COVID-19-ANALYTICS WITH COGNOS

Phase-2 document submission

Project name: covid-19-Analysis

Phase-2: Innovation

Consider incorporating data segmentation by time periods or countries for deeper insights

INTRODUCTION:

1.Data Loading and Exploration: Start by loading the dataset into a data analysis tool like Python's Pandas. Begin with a preliminary exploration of the dataset to understand its structure and contents.

2**. Data Filtering:** Filter the dataset to focus only on the relevant data for the EU/EUA. This involves selecting rows where the country belongs to the EU/EUA.

3. Data Segmentation by Time Periods: Create time-based segments for data analysis. This could be done by grouping the data into time periods such as months or years. Use the date, month, and year columns to achieve this segmentation.

4. Data Segmentation by Countries: Segment the data by individual countries or territories within the EU/EUA. Group the data by the "countries" column to create country-specific segments.

5**. Calculate Mean and Standard Deviation:** Calculate the mean and standard deviation of COVID-19 cases and associated deaths for each segmented dataset. This step results in separate mean and standard deviation values for each time period and country.

6. Data Visualization: Create visualizations to represent the calculated mean and standard deviation values. Consider using line plots, bar charts, or box plots to visualize the trends and variations in cases and deaths over time and across countries.

7. Comparisons and Contrasts: Analyze the visualizations to compare and contrast the data. Identify any notable trends, variations, or anomalies in the COVID-19

cases and associated deaths. Look for differences in how the pandemic evolved over time and across different countries.

8. Interpretation and Insights: Provide interpretations of the findings. Explain what the data reveals about the patterns and variations in COVID-19 cases and deaths within the EU/EUA. Consider the impact of time, geography, public health measures, and other factors on these trends.

PROGRAM:

importing the dataset and processing it

First we have to download the datset which is taken from the koggle using the given link and imort the dataset using pandas

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv("/content/drive/MyDrive/Certification/covid.csv")
df
```

	dateRep	day	month	year	cases	deaths	countriesAndTerritories
0	31-05-2021	31	5	2021	366	5	Austria
1	30-05-2021	30	5	2021	570	6	Austria
2	29-05-2021	29	5	2021	538	11	Austria
3	28-05-2021	28	5	2021	639	4	Austria
4	27-05-2021	27	5	2021	405	19	Austria
...	...	...	...	...	...	...	...
2725	06-03-2021	6	3	2021	3455	17	Sweden
2726	05-03-2021	5	3	2021	4069	12	Sweden
2727	04-03-2021	4	3	2021	4884	14	Sweden
2728	03-03-2021	3	3	2021	4876	19	Sweden
2729	02-03-2021	2	3	2021	6191	19	Sweden

2730 rows × 7 columns

finding the missing values

Now we have to check the null or missing values in each columns

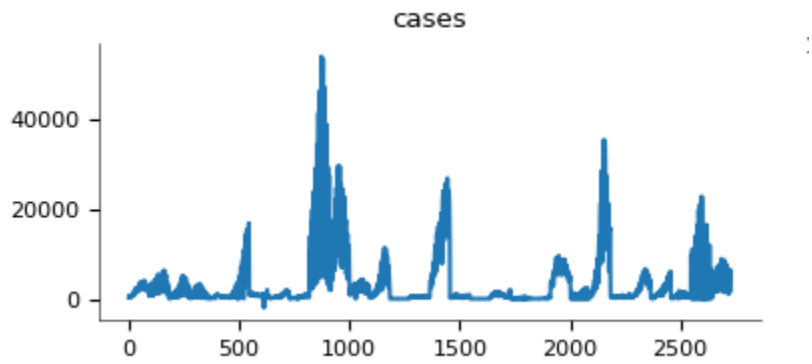
By using the following code.

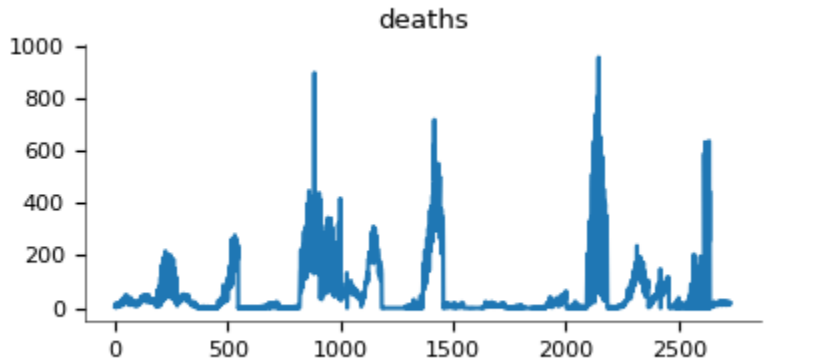
```
df.isnull().sum()
```

```
dateRep                    0
day                        0
month                      0
year                       0
cases                      0
deaths                     0
countriesAndTerritories    0
dtype: int64
```

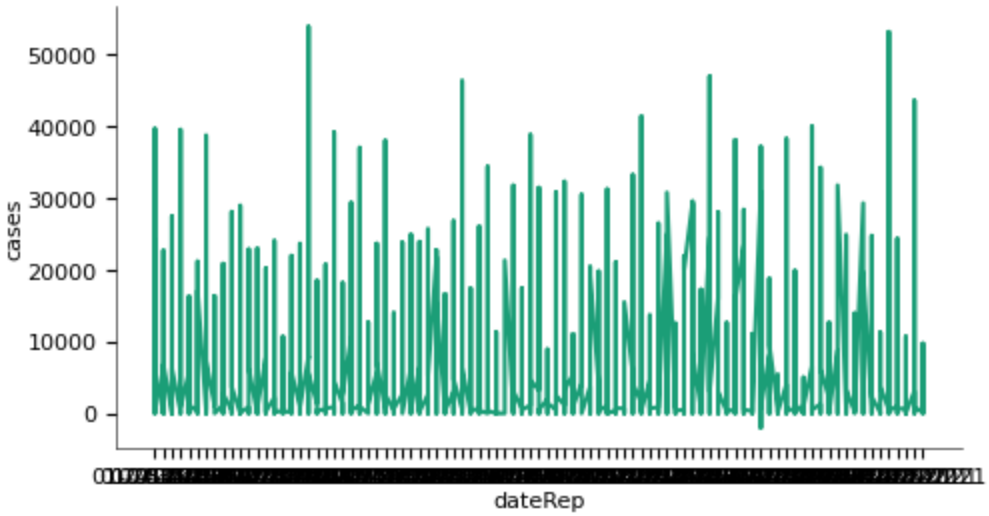
There is no null values or missing values in every columns in the given dataset.

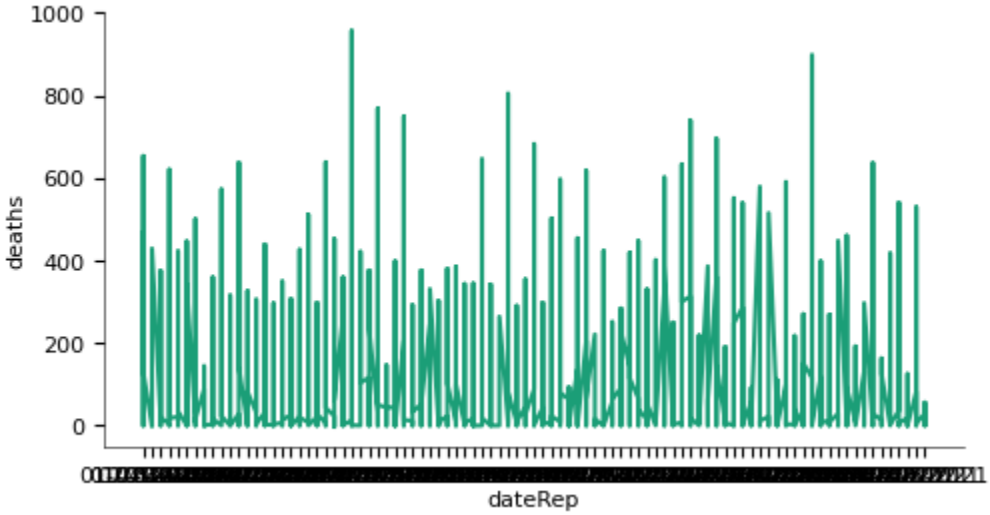
#visualizing the values of cases and deaths





Analysing the time series of the datas like cases,deaths and months





data segmentation by time periods

```
grouped_data = df.groupby(['year', 'month'])
```

We can group the data according to the time period or based upon the countries

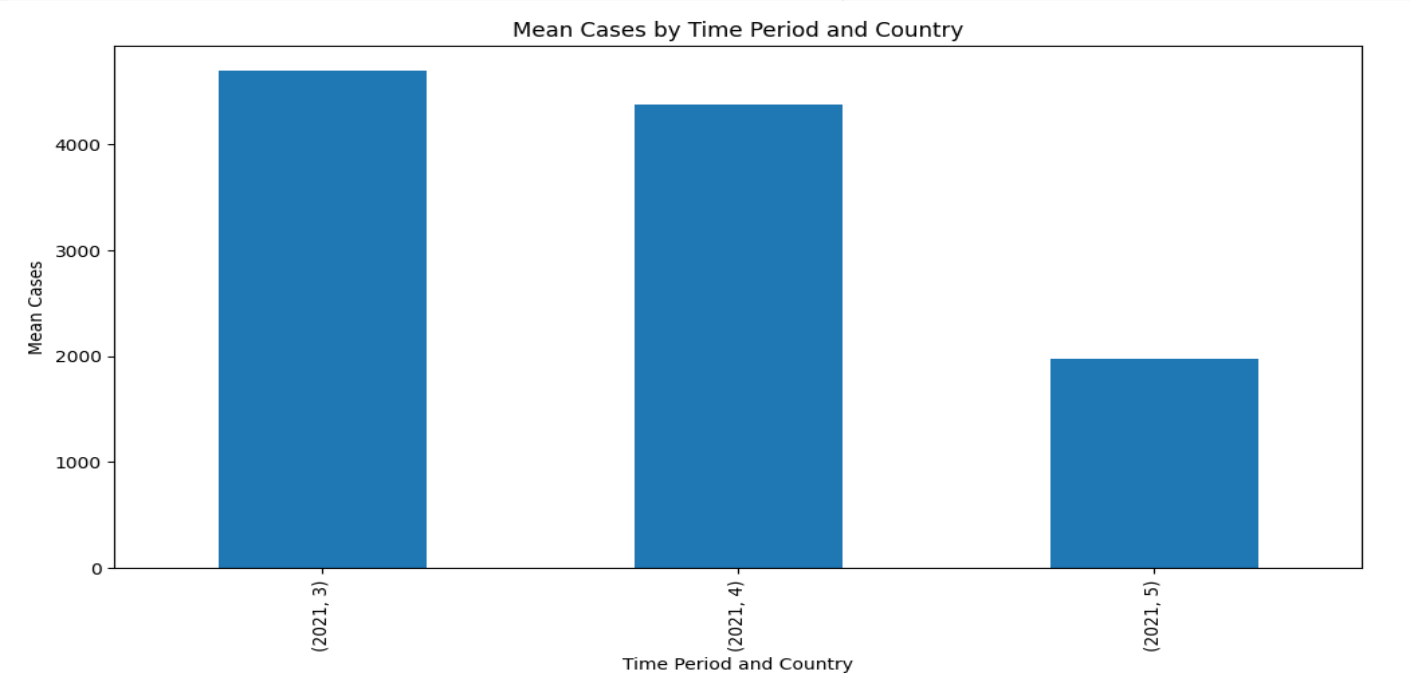
finding the mean and standard deviations for the cases and deaths segmented data by time period

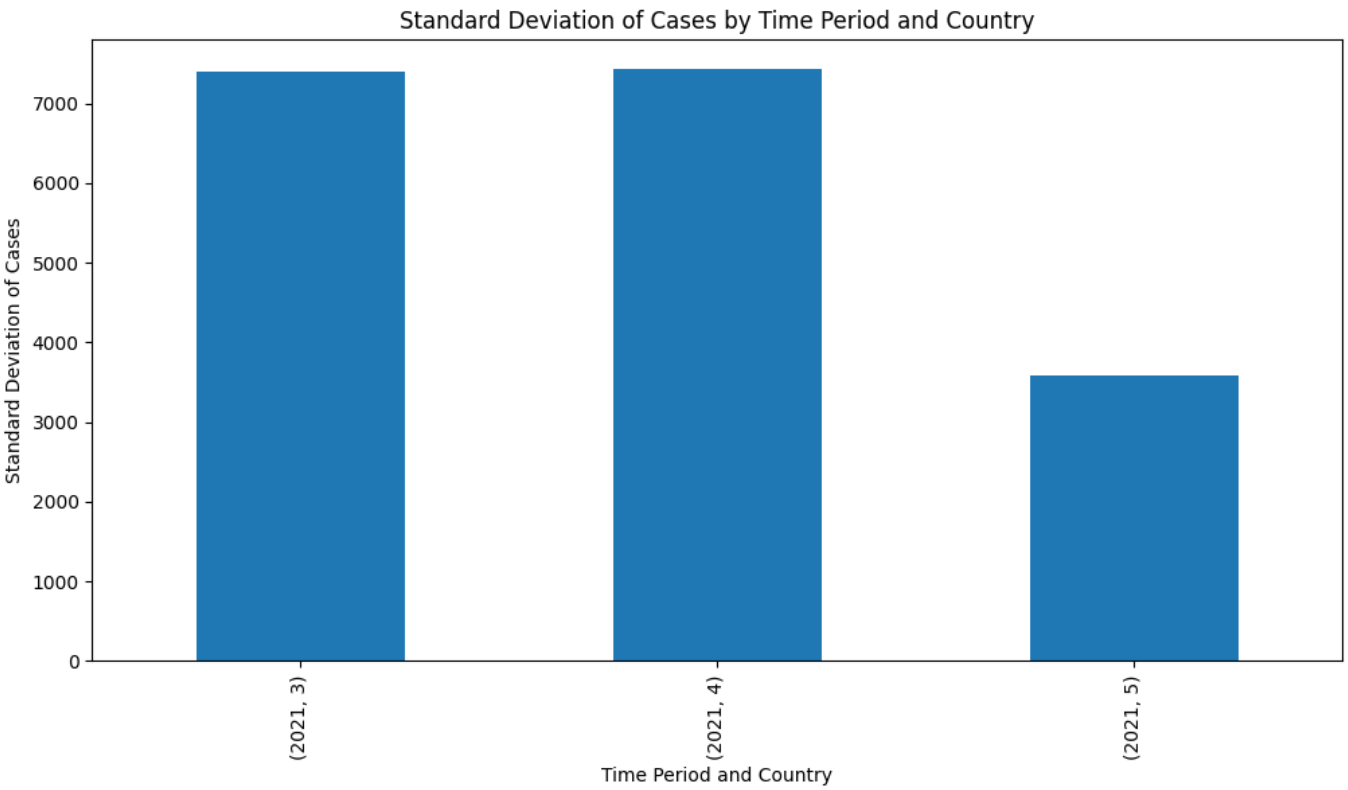
```
mean_cases = grouped_data['cases'].mean()
std_cases = grouped_data['cases'].std()

mean_deaths = grouped_data['deaths'].mean()
std_deaths = grouped_data['deaths'].std()
```

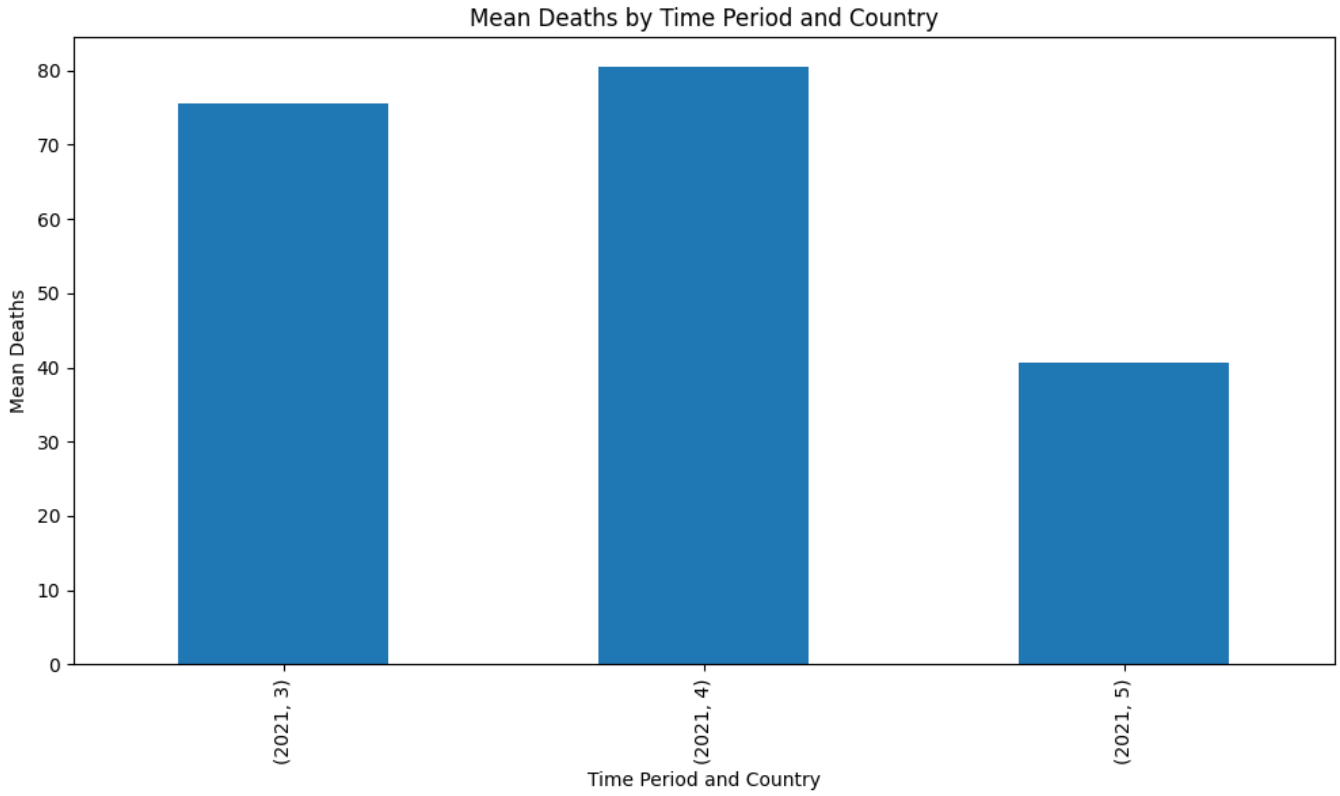
#plot the meanand standard deviation of cases and deaths overtime period

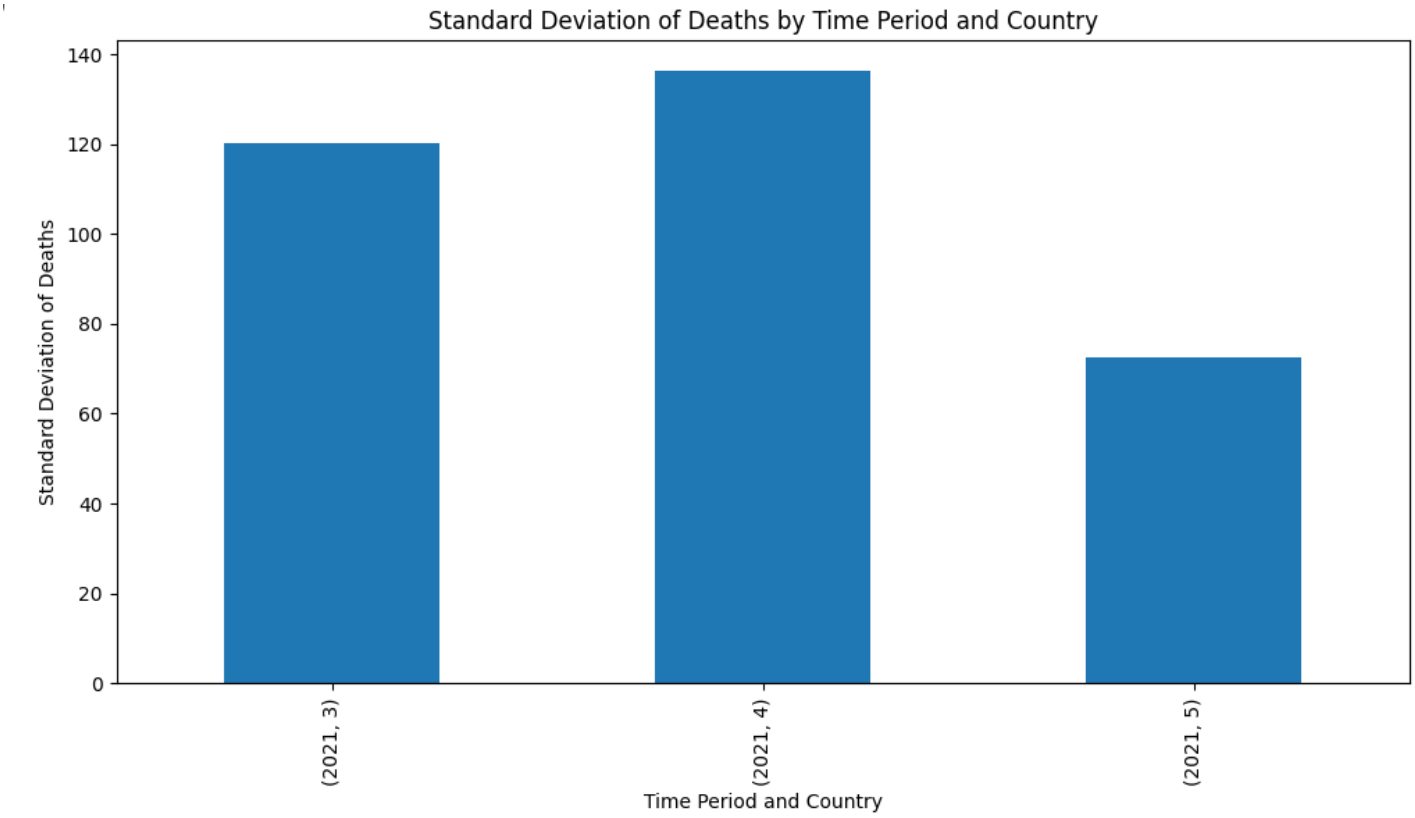
```
plt.figure(figsize=(12, 6))
mean_cases.plot(kind='bar', title='Mean Cases by Time Period and Country')
plt.xlabel('Time Period and Country')
plt.ylabel('Mean Cases')
plt.show()

plt.figure(figsize=(12, 6))
std_cases.plot(kind='bar', title='Standard Deviation of Cases by Time Period and Country')
plt.xlabel('Time Period and Country')
plt.ylabel('Standard Deviation of Cases')
plt.show()

plt.figure(figsize=(12, 6))
mean_deaths.plot(kind='bar', title='Mean Deaths by Time Period and Country')
plt.xlabel('Time Period and Country')
plt.ylabel('Mean Deaths')
plt.show()

plt.figure(figsize=(12, 6))
std_deaths.plot(kind='bar', title='Standard Deviation of Deaths by Time Period and Country')
plt.xlabel('Time Period and Country')
plt.ylabel('Standard Deviation of Deaths')
plt.show()
```

Output:




Standard Deviation of Cases by Time Period and Country
Standard Deviation of Cases
0
1000
2000
3000
4000
5000
6000
7000
(2021, 3)
(2021, 4)
(2021, 5)
Time Period and Country


Mean Deaths by Time Period and Country
Mean Deaths
0
10
20
30
40
50
60
70
80
(2021, 3)
(2021, 4)
(2021, 5)
Time Period and Country



Conclusion:

- From the above bar chart of the both cases and deaths ,mean and standard deviations of both cases and deaths . we can observe that the cases and deaths are getting decreased over the time period .
- And also we can observe when cases decreased corresponding deaths are also decreased so we can conclude that tha death is directly proportional to cases.