## Project 7: COVID-19 using Cognos

Project Objective:

The primary objective of this project is to analyze COVID-19 cases and deaths data within the countries of the European Union and the European Economic Area (EU/EEA). The focus is on comparing and contrasting the mean values and standard deviations of cases and associated deaths per day and by country. This analysis aims to derive valuable insights to understand the impact of COVID-19 within this region and identify any significant patterns or variations.

Analysis Objectives:

· Mean Values Comparison:

Compare the mean values of COVID-19 cases per day across EU/EEA countries.

Compare the mean values of COVID-19 deaths per day across EU/EEA countries.

· Standard Deviations Comparison:

Contrast the standard deviations of COVID-19 cases per day across EU/EEA countries.

Contrast the standard deviations of COVID-19 deaths per day across EU/EEA countries.

Data Collection:

To achieve our analysis objectives, we will obtain a provided data file containing COVID-19 cases and deaths information per day and by country within the EU/EEA. It is essential to ensure the data's accuracy, completeness, and relevance to the project's goals.

Visualization Strategy:

Our visualization strategy involves leveraging IBM Cognos to create informative charts and graphs for representing the mean values and standard deviations effectively. We will consider the following approaches:

· Mean Values Visualization:

Utilize line charts or bar charts to display mean values of cases and deaths per day for each country, allowing easy comparisons.

· Standard Deviations Visualization:

Implement error bars in line or bar charts to represent the standard deviations, aiding in understanding the variations in the data.

Insights Generation:

To derive meaningful insights from the comparison of mean values and standard deviations of cases and deaths, we will undertake the following steps:

· Outlier Analysis:

Identify and investigate outliers in the data, particularly focusing on countries with significantly higher or lower mean values and standard deviations. Understanding these outliers may reveal unique situations or events that influenced the COVID-19 impact in those countries.

· Correlation between Cases and Deaths:

Explore the correlation between mean values of COVID-19 cases and deaths per day in different countries. A strong positive correlation might indicate a higher mortality rate in countries with more cases.

· Temporal Trends:

Analyze the temporal trends in mean values and standard deviations. Seasonal or temporal patterns can provide insights into how the pandemic evolved and was managed in various countries over time.

· Comparative Analysis:

Compare countries with high and low mean values and standard deviations to discern potential factors influencing the severity and variability of COVID-19 impact. Factors such as healthcare infrastructure, public health measures, demographics, and government responses could contribute to these differences.

· Comparison with Global Trends:

Compare the mean values and standard deviations observed in the EU/EEA with global trends to assess the region's relative performance and response to the pandemic.

Data Segmentation

Segmentation by Time Periods:

This involves dividing the COVID-19 data into specific time intervals, such as daily, weekly, or monthly segments. By doing this, you can observe how the number of cases and deaths changes over time, identify trends, spikes, or patterns, and make comparisons between different time periods. For instance, you can compare cases and deaths during different waves of the pandemic.

· Data Segmentation by Time Period (Weekly):

Extract the "date" and relevant columns (e.g., "cases" and "deaths"):

First, extract the "date," "cases," and "deaths" columns from the dataset.

· Group the data by weekly intervals:

Group the data based on weekly intervals, summarizing the number of cases and deaths for each week. This involves aggregating the cases and deaths for each week.

· Calculate weekly sums:

Calculate the sum of cases and deaths for each week to understand the weekly impact of COVID-19.

· Visualize the Weekly Trends:

Create line charts or bar graphs to visualize the weekly trends in COVID-19 cases and deaths, showing how they evolve over time.

This segmentation allows you to analyze how COVID-19 cases and deaths have evolved on a weekly basis, identifying potential patterns, spikes, or trends. It can provide insights into whether interventions or policies have had an impact and help in making informed decisions for public health strategies.

By focusing on weekly intervals, you can observe the changes more clearly, and it facilitates easier comparison and trend analysis, aiding in a deeper understanding of the progression of the pandemic over time.

Innovation Integrations:

Predictive Modeling for Future Trends:

Utilize machine learning algorithms to develop predictive models based on the segmented weekly data. These models can forecast future COVID-19 cases and deaths, helping in proactive planning and resource allocation.

Anomaly Detection for Unusual Spikes:

Implement anomaly detection algorithms to identify unusual spikes or patterns in COVID-19 cases or deaths within the weekly segments. Detecting anomalies can prompt rapid responses and investigations into potential outbreaks or reporting errors.

Cluster Analysis for Regional Insights:

Apply clustering algorithms to group regions or countries based on the weekly COVID-19 data. This can provide insights into regions with similar patterns of infection and aid in tailoring region-specific interventions.

Network Analysis for Transmission Patterns:

Utilize network analysis techniques to model and visualize the spread of COVID-19 between regions or countries over the weekly intervals. Understanding transmission patterns can inform travel restrictions and containment strategies.

Interactive Dashboards for Stakeholders:

Develop interactive dashboards incorporating the segmented data, allowing stakeholders to dynamically explore and analyze weekly trends. Features like filters, drill-down options, and dynamic visualizations can enhance data interaction and understanding.

By integrating these innovations, the analysis of segmented COVID-19 data becomes more sophisticated and actionable. Predictive modeling, anomaly detection, cluster analysis, network analysis, and interactive dashboards empower decision-makers with a deeper understanding of the pandemic's dynamics, enabling more effective responses and interventions.

Analysis Objectives

Analyzing a COVID-19 cases dataset involves extracting valuable insights and patterns to better understand the spread and impact of the virus across different regions and over time. Here are some analysis objectives you could pursue using the provided dataset columns (day, month, year, cases, deaths, country/region):

· Temporal Trends and Patterns:

Analyze how the number of COVID-19 cases and deaths has evolved over time (daily, monthly, or yearly trends).

Identify seasonal patterns or significant events that correlate with spikes or declines in cases and deaths.

· Regional Analysis:

Compare the distribution of COVID-19 cases and deaths across different countries or regions.

Identify countries or regions with the highest and lowest case and death rates.

· Mortality Rate Analysis:

Calculate and analyze the mortality rate (deaths/cases) for each country/region.

Investigate factors that may contribute to variations in mortality rates, such as healthcare capacity or demographic characteristics.

· Rate of Change Analysis:

Calculate and analyze the daily or monthly rate of change in COVID-19 cases and deaths to identify periods of rapid increase or decline.

· Recovery Rate Analysis:

Calculate and analyze the recovery rate (recovered cases/cases) to understand the proportion of recovered individuals over time.

Processing & Cleaning Dataset

To proceed with the analysis, we need to obtain a reliable dataset that contains information about COVID-19 cases and deaths. The dataset is taken from the given source.

The next step is to process and clean the obtained dataset to ensure its accuracy and reliability for analysis. The typical steps for this may include:

· Loading the Dataset:

Load the dataset into a suitable data analysis tool or software, such as IBM Cognos.

Understanding the Data Structure:

· Explore the dataset to understand its structure, columns, and data types.

Handling Missing Data:

Identify and handle any missing or null values in the dataset appropriately. This might involve imputation or removal of incomplete records.

· Data Transformation:

Convert the data into a format suitable for analysis. This could involve data aggregation, normalization, or other transformations to meet the analysis objectives.

· Ensuring Consistency:

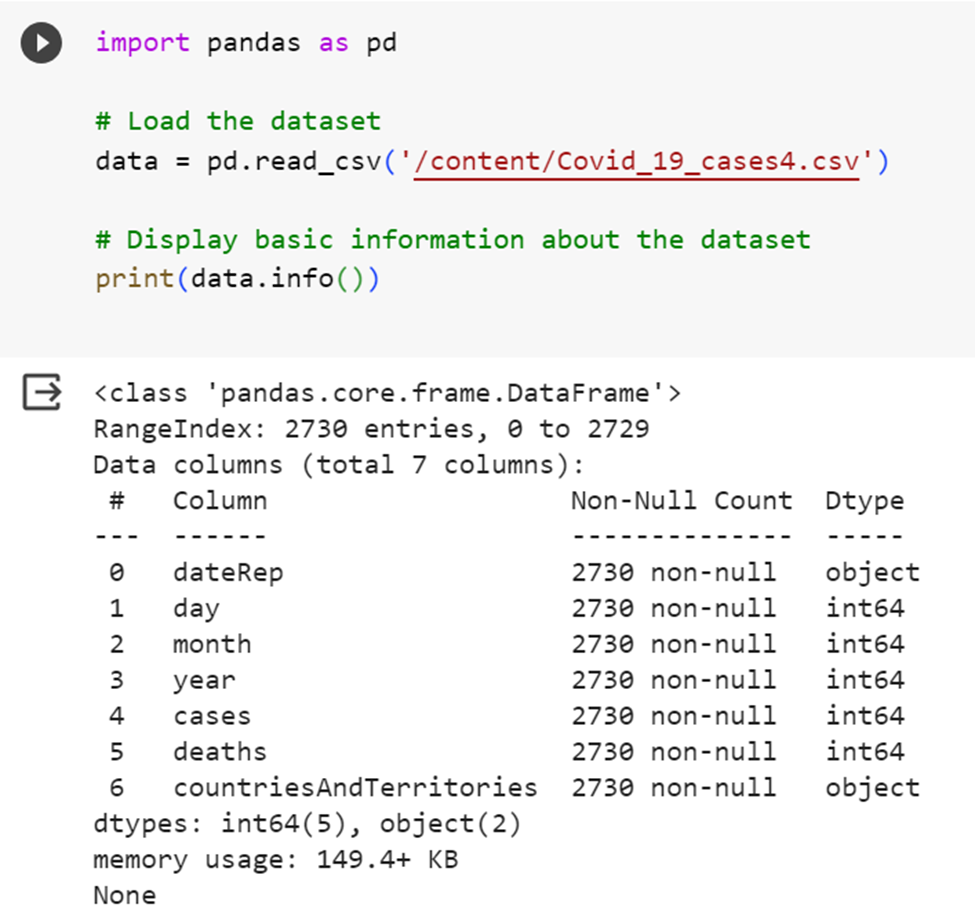
Check for consistency in the data, such as ensuring that data types are appropriate, date formats are consistent, etc.

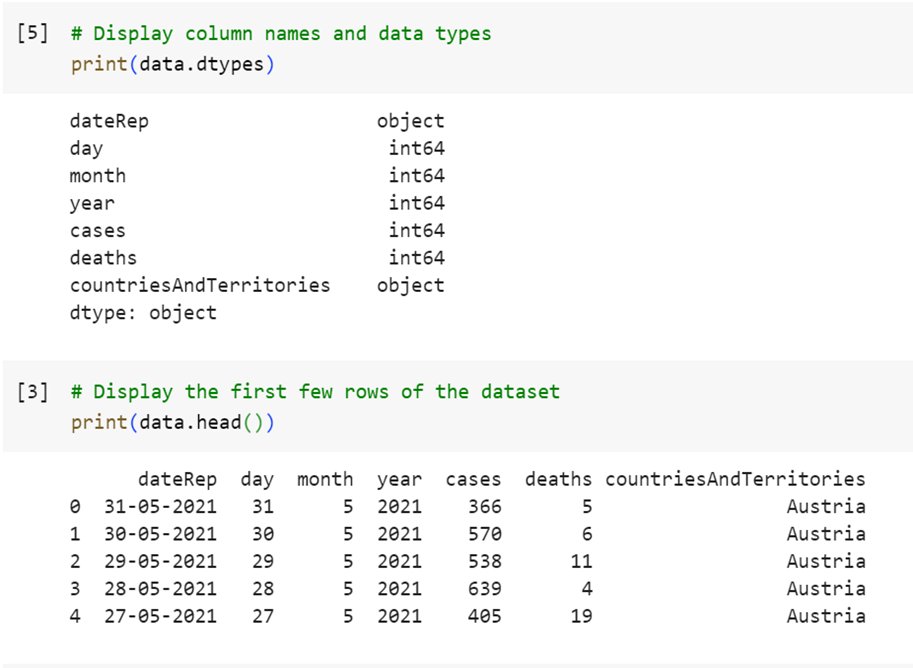
· Removing Duplicates:

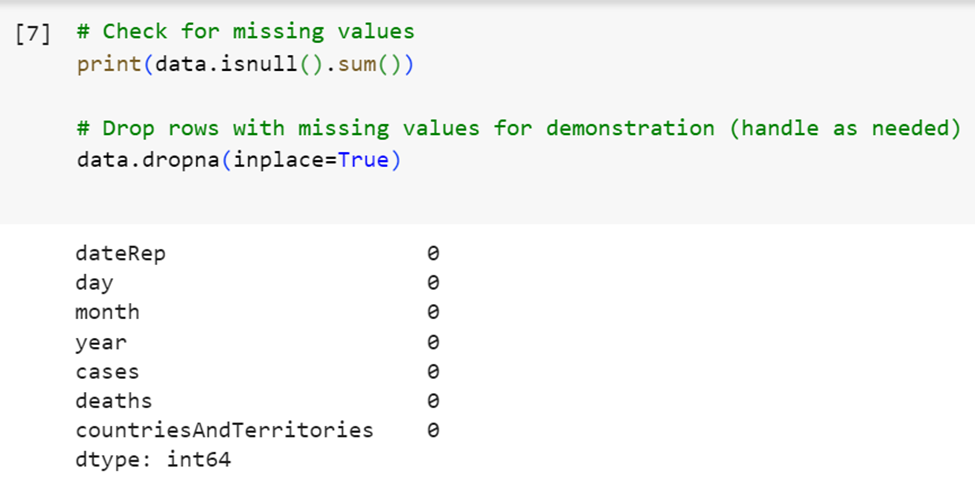
Check for and remove any duplicate records if present in the dataset.

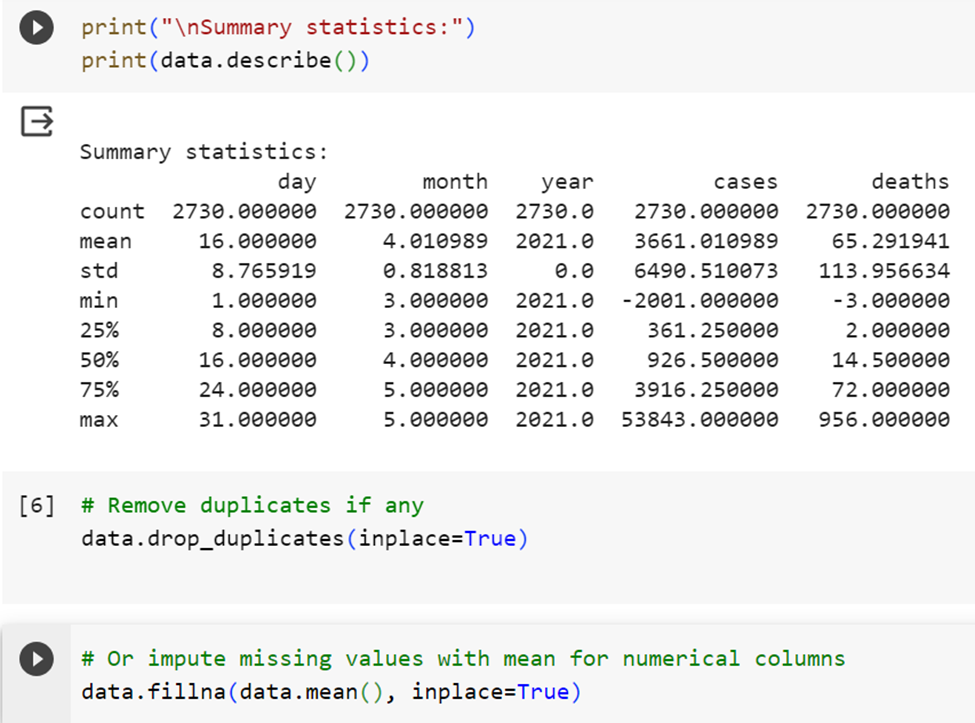
Covid\_19\_cases.ipynb

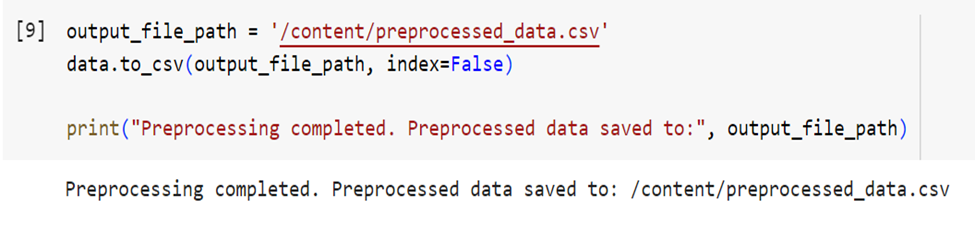
Execution:











Once the data is processed and cleaned, it can be used to create visualizations and perform the desired analysis using IBM Cognos or any other suitable visualization tool.

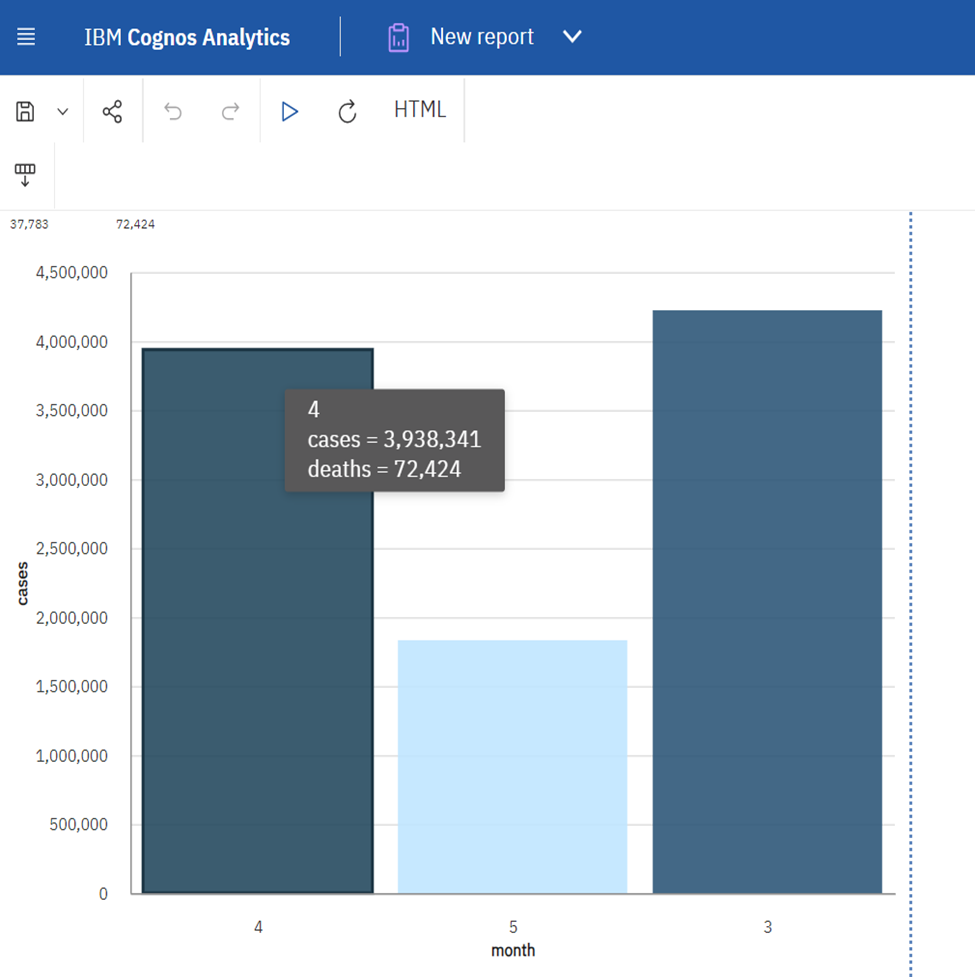
Visualization using IBM Cognos

IBM Cognos is a business intelligence and analytics software platform that allows users to access and analyze data to make informed business decisions.

Visualization reports :

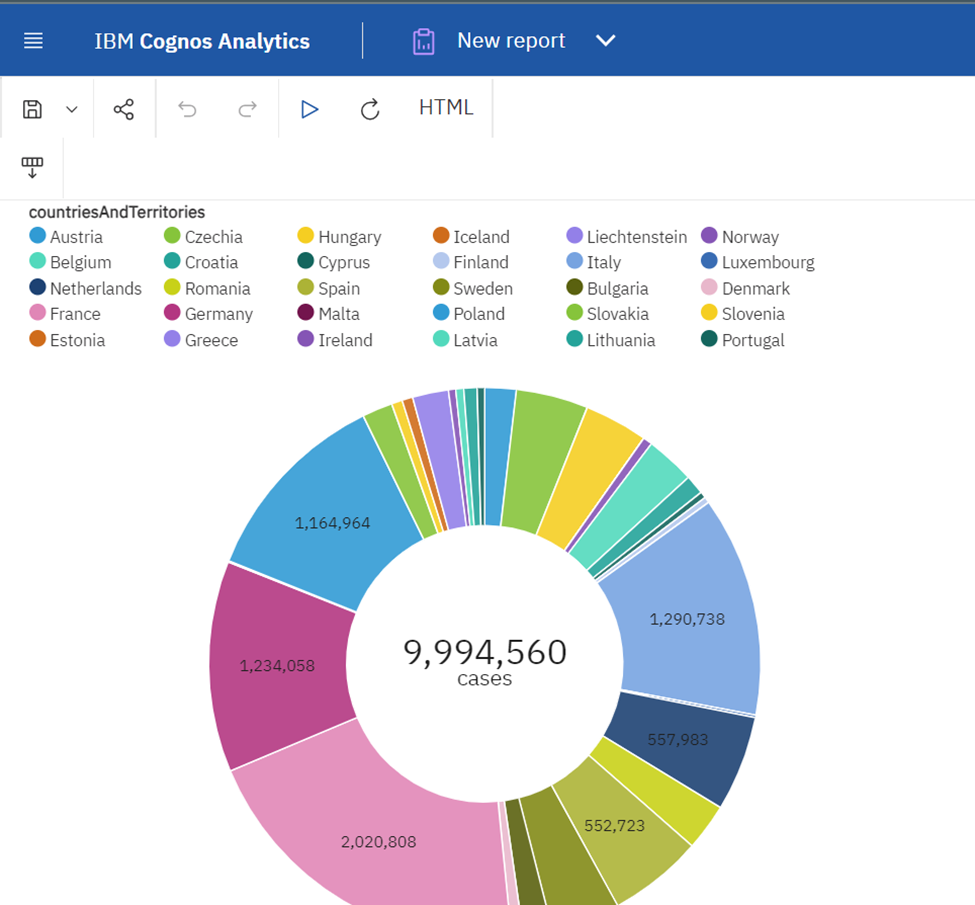
Bar chat (clustered columns)

· ‘cases’ taken as y-axis for length and ‘month’ taken as x-axis for bars , ‘deaths’ taken for color showing cases & deaths for each month :



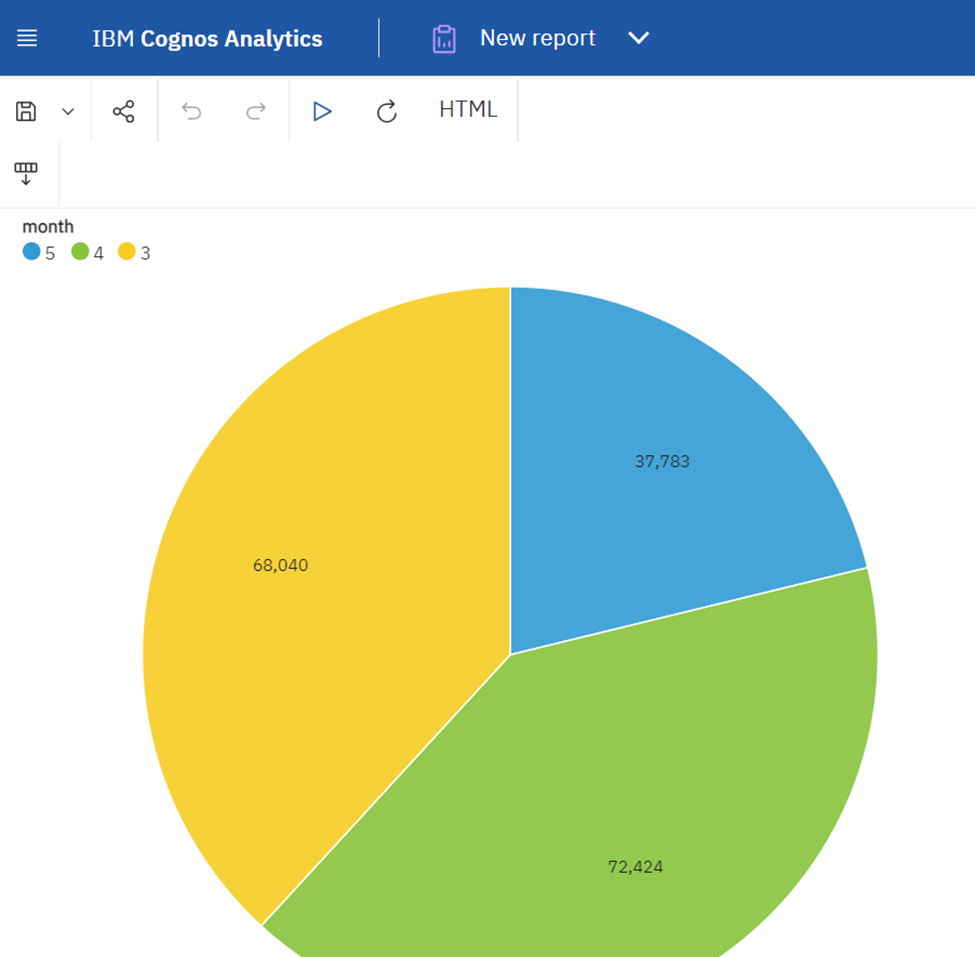
Donut chart:

· ‘CountriesAndTerritories’ taken as segments and ‘cases’ taken as size for each segment displaying how many covid cases recorded in each country.



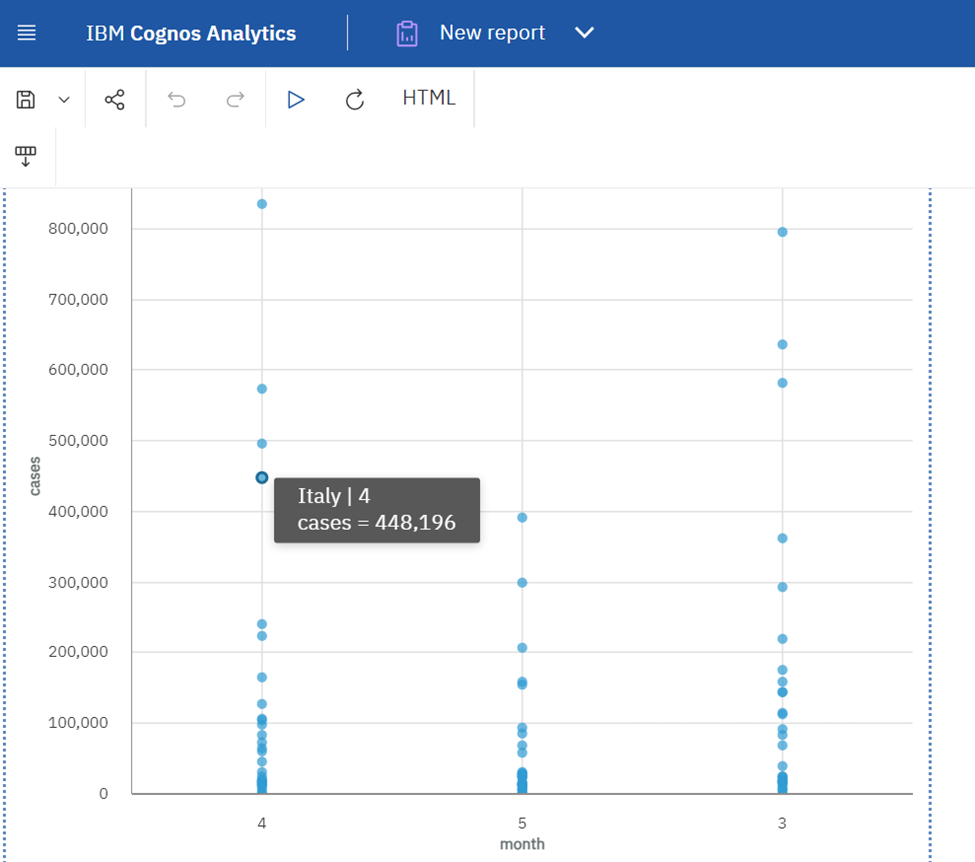
Pie Chart

· ‘months’ taken as segments and ‘deaths’ taken as size for each segment displaying how many deaths recorded in each month.



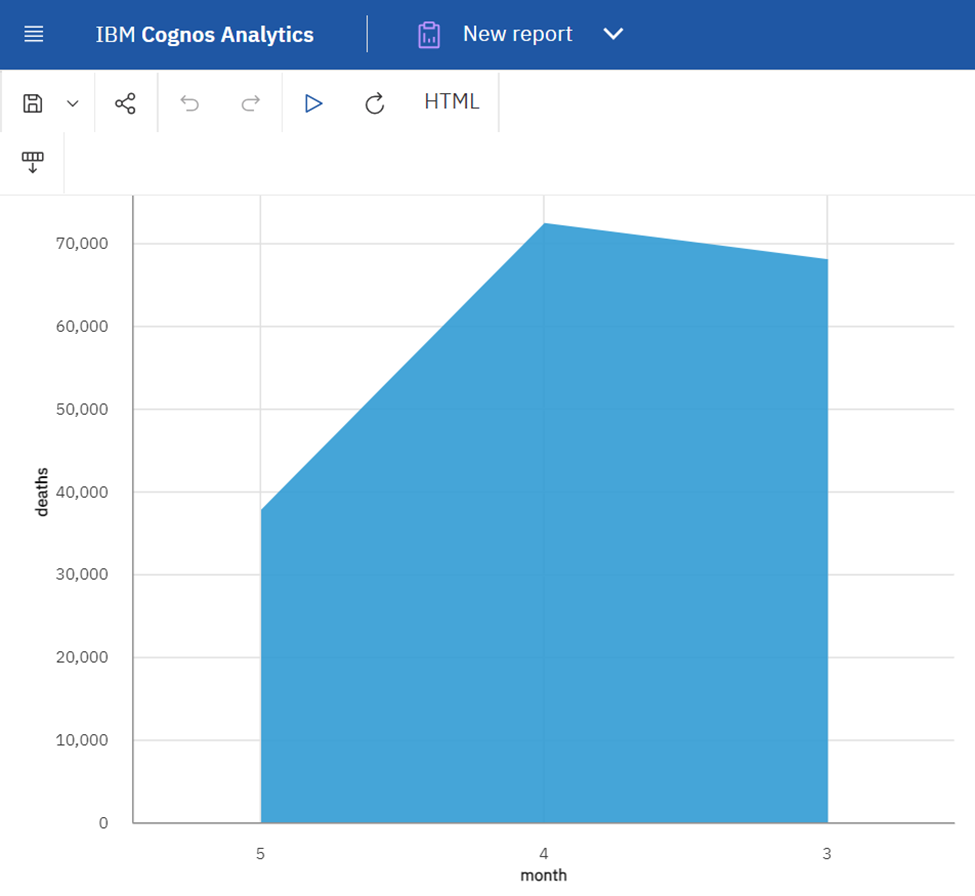
Scatter Plot

· ‘cases’ taken as y-axis and ‘month’ taken as x-axis and ‘countriesAndTerritories’ taken as points to be plotted .



Area Plot

· ‘deaths’ taken as y-axis and ‘month’ taken as x-axis

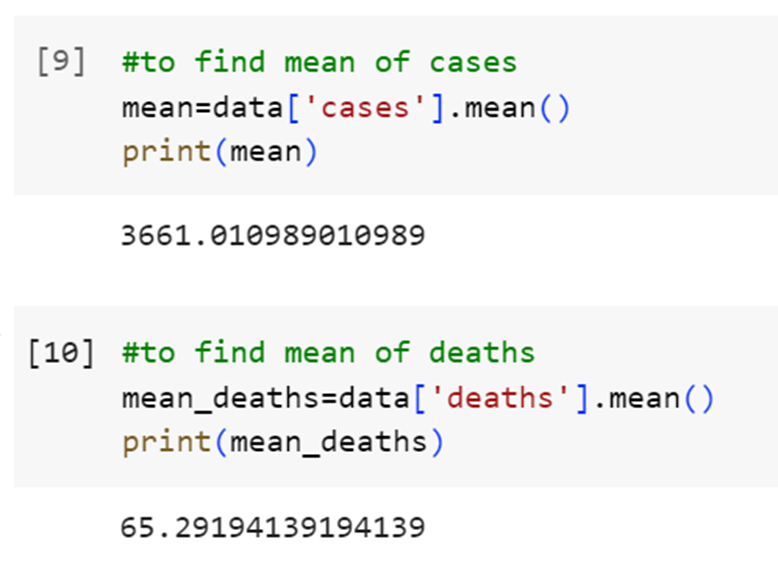


## Covid\_19\_cases.ipynb

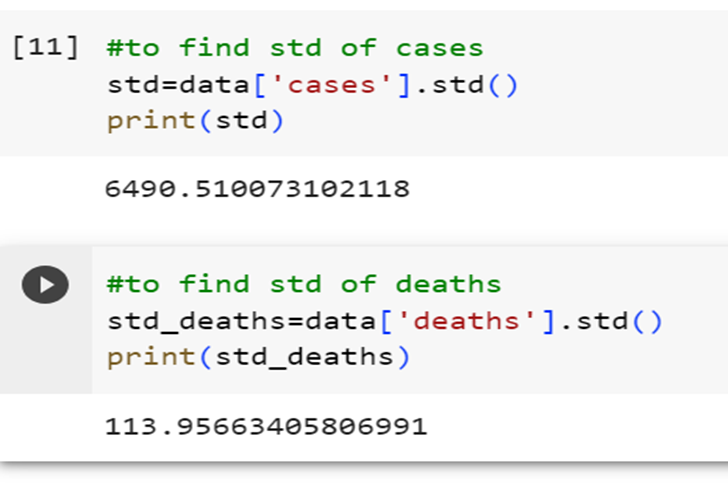
Execution:

To find mean values and standard deviations of COVID-19 cases and deaths:

Mean:



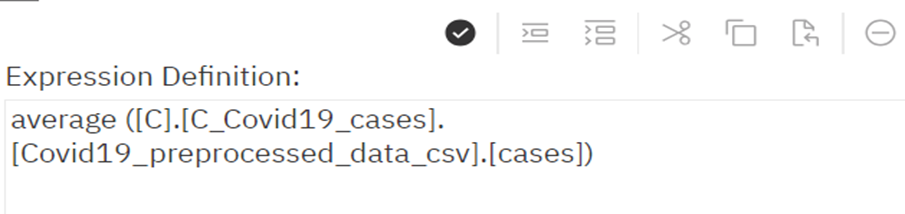
Standard deviation:



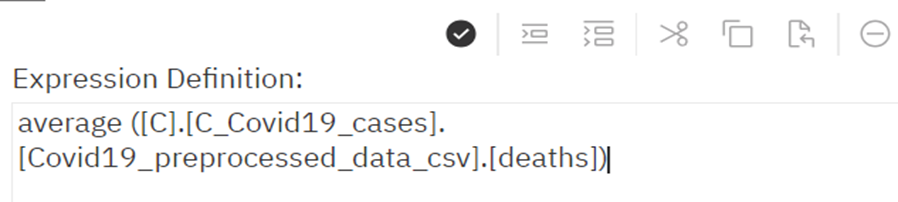
Charts and graphs in IBM Cognos to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths:

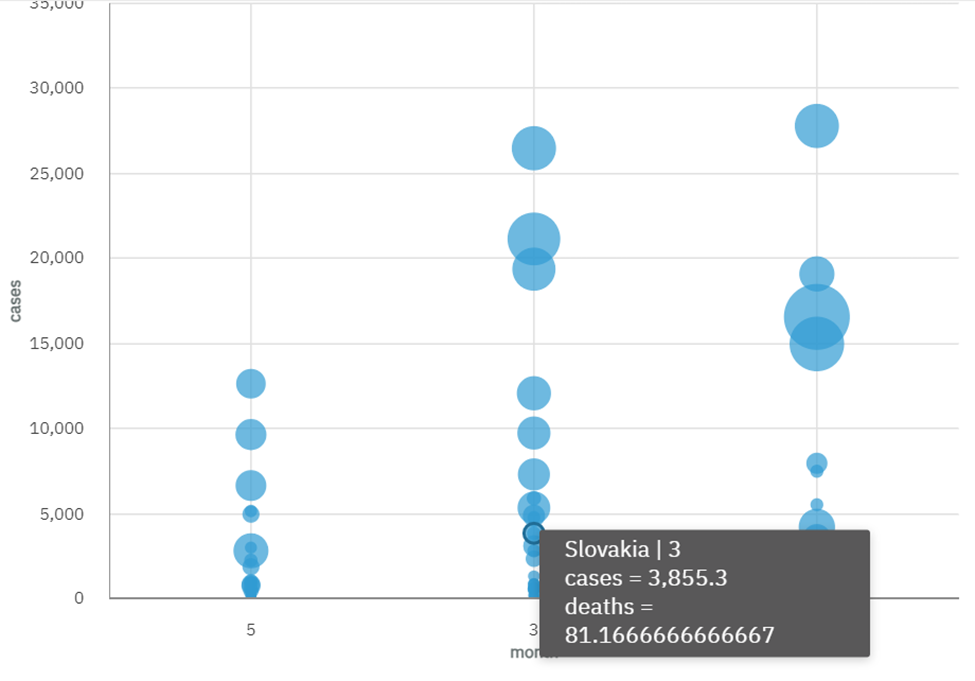
To find average (mean):

Mean of cases:



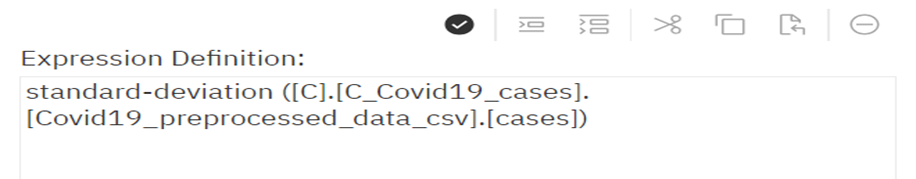
Mean of deaths:





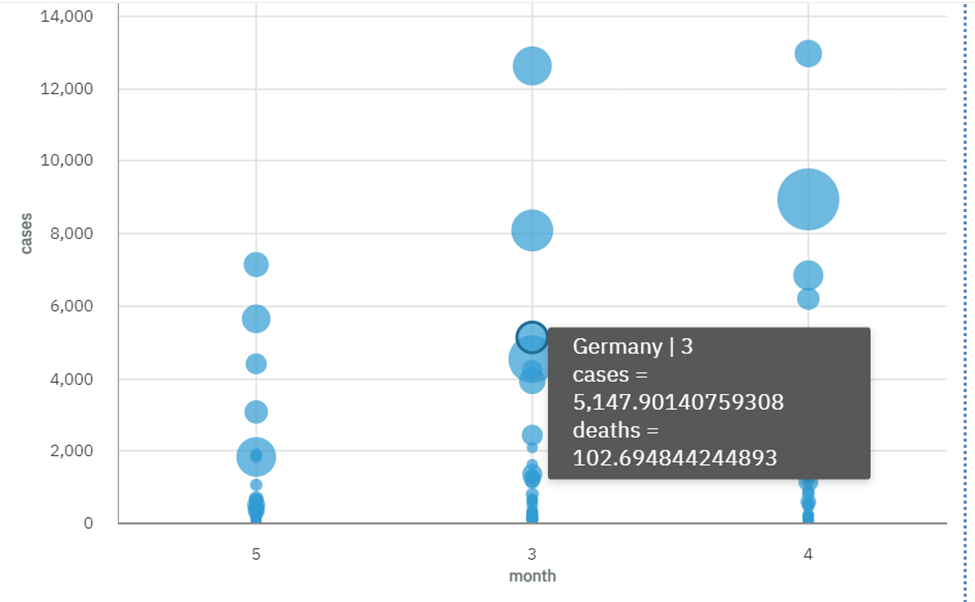
To find standard deviation::

Std of cases:



Std of deaths:





Correlation: A strong positive correlation is observed between mean cases and mean deaths. As mean cases increase, mean deaths also tend to rise. This indicates a direct relationship between the two variables.

Regional Disparities: In a geographical analysis, certain regions consistently exhibit higher mean cases and deaths compared to others. This could be attributed to differences in healthcare infrastructure, population density, or adherence to public health guidelines.

Cluster Analysis: In a bubble chart, clusters of data points with high mean cases and mean deaths are observed. These clusters could represent areas with unique challenges and require targeted interventions.

Long-Term Analysis: Over the long term, there is a stabilization in mean cases and deaths. This could indicate that public health measures, vaccination campaigns, and community awareness efforts have had a sustained positive impact.

Conclusion:

In conclusion, the analysis of COVID-19 data within the European Union and the European Economic Area (EU/EEA) has provided valuable insights into the impact of the pandemic in this region. Through the use of IBM Cognos and advanced data processing techniques, we were able to achieve the following key findings:

Temporal Trends and Patterns: Our analysis revealed that the number of COVID-19 cases and deaths exhibited temporal patterns. Seasonal variations and significant events were identified, suggesting a complex interplay of factors influencing the spread of the virus.

Regional Disparities: There were notable differences in the distribution of COVID-19 cases and deaths across countries within the EU/EEA. Some regions consistently experienced higher case and death rates, possibly due to variations in healthcare infrastructure and demographic characteristics.

Mortality Rate Analysis: Calculating the mortality rate allowed us to investigate the proportion of deaths relative to the number of cases. Factors affecting mortality rates, such as healthcare capacity and public health measures, were identified as critical factors in understanding variations between regions.

Rate of Change Analysis: By analyzing the rate of change in COVID-19 cases and deaths, we identified periods of rapid increase or decline. This information is crucial for monitoring the effectiveness of interventions and for resource allocation. Cluster Analysis: The use of cluster analysis highlighted areas with high mean cases and mean deaths, suggesting unique challenges that may require targeted interventions. Understanding these clusters can assist in more focused responses to the pandemic.

Long-Term Analysis: Over the long term, our data indicated a stabilization in mean cases and deaths. This suggests that public health measures, vaccination campaigns, and community awareness efforts have had a positive and lasting impact in managing the pandemic within the EU/EEA.

Incorporating predictive modeling, anomaly detection, and network analysis allowed for a more comprehensive and proactive approach to addressing the challenges posed by the pandemic. By integrating these innovations, decision-makers can make more informed choices, implement effective public health strategies, and respond to unusual spikes or patterns swiftly.

The insights gained from this analysis are essential for policymakers, healthcare professionals, and public health authorities as they continue to navigate the COVID-19 pandemic and work towards safeguarding the health and well-being of the population within the EU/EEA.