PHASE4 Submission document

Project title:COVID 19 cases analysis

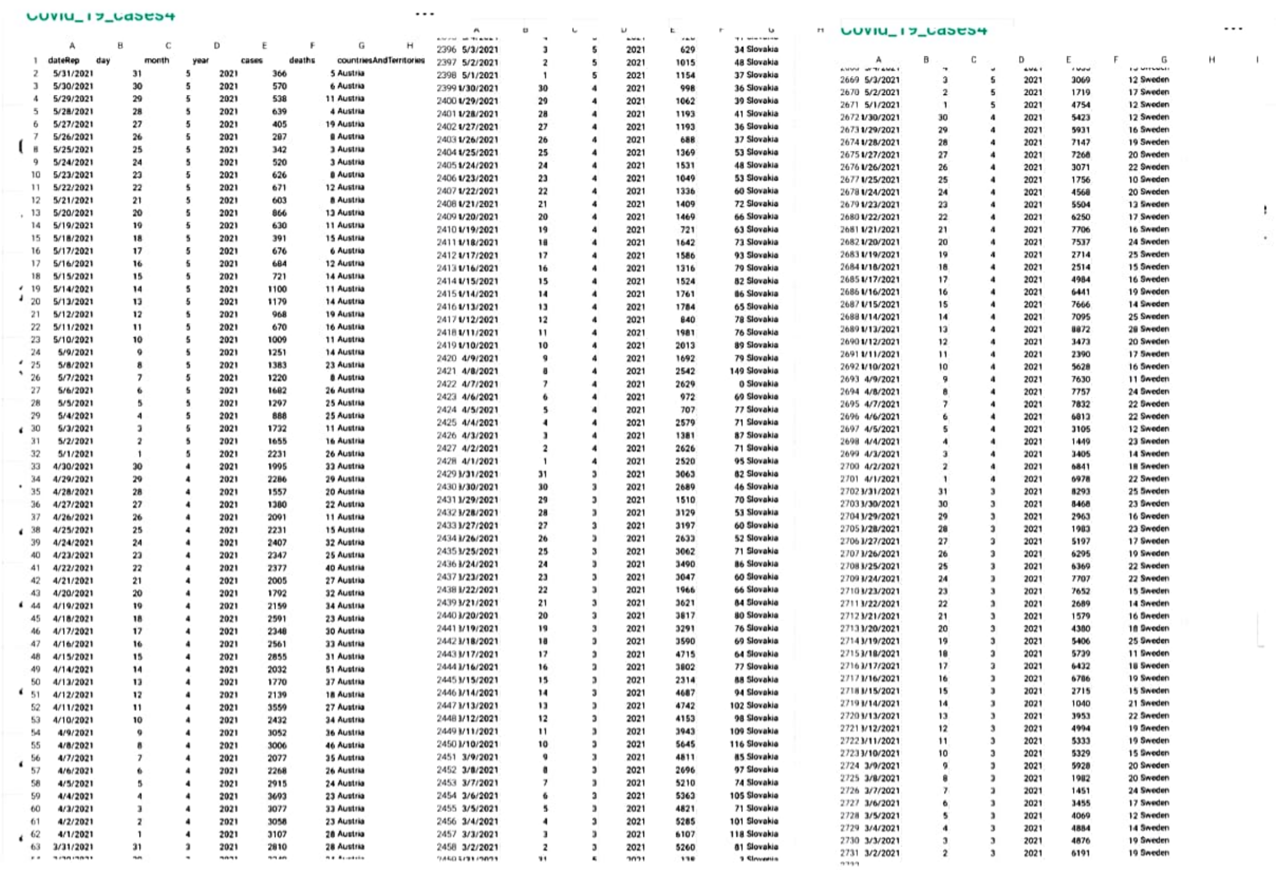
Phase 4: Development part2

COVID 19 cases analysis

Introduction:

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has left an indelible mark on the world. Since its emergence in late 2019, this global health crisis has reshaped societies, strained healthcare systems, and sparked unprecedented public health responses. Analyzing COVID-19 cases is of paramount importance in comprehending the pandemic's multifaceted impact.In just a span of a few years, COVID-19 has swept across continents, infecting millions and impacting every facet of human life. As we grapple with the evolving nature of this virus, data-driven insights have become our guiding light. This one-page submission seeks to provide a glimpse into our findings, shedding light on the virus's spread, its toll on healthcare systems, and the effectiveness of public health measures. In a world grappling with an ongoing crisis, data-driven insights are not just invaluable; they are imperative for informed decision-making and a collective path forward.Our analysis delves into the numbers and trends that underpin this pandemic, offering a snapshot of its complex and ever-changing landscape. As we navigate these uncharted waters, the story of COVID-19 cases told through data is more than an academic exercise; it's a vital chapter in our ongoing battle to safeguard public health and well-being.

Given data set:



5000 Rows x 7 Columns

Overview of the process:

1.Data-Driven Insight:

The analysis is rooted in data collected from various sources, including government reports, healthcare databases, and scientific research. It provides a factual and evidence-based understanding of the COVID-19 pandemic.

2.Pandemic Impact Assessment:

This analysis focuses on assessing the impact of the pandemic, with an emphasis on how the virus has spread across regions, influenced healthcare systems, and affected populations.

3.Healthcare System Strain:

The analysis highlights the strain on healthcare systems due to the surge in COVID-19 cases. It delves into aspects such as hospitalization rates, ICU bed occupancy, and the use of critical resources like ventilators.

4.Effectiveness of public health Measures:

It evaluates the effectiveness of various public health measures implemented to control the pandemic. These measures include social distancing, mask mandates, lockdowns, and vaccination campaigns.

5.Decision-Making and recommendations:

The analysis concludes with data-driven recommendations for future actions. It emphasizes the importance of vaccination, preparedness for future outbreaks, and informed decision-making in addressing the ongoing and potential health crises.

PROCEDURE:

Feature selection:

1.Gather a comprehensive dataset that includes various potential features related to COVID-19 cases. This may include demographic data, testing and diagnostic information, geographical variables, healthcare resources, and public health measures.

2.Clean the data by handling missing values, outliers, and inconsistencies. Ensure the dataset is in a usable format for analysis.

3.Conduct EDA to gain insights into the data. Visualizations and statistical techniques can help identify initial patterns and relationships among features.

4.Utilize statistical techniques to evaluate the importance of each feature. Common methods include:Correlation analysis to identify relationships with the target variable (e.g., case counts or mortality rates).Feature ranking through techniques like mutual information, ANOVA, or feature importance from machine learning models.

5.Employ various feature selection methods to choose the most relevant features:Filter Methods: These methods use statistical metrics to score features and select the top-ranked ones.Wrapper Methods: They involve training a machine learning model and selecting features based on their impact on model performance (e.g., recursive feature elimination).Embedded Methods: Some machine learning algorithms have built-in feature selection techniques (e.g., Lasso regression).

Feature Selection:

Open a new Syntax Editor session in SPSS Statistics by selecting File > New > Syntax.

Copy the following syntax into the Syntax Editor dialog box.

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

print('Modules are imported.')
```

Modules are imported.

importing "Covid19_Confirmed_dataset.csv" from "./Dataset" folder.

In[2]

```
df=pd.read_csv("../input/covid19/covid19_Confirmed_datas

et.csv")

df.head()
```

Out[2]:

Province/State□Country/Region□Lat□Long□1/22/20□1/23/20□1/24/20□1/25/20

1/26/20□1/27/20□...□4/21/20□4/22/20□4/23/20□4/24/20□4/25/20□4/26/20

4/27/20□4/28/20□4/29/20□4/30/20

0□NaN□Afghanistan□33.0000□65.0000□0□0□0□0□0□0□...□1092□1176□1279□1351

1463□1531□1703□1828□1939□2171

1□NaN□Albania□41.1533□20.1683□0□0□0□0□0□0□...□609□634□663□678□712□726

736□750□766□773

2□NaN□Algeria□28.0339□1.6596□0□0□0□0□0□0□...□2811□2910□3007□3127□3256

3382□3517□3649□3848□4006

3□NaN□Andorra□42.5063□1.5218□0□0□0□0□0□0□...□717□723□723□731□738□738

743□743□743□745

4□NaN□Angola□-11.2027□17.8739□0□0□0□0□0□0□...□24□25□25□25□25□26□27□27

27□27

Let's check the shape of the dataframe

In[3]:

df.shape

Out[3]:

(266, 104)

Model training:

Delete the useless columns
In[4]:
df.drop(["Lat","Long"],axis=1,inplace=True)
In[5]:
df.head()

Out[5]:

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 1/29/20 ... 4/21/20 4/22/20 4/23/20 4/24/20 4/25/20 4/26/20 4/27/20

4/28/20 4/29/20 4/30/20

0 NaN Afghanistan 0 0 0 0 0 0 0 0 ... 1092 1176 1279 1351 1463 1531 1703

1828 1939 2171

1 NaN Albania 0 0 0 0 0 0 0 0 ... 609 634 663 678 712 726 736 750 766 773

2 NaN Algeria 0 0 0 0 0 0 0 0 ... 2811 2910 3007 3127 3256 3382 3517 3649

3848 4006

3 NaN Andorra 0 0 0 0 0 0 0 0 ... 717 723 723 731 738 738 743 743 743 745

4 NaN Angola 0 0 0 0 0 0 0 0 ... 24 25 25 25 25 26

Aggregating the rows by the country

In[6]:

aggregating=df.groupby("Country/Region").sum()

In[7]:

aggregating.head()

Out[7]:

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□1/31/20□...□4/21/20□4/22/20□4/23/20□4/24/20□4/25/20□4/26/20

4/27/20□4/28/20□4/29/20□4/30/20

Country/Region

Afghanistan□0□0□0□0□0□0□0□0□0□0□...□1092□1176□1279□1351□1463□1531

1703□1828□1939□2171

Albania□0□0□0□0□0□0□0□0□0□0□...□609□634□663□678□712□726□736□750

766□773

Algeria□0□0□0□0□0□0□0□0□0□0□...□2811□2910□3007□3127□3256□3382

3517□3649□3848□4006

Andorra□0□0□0□0□0□0□0□0□0□0□...□717□723□723□731□738□738□743□743

743□745

Angola□0□0□0□0□0□0□0□0□0□0□...□24□25□25□25□25□26□27□27□27□27

In[8]:

```
aggregating.shape
```

Out[8]:

(187, 100)

Visualizing data related to a country for example China

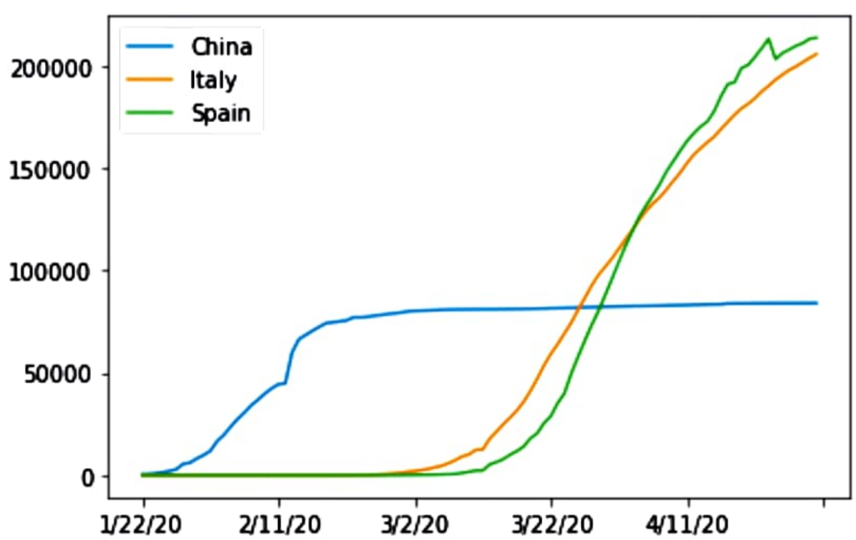
visualization always helps for better understanding of our data.

In[9]:

```
aggregating.loc["China"].plot()
aggregating.loc["Italy"].plot()
aggregating.loc["Spain"].plot()
plt.legend()
```

Out[9]:

<matplotlib.legend.Legend at 0x7f482e1e3990>



Calculating a good measure

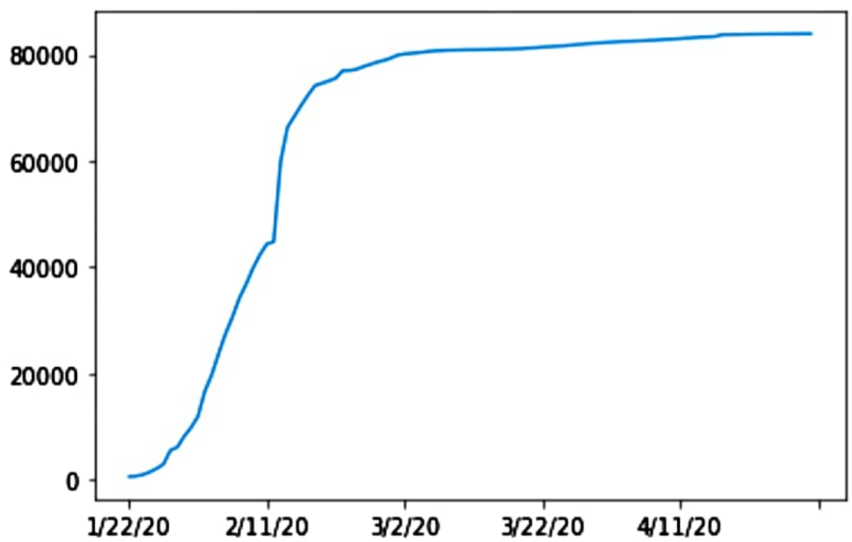
we need to find a good measure reperestend as a number, describing the spread of the virus in a country.

In[10]:

```
aggregating.loc['China'].plot()
```

Out[10]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482df94d90>

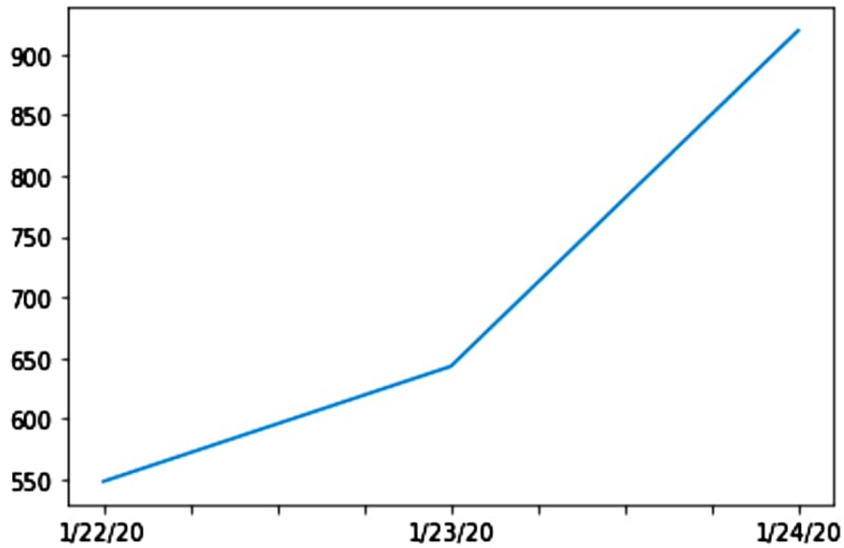


In[11]:

```
aggregating.loc['China'][:3].plot()
```

Out[11]:

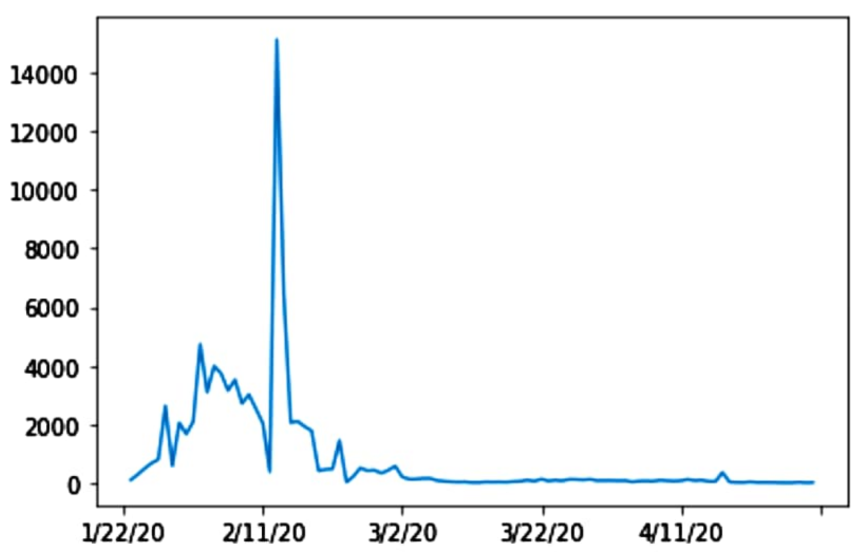
<matplotlib.axes._subplots.AxesSubplot at 0x7f482df83990>



caculating the first derivative of the curve

In[12]:
aggregating.loc['China'].diff().plot()

Out[12]:
<matplotlib.axes._subplots.AxesSubplot at 0x7f482df09290>



find maxmimum infection rate for China

In[13]:

```
aggregating.loc['China'].diff().max()
```

Out[13]:

15136.0

In[14]:

```
aggregating.loc['Italy'].diff().max()
```

Out[14]:

6557.0

In[15]:

```
aggregating.loc['Spain'].diff().max()
```

Out[15]:

9630.0

find maximum infection rate for all of the countries.

In[16]:

```
countries=list(aggregating.index)
max_infection_rates=[]
for c in countries:
   max_infection_rates.append(aggregating.loc[c].diff().max())
aggregating["max_infection_rates"]=max_infection_rates
```

In[17]:

```
aggregating.head()
```

Out[17]:

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

1/31/20 ... 4/22/20 4/23/20 4/24/20 4/25/20 4/26/20 4/27/20 4/28/20 4/29/20

4/30/20 max_infection_rates

Country/Region

Afghanistan 0 0 0 0 0 0 0 0 0 0 ... 1176 1279 1351 1463 1531 1703 1828 1939 2171

232.

Albania 0 0 0 0 0 0 0 0 0 0 ... 634 663 678 712 726 736 750 766 773 34.

Algeria 0 0 0 0 0 0 0 0 0 0 ... 2910 3007 3127 3256 3382 3517 3649 3848 4006 199.

Andorra 0 0 0 0 0 0 0 0 0 0 ... 723 723 731 738 738 743 743 743 745 43.

Angola 0 0 0 0 0 0 0 0 0 0 ... 25 25 25 25 26 27 27 27 27 5.0

5 rows × 101 columns

create a new dataframe with only needed column

In[18]:

```
data=pd.DataFrame(aggregating["max_infection_rates"])
```

In[19]:

```
data.head()
```

Out[19]:

	max_infection_rates
Country/Region	
Afghanistan	232.0
Albania	34.0
Algeria	199.0
Andorra	43.0
Angola	5.0

Importing the WorldHappinessReport.csv dataset

selecting needed columns for our analysis

join the datasets

calculate the correlations as the result of our analysis importing the dataset

In[20]:

```
happiness=pd.read_csv("../input/covid19/worldwide_happiness_report.csv")
```

In[21]:

```
happiness.head()
```

Out[21]:

	Overall rank	Country or region	Score	GDP per capita	Social suppor	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

let's drop the useless columns

In[22]:

```
cols=["Overall rank","Score","Generosity","Perceptions of corruption"]
```

In[23]:

```
happiness.drop(cols,axis=1,inplace=True)
```

```
happiness.head()
```

Out[23]:

	Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
0	Finland	1.340	1.587	0.986	0.596
1	Denmark	1.383	1.573	0.996	0.592
2	Norway	1.488	1.582	1.028	0.603
3	Iceland	1.380	1.624	1.026	0.591
4	Netherlands	1.396	1.522	0.999	0.557

changing the indices of the dataframe
In[24]:
```
happiness.set_index("Country or region",inplace=True)
happiness.head()
```
Out[24]:

	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Country or region				
Finland	1.340	1.587	0.986	0.596
Denmark	1.383	1.573	0.996	0.592
Norway	1.488	1.582	1.028	0.603
Iceland	1.380	1.624	1.026	0.591
Netherlands	1.396	1.522	0.999	0.557

now let's join two dataset we have prepared
Corona Dataset :
In[25]:
```
    data.head()
```
Out[25]:

	max_infection_rates
Country/Region	
Afghanistan	232.0
Albania	34.0
Algeria	199.0
Andorra	43.0
Angola	5.0

wolrd happiness report Dataset :

In[26]:

happiness.head()

Out[26]:

	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
2Country or region				
Finland	1.340	1.587	0.986	0.596
Denmark	1.383	1.573	0.996	0.592
Norway	1.488	1.582	1.028	0.603
Iceland	1.380	1.624	1.026	0.591
Netherlands	1.396	1.522	0.999	0.557

In[27]:

final=data.join(happiness,how="inner")

final.head()

Out[27]:

	max_infection_rates	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Afghanistan	232.0	0.350	0.517	0.361	0.000
Albania	34.0	0.947	0.848	0.874	0.383
Algeria	199.0	1.002	1.160	0.785	0.086
Argentina	291.0	1.092	1.432	0.881	0.471
Armenia	134.0	0.850	1.055	0.815	0.283

correlation matrix

In[28]:

final.corr()

Out[28]:

	max_infection_rates	GDP per capita	Healthy life expectancy	Freedom to make life choices
max_infection_rates	1.000000	0.250118	0.289263	0.078196
GDP per capita	0.250118	1.000000	0.863062	0.394603
Social support	0.191958	0.759468	0.765286	0.456246
Healthy life expectancy	0.289263	0.863062	1.000000	0.427892
Freedom to make life choices	0.078196	0.394603	0.427892	1.000000

Visualization of the results

our Analysis is not finished unless we visualize the results in terms figures and graphs so that everyone can understand what you get out of our analysis

In[29]:

```
final.head()
```

Out[29]:

	max_infection_rates	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Afghanistan	232.0	0.350	0.517	0.361	0.000
Albania	34.0	0.947	0.848	0.874	0.383
Algeria	199.0	1.002	1.160	0.785	0.086
Argentina	291.0	1.092	1.432	0.881	0.471
Armenia	134.0	0.850	1.055	0.815	0.283

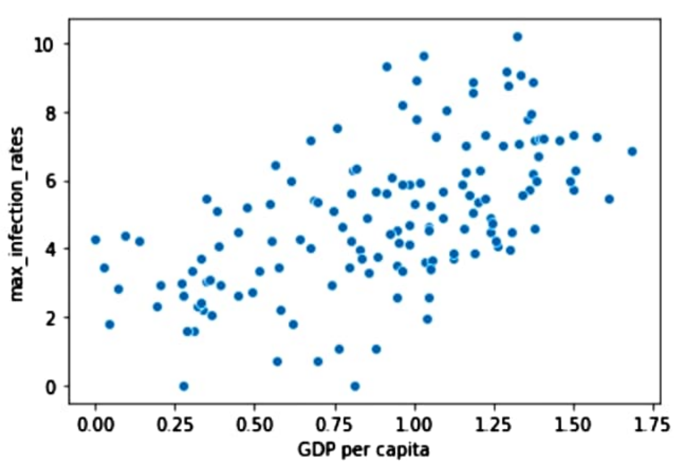
Plotting GDP vs maximum Infection rate

1. In[30]:

```
x=final["GDP per capita"]
y=final["max_infection_rates"]
sns.scatterplot(x,np.log(y))
```

Out [30]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482de36590>

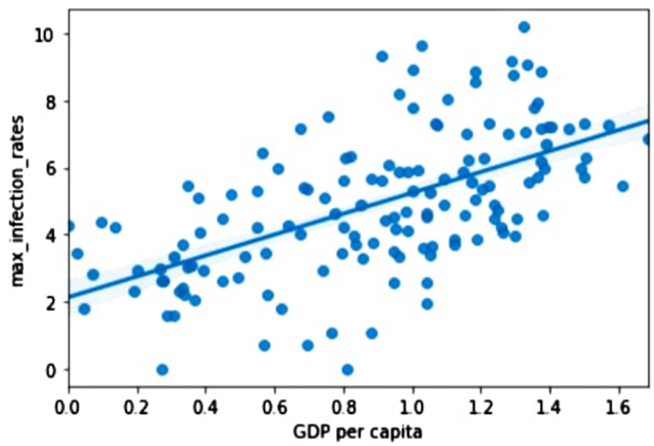


In[31]:

```
sns.regplot(x,np.log(y))
```

Out [31]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482dd8b3d0>



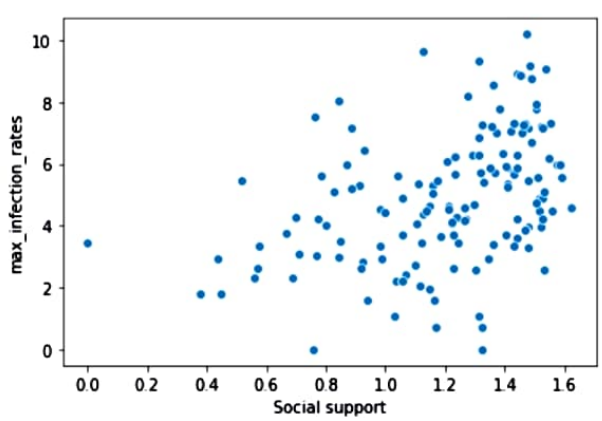
Plotting Social support vs maximum Infection rate

In[32]:

```
x=final["Social support"]
y=final["max_infection_rates"]
sns.scatterplot(x,np.log(y))
```

Out [32]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482de1b210>

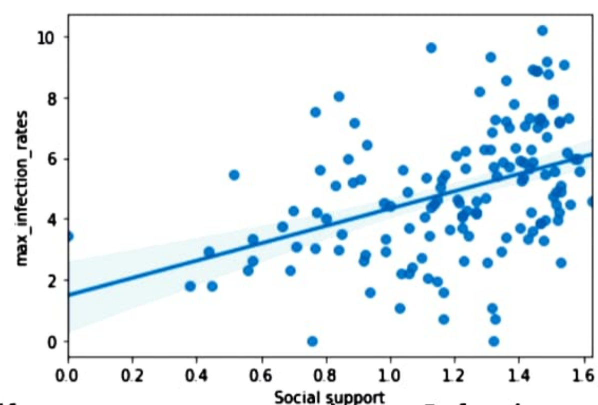


In[33]:

```
sns.regplot(x,np.log(y))
```

Out [33]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482b49a610>



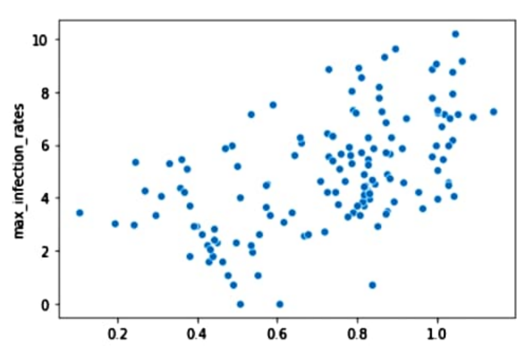
Plotting Healthy life expectancy vs maximum Infection rate

In[34]:

```
x=final["Healthy life expectancy"]
y=final["max_infection_rates"]
sns.scatterplot(x,np.log(y))
```

Out [34]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482b3d8650>

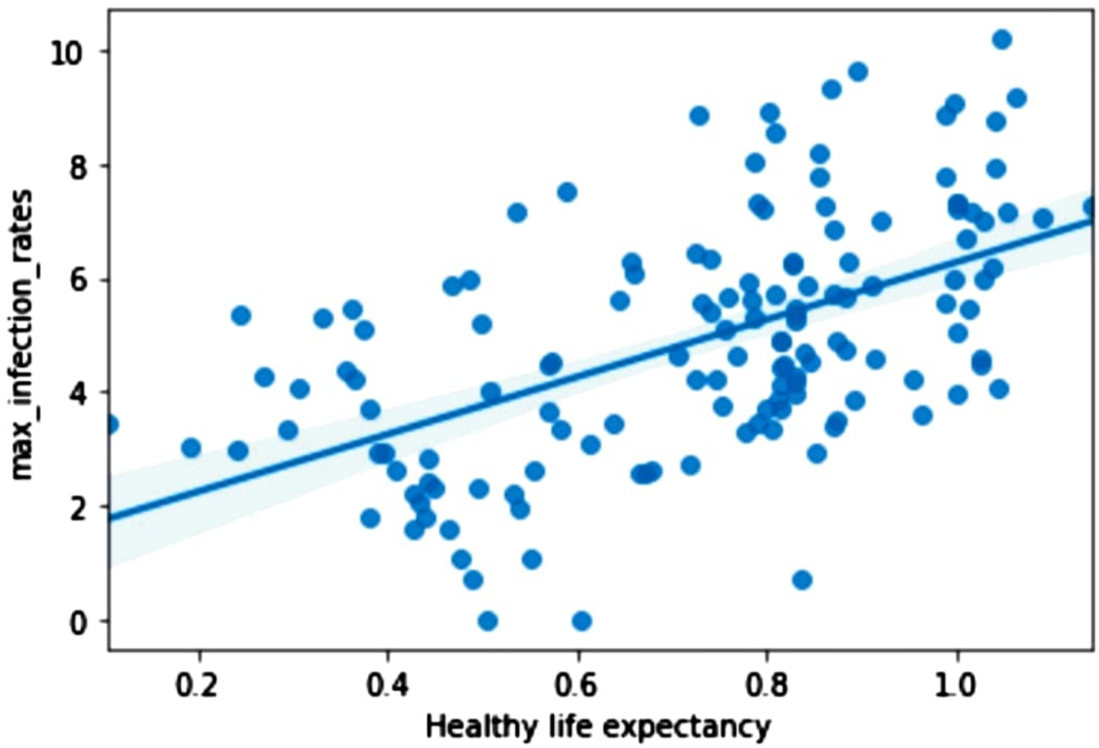


In[35]:

```
sns.regplot(x,np.log(y))
```

Out [35]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f482b3be950>
```



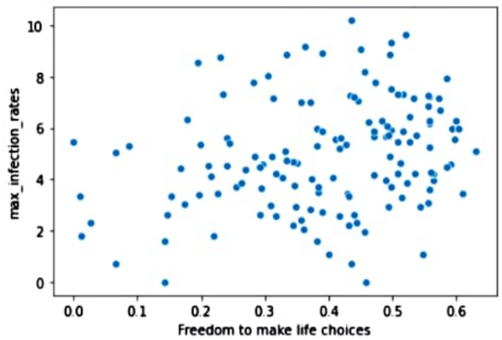
Plotting Freedom to make life choices vs maximum Infection rate

In[36]:

```
x=final["Freedom to make life choices"]
y=final["max_infection_rates"]
sns.scatterplot(x,np.log(y))
```

Out [36]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482b328c90>

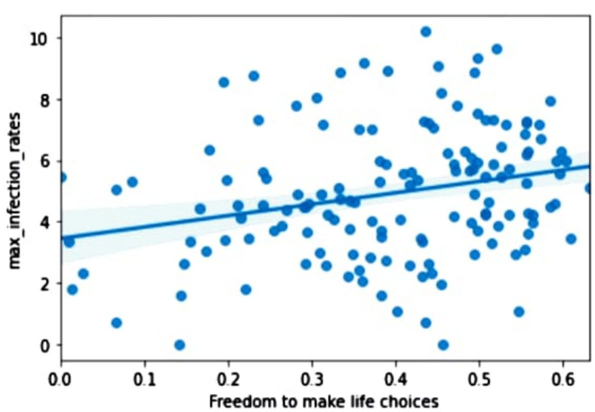


In[37]:

```
sns.regplot(x,np.log(y))
```

Out [37]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482b2a2450>



CONCLUSION:

In conclusion, the analysis of COVID-19 cases has revealed the complex and multifaceted nature of this global pandemic. It has highlighted the importance of public health measures, such as vaccination, mask-wearing, and social distancing, in controlling the spread of the virus. Furthermore, the pandemic has exposed health disparities and underscored the need for equitable access to healthcare and vaccines. While we have made significant progress in understanding and managing the virus, the ongoing vigilance and cooperation of individuals, communities, and nations are essential to overcoming this unprecedented challenge. COVID-19 has reshaped our world, emphasizing the value of science, public health, and international collaboration in confronting global health crises. As we move forward, the lessons learned from this pandemic will guide our efforts to better prepare for and respond to future threats.