**Project Title:** COVID-19 using Cognos

**Phase 3: Development Part 1**

**1. Define Analysis Objectives:**

Clearly define the objectives of your COVID-19 analysis. What insights are you trying to gain from the data? Common objectives include tracking the spread of the virus, understanding its impact on different regions, and identifying trends and patterns in COVID-19 cases and deaths.

**2. Obtain COVID-19 Data:**

To perform the analysis, you'll need a reliable source of COVID-19 cases and deaths data. This data is often provided by government health agencies or other authoritative sources. Ensure that you have the necessary permissions to use and analyze the data.

**3. Data Preprocessing:**

Before loading the data into IBM Cognos for visualization, you should preprocess it to ensure its accuracy and reliability. This involves tasks such as cleaning, transforming, and structuring the data. Here are some common preprocessing steps:

**a. Data Cleaning:**

- Remove duplicate records, if any.

- Handle missing values appropriately (e.g., through imputation or removal).

- Check for data consistency and correctness.

- Remove outliers that may skew the analysis.

**b. Data Transformation:**

- Convert data types as needed (e.g., dates, numerical values).

- Aggregate data if necessary (e.g., daily counts to weekly or monthly).

- Normalize data if you have data from multiple sources or with different units.

**c. Data Integration:**

- Merge data from different sources if you have multiple datasets (e.g., cases and deaths data from different agencies).

- Ensure that the data is in a format suitable for analysis in IBM Cognos.

**4. Data Loading in IBM Cognos:**

After cleaning and preprocessing the data, you can load it into IBM Cognos for visualization. Depending on your specific version of Cognos, the steps may vary, but generally, you'll need to create a data source connection and define data modules or packages.

**5. Visualization:**

Use the data loaded into IBM Cognos to create visualizations that address your analysis objectives. Common types of visualizations for COVID-19 analysis include line charts, bar charts, heatmaps, and geographic maps to show the spread of the virus over time.

**6. Analysis and Exploration:**

Explore the data through interactive dashboards and reports in Cognos. Use filters and parameters to allow users to drill down into the data and extract meaningful insights.

**7. Communication and Sharing:**

Share the COVID-19 analysis results with stakeholders. You can publish reports, dashboards, or interactive visualizations for others to access and use.

**PROGRAM**

import pandas as pd

# Read the dataset

df = pd.read\_csv('Covid\_19\_cases4.csv')

print(df)

df = pd.read\_csv('Covid\_19\_cases4.csv')

# Check for missing values

print(df.isnull().sum())

# Drop rows with missing values

df = df.dropna()

# Remove duplicate rows

df = df.drop\_duplicates()

# Save the cleaned dataset

df.to\_csv('Tiger.csv', index=False)

OUTPUT

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

dateRep 0

day 0

month 0

year 0

cases 0

deaths 0

countriesAndTerritories 0

dtype: int64

**Data visualization for Covid\_19\_cases analysis**

**Program:**

import pandas as pd

import matplotlib.pyplot as plt

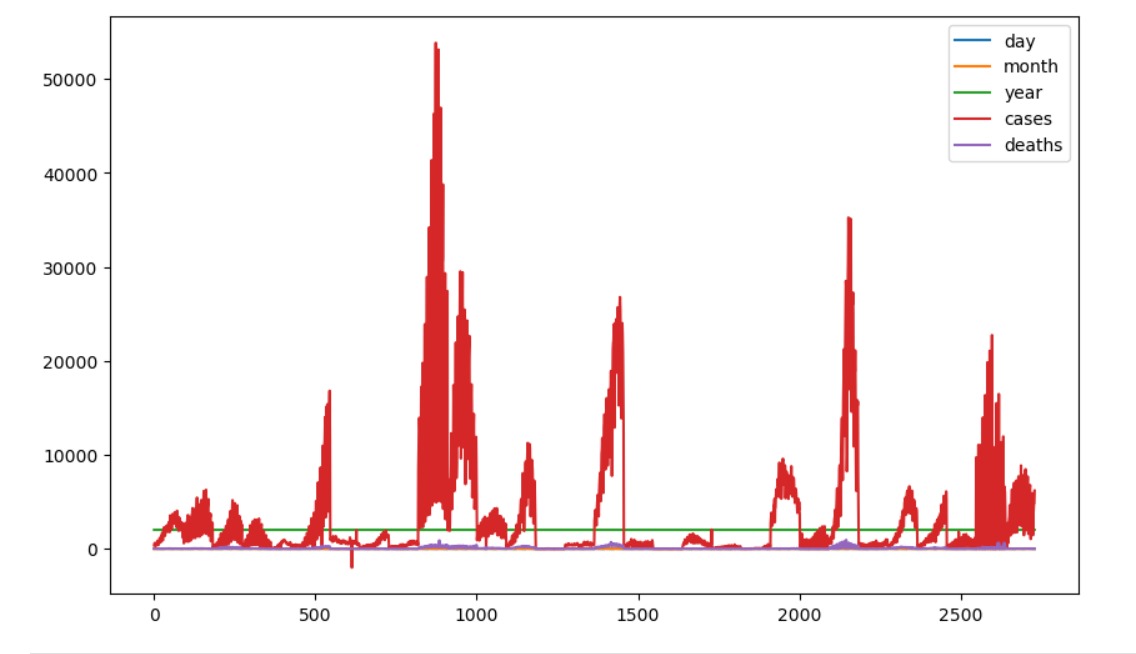
df=pd.read\_csv("Covid\_19\_cases4.csv")

df.head()

df.info()

df.plot(figsize=(10,6))

**OUTPUT:**

****