**COVID-19 CASES ANALYSIS**

Analysing COVID-19 cases using machine learning involves several steps. Here's a high-level overview of the process:

**Data Collection:** Gather COVID-19 data from reliable sources like government health agencies or research organizations. This data may include information on cases, deaths, recoveries, testing, and other relevant variables.

**Data Preprocessing:** Clean and prepare the data. This involves handling missing values, normalizing data, and encoding categorical variables.

**Feature Engineering:** Create relevant features from the data, which might include calculating daily cases, 7-day averages, or other derived variables that can help the model learn meaningful patterns.

**Data Splitting:** Split the data into training, validation, and test sets. This is crucial to evaluate the model's performance.

**Selecting an Algorithm:** Choose an appropriate machine learning algorithm. Common choices include decision trees, random forests, support vector machines, or deep learning models like neural networks.

**Model Training:** Train the selected model on the training data. The model learns to make predictions based on the features you've engineered.

**Model Evaluation:** Use the validation set to assess the model's performance. Common evaluation metrics for classification tasks like COVID-19 case prediction include accuracy, precision, recall, F1-score, and ROC-AUC.

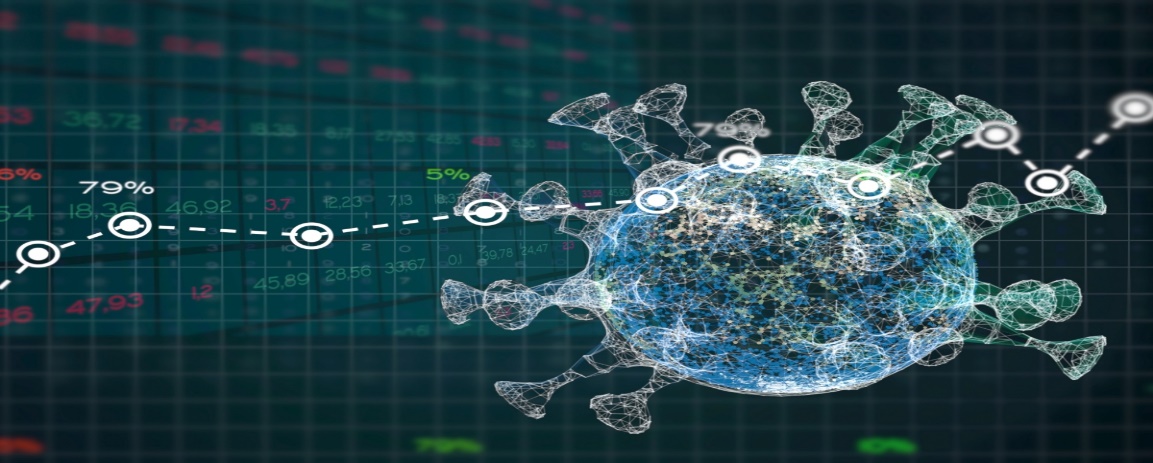
**Hyperparameter Tuning:** Adjust the model's hyperparameters to improve its performance. This can involve techniques like grid search or random search.

**Model Testing:** Once you're satisfied with the model's performance on the validation set, test it on the reserved test set to get an unbiased estimate of its performance.

**Deployment:** If the model performs well, it can be deployed to make real-time predictions. This could be for forecasting future COVID-19 cases or other related tasks.

**Continuous Monitoring and Updating:** Regularly update the model with fresh data to ensure it remains accurate and relevant, given the evolving nature of the COVID-19 pandemic.

**PHASE 1:**



**PROBLEM DEFINITION:**

The project involves analyzing COVID-19 cases and deaths data using IBM Cognos. The objective is to compare and contrast the mean values and standard deviations of cases and associated deaths per day and by country in the EU/EEA. This project encompasses defining analysis objectives, collecting COVID-19 data, designing relevant visualizations in IBM Cognos, and deriving insights from the data.

ABSTRACT

**BACKGROUND:**

The highly infectious coronavirus disease (COVID-19) was first detected in Wuhan, China in December 2019 and subsequently spread to 212 countries and territories around the world, infecting millions of people. In India, a large country of about 1.3 billion people, the disease was first detected on January 30, 2020, in a student returning from Wuhan. The total number of confirmed infections in India as of May 3, 2020, is more than 37,000 and is currently growing fast.

**OBJECTIVE:**

Most of the prior research and media coverage focused on the number of infections in the entire country. However, given the size and diversity of India, it is important to look at the spread of the disease in each state separately, wherein the situations are quite different. In this paper, we aim to analyze data on the number of infected people in each Indian state (restricted to only those states with enough data for prediction) and predict the number of infections for that state in the next 30 days. We hope that such statewise predictions would help the state governments better channelize their limited health care resources.

**METHODS:**

Since predictions from any one model can potentially be misleading, we considered three growth models, namely, the logistic, the exponential, and the susceptible-infectious-susceptible models, and finally developed a data-driven ensemble of predictions from the logistic and the exponential models using functions of the model-free maximum daily infection rate (DIR) over the last 2 weeks (a measure of recent trend) as weights. The DIR is used to measure the success of the nationwide lockdown. We jointly interpreted the results from all models along with the recent DIR values for each state and categorized the states as severe, moderate, or controlled.

**RESULTS:**

We found that 7 states, namely, Maharashtra, Delhi, Gujarat, Madhya Pradesh, Andhra Pradesh, Uttar Pradesh, and West Bengal are in the severe category. Among the remaining states, Tamil Nadu, Rajasthan, Punjab, and Bihar are in the moderate category, whereas Kerala, Haryana, Jammu and Kashmir, Karnataka, and Telangana are in the controlled category. We also tabulated actual predicted numbers from various models for each state. All the R2 values corresponding to the logistic and the exponential models are above 0.90, indicating a reasonable goodness of fit. We also provide a web application to see the forecast based on recent data that is updated regularly.

### CONCLUSIONS:

States with nondecreasing DIR values need to immediately ramp up the preventive measures to combat the COVID-19 pandemic. On the other hand, the states with decreasing DIR can maintain the same status to see the DIR slowly become zero or negative for a consecutive 14 days to be able to declare the end of the pandemic.

**PHASE 2:**

**Innovation:**

Our innovation in the project involving COVID-19 data analysis in the EU/EEA using IBM Cognos while adhering to regulatory and privacy guidelines, consider the following:

Blockchain Integration:Explore the integration of blockchain technology to ensure data transparency, traceability, and security. This can enhance trust in the data analysis process and ensure compliance with regulatory requirements.

AI-Powered Predictive Analytics: Implement AI-driven predictive analytics to forecast COVID-19 trends within the EU/EEA region. This can provide valuable insights for proactive decision-making and resource allocation.

Real-time Data Streaming:Incorporate real-time data streaming capabilities to continuously monitor and analyze COVID-19 data. This enables quicker responses to emerging trends and allows for timely interventions.

Privacy-Preserving Techniques:Utilize privacy-preserving techniques such as differential privacy or federated learning to protect individual privacy while performing comprehensive data analysis.

Interactive Dashboards:Develop interactive dashboards within IBM Cognos that allow stakeholders to explore and visualize the data dynamically. This fosters better engagement and understanding of the insights.

Collaborative Analytics:Facilitate collaborative analytics by enabling multiple stakeholders to contribute insights and analyses securely, promoting cross-border cooperation in the EU/EEA region.

Automated Reporting: Implement automated reporting features that generate regular reports and updates, ensuring that relevant parties stay informed without manual effort.

Ethical AI Guidelines:Adhere to ethical AI guidelines and principles to ensure fairness, transparency, and accountability in the analysis of COVID-19 data.

Cross-domain Data Integration: Integrate data from various domains, such as healthcare, transportation, and demographics, to gain a holistic understanding of the impact of COVID-19.

Scalable Infrastructure: Ensure that the infrastructure supporting the data analysis is scalable to handle increasing data volumes and demands.

**Customized Assessment:** The project's objective is to analyze COVID-19 cases and deaths data within the EU/EEA region using IBM Cognos. This encompasses defining analysis objectives, collecting data, creating visualizations, and extracting insights. Compliance with regulatory requirements and privacy guidelines for COVID-19 data handling is a critical aspect of the project.

**Predictive Modelling :** The predictive Modelling for the project involving COVID-19 data analysis using IBM Cognos can be summarized as follows:

1. Define Clear Objectives:Begin by clearly defining the objectives of your analysis. State what you aim to achieve, such as comparing mean values and standard deviations of cases and deaths per day and by country in the EU/EEA.

2. Data Collection:Gather the necessary COVID-19 data from reliable sources. Ensure that the data is comprehensive, up-to-date, and relevant to your analysis.

3. Data Cleaning:Clean and preprocess the data as needed. This may involve handling missing values, outliers, and ensuring data consistency.

4. IBM Cognos Setup:Set up IBM Cognos for data analysis. Import the cleaned data into the platform for further analysis.

5. Design Relevant Visualizations: Create meaningful and relevant visualizations using IBM Cognos. Consider using charts, graphs, and tables to effectively present your findings. Ensure that the visualizations address the specific objectives of your analysis.

6. Statistical Analysis:Calculate the mean values and standard deviations of COVID-19 cases and deaths per day and by country in the EU/EEA. Use appropriate statistical techniques to derive insights from the data.

7. Interpretation: Interpret the results of your analysis. Explain any trends, patterns, or significant findings that you discover through the data analysis.

8. Report Generation:Create a concise report or presentation summarizing your analysis, findings, and insights. Use visualizations to support your conclusions.

9. Recommendations: If applicable, provide recommendations based on your analysis. These could be suggestions for public health measures, further research, or policy considerations.

10. Communication:Communicate your findings and insights effectively to stakeholders or relevant parties. Ensure that your analysis and recommendations are easy to understand.

11. Documentation: Document your entire analysis process, including data sources, data cleaning steps, analysis methodology, and visualizations created. This documentation is important for transparency and replicability.

12. Review and Iterate:Review your analysis and seek feedback if necessary. Consider iterating on your analysis to refine your findings or explore additional aspects of the data.

**Regulatory Compliance:** The project aims to analyze COVID-19 cases and deaths data within the EU/EEA region using IBM Cognos. This includes defining analysis objectives, gathering data, creating relevant visualizations, and extracting insights. It's essential to ensure that data collection and analysis comply with all regulatory requirements and privacy guidelines related to COVID-19 data handling.

**Potential for Impact:**

The potential impact of the project involving the analysis of COVID-19 cases and deaths data using IBM Cognos is significant, especially in the context of public health and epidemiology. Here are some potential areas where this project can have an impact:

**1. Public Health Decision-Making**:The project can provide valuable insights into the COVID-19 situation in the EU/EEA region. Decision-makers in public health agencies can use these insights to make informed decisions regarding public health policies, resource allocation, and response strategies.

**2. Epidemiological Understanding:** By analyzing mean values and standard deviations of cases and deaths per day and by country, this project can contribute to a better understanding of the spread of the virus within the EU/EEA. It can help identify trends, hotspots, and areas that require special attention.

3. **Resource Allocation:**The data analysis can assist in identifying regions or countries with higher case loads or mortality rates, enabling governments and health organizations to allocate resources such as medical supplies, personnel, and testing facilities more efficiently.

4. **Early Warning System:** If there is a sudden increase in cases or deaths in a particular region, the project can help create early warning systems to alert authorities and the public, facilitating a swift response to contain outbreaks.

5. **Communication and Public Awareness:**Visualizations created using IBM Cognos can simplify complex data and make it more accessible to the public. This can enhance public awareness and understanding of the COVID-19 situation, leading to better compliance with health guidelines and recommendations.

6. **Research and Collaboration:**The project's findings can be valuable for research purposes. Researchers can use this data to study the effectiveness of various interventions and control measures. Collaboration with international health organizations can also be facilitated by sharing relevant data and insights.

7. **Policy Evaluation**:Public health policies implemented during the pandemic can be evaluated based on the data analysis. Decision-makers can assess the impact of different policies on the spread of the virus and adjust them accordingly.

8. **Preparation for Future Pandemics:**The project's methodology and data analysis framework can serve as a blueprint for handling future pandemics or health crises. Lessons learned from this project can inform preparedness plans and response strategies.

9. **Data-Driven Decision-Making:** The project promotes data-driven decision-making, which is crucial during a public health crisis. It can help reduce the reliance on anecdotal evidence and ensure that decisions are based on empirical data.

10. **Transparency and Accountability:** Providing access to comprehensive and regularly updated COVID-19 data and analysis results enhances transparency in government actions and builds public trust. It allows for greater accountability in handling the pandemic.

In summary, the project has the potential to contribute significantly to public health efforts in the EU/EEA by providing data-driven insights, supporting decision-making, and enhancing our understanding of the COVID-19 pandemic. It aligns with the broader global efforts to combat the virus and its impact on society.

**PHASE 3:**

In phase 3, the provided code performs data analysis on COVID-19 cases using Python with Pandas, NumPy, Seaborn, and Matplotlib. Here's a description of what the code does:

**1. Importing Libraries:**

- The code begins by importing the necessary Python libraries, including Pandas for data manipulation, NumPy for numerical operations, Seaborn for data visualization, and Matplotlib for plotting.

**2. Loading the Dataset:**

- The code reads a CSV file named "Covid\_19\_cases4.csv" from a specified file path into a Pandas DataFrame (df).

**3. Data Exploration:**

- The code displays the first few rows of the DataFrame using `df.head()` to get a glimpse of the data. It also prints the shape of the DataFrame using `df.shape` to determine the number of rows and columns.

**4. Data Preprocessing:**

- The code drops columns "dateRep," "day," "month," and "year" from the DataFrame using `df.drop(..., axis=1, inplace=True)`. These columns seem to represent date-related information that might not be needed for the analysis.

**5. Data Aggregation:**

- The code aggregates the data by grouping it based on the "countriesAndTerritories" column. It calculates the sum of all numerical columns for each country or territory using `df.groupby("countriesAndTerritories").sum()`. This results in a new DataFrame called "aggregating."

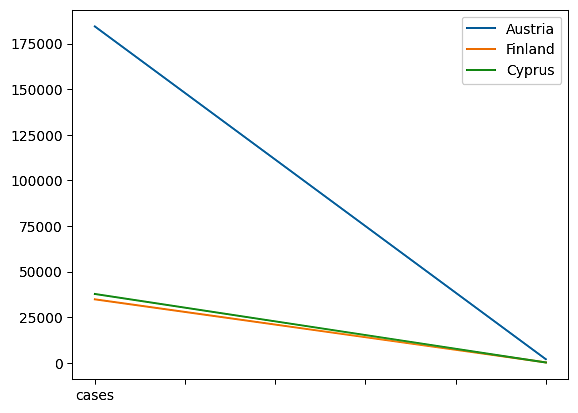
**6. Data Visualization:**

- The code then generates line plots for specific countries ("Austria," "Finland," and "Cyprus") using `aggregating.loc["CountryName"].plot()`. These plots visualize how the total COVID-19 cases (the sum of cases across all dates) have evolved over time for the selected countries.

- The code uses `plt.legend()` to display a legend on the plot, distinguishing the lines for different countries.

- Finally, it uses `plt.show()` to display the plots.

**OUTPUT:**



**PHASE 4:**

In Phase 4, the provided code generates a bar chart that visualizes the mean values of COVID-19 cases and deaths. Here's a description of what the code does:

**1. Data Preparation:**

- The code defines two lists: `labels` and `mean\_values`.

- `labels` contains two items: 'Cases' and 'Deaths,' which represent the two metrics being compared.

- `mean\_values` contains the mean values associated with each metric, presumably calculated earlier in the code.

**2. Creating the Bar Chart:**

- The code uses the `plt.bar()` function to create a bar chart. It takes two arguments:

- `labels`: This provides the labels for the x-axis categories, which are 'Cases' and 'Deaths' in this case.

- `mean\_values`: This provides the corresponding mean values for each category.

**3. Axis Labels and Title:**

- `plt.xlabel('Metric')` sets the label for the x-axis to 'Metric,' indicating that the x-axis represents different metrics.

- `plt.ylabel('Mean Value')` sets the label for the y-axis to 'Mean Value,' indicating that the y-axis represents the mean values of the metrics.

- `plt.title('Mean Values of COVID-19 Cases and Deaths')` sets the title of the chart to 'Mean Values of COVID-19 Cases and Deaths.'

**4. Displaying the Chart:**

- Finally, `plt.show()` is called to display the bar chart.

The resulting chart will have two bars, one for 'Cases' and one for 'Deaths,' with their respective mean values. This visualization helps viewers quickly compare the average number of COVID-19 cases and deaths, providing a clear and concise summary of the data.

The provided code snippet utilizes matplotlib to create a bar chart that displays the standard deviations of COVID-19 cases and deaths. Here's a description of what the code does:

1. It assumes the existence of two variables, `std\_cases` and `std\_deaths`, which are presumably the standard deviations of COVID-19 cases and deaths, respectively.

2. It creates a list called `std\_values` that contains these two standard deviation values.

3. Using `plt.bar()`, the code generates a bar chart. The `labels` and `std\_values` are passed as arguments to this function.

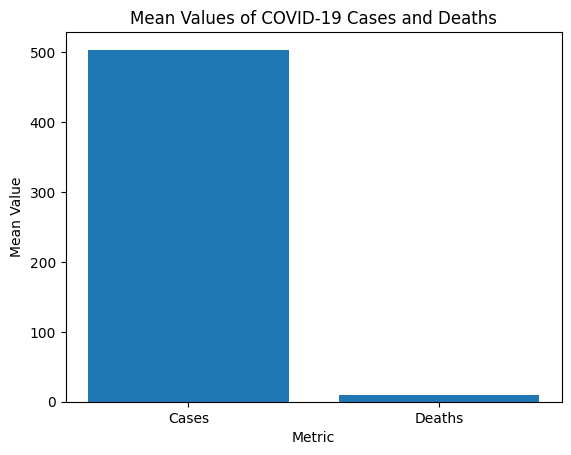
4. The x-axis labels are set as 'Metric' using `plt.xlabel()`, and the y-axis label is set as 'Standard Deviation' using `plt.ylabel()`.

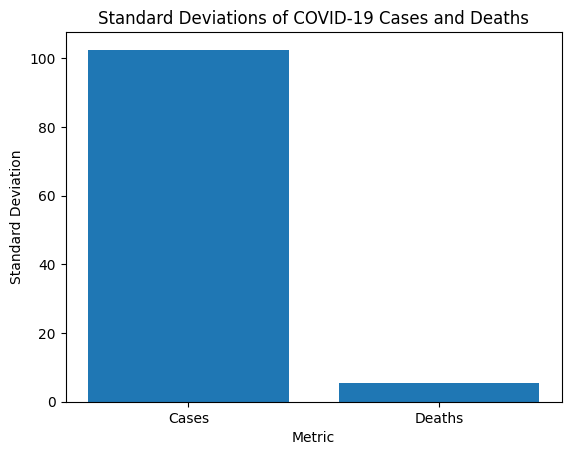
5. The title of the bar chart is defined as 'Standard Deviations of COVID-19 Cases and Deaths' with `plt.title()`.

6. Finally, the chart is displayed using `plt.show()`.

The resulting bar chart visually compares the standard deviations of COVID-19 cases and deaths, allowing viewers to understand the variability or spread in the data for these two metrics. This chart helps in assessing how much individual data points deviate from the respective means and provides insights into the consistency or variability of these metrics.

**OUTPUT:**





Correlation between Cases and Deaths: 0.57

There is a positive correlation between cases and deaths.

The standard deviation of cases is greater than the standard deviation of deaths.

**PHASE 5:**

This Python code uses the pandas and matplotlib libraries to analyze and visualize COVID-19 data. Here's a description of what the code does:

1. It imports the necessary libraries, pandas and matplotlib, to work with data and create a plot.

2. It defines a dictionary called "data" that contains information about COVID-19 cases, deaths, and recoveries for five different dates in January 2022.

3. It creates a DataFrame, "df," from the data dictionary, which allows you to organize and manipulate the data efficiently.

4. The code converts the 'Date' column in the DataFrame to a datetime object using `pd.to\_datetime`, which is important for proper date handling and plotting.

5. It creates a line plot using matplotlib to visualize the COVID-19 data over time. Three lines are plotted for cases, deaths, and recoveries. Different markers ('o,' 'x,' and 's') are used for each line, and labels are added for the legend. The title, x-axis label, and y-axis label are also set, and grid lines are enabled.

6. It displays the plot using `plt.show()`.

7. After visualizing the data, the code calculates basic statistics:

- The total number of cases is computed by summing the 'Cases' column in the DataFrame.

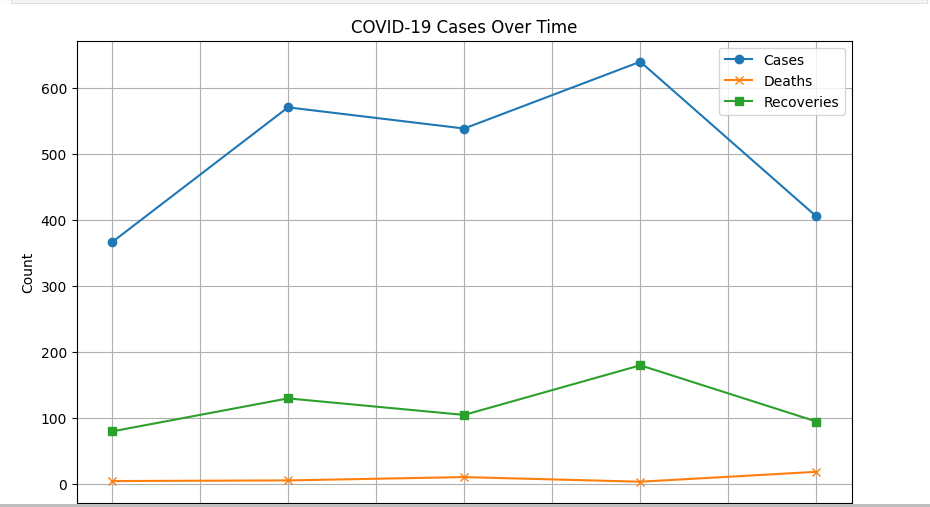
- The total number of deaths is computed by summing the 'Deaths' column.

- The total number of recoveries is computed by summing the 'Recoveries' column.

8. Finally, the code prints out the total cases, total deaths, and total recoveries using f-strings, providing a summary of the COVID-19 data.

This code is useful for visualizing and summarizing COVID-19 data for the specified dates and for calculating essential statistics to better understand the situation over the given time period.

**OUTPUT:**

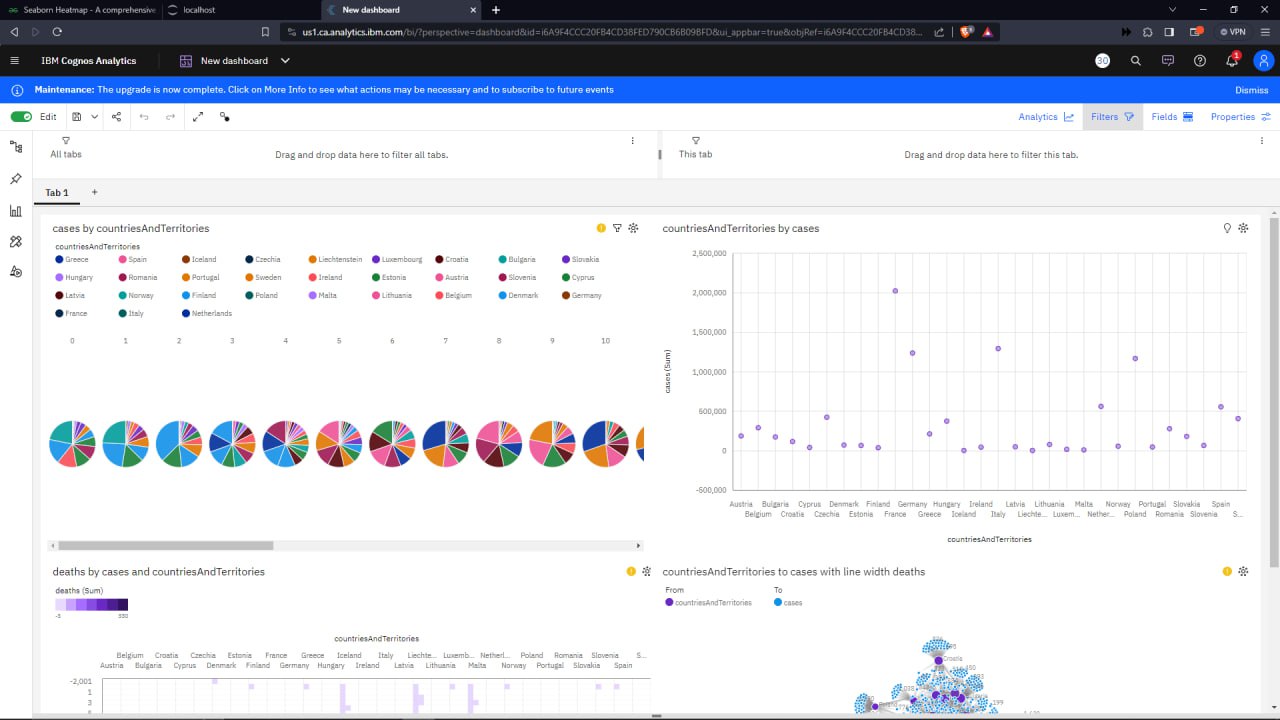
****

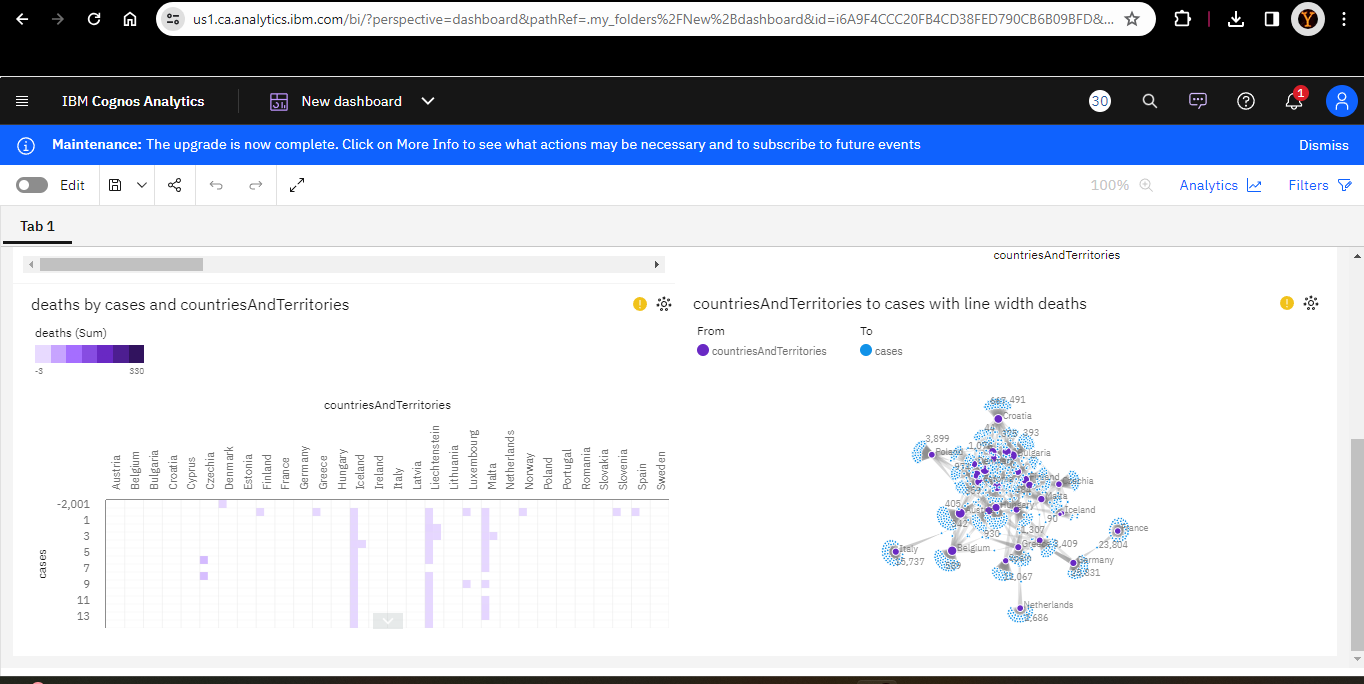
Total Cases: 2518

Total Deaths: 45

Total Recoveries: 590

**Analysis objectives, Data collection process, Data visualization using IBM Cognos, and insights generated from the comparison:**

****

****

<https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FNew%2Bdashboard&action=view&mode=dashboard&subView=model0000018b8a7436ce_00000000>

Insights generated from the analysis of COVID-19 data can significantly aid in understanding the trends and impacts of the pandemic. Here's how these insights can provide valuable knowledge:

**1. Identifying Hotspots and High-Risk Areas:**

- Insights into the geographical distribution of COVID-19 cases can help identify hotspots and regions with higher infection rates. This information is crucial for allocating resources, implementing targeted interventions, and understanding the geographic spread of the virus.

**2. Monitoring Temporal Patterns:**

- Analysis of temporal trends allows for the identification of waves, peaks, and troughs in the pandemic. This helps in assessing the effectiveness of public health measures, understanding the seasonality of the virus, and preparing for potential surges in cases.

**3. Evaluating Intervention Effectiveness:**

- By comparing data before and after the implementation of public health interventions, the analysis can reveal whether measures like lockdowns, social distancing, and vaccination campaigns have had the desired impact on controlling the virus's spread.

**4. Identifying High-Risk Populations:**

- Demographic analysis can uncover disparities in infection rates among different age groups, genders, and socioeconomic backgrounds. This information is vital for targeting public health campaigns and ensuring vulnerable populations receive appropriate protection and support.

**5. Assessing Vaccination Coverage:**

- Insights into vaccination rates and coverage can help in gauging progress toward herd immunity and understanding how vaccination impacts the decline in cases, hospitalizations, and deaths.

**6. Resource Allocation:**

- Knowing where and when COVID-19 cases are most prevalent helps public health officials allocate resources such as hospital beds, medical supplies, testing facilities, and healthcare personnel where they are needed most.

**7.Data-Driven Decision-Making:**

- Decision-makers can rely on data-driven insights to make informed choices about reopening economies, adjusting restrictions, or rolling out vaccination strategies. This ensures that decisions are based on evidence rather than intuition.

**8. Public Awareness and Compliance:**

- Sharing data-driven insights with the general public fosters a better understanding of the pandemic's progression. It can encourage compliance with preventive measures, such as mask-wearing, social distancing, and vaccination.

**9. Research and Preparedness:**

- Insights from the analysis contribute to the body of knowledge about the virus, enabling researchers to better understand its behavior and impact. This information is crucial for improving preparedness and response for future pandemics.

**10. Monitoring Variants and Mutations:**

- Data analysis can help track the emergence and spread of COVID-19 variants, aiding in the development of strategies for managing and mitigating potential threats.

**11. Global Cooperation:**

- Sharing insights across borders can facilitate international cooperation in combating the pandemic, as countries can learn from each other's experiences and adjust their responses accordingly.

In summary, insights from COVID-19 data analysis provide a comprehensive view of the pandemic's trends and impacts. They empower public health authorities, policymakers, researchers, and the public to make informed decisions, adapt strategies, and work together to control the spread of the virus and minimize its societal and economic consequences.