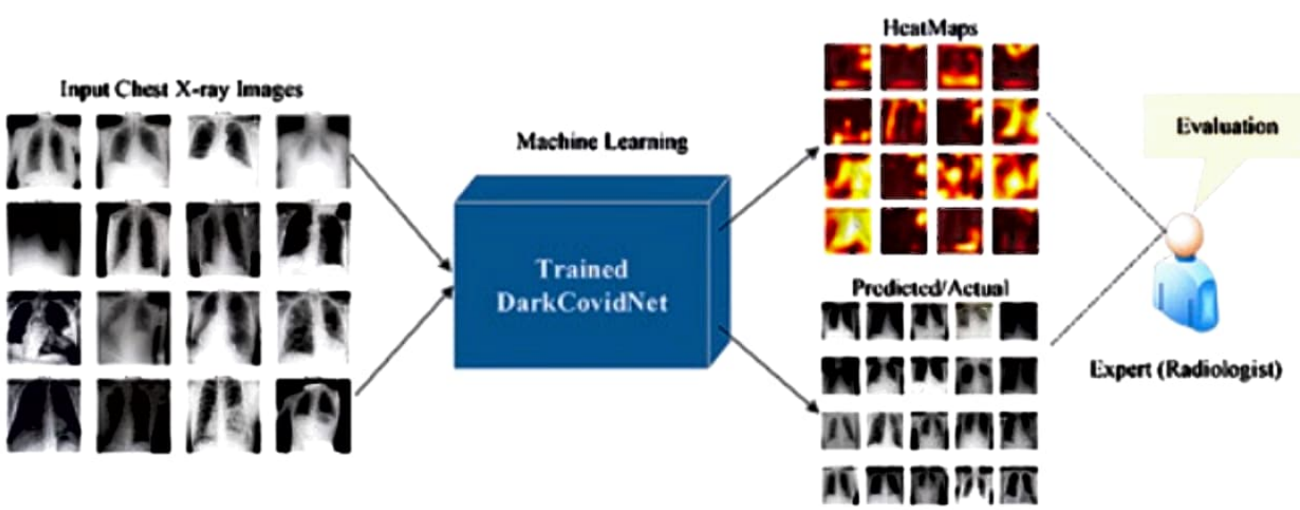
Phase 5 submission document

Project Title: COVID 19 cases analysis

Phase 5: Project Documentation & Submission

Topic: In this section we will document the complete project and prepare it for submission.

Graphical abstract



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.

Dataset Link: https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases

Phase 1: Project Definition and Design Thinking

Project Definition: The COVID-19 pandemic has had a profound impact on global health. This study aims to assess the spread and impact of COVID-19 cases. This document outlines the procedures for analyzing COVID-19 case data, including setting analysis objectives, gathering relevant data, creating informative visualizations, and extracting valuable insights. Ultimately, the project aims to provide a comprehensive understanding of the pandemic's effects and contribute to effective solutions.

Design Thinking:

Monitoring and analyzing COVID-19 cases are crucial for pandemic management. Assessment is the process by which case data is transformed into actionable information. This information is essential for policymakers and health authorities to make informed decisions.

1. Analyzing Objectives:

COVID-19 cases are analyzed using various metrics, including infection rates, mortality rates, and vaccination coverage. A critical tool for this purpose is the COVID-19 Severity Index (CSI). This index summarizes complex case data into understandable terms, aiding decision-making. The objective is to provide clear, actionable information to the public.

2. Data Collection:

We obtain COVID-19 case data, including infection counts, mortality figures, testing rates, and vaccination data.

3. Visualization Strategy:

The COVID-19 Severity Index (CSI) model is employed to assess the severity of the pandemic. CSI involves four stages: (1) selection of relevant case parameters, (2) generation of sub-indices for each parameter, (3) assignment of parameter weight values, and (4) computation of the overall severity index. This approach categorizes the pandemic's severity based on the index value, providing a clear picture of the situation.

4. Predictive Modeling:

Predictive modeling plays a crucial role in anticipating and managing the course of the COVID-19 pandemic. This phase involves using historical case data and various statistical and machine-learning techniques to make informed projections about future trends. By analyzing factors such as vaccination rates, public health measures, and population demographics, predictive models can help policymakers prepare for potential surges in cases and optimize resource allocation.

Topic: Innovation in COVID-19 Cases Analysis

In the face of unprecedented challenges posed by the COVID-19 pandemic, innovation in data analysis is not only beneficial but essential for understanding, managing, and mitigating the impact of the virus. This section will explore various innovative approaches that can enhance the quality and depth of COVID-19 cases analysis:

1. Advanced Data Visualization:

- Cutting-edge data visualisation techniques have the power to transform complex COVID-19 data into accessible and engaging formats.
- Techniques such as 3D visualisations, virtual reality (VR), or augmented reality (AR) applications provide an immersive understanding of data trends.
- These advanced visualisations can help convey critical information to the public and decision-makers with unprecedented clarity, enabling more informed decision-making and timely interventions.

2. Predictive Analytics:

- Leveraging advanced machine learning models and predictive analytics, we can go beyond historical trends to forecast future COVID-19 developments.
- Models such as regression analysis, time series forecasting, or deep learning algorithms can predict case numbers, hospitalizations, and vaccination rates.
- These forecasts are invaluable for public health planning, enabling the allocation of resources where and when they are most needed, thus minimising the impact of the virus.

3. Sentiment Analysis:

- An innovative approach to understanding the pandemic's impact on the public is sentiment analysis.
- This involves analysing social media data to gauge public sentiment, emotions, and reactions in response to COVID-19 developments.
- This analysis helps uncover trends in public opinion and behaviour, which can inform public health messaging, community engagement strategies, and interventions that resonate with the public.

4. Genomic Analysis:

- The genomic analysis focuses on understanding the genetic variants of the COVID-19 virus and their implications for transmission, virulence, and vaccine effectiveness.
- Innovations in this area can provide real-time insights into the emergence and spread of variants, aiding in vaccine development and public health responses.

5. Healthcare Resource Allocation:

- Innovative resource allocation techniques, driven by optimization algorithms, ensure that healthcare resources are distributed efficiently based on COVID-19 case projections.

- These algorithms can consider factors such as ICU bed availability, ventilator requirements, and vaccine distribution.
- This innovation optimizes patient care and supports healthcare systems in providing the best care possible during the pandemic.

6. Community Engagement Apps:

- Innovations in mobile applications engage the community in reporting symptoms, vaccination status, and potential exposure to the virus.
- These apps provide real-time data for analysis.
- Through the use of these apps, individuals can actively participate in data collection and receive important information, contributing to a more informed and engaged public.

7. Blockchain for Data Integrity:

- Blockchain technology, with its inherent security and transparency, safeguards the integrity of COVID-19 data.
- It ensures data is tamper-proof, providing a reliable source of statistics.
- Beyond data integrity, blockchain can also be used to secure digital vaccine passports and other applications in the fight against the virus.

8. AI Chatbots:

- AI-powered chatbots offer real-time information, answer common questions, and assist users in finding testing and vaccination centres.
- These chatbots are available around the clock, providing immediate assistance.
- Their scalability and availability have proven to be a valuable asset in disseminating critical information and easing the burden on healthcare hotlines.

9. Data Fusion:

- Data fusion techniques combine diverse datasets, including weather data, mobility data, and economic indicators, with COVID-19 data.
- This approach identifies correlations and causations that may not be evident from individual datasets.
- These insights have profound policy implications, enabling targeted interventions and lockdown measures.

10. Remote Monitoring:

- Innovations in remote monitoring include wearable devices and IoT solutions that collect vital health data from COVID-19 patients or individuals in quarantine.
- These devices send alerts when critical health conditions are detected.
- They enable early intervention and reduce the burden on healthcare facilities, ultimately saving lives.

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□72

736□750□766□773

2□NaN□Algeria□28.0339□1.6596□0□0□0□0□0□0□...□2811□2910□3007□3127□325

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

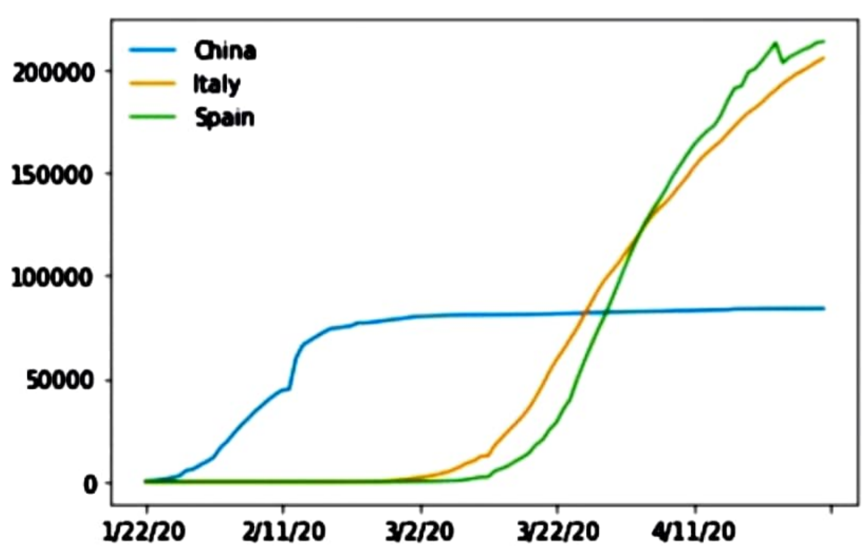
visualization always helps for better understanding of our data.

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

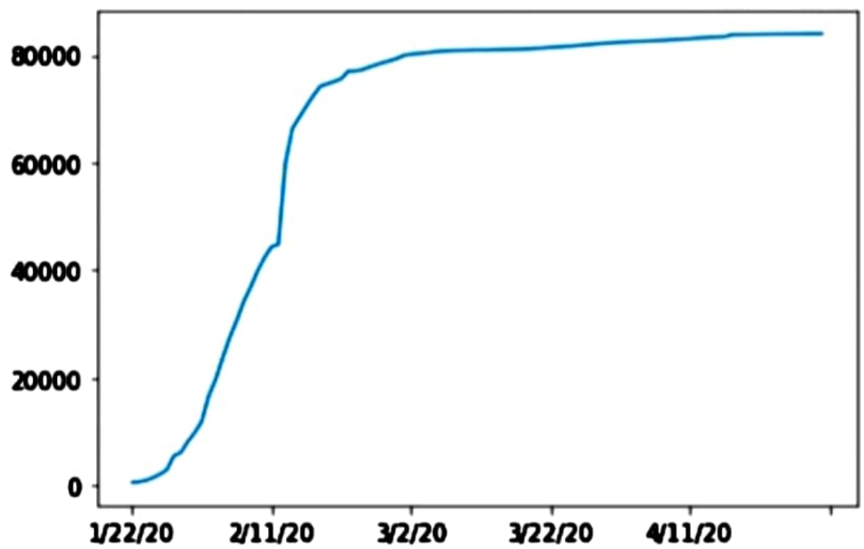
we need to find a good measure reperestend as a number, describing the spread of the virus in a country.

In[10]:

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

Out[10]:

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

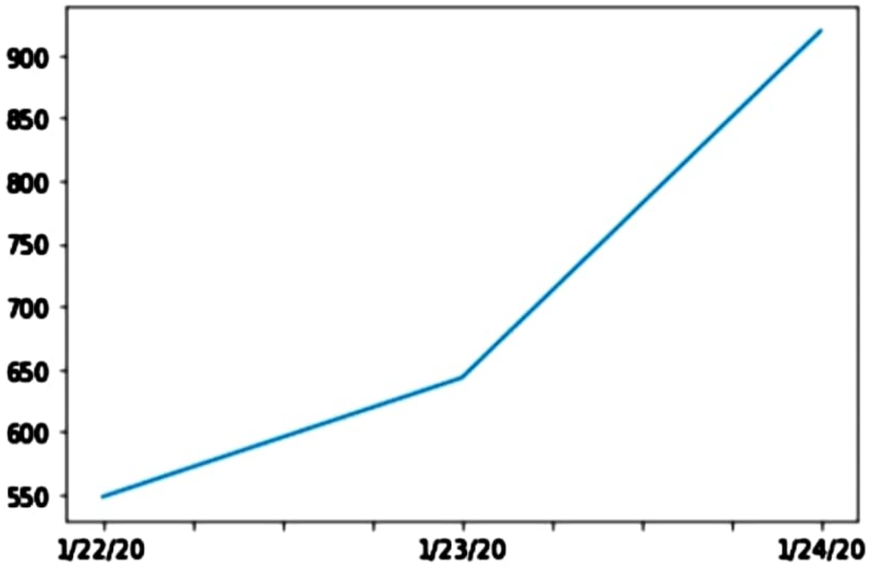


In[11]:

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

Out[11]:

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



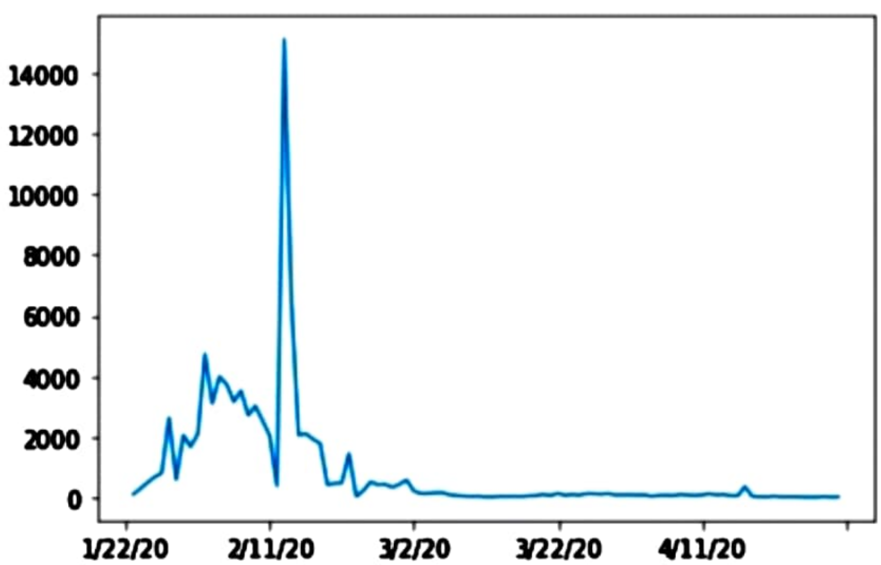
caculating the first derivative of the curve

In[12]:

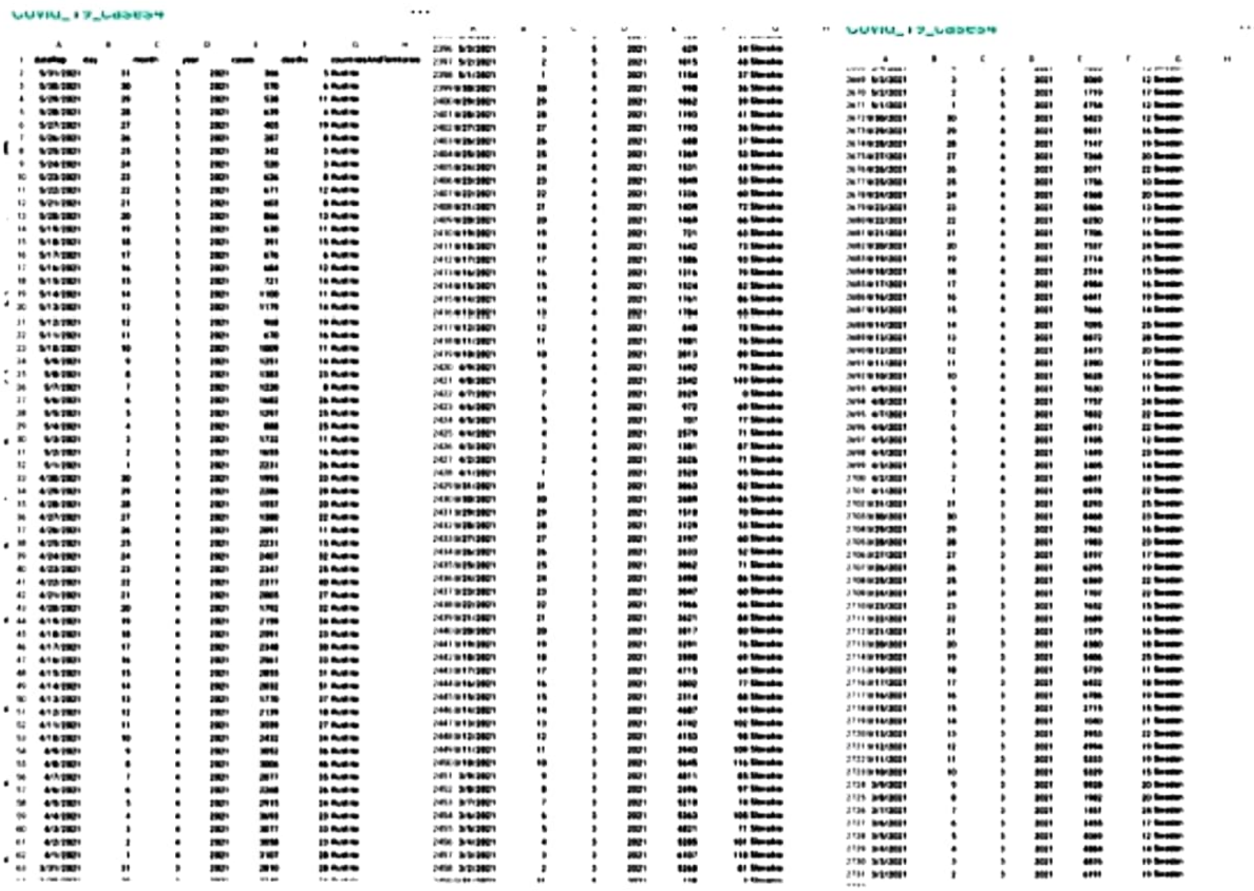
```
aggregating.loc['China'].diff().plot()
```

Out[12]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f482df09290>



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.

Page 14

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.

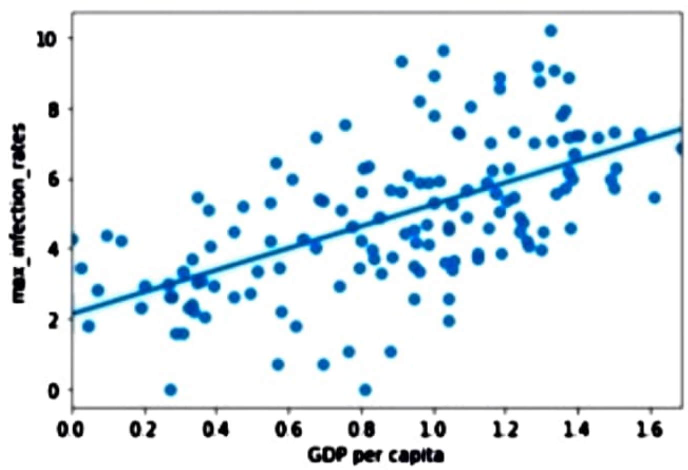
5.Employ various feature selection methods to choose the most relevant features:Filter Methods: These methods use statistical metrics to score features and select the top-ranked ones.Wrapper Methods: They involve training a machine learning model and selecting features based on their impact on model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

In[31]:

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

Out [31]:

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



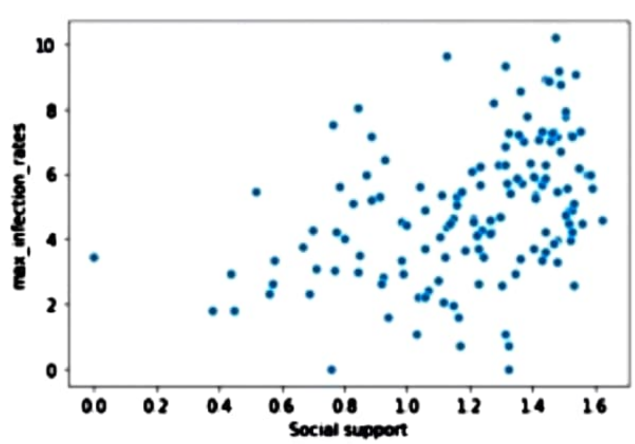
Plotting Social support vs maximum Infection rate

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>

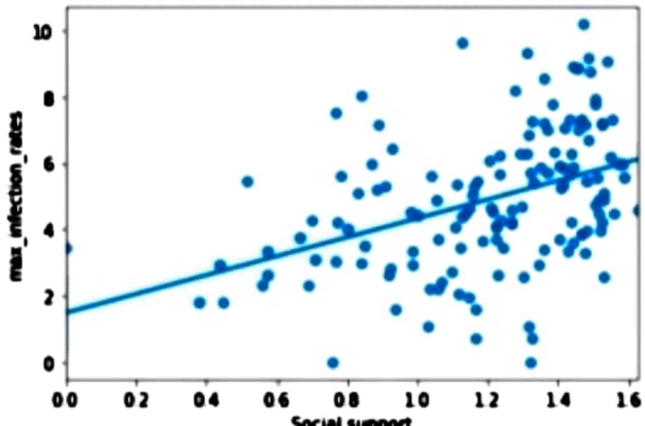


In[33]:

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

Out [33]:

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



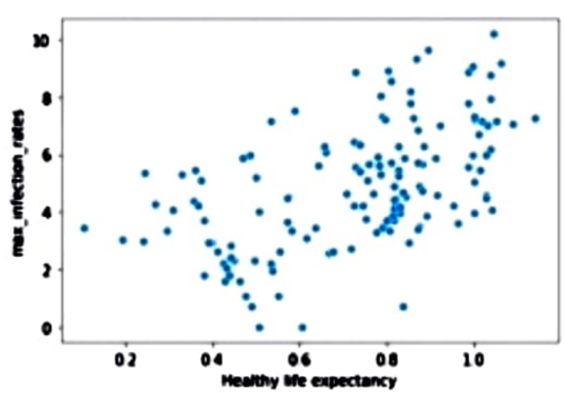
Plotting Healthy life expectancy vs maximum Infection rate

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>

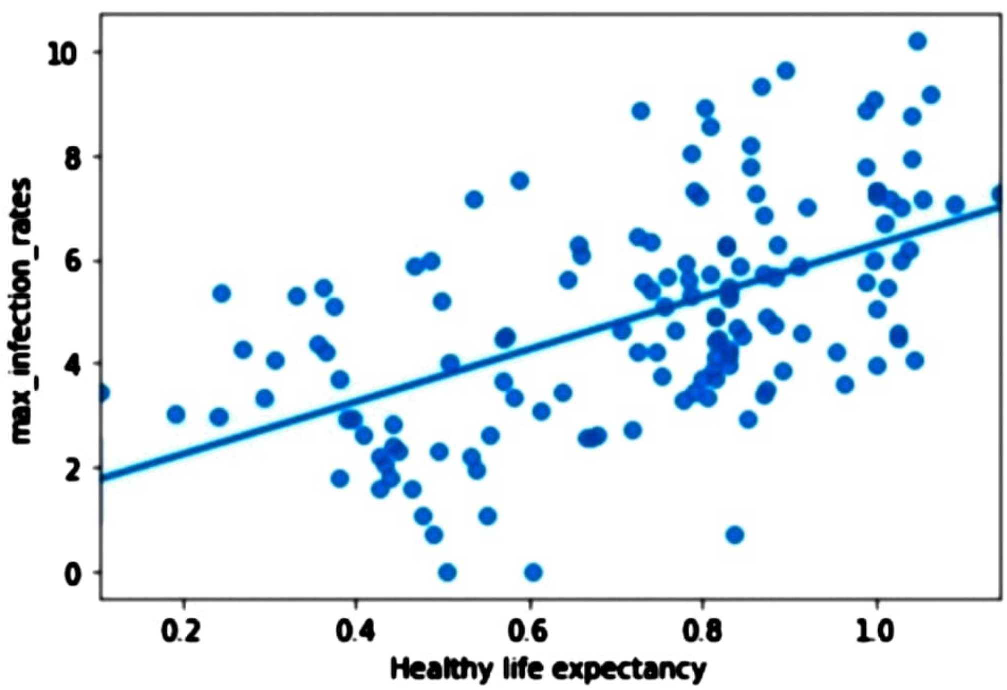


In[35]:

```
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))
```

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.