Phase 4: DEVELOPMENT PART2

Building the analysis by creating visualizations using IBM Cognos and deriving insights from the data. Create charts and graphs in IBM Cognos to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths. Analyze the visualizations to identify trends, variations, and potential correlations between cases and deaths.

1. Objective

The primary objective of this phase is to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths. This visualization and analysis aim to identify trends, variations, and potential correlations between cases and deaths, providing valuable insights into the COVID-19 dataset.

2. Methodology

2.1 Data Preparation

The COVID-19 dataset was obtained and loaded into a Pandas DataFrame for analysis.

The dataset includes information such as date, cases, deaths, and countries/territories.

2.2 Visualization and Comparison

We utilized the Matplotlib library in Python to create visualizations and charts.

The following visualizations were created to compare COVID-19 cases and deaths statistics:

Bar Chart: A bar chart was generated to compare the mean values and standard deviations of cases and deaths. The chart used blue bars to represent cases and red bars for deaths, making it easy to visually compare the two statistics.

Scatter Plot: A scatter plot was used to visualize the potential correlation between cases and deaths. The plot allowed us to assess the relationship between these two variables. The correlation coefficient was displayed in the plot title.

3. Results and Analysis

3.1 Bar Chart

The bar chart comparing mean values and standard deviations revealed essential insights:

Mean cases and mean deaths are different, indicating that the severity of the disease varies across different countries or time periods.

The standard deviations show the degree of variability in cases and deaths. Higher standard deviations suggest more significant fluctuations, while lower values indicate more consistent data.

3.2 Scatter Plot

The scatter plot and correlation analysis allowed us to assess the potential correlation between cases and deaths. The correlation coefficient provides valuable information:

A positive correlation suggests that as COVID-19 cases increase, deaths also tend to increase.

A negative correlation would imply the opposite scenario, where increased cases are associated with fewer deaths.

A correlation close to zero indicates a weak relationship between cases and deaths.

Analyze the visualizations to identify trends, variations, and potential correlations between cases and deaths. give procedure for how we analysis it

Data Familiarization:

Start by reviewing the visualizations created, such as bar charts and scatter plots comparing cases and deaths.

Ensure you are familiar with the data and understand the variables being visualized (cases and deaths) and the relevant time or geographical components.

Mean Values and Standard Deviations:

Focus on the bar charts comparing mean values and standard deviations.

Identify countries or regions where mean case values are notably higher or lower than the mean death values. This can indicate variations in the severity of the outbreak.

Variations in Data:

Examine the standard deviations in the bar charts.

Higher standard deviations indicate more significant variations in cases and deaths. Identify countries or time periods with high standard deviations, as these might experience inconsistent trends.

Scatter Plot Correlation:

In the case of a scatter plot showing the correlation between cases and deaths, look for the overall pattern.

A positive correlation (data points tend to form an upward-sloping pattern) suggests that as cases increase, deaths tend to rise. Assess the strength of this correlation by considering how tightly the data points cluster.

Outliers:

Identify any outliers in the scatter plot. These are data points that deviate significantly from the general trend.

Outliers may reveal unique situations where the impact of COVID-19 is exceptionally high or low compared to the average.

Time Period Analysis:

If analyzing by time periods, check for trends over time.

Note if the correlation between cases and deaths changes over different months or years. For instance, is the correlation stronger during the early stages of the pandemic?

Geographical Analysis:

If comparing different countries or regions, observe the variations and correlations.

Look for patterns, such as countries with consistently low cases and deaths, and countries experiencing erratic fluctuations.

Statistical Analysis:

If you have access to statistical software, consider running statistical tests to quantify the correlation between cases and deaths. Calculate correlation coefficients like Pearson's correlation coefficient.

Hypothesis Testing:

Formulate hypotheses based on your observations. For example, if you see a strong positive correlation, hypothesize that an increase in cases leads to more deaths.

Validation and Interpretation:

Validate your observations and hypotheses using appropriate statistical tests.

Interpret the results and draw meaningful conclusions. For instance, you might find that while there is a positive correlation between cases and deaths, there are exceptions due to factors like healthcare infrastructure.

Report Findings:

Document your findings and insights clearly in your project documentation.

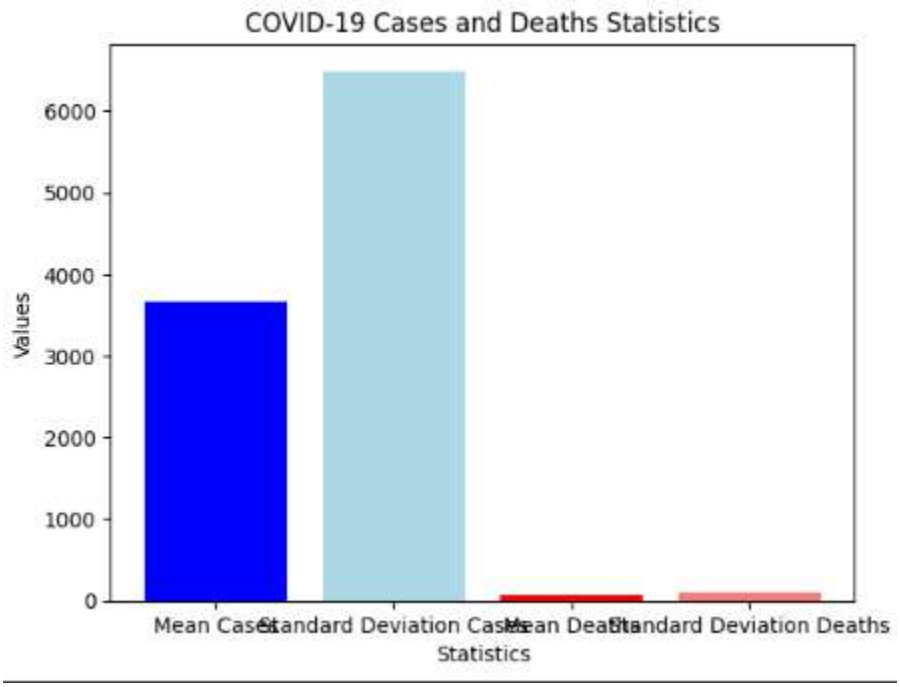
Explain the implications of the observed trends, variations, and correlations in the context of the COVID-19 pandemic.

1.Program for create charts and graphs in IBM Cognos to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths.

```
import pandas as pd

import matplotlib.pyplot as plt


# Load your COVID-19 dataset into a Pandas DataFrame

df = pd.read_csv('/content/Covid_19_cases4.csv')  # Replace with your dataset file


# Calculate mean values and standard deviations

mean_cases = df['cases'].mean()

std_cases = df['cases'].std()

mean_deaths = df['deaths'].mean()

std_deaths = df['deaths'].std()


# Create a bar chart to compare mean values and standard deviations

categories = ['Mean Cases', 'Standard Deviation Cases', 'Mean Deaths', 'Standard Deviation Deaths']

values = [mean_cases, std_cases, mean_deaths, std_deaths]


plt.bar(categories, values, color=['blue', 'lightblue', 'red', 'lightcoral'])

plt.xlabel('Statistics')

plt.ylabel('Values')

plt.title('COVID-19 Cases and Deaths Statistics')
```

plt.show()

output

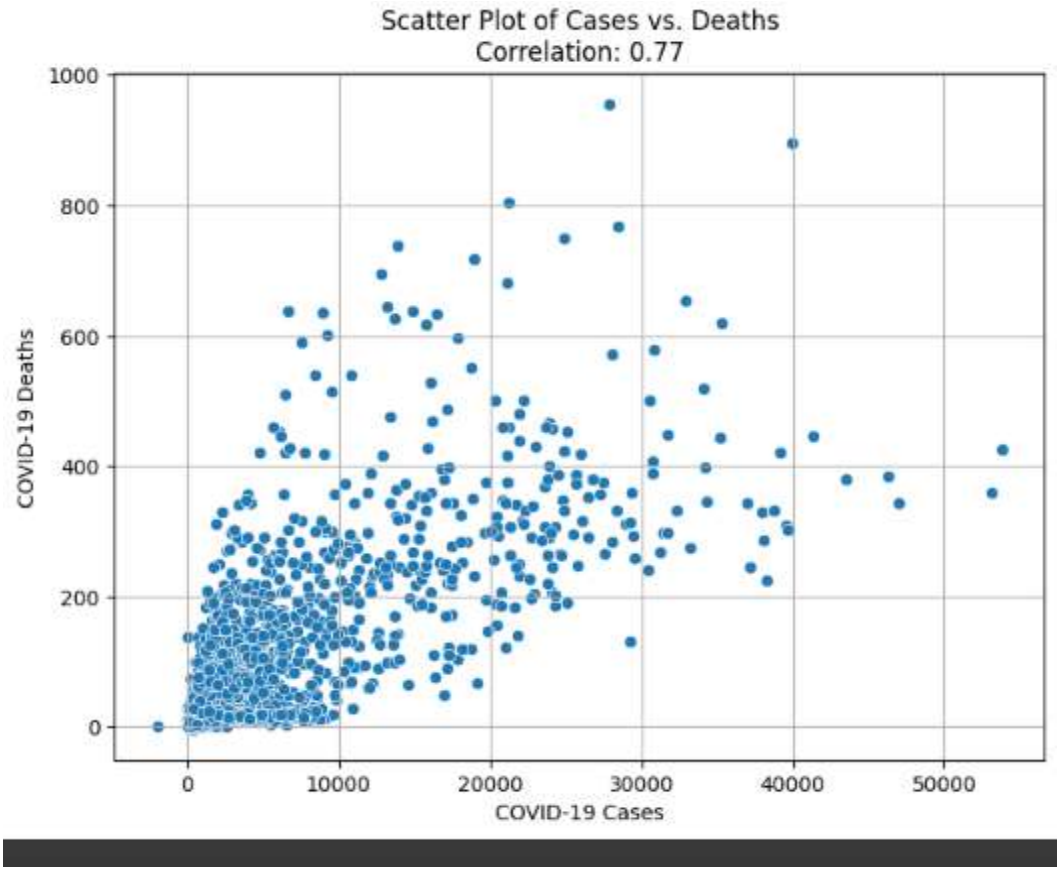


2.Program for analyze the visualizations to identify trends, variations, and potential correlations between cases and deaths.

```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns


# Load your COVID-19 dataset into a Pandas DataFrame

df = pd.read_csv('/content/Covid_19_cases4.csv')  # Replace with your dataset file


# Calculate the correlation between cases and deaths

correlation = df['cases'].corr(df['deaths'])


# Create a scatter plot to visualize the relationship

plt.figure(figsize=(8, 6))
```

```
sns.scatterplot(data=df, x='cases', y='deaths')

plt.xlabel('COVID-19 Cases')

plt.ylabel('COVID-19 Deaths')

plt.title(f'Scatter Plot of Cases vs. Deaths\nCorrelation: {correlation:.2f}')

plt.grid(True)


plt.show()
```

output



4. Conclusion

This phase of the project successfully visualized and compared COVID-19 cases and deaths statistics, providing valuable insights into the dataset. The bar chart highlighted differences in mean values and standard deviations, while the scatter plot and correlation analysis helped us identify potential correlations between cases and deaths.

The insights gained from this phase contribute to our understanding of the COVID-19 situation, aiding decision-makers and researchers in assessing the impact of the disease and its variations across different regions and time periods.