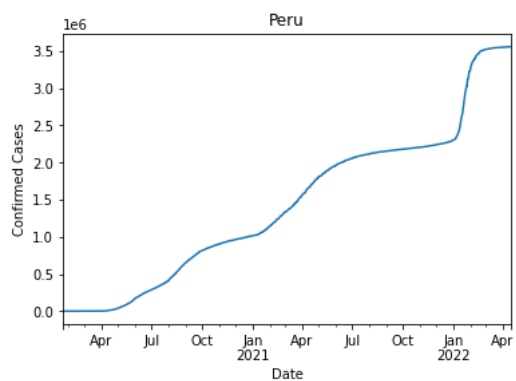


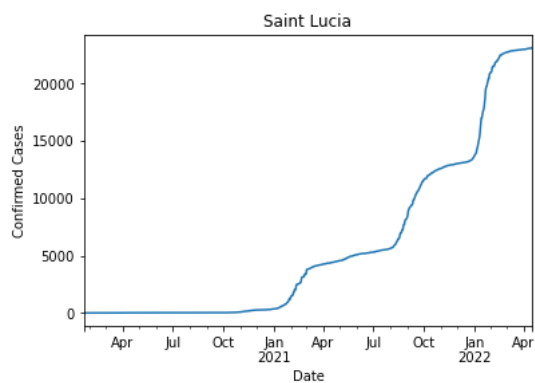
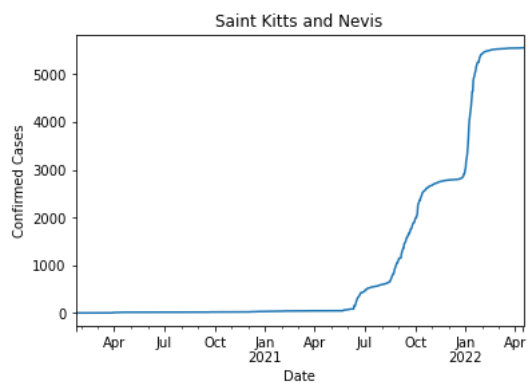
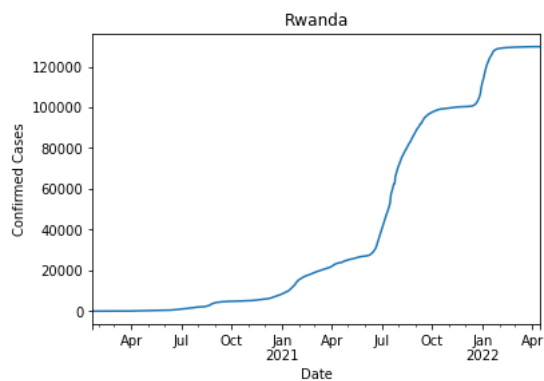
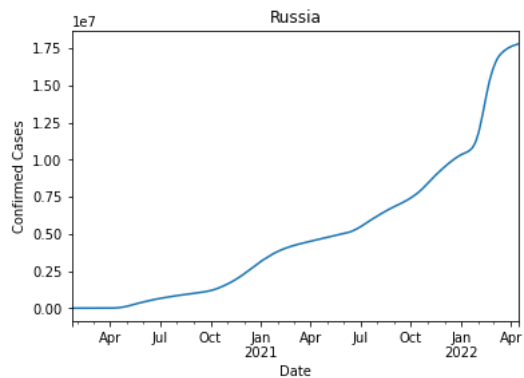
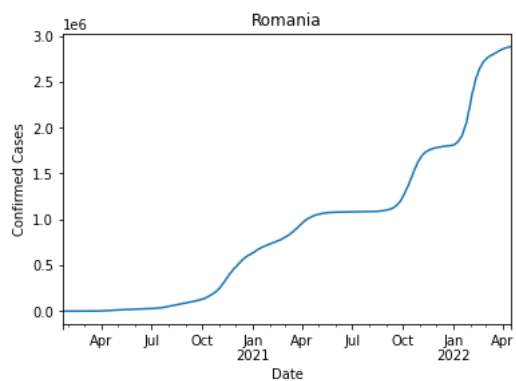


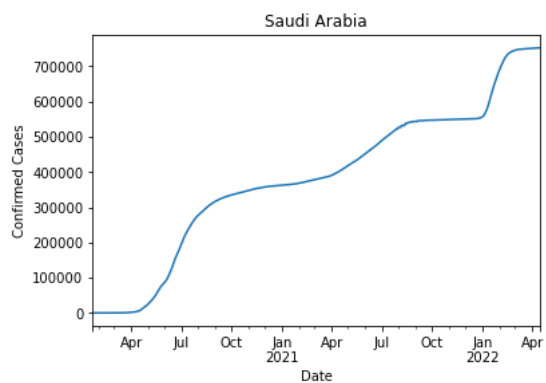
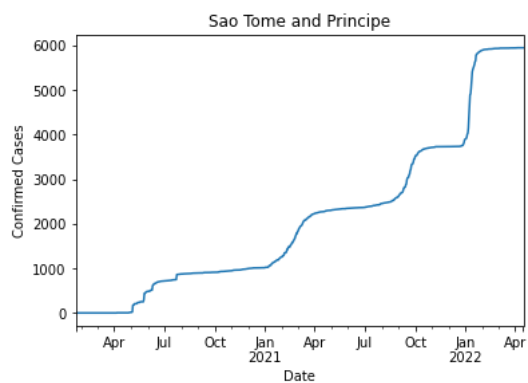
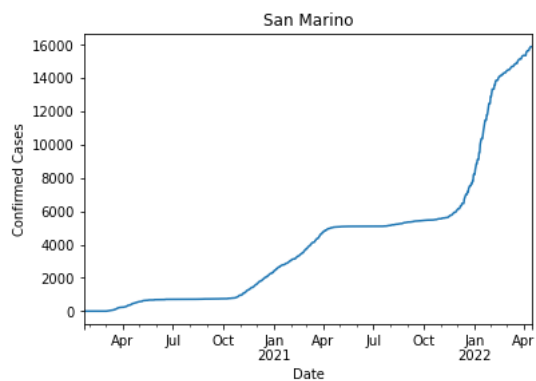
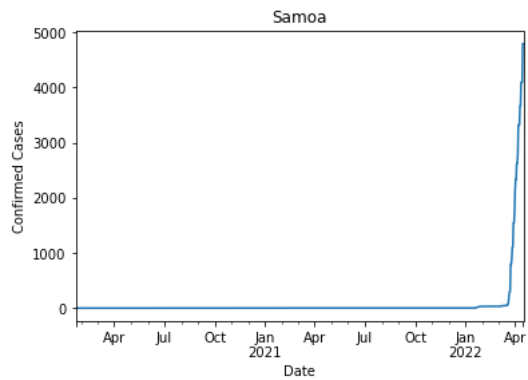
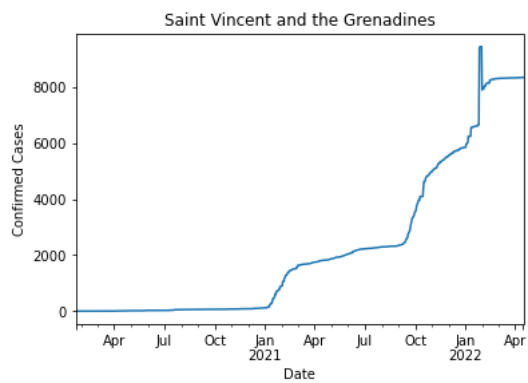


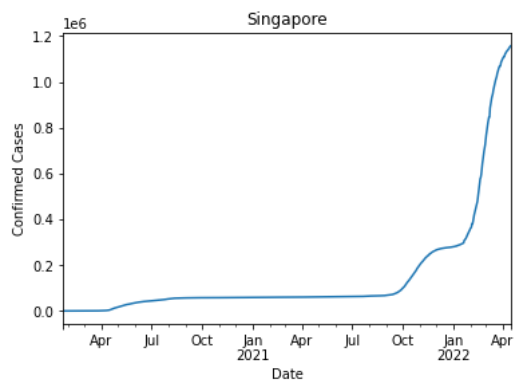
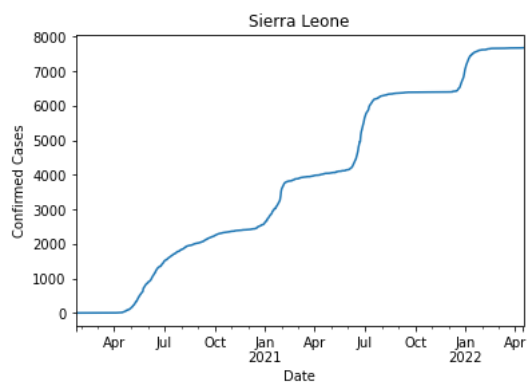
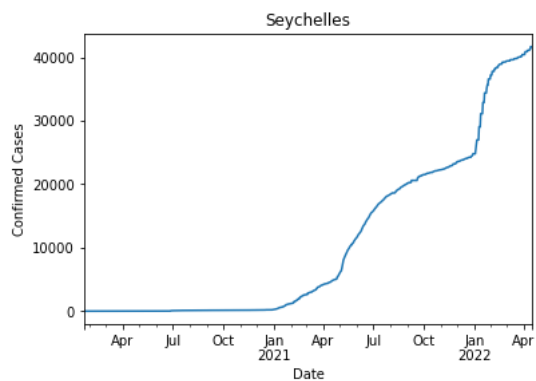
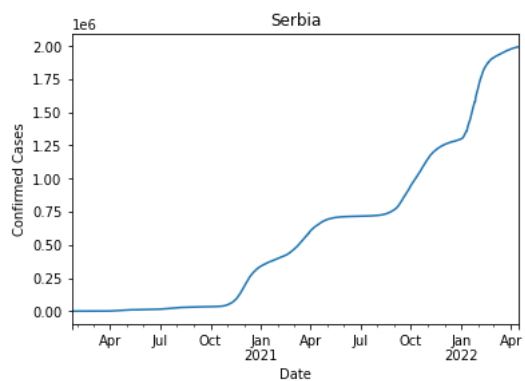
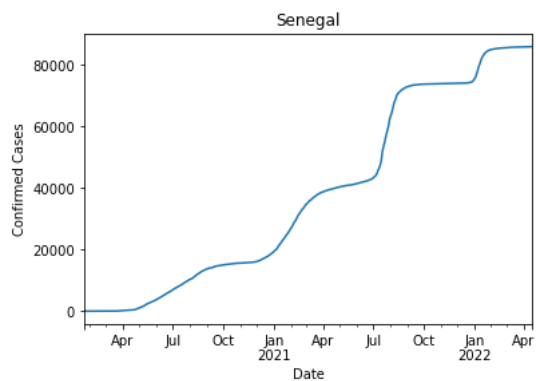


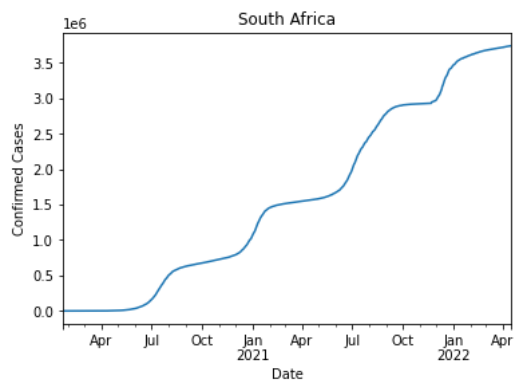
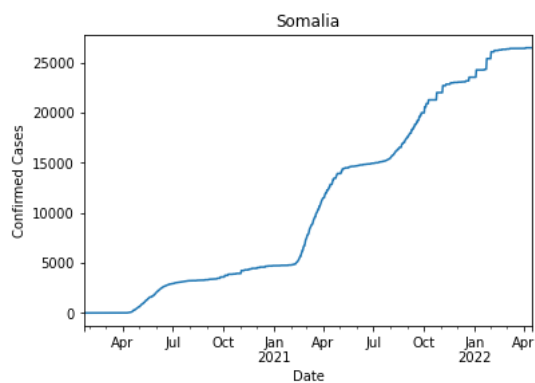
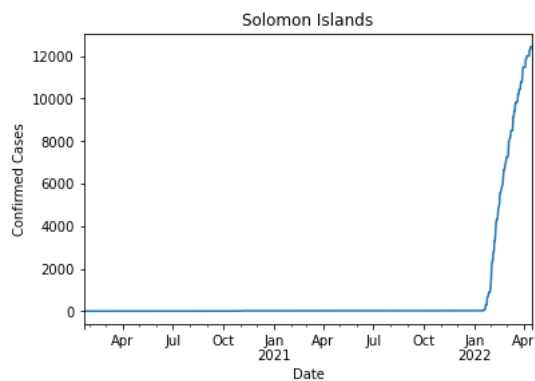
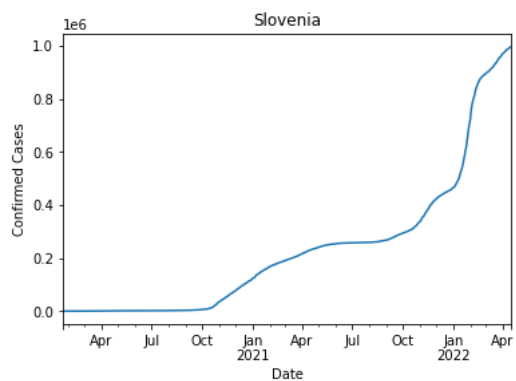
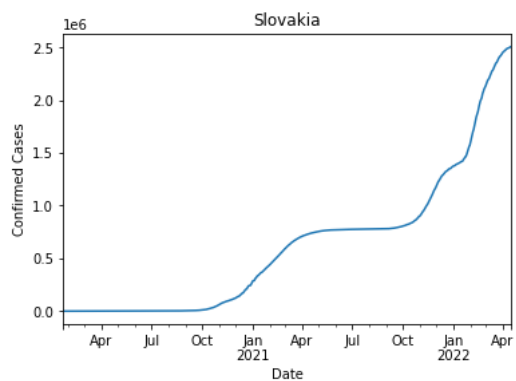


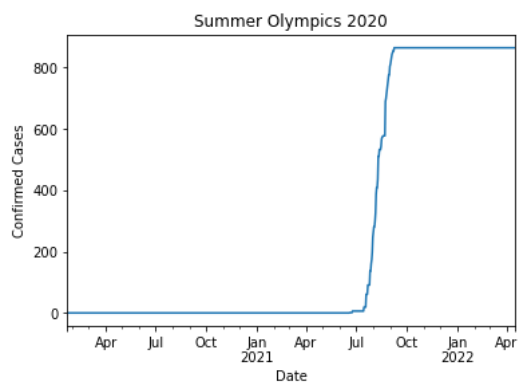
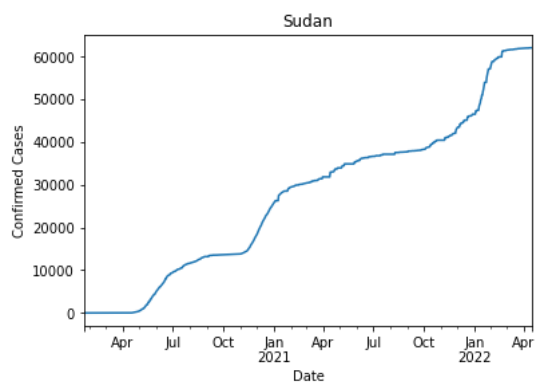
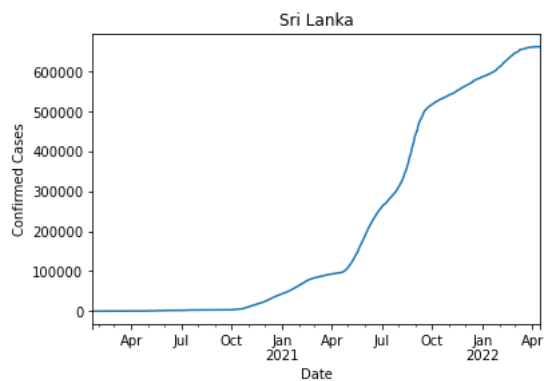
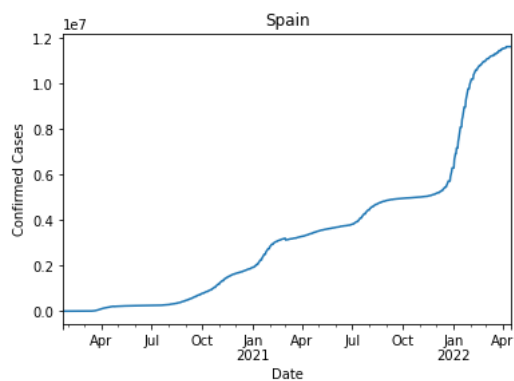
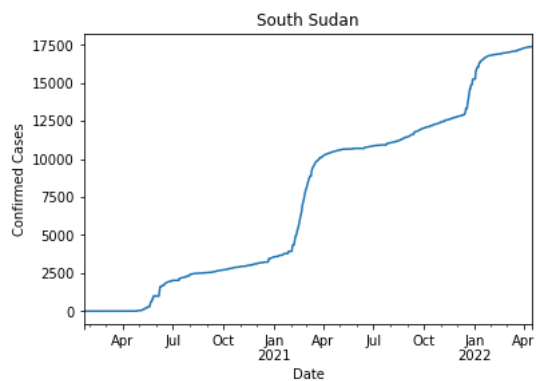


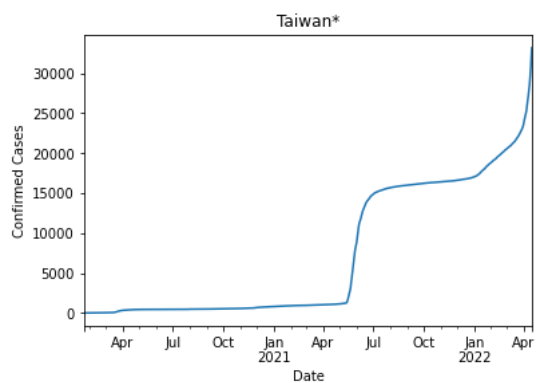
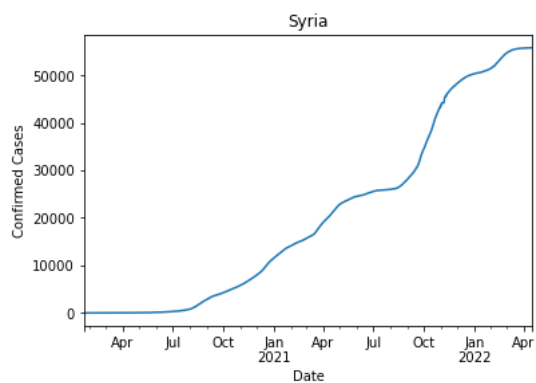
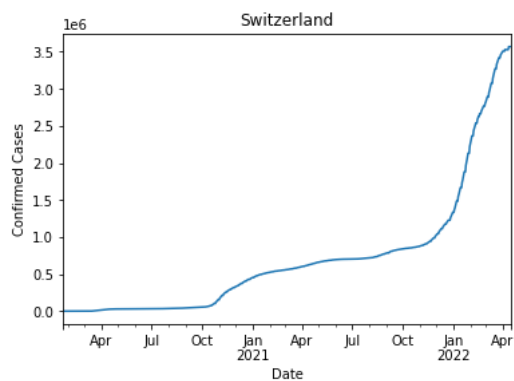
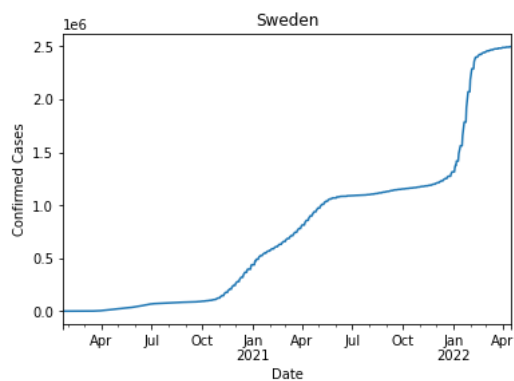
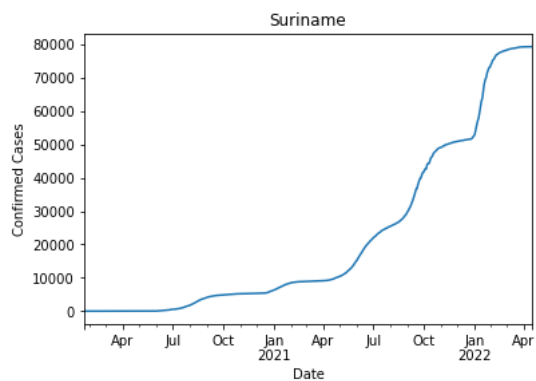


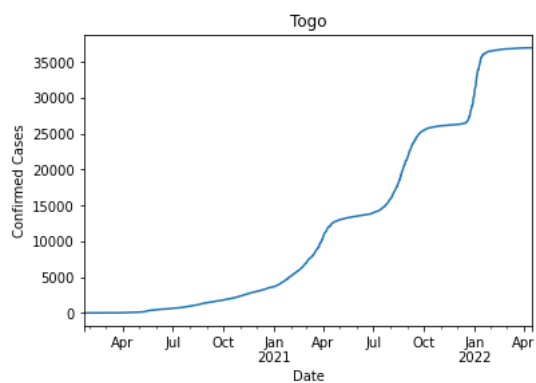
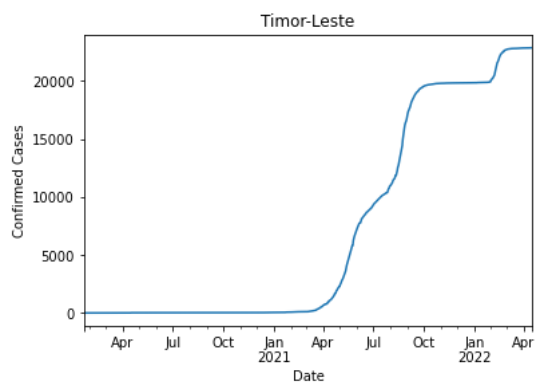
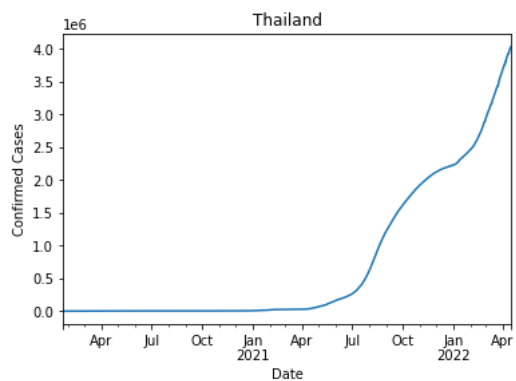
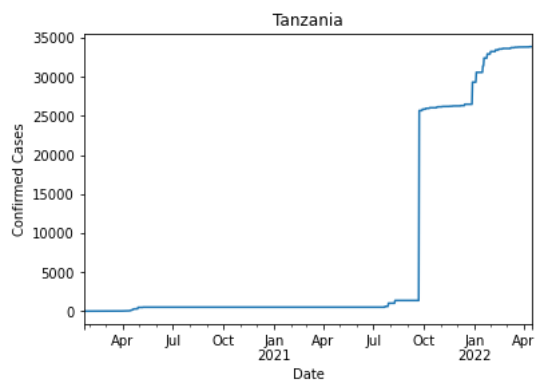
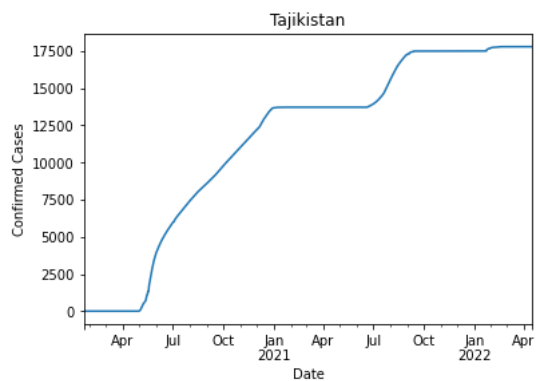


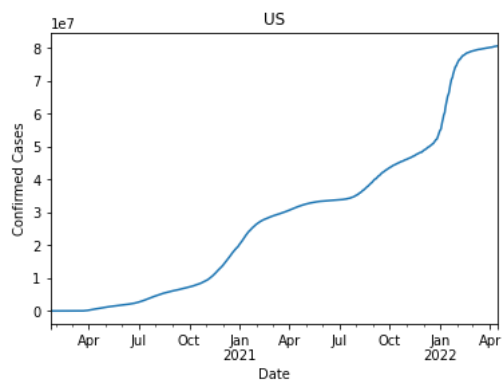
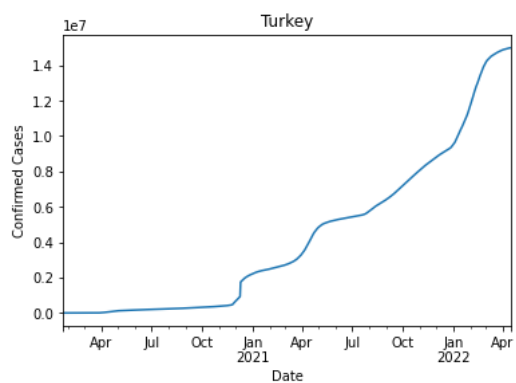
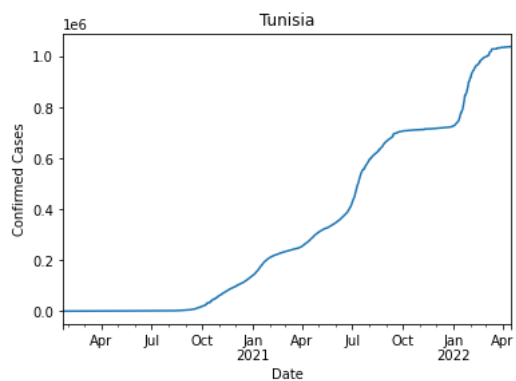
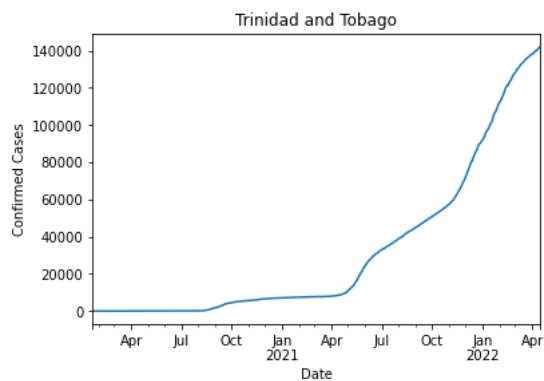
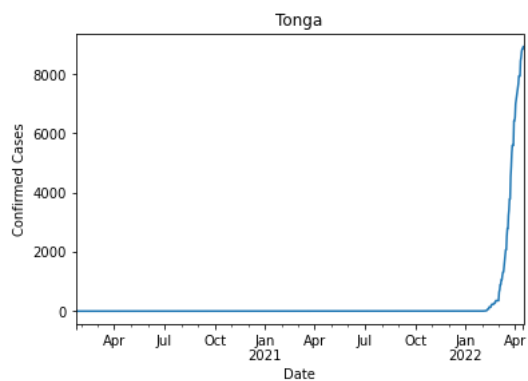


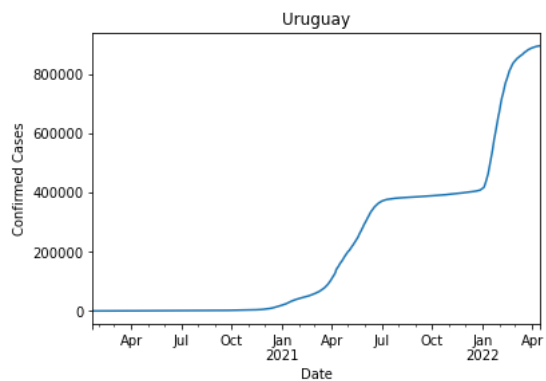
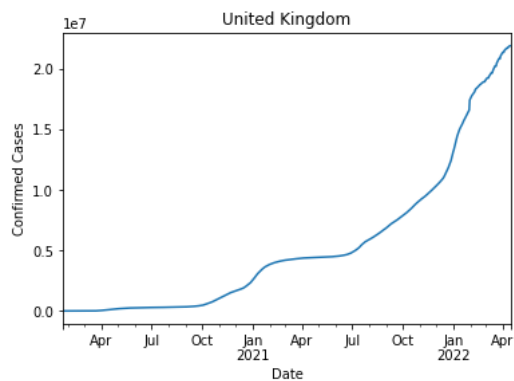
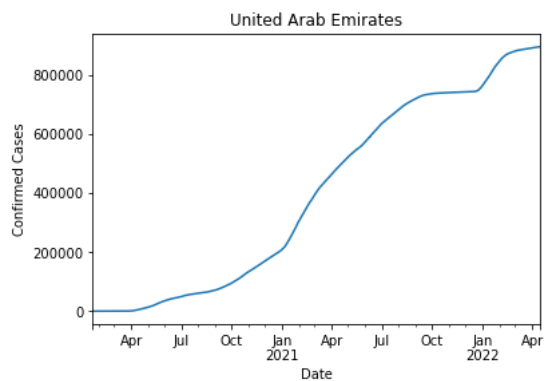
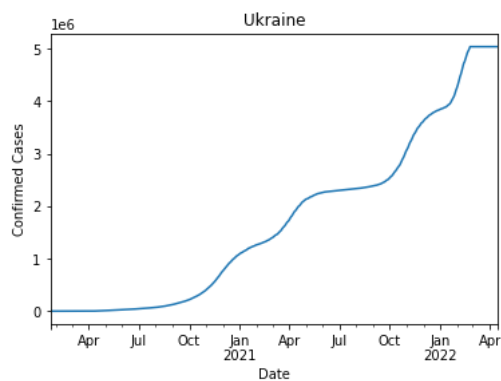
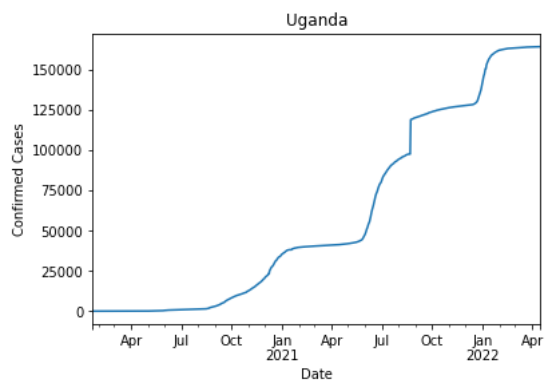


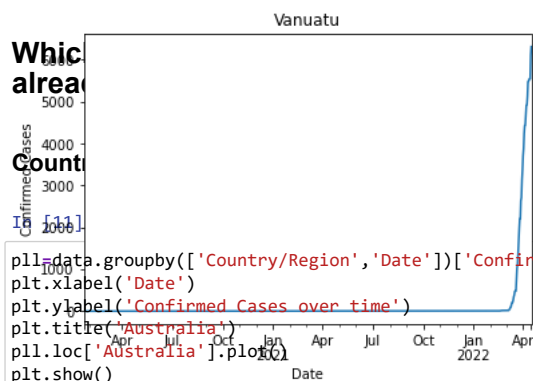
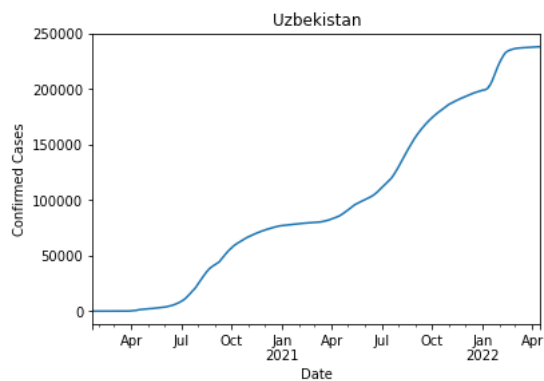








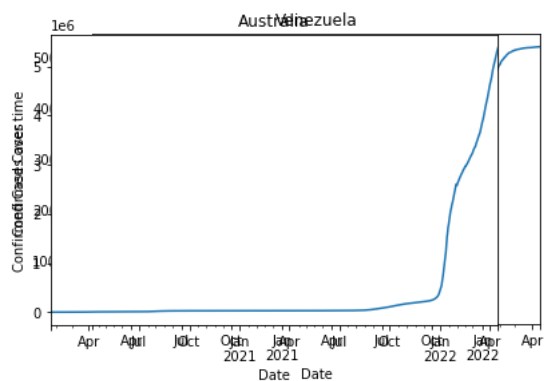




```

In [11]:
p11=data.groupby(['Country/Region','Date'])['Confirmed'].sum()
plt.xlabel('Date')
plt.ylabel('Confirmed Cases over time')
plt.title('Australia')
p11.loc['Australia'].plot()
plt.show()

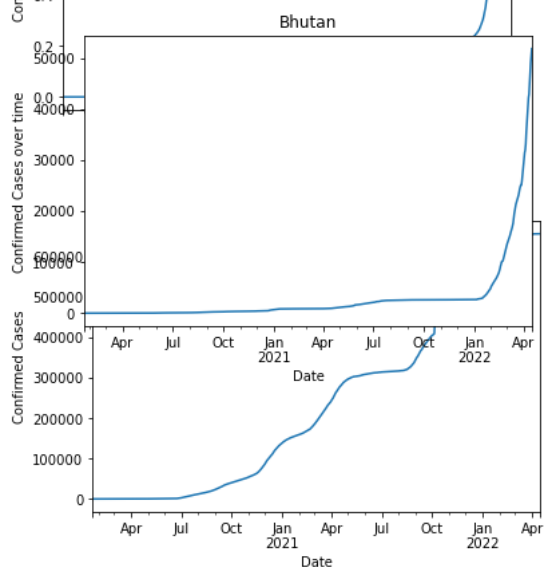
```



```

In [12]:
p11=data.groupby(['Country/Region','Date'])['Confirmed'].sum()
plt.xlabel('Date')
plt.ylabel('Confirmed Cases over time')
plt.title('Bhutan')
p11.loc['Bhutan'].plot()
plt.show()

```

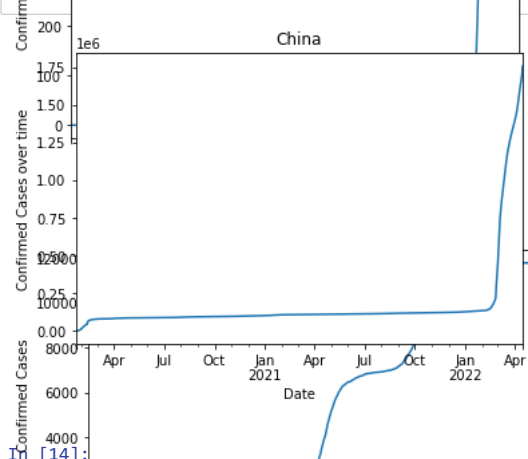


growth in the number of cases and which countries are

In [13]:

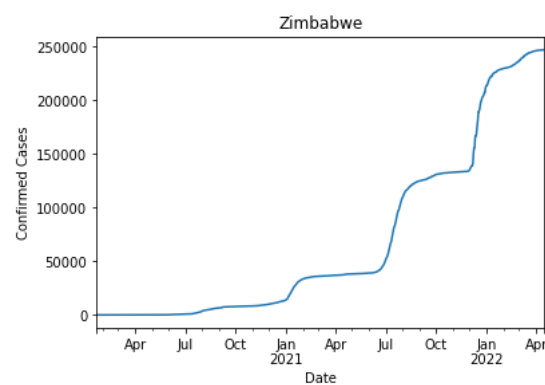
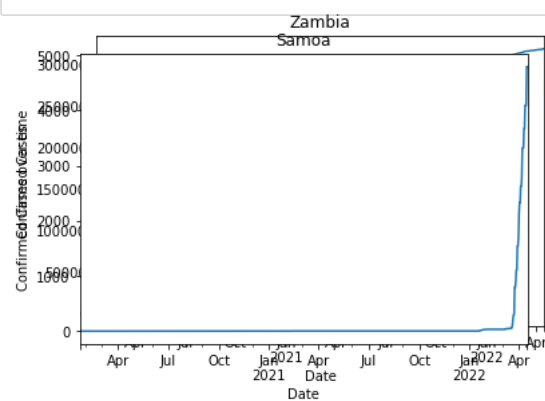
Winter Olympics 2022

```
p11=data.groupby(['Country/Region','Date'])['Confirmed'].sum()
plt.xlabel('Date')
plt.ylabel('Confirmed Cases over time')
plt.title('China')
p11.loc['China'].plot()
plt.show()
```



In [14]:

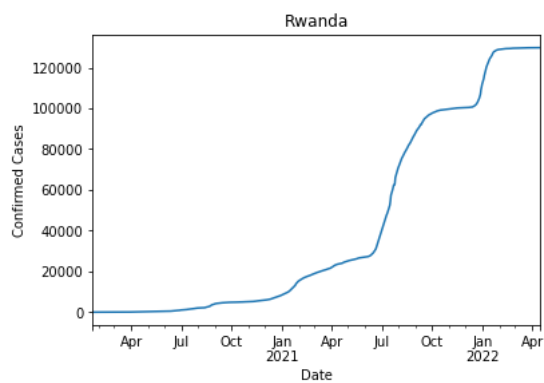
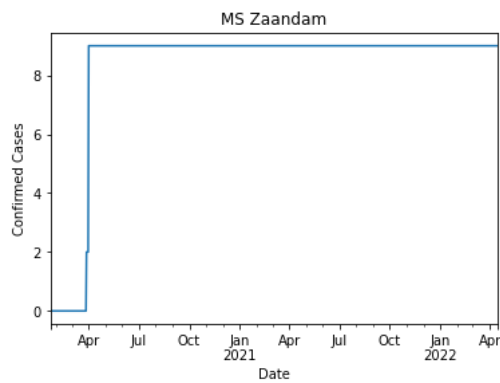
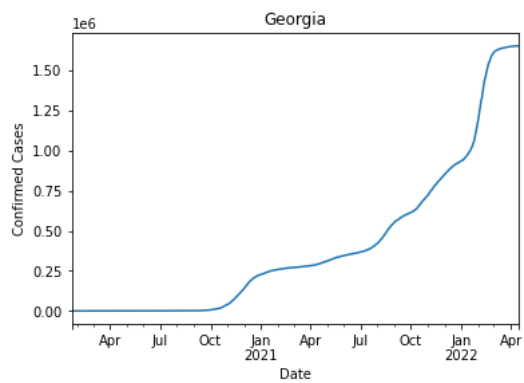
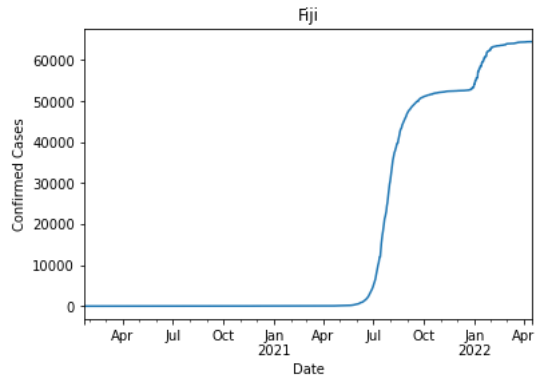
```
p11=data.groupby(['Country/Region','Date'])['Confirmed'].sum()
plt.xlabel('Date')
plt.ylabel('Confirmed Cases over time')
plt.title('Samoa')
p11.loc['Samoa'].plot()
plt.show()
```



Countries already leaving exponential growth

In [15]:

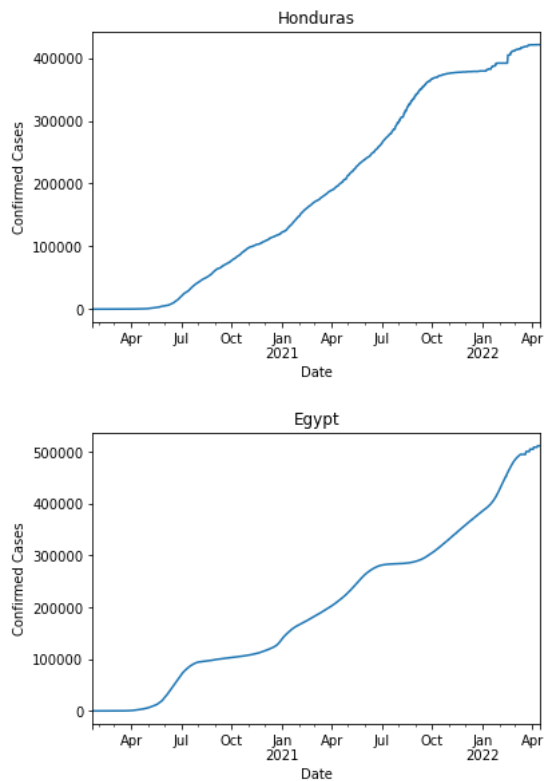
```
l=['Fiji','Georgia','MS Zaandam','Rwanda']  
for i in l:  
    plt.xlabel('Date')  
    plt.ylabel('Confirmed Cases')  
    plt.title(i)  
    pl.loc[i].plot()  
    plt.show();
```



Countries showing linear growth

In [16]:

```
l=['Honduras','Egypt']
for i in l:
    plt.xlabel('Date')
    plt.ylabel('Confirmed Cases')
    plt.title(i)
    pl.loc[i].plot()
    plt.show();
```



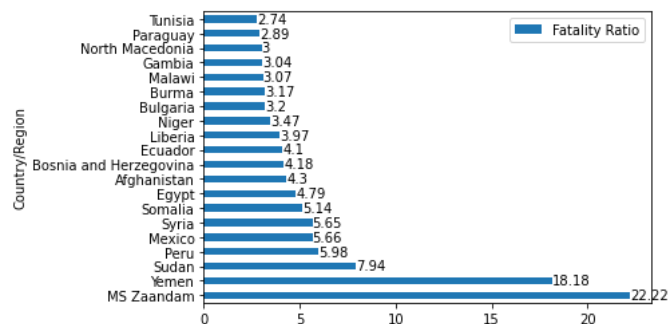
Create a bar plot that shows the number of deaths per 100 confirmed cases (observed case-fatality ratio) for the 20 most affected countries.

In [17]:

```
res['fatality_ratio'] = ((res['Deaths']/res['Confirmed'])*100).round(2) #calculating fatality ratio
res1=res
res1.drop(['Confirmed','Deaths'], axis =1) #dropping unnecessary columns
top = res1.nlargest(20,'fatality_ratio')['fatality_ratio'] #20 most affected countries
```

In [18]:

```
ax = top.plot.barh(x='Fatality Ratio', y='Country/Region') #plotting horizontal bar graph
ax.bar_label(ax.containers[0])
plt.legend(['Fatality Ratio']);
```



Compute the ratio between the total number of confirmed cases and the population size of each country. The file worldpopulation.json contains data on the population size of each country. Note that countries may have different names in different data sets (e.g., United Kingdom/U.K., US/USA/U.S.). What are the 10 countries with the highest number of confirmed COVID-19 cases per capita?

In [19]:

```
# reading the world population data from a csv and convert it to a DataFrame
population = pd.read_json('worldpopulation.json')
population.head()
```

Out[19]:

| | Rank | country | population | World |
|---|------|-----------|------------|-------|
| 0 | 1 | China | 1388232693 | 0.185 |
| 1 | 2 | India | 1342512706 | 0.179 |
| 2 | 3 | U.S. | 326474013 | 0.043 |
| 3 | 4 | Indonesia | 263510146 | 0.035 |
| 4 | 5 | Brazil | 211243220 | 0.028 |

In [20]:

```
# rename the column name in population so that it can be used while merging/ joining
population.rename(columns = {'country':'Country/Region'}, inplace = True)

# create a dictionary of all the country names which have to be modified and replace
population['Country/Region']=population['Country/Region'].replace({'U.S.': 'US', 'Viet Nam' : 'Vietnam', 'U.K.': 'United Kingdom', 'Myanmar' : 'Burma', 'South Korea': 'Korea, South', 'Côte d'Ivoire': 'Cote d'Ivoire', 'Czech Republic': 'Czechia', 'DR Congo': 'Congo (Kinshasa)', 'Congo': 'Congo (Brazzaville)', 'TFYR Macedonia': 'North Macedonia', 'Swaziland': 'Eswatini', 'St. Vincent & Grenadines': 'Saint Vincent and the Grenadines', 'State of Palestine': 'West Bank and Gaza'})

population
```

Out[20]:

| | Rank | Country/Region | population | World |
|-----|------|----------------|------------|-------|
| 0 | 1 | China | 1388232693 | 0.185 |
| 1 | 2 | India | 1342512706 | 0.179 |
| 2 | 3 | US | 326474013 | 0.043 |
| 3 | 4 | Indonesia | 263510146 | 0.035 |
| 4 | 5 | Brazil | 211243220 | 0.028 |
| ... | ... | ... | ... | ... |
| 190 | 191 | San Marino | 32104 | 0.000 |
| 191 | 192 | Palau | 21726 | 0.000 |
| 192 | 193 | Nauru | 10301 | 0.000 |
| 193 | 194 | Tuvalu | 9975 | 0.000 |
| 194 | 195 | Holy See | 801 | 0.000 |

195 rows × 4 columns

In [21]:

```
merge_table = pd.merge(res, population, on = "Country/Region", how = "inner")

#create a new column confirmed per capita and update the column values
merge_table['confirmed/capita']=merge_table['Confirmed']/merge_table['population']
merge_table
```

Out[21]:

| | Country/Region | Confirmed | Deaths | fatality_ratio | Rank | population | World | confirmed/capita |
|-----|--------------------|-----------|--------|----------------|------|------------|-------|------------------|
| 0 | Afghanistan | 178387 | 7676 | 4.30 | 40 | 34169169 | 0.005 | 0.005221 |
| 1 | Albania | 274462 | 3496 | 1.27 | 136 | 2911428 | 0.000 | 0.094271 |
| 2 | Algeria | 265739 | 6874 | 2.59 | 35 | 41063753 | 0.005 | 0.006471 |
| 3 | Andorra | 40709 | 153 | 0.38 | 186 | 68728 | 0.000 | 0.592320 |
| 4 | Angola | 99194 | 1900 | 1.92 | 50 | 26655513 | 0.004 | 0.003721 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 186 | Vietnam | 10417887 | 42934 | 0.41 | 14 | 95414640 | 0.013 | 0.109185 |
| 187 | West Bank and Gaza | 656617 | 5656 | 0.86 | 119 | 4928225 | 0.001 | 0.133236 |
| 188 | Yemen | 11817 | 2148 | 18.18 | 49 | 28119546 | 0.004 | 0.000420 |
| 189 | Zambia | 318467 | 3973 | 1.25 | 65 | 17237931 | 0.002 | 0.018475 |
| 190 | Zimbabwe | 247237 | 5462 | 2.21 | 69 | 16337760 | 0.002 | 0.015133 |

191 rows × 8 columns

In [22]:

```
# top 10 countries with confirmed cases per capita
merge_table.nlargest(10,'confirmed/capita')[['Country/Region','confirmed/capita']]
```

Out[22]:

| | Country/Region | confirmed/capita |
|-----|----------------|------------------|
| 3 | Andorra | 0.592320 |
| 47 | Denmark | 0.550374 |
| 77 | Iceland | 0.550321 |
| 147 | San Marino | 0.494456 |
| 83 | Israel | 0.484074 |
| 156 | Slovenia | 0.481270 |
| 122 | Netherlands | 0.481126 |
| 106 | Maldives | 0.474423 |
| 9 | Austria | 0.470859 |
| 155 | Slovakia | 0.461321 |

In this part we would like to test the hypothesis that the spread of the virus is slowed down by warm weather. Plot a graph of the monthly number of confirmed cases vs. the average monthly temperature for a few selected countries, and analyze the correlation between these two factors. You may use the file `climate.json` which contains monthly climate data from over 100 stations around the world, or use your own data sources.

In [23]:

```
import json
# read the json file
with open('climate.json') as f:
    climate_rec = json.load(f)

# create an empty data frame
climate = pd.DataFrame()

# create columns and fill them from the file
climate['city'] = [record['city'] for record in climate_rec]
climate['Country/Region'] = [record['country'] for record in climate_rec]

# create 12 months columns for each country and fill them with their avg temperature every month
for i in range(12):
    climate[f'month_{i+1}_avg'] = [(record['monthlyAvg'][i]['high'] + record['monthlyAvg'][i]['low'])/2 for record in climate_rec]

country_avg = climate.groupby('Country/Region').mean(numeric_only=True).round(2).reset_index()
country_avg
```

Out[23]:

| | Country/Region | month_1_avg | month_2_avg | month_3_avg | month_4_avg | month_5_avg | month_6_avg | month_7_avg | month_8_avg | month_9_avg | mo |
|----|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----|
| 0 | Argentina | 25.00 | 24.00 | 22.50 | 18.50 | 15.50 | 12.50 | 11.50 | 13.00 | 15.00 | |
| 1 | Australia | 22.00 | 22.25 | 20.12 | 17.00 | 14.12 | 11.62 | 10.62 | 11.75 | 13.88 | |
| 2 | Austria | 0.00 | 2.00 | 5.50 | 10.50 | 15.00 | 18.50 | 20.00 | 20.00 | 15.50 | |
| 3 | Belgium | 3.50 | 4.50 | 7.50 | 9.50 | 14.00 | 16.00 | 18.50 | 18.50 | 15.00 | |
| 4 | Brazil | 26.17 | 26.33 | 25.83 | 24.83 | 22.17 | 21.33 | 20.83 | 22.33 | 23.17 | |
| 5 | Bulgaria | -0.50 | 1.50 | 6.00 | 11.00 | 15.00 | 19.00 | 21.00 | 21.00 | 16.50 | |
| 6 | Canada | -6.14 | -4.86 | -0.57 | 5.71 | 11.29 | 16.36 | 18.93 | 18.57 | 15.14 | |
| 7 | Chile | 21.50 | 21.00 | 19.00 | 15.50 | 12.50 | 10.50 | 9.00 | 11.00 | 12.50 | |
| 8 | China | 0.50 | 3.50 | 8.75 | 15.25 | 20.75 | 25.00 | 28.00 | 27.50 | 22.50 | |
| 9 | Czech Republic | 0.00 | 1.50 | 5.00 | 9.00 | 14.50 | 17.50 | 19.50 | 19.00 | 14.00 | |
| 10 | Denmark | 1.50 | 2.00 | 4.00 | 8.00 | 12.50 | 15.50 | 18.00 | 18.00 | 13.50 | |
| 11 | France | 6.00 | 6.50 | 9.50 | 12.00 | 16.25 | 20.00 | 21.50 | 21.50 | 18.50 | |
| 12 | Germany | 1.00 | 0.50 | 4.50 | 8.50 | 14.00 | 16.50 | 19.00 | 19.00 | 15.00 | |
| 13 | Greece | 9.50 | 9.50 | 11.00 | 14.50 | 18.50 | 22.50 | 25.00 | 25.00 | 22.00 | |
| 14 | Hawaii | 22.50 | 22.50 | 23.50 | 24.00 | 24.50 | 26.50 | 26.50 | 26.50 | 26.50 | |
| 15 | Hong Kong | 15.00 | 16.00 | 18.50 | 22.50 | 26.50 | 28.50 | 29.00 | 29.00 | 28.00 | |
| 16 | Hungary | 0.00 | 2.00 | 6.50 | 12.00 | 17.00 | 20.50 | 22.00 | 22.00 | 17.00 | |
| 17 | Iceland | 1.00 | 0.00 | 1.00 | 3.50 | 7.00 | 10.00 | 11.50 | 11.00 | 8.50 | |
| 18 | India | 18.17 | 20.67 | 24.17 | 27.17 | 28.83 | 28.83 | 27.67 | 27.33 | 27.00 | |
| 19 | Indonesia | 27.75 | 27.50 | 28.00 | 28.00 | 28.00 | 27.50 | 27.00 | 27.00 | 22.50 | |
| 20 | Ireland | 4.50 | 5.00 | 6.00 | 7.50 | 11.00 | 13.00 | 15.00 | 14.50 | 12.50 | |
| 21 | Israel | 10.75 | 11.25 | 13.50 | 17.50 | 21.00 | 23.75 | 26.00 | 26.00 | 24.50 | |
| 22 | Italy | 6.17 | 7.17 | 10.67 | 13.17 | 18.50 | 22.33 | 25.00 | 25.00 | 20.50 | |
| 23 | Japan | 5.25 | 6.00 | 9.00 | 14.50 | 19.00 | 23.00 | 27.00 | 28.25 | 24.25 | |
| 24 | Malaysia | 28.00 | 28.50 | 29.00 | 28.50 | 29.00 | 28.50 | 28.50 | 28.50 | 28.50 | |
| 25 | Mexico | 14.00 | 16.00 | 17.50 | 19.50 | 20.00 | 19.50 | 18.50 | 18.50 | 18.00 | |
| 26 | Morocco | 15.00 | 16.50 | 18.50 | 19.00 | 20.00 | 22.50 | 24.00 | 24.00 | 23.00 | |
| 27 | Netherlands | 5.00 | 4.50 | 8.00 | 9.00 | 13.50 | 14.00 | 16.00 | 16.00 | 13.50 | |
| 28 | New Zealand | 17.67 | 18.00 | 16.50 | 14.17 | 11.83 | 9.67 | 8.83 | 9.50 | 11.17 | |
| 29 | Norway | -2.00 | -1.00 | 1.00 | 6.00 | 11.50 | 15.50 | 17.50 | 17.50 | 12.50 | |
| 30 | Poland | -1.00 | -0.50 | 3.00 | 8.50 | 14.00 | 16.50 | 19.50 | 18.50 | 13.50 | |
| 31 | Portugal | 11.00 | 11.50 | 13.00 | 13.75 | 15.50 | 17.50 | 18.75 | 19.25 | 18.25 | |
| 32 | Russia | -5.75 | -6.75 | -1.50 | 5.50 | 10.75 | 15.75 | 18.25 | 16.25 | 10.75 | |
| 33 | Singapore | 27.00 | 28.00 | 28.50 | 28.50 | 28.50 | 28.50 | 28.00 | 28.00 | 28.50 | |
| 34 | South Africa | 21.00 | 21.25 | 20.00 | 17.50 | 14.50 | 12.00 | 11.75 | 13.25 | 16.00 | |
| 35 | South Korea | -2.00 | 0.50 | 6.50 | 13.00 | 18.00 | 22.50 | 25.50 | 26.00 | 21.50 | |
| 36 | Spain | 8.70 | 10.00 | 12.20 | 13.60 | 16.90 | 21.10 | 24.00 | 24.20 | 21.30 | |
| 37 | Sweden | -1.50 | -2.00 | 1.00 | 5.50 | 10.50 | 15.00 | 17.00 | 16.00 | 11.50 | |
| 38 | Switzerland | 1.00 | 2.00 | 6.00 | 9.00 | 14.00 | 17.00 | 19.00 | 19.00 | 14.50 | |
| 39 | Thailand | 25.33 | 26.50 | 28.33 | 30.00 | 29.33 | 29.00 | 28.50 | 28.17 | 28.17 | |
| 40 | Turkey | 5.00 | 6.00 | 9.00 | 13.00 | 17.50 | 22.50 | 24.50 | 25.00 | 20.50 | |
| 41 | United Arab Emirates | 19.50 | 20.50 | 23.00 | 27.00 | 31.50 | 34.00 | 35.50 | 36.00 | 33.50 | |
| 42 | United Kingdom | 7.00 | 7.50 | 9.00 | 11.00 | 14.00 | 17.00 | 19.00 | 19.00 | 16.50 | |
| 43 | United States | 5.21 | 6.43 | 10.31 | 14.55 | 18.98 | 23.07 | 25.36 | 24.88 | 21.81 | |
| 44 | Vietnam | 22.25 | 22.75 | 24.75 | 27.75 | 28.75 | 29.50 | 29.00 | 28.75 | 28.25 | |

In [24]:

```
can = ts.loc['Australia'].reset_index()
can
```

Out[24]:

| | Date | Confirmed | Recovered | Deaths |
|-----|------------|-----------|-----------|--------|
| 0 | 2020-01-22 | 0 | 0.0 | 0 |
| 1 | 2020-01-23 | 0 | 0.0 | 0 |
| 2 | 2020-01-24 | 0 | 0.0 | 0 |
| 3 | 2020-01-25 | 0 | 0.0 | 0 |
| 4 | 2020-01-26 | 4 | 0.0 | 0 |
| ... | ... | ... | ... | ... |
| 811 | 2022-04-12 | 5207650 | 0.0 | 6648 |
| 812 | 2022-04-13 | 5262359 | 0.0 | 6693 |
| 813 | 2022-04-14 | 5308858 | 0.0 | 6727 |
| 814 | 2022-04-15 | 5345438 | 0.0 | 6755 |
| 815 | 2022-04-16 | 5384615 | 0.0 | 6779 |

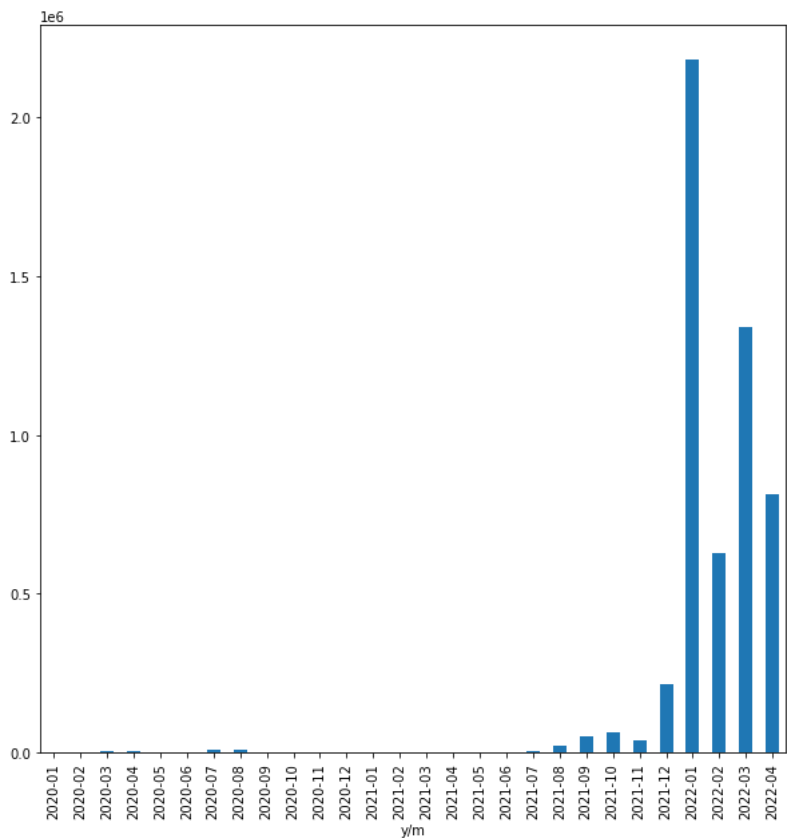
816 rows × 4 columns

In [25]:

```
can['y/m'] = can['Date'].dt.to_period('M')
z = pd.DataFrame(can.groupby('y/m')['Confirmed'].max())
a = z.diff().fillna(z.loc['2020-01'])
a = a.reset_index()
a = a.set_index('y/m')
a['Confirmed'].plot(kind = 'bar',figsize=(10, 10))
```

Out[25]:

<AxesSubplot: xlabel='y/m'>



In [26]:

```
country_avg[country_avg['Country/Region'] == 'Canada']
```

Out[26]:

| | Country/Region | month_1_avg | month_2_avg | month_3_avg | month_4_avg | month_5_avg | month_6_avg | month_7_avg | month_8_avg | month_9_avg | mon |
|---|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----|
| 6 | Canada | -6.14 | -4.86 | -0.57 | 5.71 | 11.29 | 16.36 | 18.93 | 18.57 | 15.14 | |

In [27]:

```
ts
```

Out[27]:

| | | Confirmed | Recovered | Deaths |
|----------------|------------|-----------|-----------|--------|
| Country/Region | Date | | | |
| Afghanistan | 2020-01-22 | 0 | 0.0 | 0 |
| | 2020-01-23 | 0 | 0.0 | 0 |
| | 2020-01-24 | 0 | 0.0 | 0 |
| | 2020-01-25 | 0 | 0.0 | 0 |
| | 2020-01-26 | 0 | 0.0 | 0 |
| ... | ... | ... | ... | ... |
| Zimbabwe | 2022-04-12 | 247094 | 0.0 | 5460 |
| | 2022-04-13 | 247160 | 0.0 | 5460 |
| | 2022-04-14 | 247208 | 0.0 | 5462 |
| | 2022-04-15 | 247237 | 0.0 | 5462 |
| | 2022-04-16 | 247237 | 0.0 | 5462 |

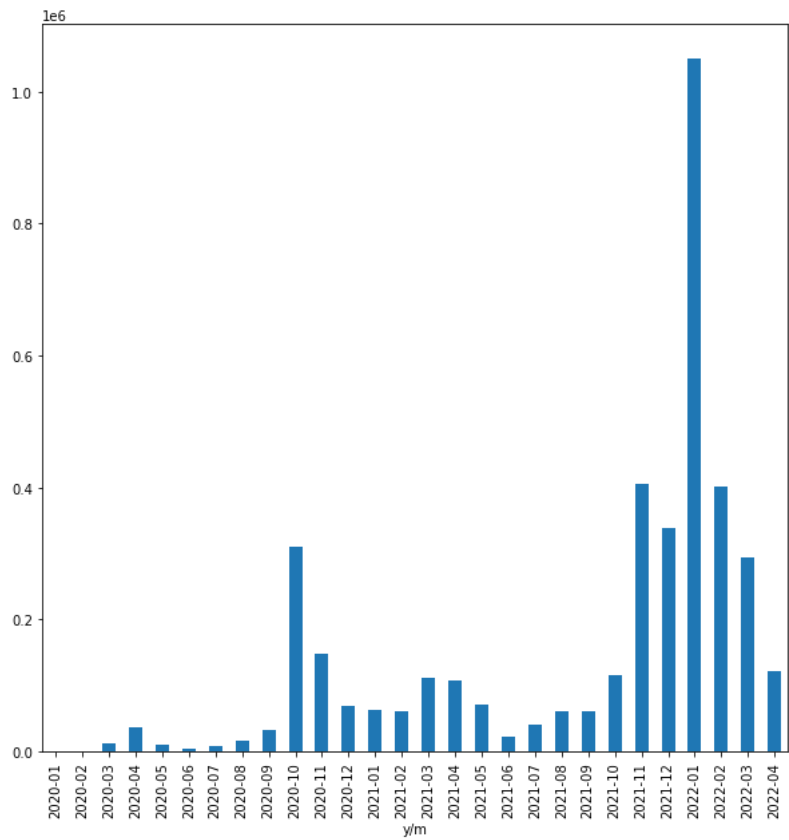
161568 rows × 3 columns

In [28]:

```
bel = 'Belgium'
belg = ts.loc[bel].reset_index()
belg['y/m'] = belg['Date'].dt.to_period('M')
z1 = pd.DataFrame(belg.groupby('y/m')['Confirmed'].max())
a1 = z1.diff().fillna(z1.loc['2020-01'])
a1['Confirmed'].plot(kind = 'bar',figsize=(10, 10))
```

Out[28]:

<AxesSubplot: xlabel='y/m'>



In [29]:

```
a1.reset_index()
```

Out[29]:

| | y/m | Confirmed |
|----|---------|-----------|
| 0 | 2020-01 | 0.0 |
| 1 | 2020-02 | 1.0 |
| 2 | 2020-03 | 12774.0 |
| 3 | 2020-04 | 35744.0 |
| 4 | 2020-05 | 9862.0 |
| 5 | 2020-06 | 3046.0 |
| 6 | 2020-07 | 7324.0 |
| 7 | 2020-08 | 16485.0 |
| 8 | 2020-09 | 33216.0 |
| 9 | 2020-10 | 310777.0 |
| 10 | 2020-11 | 148116.0 |
| 11 | 2020-12 | 69151.0 |
| 12 | 2021-01 | 63657.0 |
| 13 | 2021-02 | 61358.0 |
| 14 | 2021-03 | 110942.0 |
| 15 | 2021-04 | 107776.0 |
| 16 | 2021-05 | 71772.0 |
| 17 | 2021-06 | 23130.0 |
| 18 | 2021-07 | 39584.0 |
| 19 | 2021-08 | 59949.0 |
| 20 | 2021-09 | 60290.0 |
| 21 | 2021-10 | 115696.0 |
| 22 | 2021-11 | 405385.0 |
| 23 | 2021-12 | 339308.0 |
| 24 | 2022-01 | 1049896.0 |
| 25 | 2022-02 | 402075.0 |
| 26 | 2022-03 | 293734.0 |
| 27 | 2022-04 | 121915.0 |

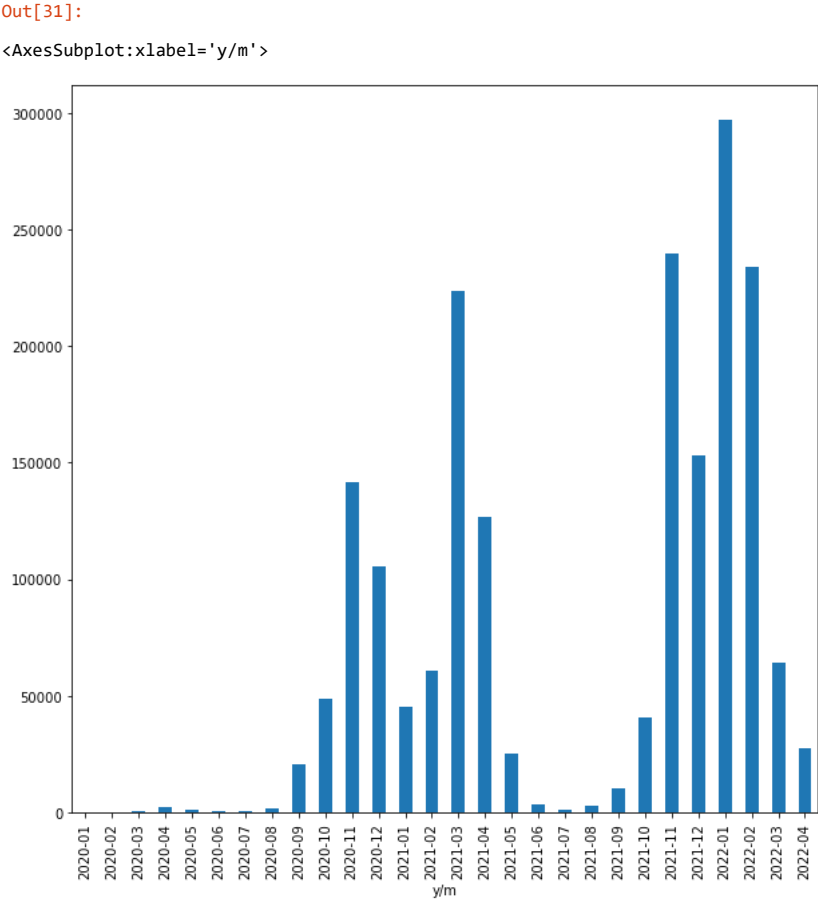
In [30]:

```
country_avg[country_avg['Country/Region'] == bel]
```

Out[30]:

| | Country/Region | month_1_avg | month_2_avg | month_3_avg | month_4_avg | month_5_avg | month_6_avg | month_7_avg | month_8_avg | month_9_avg | mon |
|---|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----|
| 3 | Belgium | 3.5 | 4.5 | 7.5 | 9.5 | 14.0 | 16.0 | 18.5 | 18.5 | 15.0 | |

```
In [31]:
h = 'Hungary'
hun = ts.loc[h].reset_index()
hun['y/m'] = hun['Date'].dt.to_period('M')
z2 = pd.DataFrame(hun.groupby('y/m')['Confirmed'].max())
a2 = z2.diff().fillna(z2.loc['2020-01'])
a2['Confirmed'].plot(kind = 'bar',figsize=(10, 10))
```



```
In [32]:
country_avg[country_avg['Country/Region'] == h]
```

Out[32]:

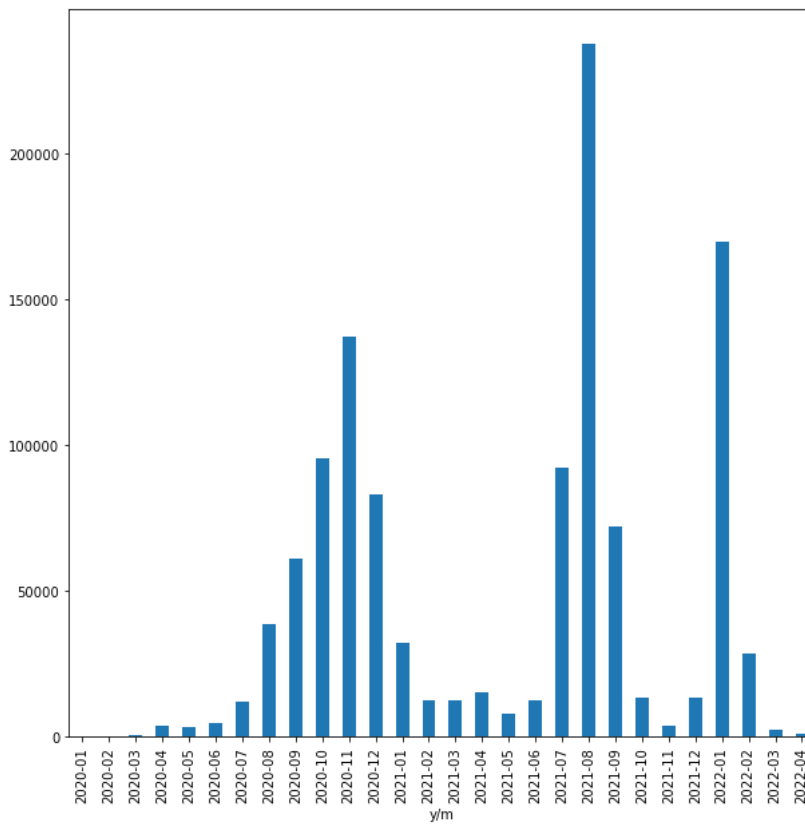
| | Country/Region | month_1_avg | month_2_avg | month_3_avg | month_4_avg | month_5_avg | month_6_avg | month_7_avg | month_8_avg | month_9_avg | mo |
|----|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----|
| 16 | Hungary | 0.0 | 2.0 | 6.5 | 12.0 | 17.0 | 20.5 | 22.0 | 22.0 | 17.0 | |

In [33]:

```
m = 'Morocco'
rocco = ts.loc[m].reset_index()
rocco['y/m'] = rocco['Date'].dt.to_period('M')
z3 = pd.DataFrame(rocco.groupby('y/m')['Confirmed'].max())
a3 = z3.diff().fillna(z3.loc['2020-01'])
a3['Confirmed'].plot(kind = 'bar',figsize=(10, 10))
```

Out[33]:

<AxesSubplot:xlabel='y/m'>



In [34]:

```
country_avg[country_avg['Country/Region'] == m]
```

Out[34]:

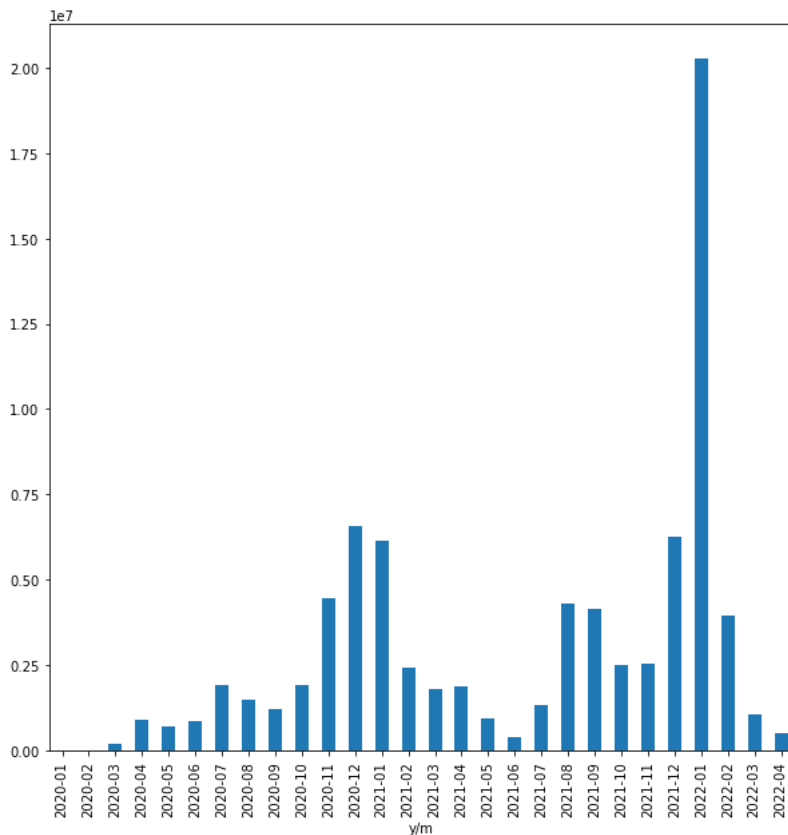
| | Country/Region | month_1_avg | month_2_avg | month_3_avg | month_4_avg | month_5_avg | month_6_avg | month_7_avg | month_8_avg | month_9_avg | mo |
|----|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----|
| 26 | Morocco | 15.0 | 16.5 | 18.5 | 19.0 | 20.0 | 22.5 | 24.0 | 24.0 | 23.0 | |

In [35]:

```
new = 'US'
land = ts.loc[new].reset_index()
land['y/m'] = land['Date'].dt.to_period('M')
z4 = pd.DataFrame(land.groupby('y/m')['Confirmed'].max())
a4 = z4.diff().fillna(z4.loc['2020-01'])
a4['Confirmed'].plot(kind = 'bar',figsize=(10, 10))
```

Out[35]:

<AxesSubplot:xlabel='y/m'>



Articulate three additional research questions related to COVID-19 data and try to provide answers to them using the given data set.

1. What are the countries having Death Percentage above 5%? Also, plot histogram

In [36]:

```
#calculating and sorting countries having Death % greater than 5
death_percentage = res[res.fatality_ratio > 5].sort_values('fatality_ratio', ascending = True)
#dropping unnecessary columns
death_percentage = death_percentage.drop(['Confirmed','Deaths'], axis =1)
death_percentage
```

Out[36]:

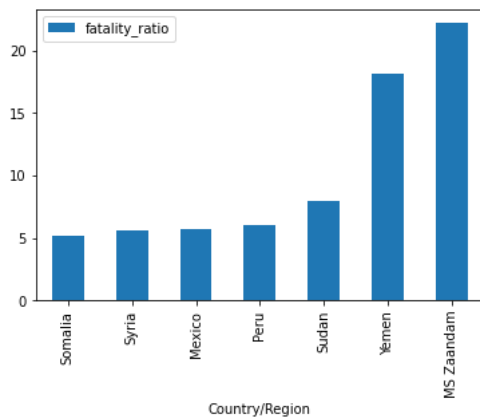
| fatality_ratio | |
|----------------|-------|
| Country/Region | |
| Somalia | 5.14 |
| Syria | 5.65 |
| Mexico | 5.66 |
| Peru | 5.98 |
| Sudan | 7.94 |
| Yemen | 18.18 |
| MS Zaandam | 22.22 |

In [37]:

```
#plotting bar graph of the data sorted above
death_percentage.plot.bar()
```

Out[37]:

<AxesSubplot:xlabel='Country/Region'>



2. Plot bar graph of the countries having lowest deaths

In [38]:

```
#Calculating countries having lowest deaths
```

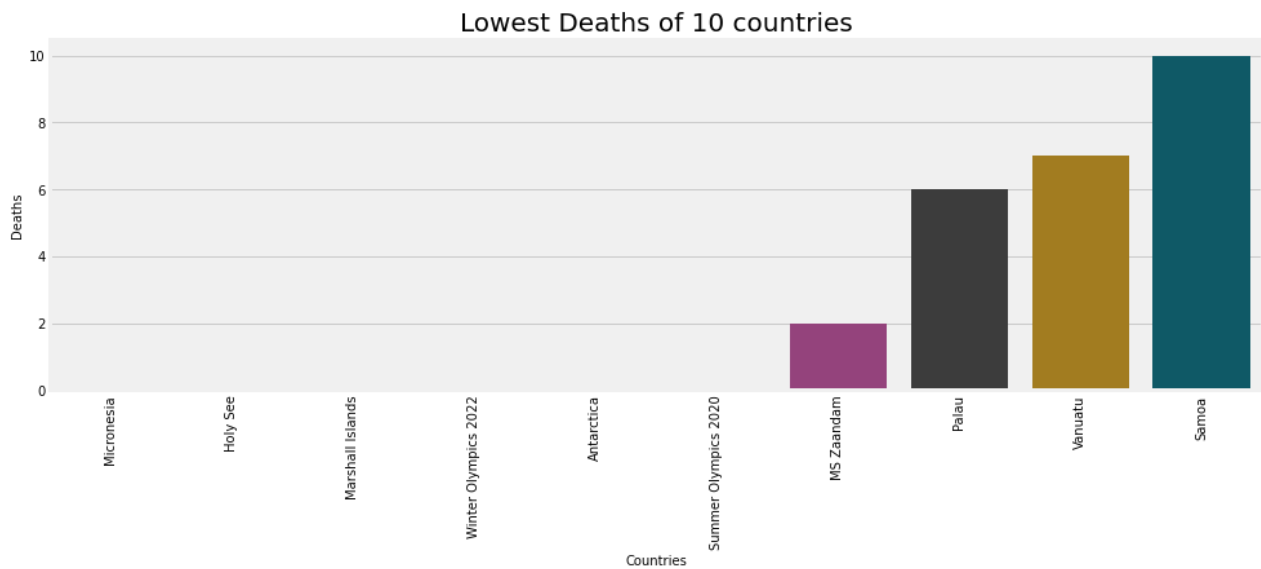
```
Top_10_countries_lowest_deaths = data.groupby('Country/Region', as_index=False)['Deaths'].max('Date').sort_values('Deaths', ascending = True)
Top_10_countries_lowest_deaths
```

Out[38]:

| | Country/Region | Deaths |
|-----|----------------------|--------|
| 117 | Micronesia | 0 |
| 76 | Holy See | 0 |
| 113 | Marshall Islands | 0 |
| 194 | Winter Olympics 2022 | 0 |
| 5 | Antarctica | 0 |
| 168 | Summer Olympics 2020 | 0 |
| 106 | MS Zaandam | 2 |
| 135 | Palau | 6 |
| 190 | Vanuatu | 7 |
| 150 | Samoa | 10 |

In [39]:

```
#plotting the dataframe concluded above
import seaborn as sns
import plotly.express as px
plt.style.use('fivethirtyeight')
%matplotlib inline
plt.rcParams['figure.figsize'] = (15, 5)
ax = sns.barplot(x = Top_10_countries_lowest_deaths['Country/Region'], y = Top_10_countries_lowest_deaths['Deaths'], palette = 'dark')
ax.set_xlabel(xlabel = 'Countries', fontsize = 10)
ax.set_ylabel(ylabel = 'Deaths', fontsize = 10)
ax.set_title(label = 'Lowest Deaths of 10 countries', fontsize = 20)
plt.xticks(rotation = 90)
plt.show()
```



3. Calculate the total number of confirmed cases of Australia till 12th August 2021.

In [40]:

```
#Locating data for that specific country and date
data.loc[(data['Country/Region']=='Australia') & (data['Date']=='2021-08-12')]
```

Out[40]:

| | Date | Country/Region | Province/State | Confirmed | Recovered | Deaths |
|-------|------------|----------------|------------------------------|-----------|-----------|--------|
| 7912 | 2021-08-12 | Australia | Australian Capital Territory | 129 | 0.0 | 3 |
| 8728 | 2021-08-12 | Australia | New South Wales | 12629 | 0.0 | 93 |
| 9544 | 2021-08-12 | Australia | Northern Territory | 199 | 0.0 | 0 |
| 10360 | 2021-08-12 | Australia | Queensland | 1948 | 0.0 | 7 |
| 11176 | 2021-08-12 | Australia | South Australia | 868 | 0.0 | 4 |
| 11992 | 2021-08-12 | Australia | Tasmania | 235 | 0.0 | 13 |
| 12808 | 2021-08-12 | Australia | Victoria | 21098 | 0.0 | 820 |
| 13624 | 2021-08-12 | Australia | Western Australia | 1059 | 0.0 | 9 |

In [41]:

```
#adding all the confirmed cases of the states present in that country
data.loc[(data['Country/Region']=='Australia') & (data['Date']=='2021-08-12')]['Confirmed'].sum()
```

Out[41]:

38165