Project Title: Understanding COVID-19 Trends and Impacts

Phase 5

TEAM MEMBERS

YUVARAJ T **2021115124**

ABINAYA S **2021115004**

AARTHI R **2021115001**

TAMIZHARASAN D **2021115325**

HARISH M **2021115326**

Objectives:

The primary objectives of this project are to analyze COVID-19 data, understand trends, and assess its impact on various factors. Specifically, the goals are to:

1. Analyze COVID-19 data to identify patterns, trends, and insights.
2. Assess the relationship between COVID-19 cases and factors like demographics, healthcare facilities, and government responses.
3. Use data visualization tools, including IBM Cognos, to create informative visualizations.
4. Generate insights from the analysis to aid in understanding the COVID-19 pandemic.

Design Thinking Process:

Empathize:

- Understand the importance of analyzing COVID-19 data for public health and decision-making.
- Identify stakeholders and their data needs, such as healthcare professionals, policymakers, and the general public.

Define:

- Clearly define the project's objectives and scope.
- Specify the data sources and types of analysis to be performed.

Ideate:

- Brainstorm analysis methods, including statistical analysis and data visualization techniques.

- Choose appropriate tools for data visualization, such as IBM Cognos, to create impactful visualizations.

Prototype:

- Develop a plan for data collection, cleaning, and preprocessing.
- Create a prototype of the analysis pipeline, including steps for generating visualizations.

Test:

- Test data collection processes to ensure data quality.
- Verify the accuracy and effectiveness of the visualization techniques.

Development Phases:

1. **Data Collection**:
    - Gather COVID-19 data from reliable sources, including infection rates, demographic information, healthcare data, and government responses.
2. **Data Cleaning and Preprocessing**:
    - Clean the data by handling missing values, outliers, and inconsistencies.
3. **Data Analysis**:
    - Perform statistical analysis to identify correlations and trends between COVID-19 cases and various factors.
4. **Data Visualization**:
    - Utilize IBM Cognos and other visualization tools to create charts, graphs, and interactive visualizations.
5. **Insights Generation**:
    - Extract insights from the analysis, such as the impact of government policies, healthcare facilities, and demographic factors on COVID-19 trends.
6. **Report and Documentation**:
    - Compile the findings into a comprehensive report, including visualizations, interpretations, and recommendations.

Data Analysis Objectives:

The objectives of the data analysis phase are as follows:

- To understand the relationships between COVID-19 trends and various factors.
- To provide actionable insights for healthcare professionals, policymakers, and the general public.

Data Collection Process:

- Data is collected from reliable sources, such as government health agencies and research institutions.
- Data includes daily or weekly updates on COVID-19 cases, vaccination rates, demographic information, and government responses.

Data Visualization Using IBM Cognos:

- IBM Cognos is utilized to create a range of visualizations, such as:
    - Line charts showing the trend of COVID-19 cases over time.
    - Choropleth maps displaying regional infection rates.
    - Stacked bar charts comparing healthcare facilities by region.

Insights:

- Insights are generated by analyzing data and visualizations.
- Example insights might include the impact of vaccination campaigns, regional disparities in healthcare access, and the effectiveness of government policies.

Aid in Understanding COVID-19 Trends and Impacts:

The insights from this analysis can help in the following ways:

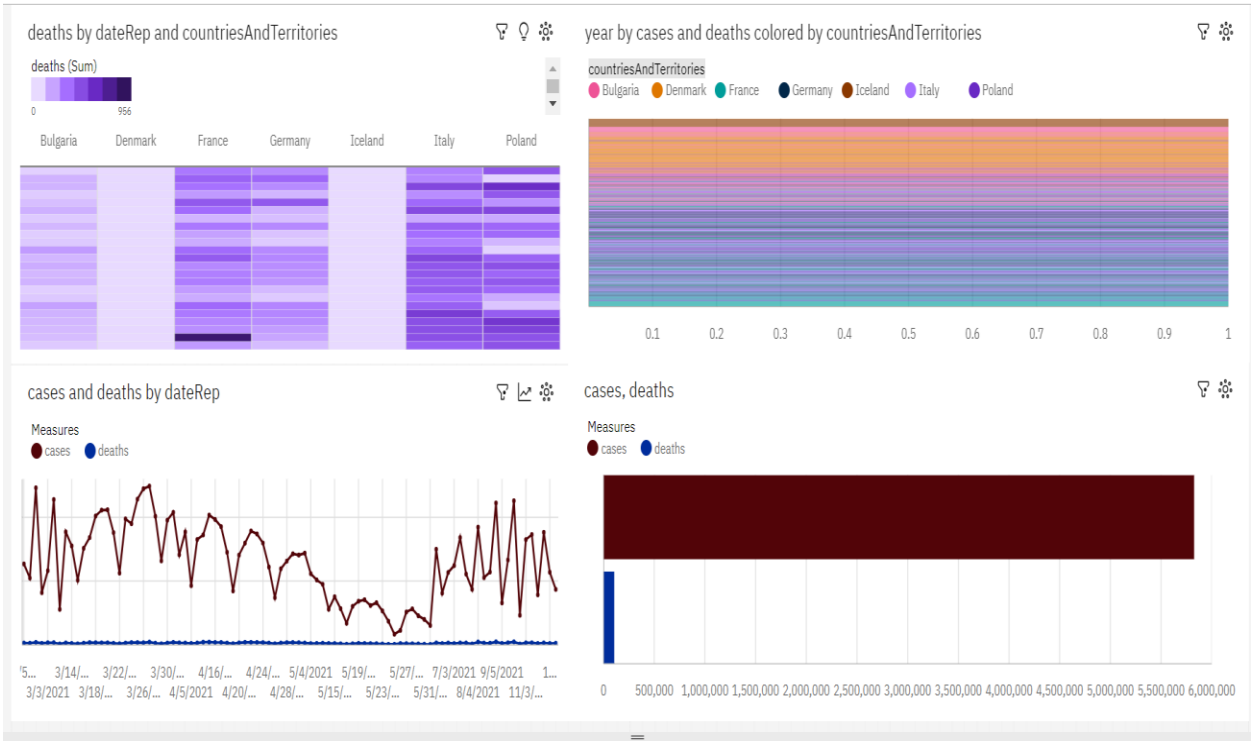
- Policymakers can use the information to make informed decisions about public health measures.
- Healthcare professionals can allocate resources more effectively.
- The general public can gain a better understanding of the pandemic's dynamics.

Replicating the Analysis:

1. Clone the GitHub repository to your local machine.
2. Follow the instructions in the README file to set up the analysis environment.
3. Run the analysis scripts as outlined in the repository.

Example Outputs:

Dashboard created using IBM Cognos:



Insights by IBM Cognos:

dateRep 2021-09-04 has the highest **Total deaths** but is ranked #**7** in **Total cases**.

dateRep 2021-01-05 has the highest **Count distinct countriesAndTerritories** but is ranked #**46** in **Total cases**.

dateRep 2021-03-27 has the highest **Total cases** but is ranked #**4** in **Total deaths**.

dateRep 2021-09-04 has the highest **Total deaths** but is ranked #**7** in **Total cases**.

From **2021-03-29** to **2021-03-30**, **France**'s **cases** increased by **937**%.

dateRep 2021-03-27 has the highest **Total cases** but is ranked #**4** in **Total deaths**.

Across all values of **dateRep**, the average of **cases** is **over nine thousand**.

dateRep 2021-03-27 has the highest **Total cases** but is ranked #**1** in **Count distinct countriesAndTerritories**.

Over all values of **dateRep**, the average of **deaths** is **167.5**.

The overall number of results for **cases** is **630**.

Insights by IBM Cognos:

deaths is unusually low when **countriesAndTerritories** is **Iceland** and **Denmark**.

☆

countriesAndTerritories Poland has the highest **Total deaths** but is ranked **#4** in **Total cases**.

☆

From **2021-09-03** to **2021-09-04**, **Poland**'s **deaths** increased by **2888%**.

☆

countriesAndTerritories France has the highest **Total cases** but is ranked **#3** in **Total deaths**.

☆

Over all values of **dateRep** and **countriesAndTerritories**, the sum of **deaths** is **nearly 106 thousand**.

☆

The total number of results for **cases**, across all **dateRep**, is **630**.

☆

The total number of results for **deaths**, across all **dateRep**, is **630**.

☆

cases ranges from **over eight thousand**, when **dateRep** is **2021-05-25**, to **nearly 125 thousand**, when **dateRep** is **2021-03-27**.

☆

deaths ranges from **185**, when **dateRep** is **2021-05-31**, to **over two thousand**, when **dateRep** is **2021-09-04**.

☆

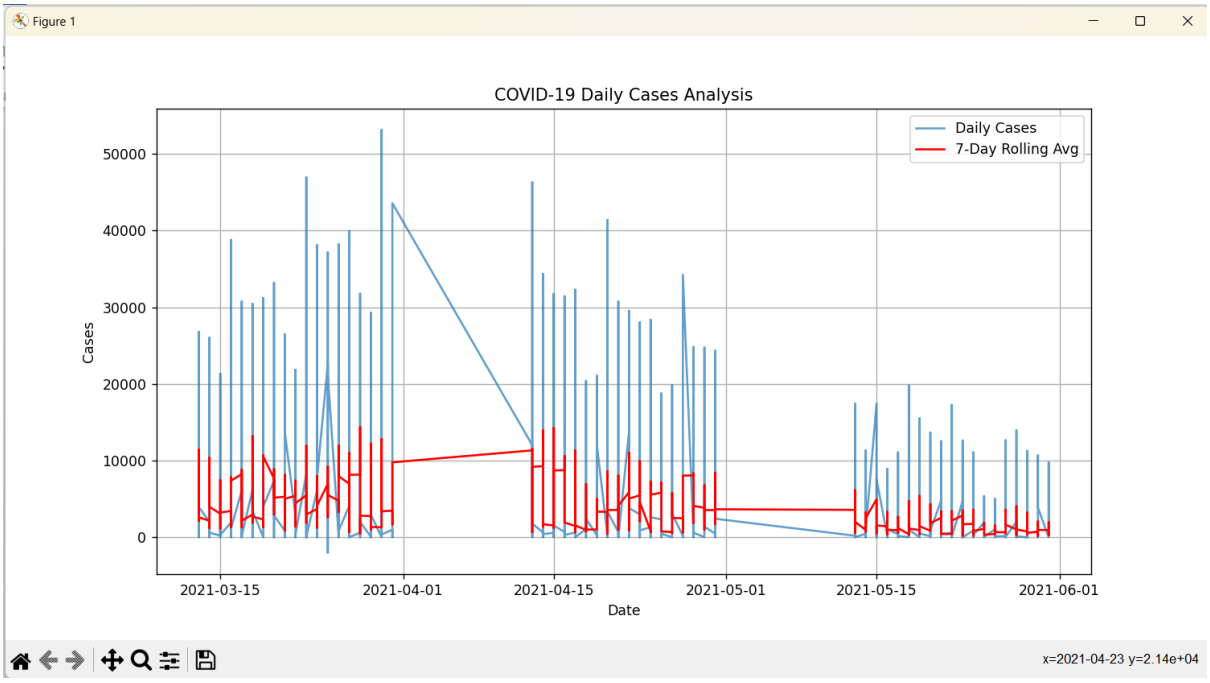
Charts created using python libraries:

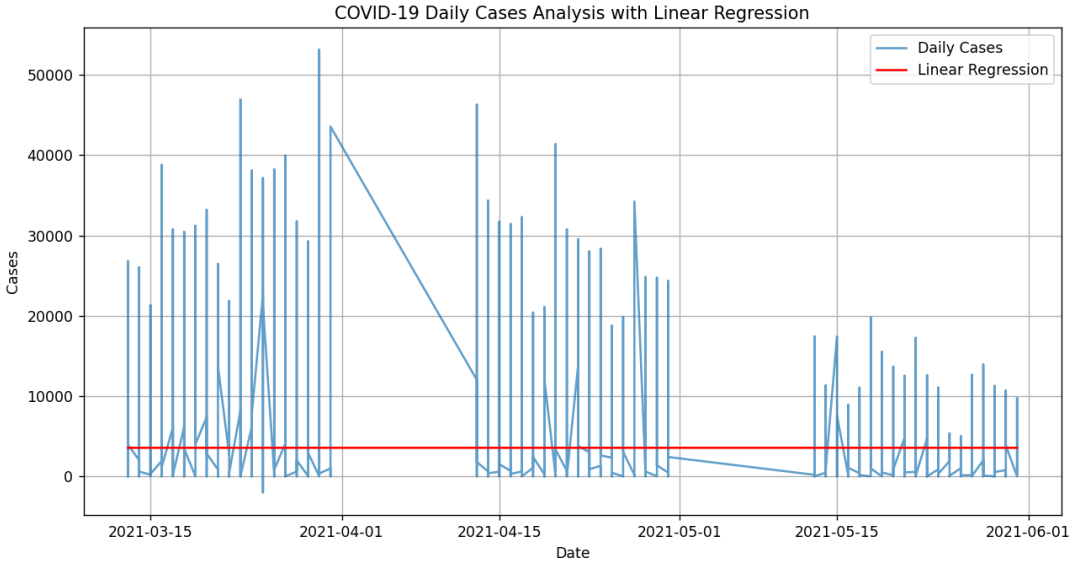
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Load the data from the CSV file
df = pd.read_csv('covidcases.csv')

# Convert 'dateRep' to a datetime object
df['dateRep'] = pd.to_datetime(df['dateRep'], format='%d-%m-%Y',errors='coerce')
```

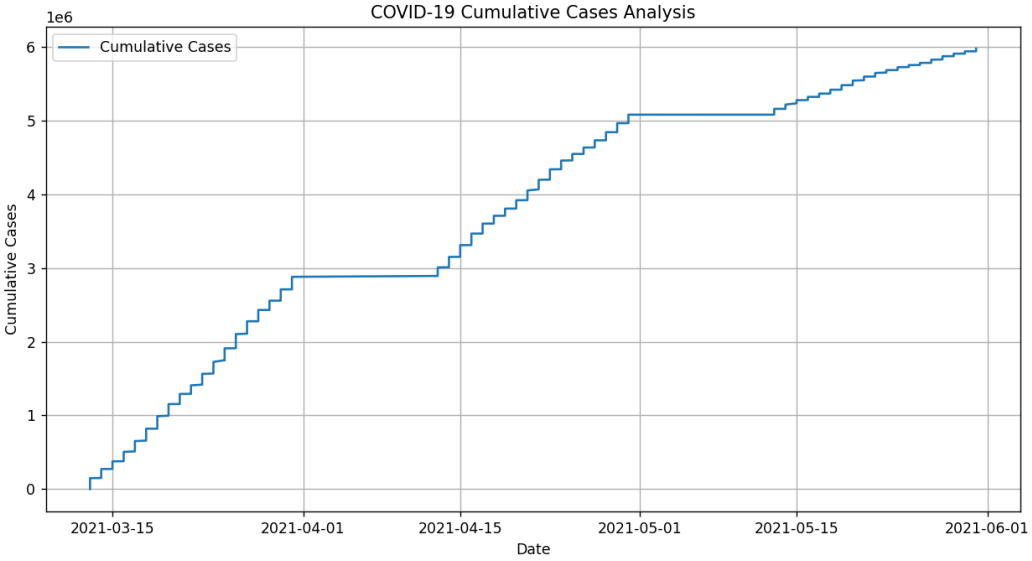
```
# Sort the data by date
df = df.sort_values(by='dateRep')

# Check for missing data
if df.isnull().values.any():
    print("There are missing data. You should handle them.")

# Data Visualization
plt.figure(figsize=(12, 6))
plt.plot(df['dateRep'], df['cases'], label='Daily Cases', alpha=0.7)
plt.xlabel('Date')
plt.ylabel('Cases')
plt.title('COVID-19 Daily Cases Analysis')
plt.legend()
plt.grid()

# Calculate and plot the rolling 7-day average
rolling_average = df['cases'].rolling(7).mean()
plt.plot(df['dateRep'], rolling_average, label='7-Day Rolling Avg', color='red')
plt.legend()

# Linear Regression Model
model = LinearRegression()
X = df['dateRep'].astype('int64').values.reshape(-1, 1)
y = df['cases'].values
model.fit(X, y)

# Predict cases using the linear regression model
y_pred = model.predict(X)

# Plot the linear regression line
plt.figure(figsize=(12, 6))
plt.plot(df['dateRep'], df['cases'], label='Daily Cases', alpha=0.7)
plt.plot(df['dateRep'], y_pred, label='Linear Regression', color='red')
plt.xlabel('Date')
plt.ylabel('Cases')
plt.title('COVID-19 Daily Cases Analysis with Linear Regression')
plt.legend()
plt.grid()

# Explore cumulative cases
df['cumulative_cases'] = df['cases'].cumsum()
plt.figure(figsize=(12, 6))
plt.plot(df['dateRep'], df['cumulative_cases'], label='Cumulative Cases')
plt.xlabel('Date')
```

```
plt.ylabel('Cumulative Cases')
plt.title('COVID-19 Cumulative Cases Analysis')
plt.legend()
plt.grid()

# Calculate and plot the daily increase in cases
df['daily_increase'] = df['cases'].diff()
plt.figure(figsize=(12, 6))
plt.plot(df['dateRep'], df['daily_increase'], label='Daily Increase')
plt.xlabel('Date')
plt.ylabel('Daily Increase')
plt.title('COVID-19 Daily Increase in Cases')
plt.legend()
plt.grid()

plt.show()
```

Output:







Conclusion:

This project underscores the critical role of data-driven decision-making during the ongoing COVID-19 pandemic. By analyzing data and generating insights, it provides a comprehensive understanding of the virus's trends and its impact on various factors. The design thinking approach guided the project, emphasizing the need for clear objectives and an effective analysis pipeline. Data collection, cleaning, and quality assurance were highlighted as essential steps, ensuring the reliability of the insights. Utilizing data visualization tools such as IBM Cognos proved invaluable for creating impactful visuals, aiding better decision-making.