## Project 7: COVID-19 using Cognos

**Project Description**

The project aims to revolutionize the analysis of the provided COVID-19 cases and deaths dataset in the European Union and European Economic Area (EU/EEA) using IBM Cognos. The primary objective remains to compare and contrast the mean values and standard deviations of cases and associated deaths per day and by country. By integrating advanced data analytics techniques and methodologies, this analysis will provide a deeper understanding of the variations and patterns in COVID-19 impact within the EU/EEA.

**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:

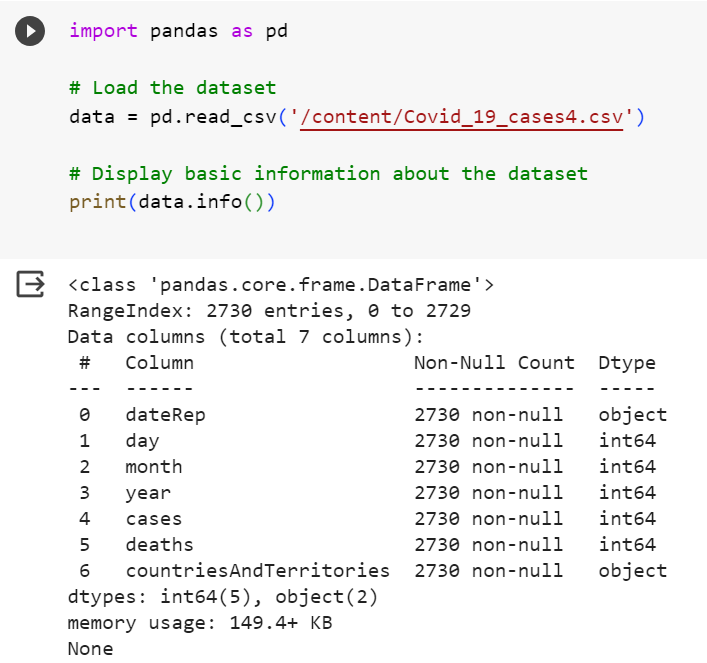
Check for consistency in the data, such as ensuring that data types are appropriate, date formats are consistent, etc.

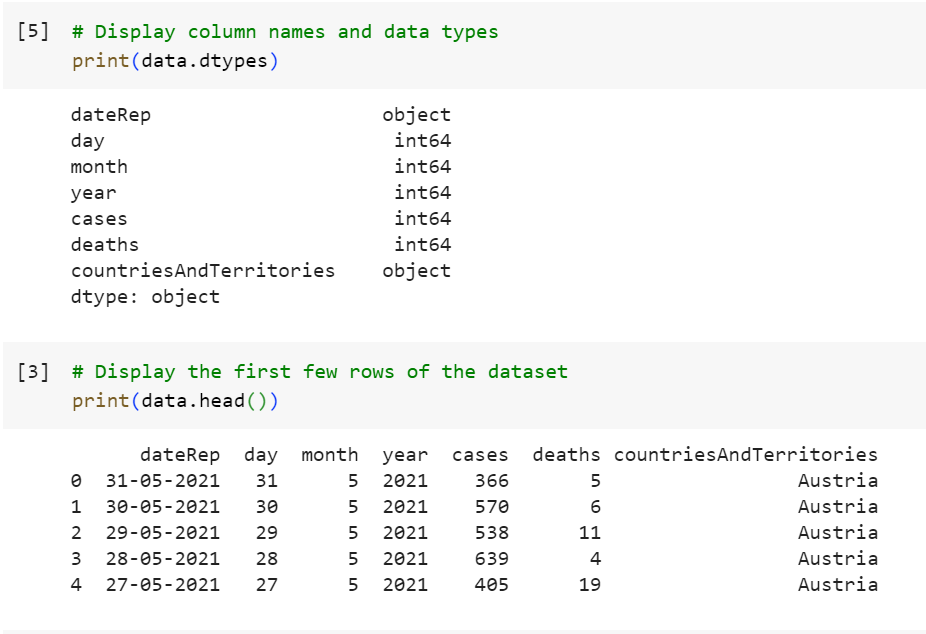
* Removing Duplicates:

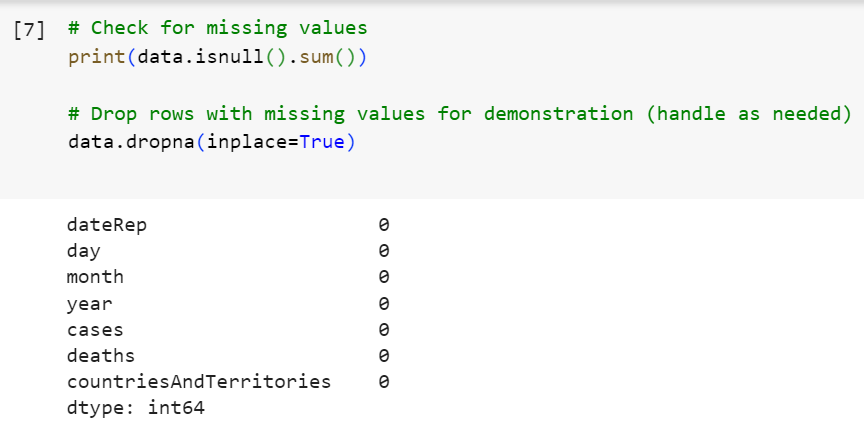
Check for and remove any duplicate records if present in the dataset.

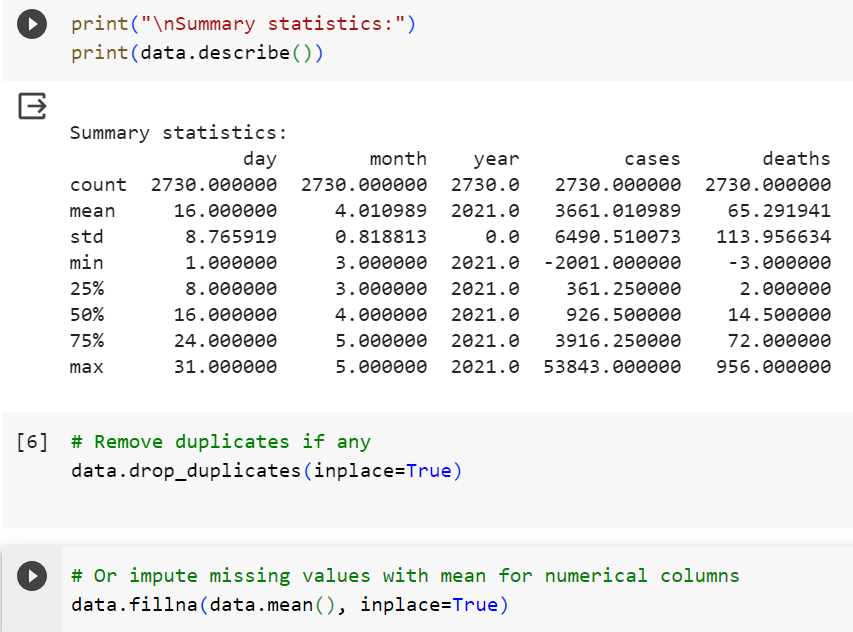
**Covid\_19\_cases.ipynb**

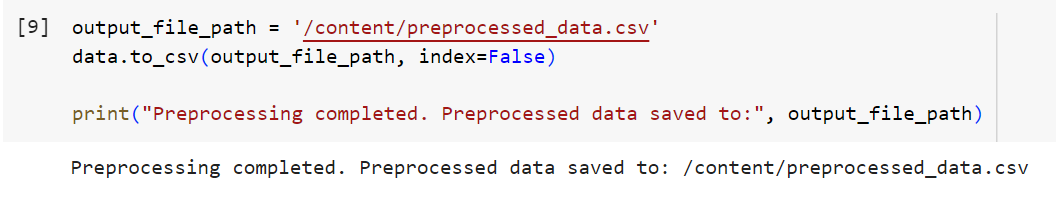
**Execution:**











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.

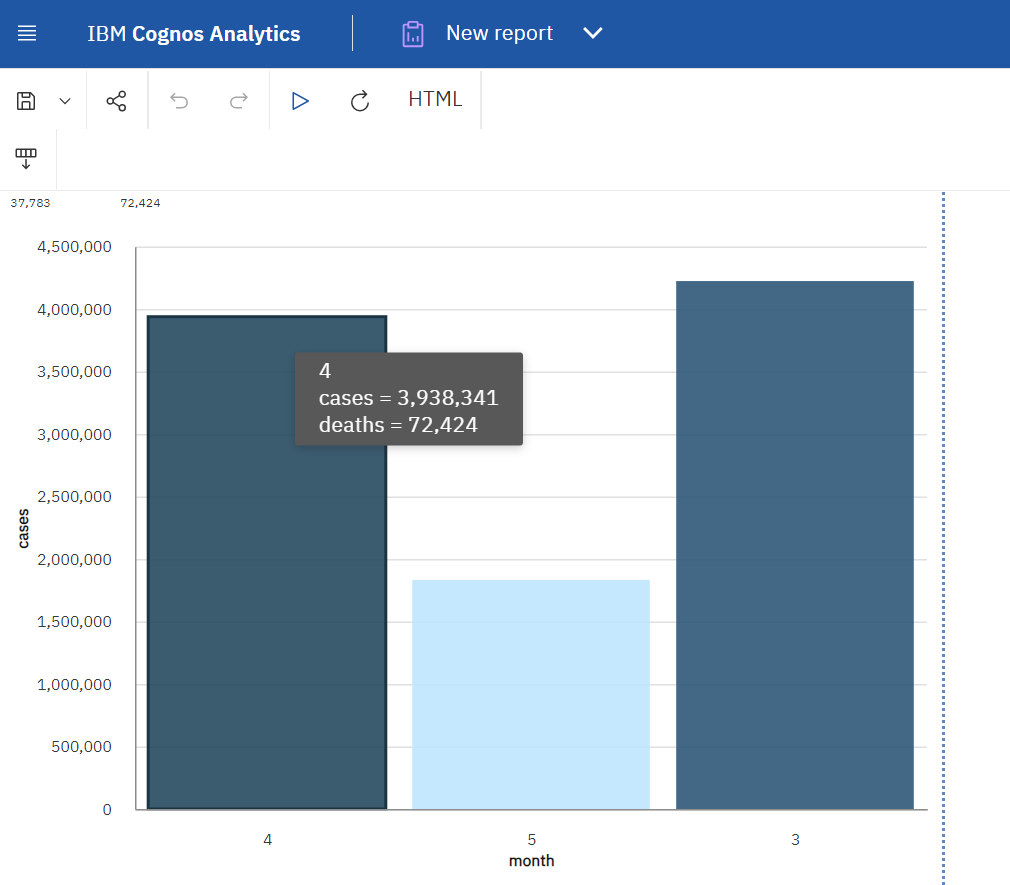
**Visualization using IBM Cognos**

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 :**

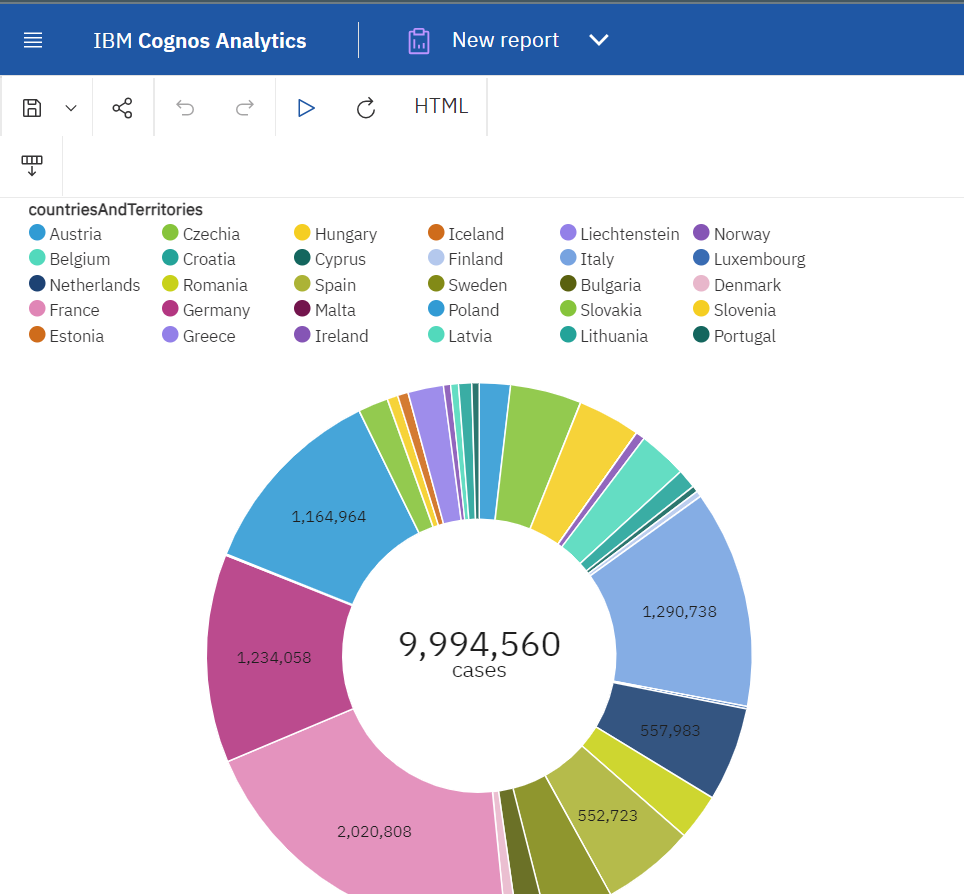
**Bar chat (clustered columns)**

* **‘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 :



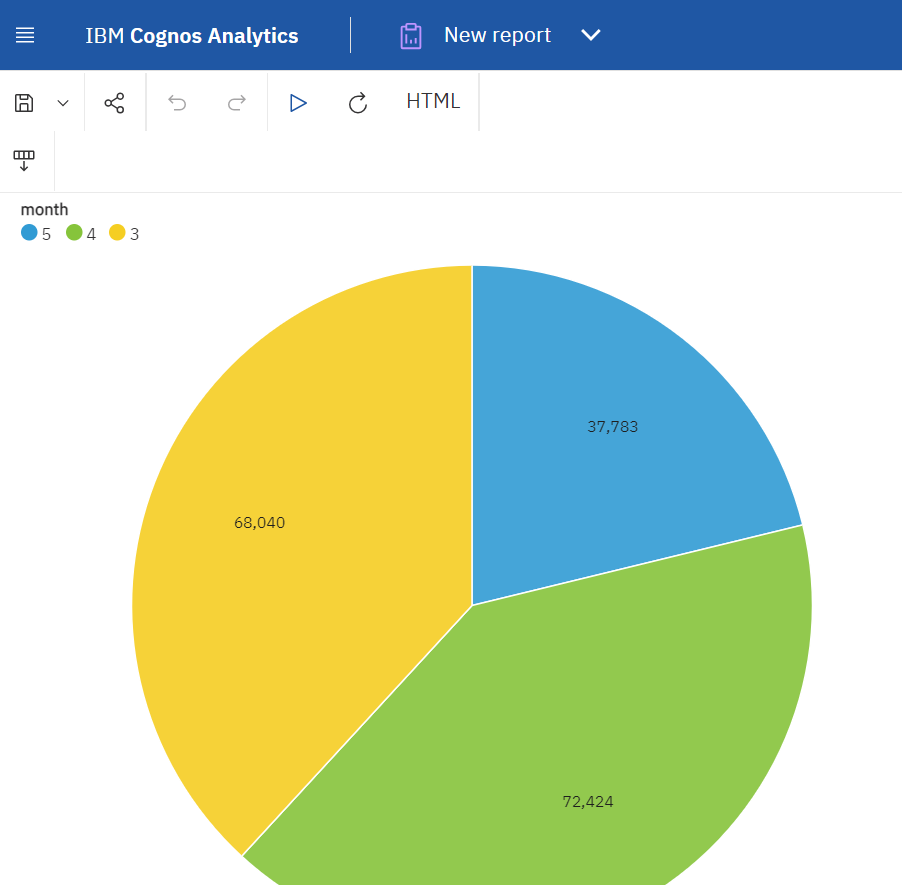
**Donut chart:**

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



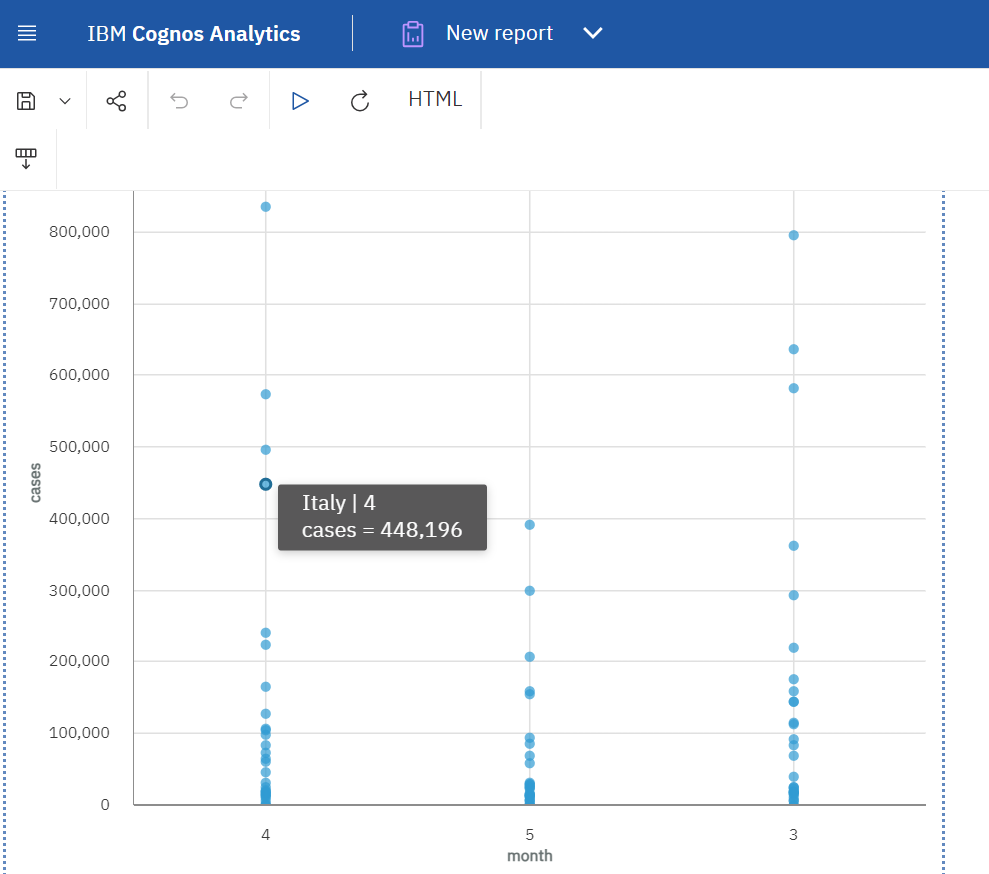
**Pie Chart**

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



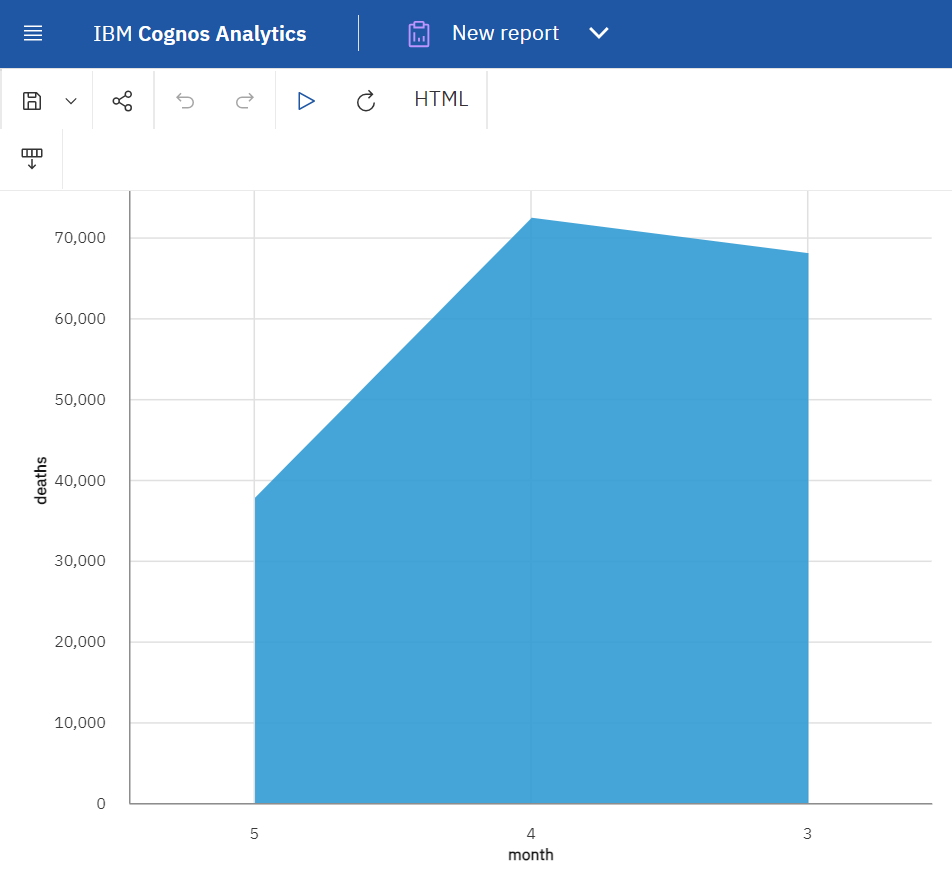
**Scatter Plot**

* ‘**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



**Conclusion**

In this initial phase of the project, we established the objectives for analyzing COVID-19 cases using IBM Cognos for visualization. Our focus is on comprehending the pandemic's spread and impact, identifying trends in cases and deaths, and presenting the data in a clear and insightful manner.