SRISHTI PANDEY - COVID 19 DATA ANALYSIS PROJECT

In [33]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [34]:
```
corona_dataset_csv = pd.read_csv("covid19_Confirmed_dataset.csv")
corona_dataset_csv.head()
```

Out[34]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1
0	NaN	Afghanistan	33.0000	65.0000	0	0	0	0	0	
1	NaN	Albania	41.1533	20.1683	0	0	0	0	0	
2	NaN	Algeria	28.0339	1.6596	0	0	0	0	0	
3	NaN	Andorra	42.5063	1.5218	0	0	0	0	0	
4	NaN	Angola	-11.2027	17.8739	0	0	0	0	0	

5 rows × 104 columns

In [35]:
```
corona_dataset_csv.shape
```

Out[35]:
```
(266, 104)
```

In [36]:
```
corona_dataset_csv.drop(["Lat" ,"Long"], axis = 1,inplace = True )
```

In [37]:
```
corona_dataset_csv.head()
```

Out[37]:

	Province/State	Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1
0	NaN	Afghanistan	0	0	0	0	0	0	0	
1	NaN	Albania	0	0	0	0	0	0	0	
2	NaN	Algeria	0	0	0	0	0	0	0	
3	NaN	Andorra	0	0	0	0	0	0	0	
4	NaN	Angola	0	0	0	0	0	0	0	

5 rows × 102 columns

Aggregating State Col with Country Col.

In [38]:
```
corona_dataset_aggregated = corona_dataset_csv.groupby("Country/Region").sum()
```

In [39]:
```
corona_dataset_aggregated.head(10)
```

Out[39]:

Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20
Afghanistan	0	0	0	0	0	0	0	0	0
Albania	0	0	0	0	0	0	0	0	0
Algeria	0	0	0	0	0	0	0	0	0
Andorra	0	0	0	0	0	0	0	0	0
Angola	0	0	0	0	0	0	0	0	0
Antigua and Barbuda	0	0	0	0	0	0	0	0	0
Argentina	0	0	0	0	0	0	0	0	0
Armenia	0	0	0	0	0	0	0	0	0
Australia	0	0	0	0	4	5	5	6	9
Austria	0	0	0	0	0	0	0	0	0

10 rows × 100 columns

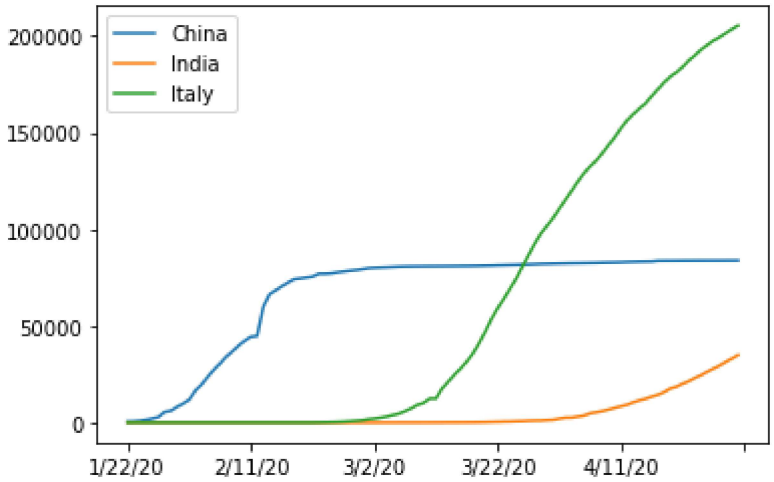
In [40]:
```
corona_dataset_aggregated.shape
```

Out[40]:
```
(187, 100)
```

Visualizing data with country name

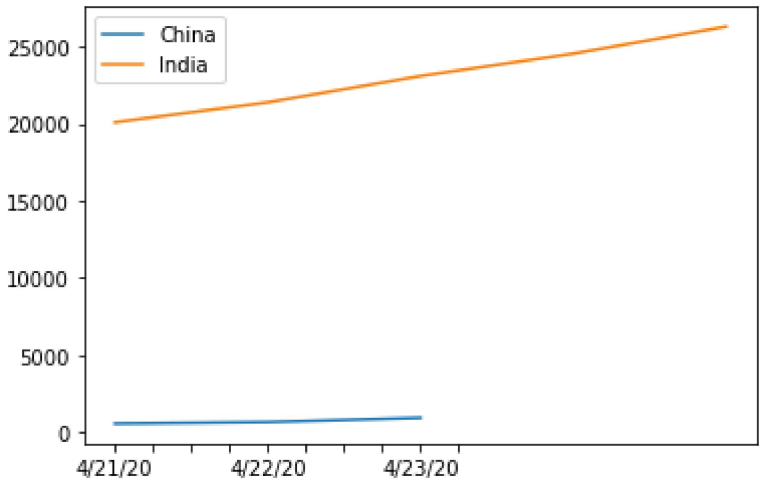
In [41]:
```
corona_dataset_aggregated.loc["China"].plot()
corona_dataset_aggregated.loc["India"].plot()
corona_dataset_aggregated.loc["Italy"].plot()
plt.legend()
```

Out[41]:
```
<matplotlib.legend.Legend at 0x2b2f7897ac0>
```



In [75]:
```
corona_dataset_aggregated.loc["China"][:3].plot()
corona_dataset_aggregated.loc["India"]["4/21/20":"4/25/20"].plot()
plt.legend()
```

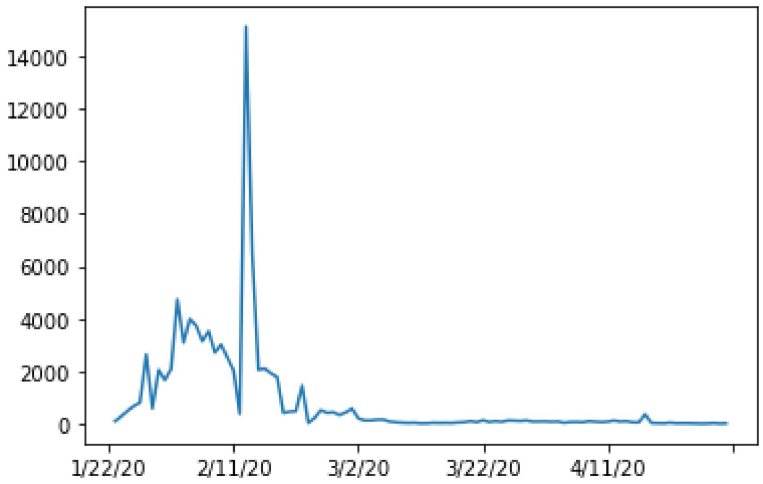
Out[75]: <matplotlib.legend.Legend at 0x2b2fa8b99a0>



plotting maximum hike in cases

In [43]:
```
corona_dataset_aggregated.loc["China"].diff().plot()
```

Out[43]: <AxesSubplot:>



Calculating maximum cases discovered in single day

In [31]:
```
corona_dataset_aggregated.loc["China"].diff().max()
```

Out[31]: 15136.0

In [32]:
```
corona_dataset_aggregated.loc["Italy"].diff().max()
```

Out[32]: 6557.0

Calculating maximum Covid cases for all countries

In [44]:
```
countries = list(corona_dataset_aggregated.index)
max_infection_rates = []
for c in countries:
    max_infection_rates.append(corona_dataset_aggregated.loc[c].diff().max())
corona_dataset_aggregated["Total COVID 19 Cases"] = max_infection_rates
corona_dataset_aggregated.head()
```

Out[44]:

Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20
Afghanistan	0	0	0	0	0	0	0	0	0
Albania	0	0	0	0	0	0	0	0	0
Algeria	0	0	0	0	0	0	0	0	0
Andorra	0	0	0	0	0	0	0	0	0
Angola	0	0	0	0	0	0	0	0	0

5 rows × 101 columns

Copying final data to new dataframe

In [76]:
```
corona_data = pd.DataFrame(corona_dataset_aggregated["Total COVID 19 Cases"])
corona_data.head(15)
```

Out[76]:

Country/Region	Total COVID 19 Cases
Afghanistan	232.0
Albania	34.0
Algeria	199.0
Andorra	43.0
Angola	5.0
Antigua and Barbuda	6.0
Argentina	291.0
Armenia	134.0
Australia	497.0
Austria	1321.0

Country/Region	Total COVID 19 Cases
Azerbaijan	105.0
Bahamas	7.0
Bahrain	301.0
Bangladesh	641.0
Barbados	12.0

Analysing data after COVID 19

In [80]:
```
happiness_report_csv = pd.read_csv("worldwide_happiness_report.csv")
happiness_report_csv.head()
```

Out[80]:

	Overall rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
0	1	Finland	7.769	1.340	1.587	0.986	0.596	0.153	0.393
1	2	Denmark	7.600	1.383	1.573	0.996	0.592	0.252	0.410
2	3	Norway	7.554	1.488	1.582	1.028	0.603	0.271	0.341
3	4	Iceland	7.494	1.380	1.624	1.026	0.591	0.354	0.118
4	5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.322	0.298

In [81]:
```
Useless_cols = ["Social support", "Generosity"]
happiness_report_csv.drop(Useless_cols,axis = 1, inplace = True)
```

In [51]:
```
happiness_report_csv.head()
```

Out[51]:

	Overall rank	Country or region	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
0	1	Finland	7.769	1.340	0.986	0.596	0.393
1	2	Denmark	7.600	1.383	0.996	0.592	0.410
2	3	Norway	7.554	1.488	1.028	0.603	0.341
3	4	Iceland	7.494	1.380	1.026	0.591	0.118
4	5	Netherlands	7.488	1.396	0.999	0.557	0.298

In [82]:
```
happiness_report_csv.drop("Overall rank", axis = 1, inplace = True)
```

In [54]:
```
happiness_report_csv.head()
```

Out[54]:

	Country or region	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption

	Country or region	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
0	Finland	7.769	1.340	0.986	0.596	0.393
1	Denmark	7.600	1.383	0.996	0.592	0.410
2	Norway	7.554	1.488	1.028	0.603	0.341
3	Iceland	7.494	1.380	1.026	0.591	0.118
4	Netherlands	7.488	1.396	0.999	0.557	0.298

In [55]:
```
happiness_report_csv.set_index("Country or region", inplace = True)
happiness_report_csv.head()
```

Out[55]:

Country or region	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Finland	7.769	1.340	0.986	0.596	0.393
Denmark	7.600	1.383	0.996	0.592	0.410
Norway	7.554	1.488	1.028	0.603	0.341
Iceland	7.494	1.380	1.026	0.591	0.118
Netherlands	7.488	1.396	0.999	0.557	0.298

Merging Corona Data and Happiness Report

In [56]:
```
corona_data.shape
```

Out[56]:
```
(187, 1)
```

In [58]:
```
happiness_report_csv.shape
```

Out[58]:
```
(156, 5)
```

In [62]:
```
data= corona_data.join(happiness_report_csv, how="inner")
data.head()
```

Out[62]:

	Total COVID 19 Cases	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Afghanistan	232.0	3.203	0.350	0.361	0.000	0.025
Albania	34.0	4.719	0.947	0.874	0.383	0.027
Algeria	199.0	5.211	1.002	0.785	0.086	0.114
Argentina	291.0	6.086	1.092	0.881	0.471	0.050
Armenia	134.0	4.559	0.850	0.815	0.283	0.064

In [85]:
```
dataLeft= corona_data.join(happiness_report_csv, how="left")
dataLeft.head()
```

Out[85]:

Country/Region	Total COVID 19 Cases	Country or region	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Afghanistan	232.0	NaN	NaN	NaN	NaN	NaN	NaN
Albania	34.0	NaN	NaN	NaN	NaN	NaN	NaN
Algeria	199.0	NaN	NaN	NaN	NaN	NaN	NaN
Andorra	43.0	NaN	NaN	NaN	NaN	NaN	NaN
Angola	5.0	NaN	NaN	NaN	NaN	NaN	NaN

In [63]:
```
data.shape
```

Out[63]: (143, 6)

In [64]:
```
#Checking correlation between all variables
```

In [65]:
```
data.corr()
```

Out[65]:

	Total COVID 19 Cases	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Total COVID 19 Cases	1.000000	0.228783	0.250118	0.289263	0.078196	0.097703
Score	0.228783	1.000000	0.793847	0.799893	0.587007	0.420437
GDP per capita	0.250118	0.793847	1.000000	0.863062	0.394603	0.311577
Healthy life expectancy	0.289263	0.799893	0.863062	1.000000	0.427892	0.314811
Freedom to make life choices	0.078196	0.587007	0.394603	0.427892	1.000000	0.446677
Perceptions of corruption	0.097703	0.420437	0.311577	0.314811	0.446677	1.000000

In [66]:
```
#Visualizating our results
```

In [67]:
```
data.head()
```

Out[67]:

	Total COVID 19 Cases	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Afghanistan	232.0	3.203	0.350	0.361	0.000	0.025
Albania	34.0	4.719	0.947	0.874	0.383	0.027
Algeria	199.0	5.211	1.002	0.785	0.086	0.114
Argentina	291.0	6.086	1.092	0.881	0.471	0.050

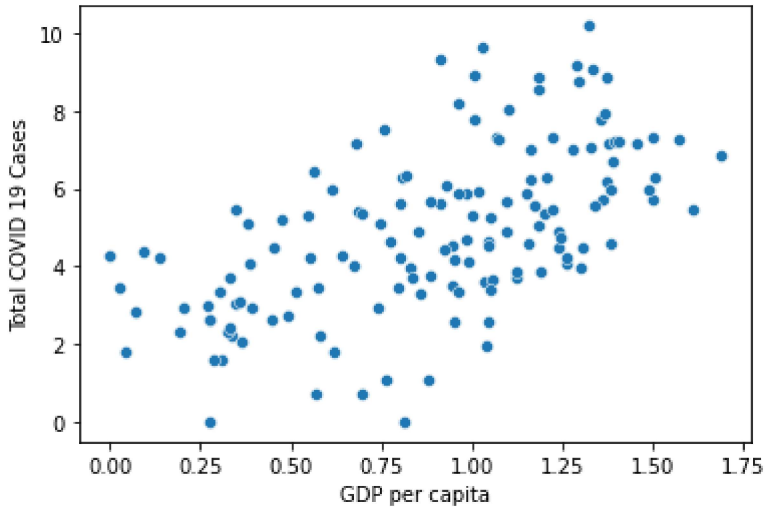
	Total COVID 19 Cases	Score	GDP per capita	Healthy life expectancy	Freedom to make life choices	Perceptions of corruption
Armenia	134.0	4.559	0.850	0.815	0.283	0.064

In [68]:
```
#Plotting GDP Vs Total Covid 19 cases
```

In [71]:
```
x = data["GDP per capita"]
y = data["Total COVID 19 Cases"]
sns.scatterplot(x,np.log(y))
```

```
C:\Users\shiva\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pas
s the following variables as keyword args: x, y. From version 0.12, the only valid posit
ional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```

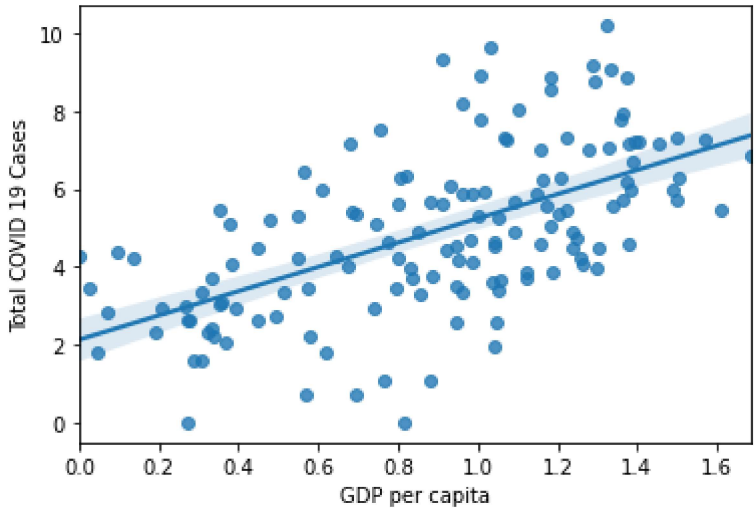
Out[71]: <AxesSubplot:xlabel='GDP per capita', ylabel='Total COVID 19 Cases'>



In [72]:
```
sns.regplot(x,np.log(y))
```

```
C:\Users\shiva\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pas
s the following variables as keyword args: x, y. From version 0.12, the only valid posit
ional argument will be `data`, and passing other arguments without an explicit keyword w
ill result in an error or misinterpretation.
  warnings.warn(
```

Out[72]: <AxesSubplot:xlabel='GDP per capita', ylabel='Total COVID 19 Cases'>


Total COVID 19 Cases
GDP per capita
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6

In [ ]: