# COVID 19 Cases Analysis Using Data Analytics Tool

| | |
|---|---|
| **DATE** | **31-10-2023** |
| **TEAM ID** | **720** |
| **PROJECT NAME** | **COVID 19 CASE ANALYSIS** |

## TABLE OF CONTENTS



## 1. Introduction

The COVID-19 pandemic has had a profound impact on societies, healthcare systems, and economies worldwide. As the pandemic continues to evolve, it is crucial to conduct a comprehensive case analysis to better understand, manage, and mitigate its effects. This problem statement outlines the key aspects of the analysis required to address the ongoing challenges posed by COVID-19

The task at hand is to design and implement a comprehensive COVID-19 case analysis system for tracking, visualizing, and deriving insights from pandemic data

## 2. Problem Statement:

Designing a project to analyze COVID-19 cases and deaths using IBM Cognos, The objective is to compare and contrast the mean and standard deviation of cases and deaths, which is a valuable undertaking. This project will involve data analysis, visualization, and deriving insights from the data.

## 3. Project Objectives:

- **Data Collection**: Gather covid19 data from reliable sources across the globe. This data will include parameters such as cases, deaths, countries/Territories, and other relevant variables.
- **Data Analysis:** Perform exploratory data analysis (EDA) to understand the distribution of covid cases, detect outliers and identify trends and patterns.
- **Visualization:** Utilize data visualization techniques to represent covid19 data geospatially and temporally. This will help identify hotspots areas and understand trends over time.
- **Identification of Highly-Affected Areas:** Determine areas with consistently high case levels and investigate the factors contributing to this covid trend.
- **Predictive Model:** Develop a predictive model, likely using machine learning techniques, to estimate case and death rate This model will be valuable for forecasting covid19 spreads, fatalities and identifying hotspot areas that require immediate attention.

## 4. Steps Involved in Model Evaluation:

### 4.1. Data Collection:

First, ensure you have access to COVID 19 case data, Gather data from reliable sources such as government health agencies, the World Health Organization (WHO), and reputable research institutions. Collect data on the number of cases, deaths, recoveries, vaccination rates, and other relevant variables.

### 4.2. import Libraries:

Start by importing the necessary libraries such as numpy, pandas for data manipulations, matplotlib and seaborn for visualisations, etc…

IMPORT LIBRARIES

```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import plotly as py
import plotly.graph_objects as go
import plotly.io as pio
import plotly.express as px
```

### 4.3. Load the Dataset:

- This step involves loading your COVID 19 dataset into your Python environment. The dataset should be in a format that Pandas can easily handle, such as a CSV file.

LOADING DATASET

```
covid_data=pd.read_csv('covid_19_cases4 (1).csv')
```

- The read.csv () function is used to load a CSV (Comma-Separated Values) file into a Pandas DataFrame. You specify the file path within the parentheses.
- The result of this operation is a DataFrame, which is a tabular data structure that's similar to a spreadsheet. It allows you to work with your data in a structured and flexible way.

## 4.4. Explore the Dataset

Before diving into data preprocessing, it's important to understand your dataset. You can use various Pandas functions to explore it:

### data.head ():

- This function displays the first few rows of your dataset, giving you a glimpse of its structure.

```
covid_data.head()
```

### Data.describe():

- It provides basic statistical information about your data, including measures like mean, standard deviation, and quartiles for numerical columns.

```
covid_data.describe()
```

### data.columns():

- This helps you see the names of all the columns in your dataset

```
covid_data.columns
```

### Data.info():

- This method prints information about a DataFrame including the index dtype and columns, non-null values and memory usage.

```
covid_data.info()
```

## 4.5. Data Pre-proccessing:

- Data preprocessing is crucial for ensuring the quality and usability of your data:

### Handle Missing Values:

- Check for missing values in your dataset and decide on an appropriate strategy to handle them. You can fill missing values using methods like forward-fill, backward-fill, mean, median, or simply remove rows with missing values.

```
covid_data.isnull().sum()
```

```
covid_data.isnull().any()
```

### Data Transformation:

- If your dataset contains date or time columns, convert them to the datetime data type for time-based analysis.

```
"# Example: Convert a date column to datetime
data['Date'] = pd.to_datetime (data['Date'])"
```

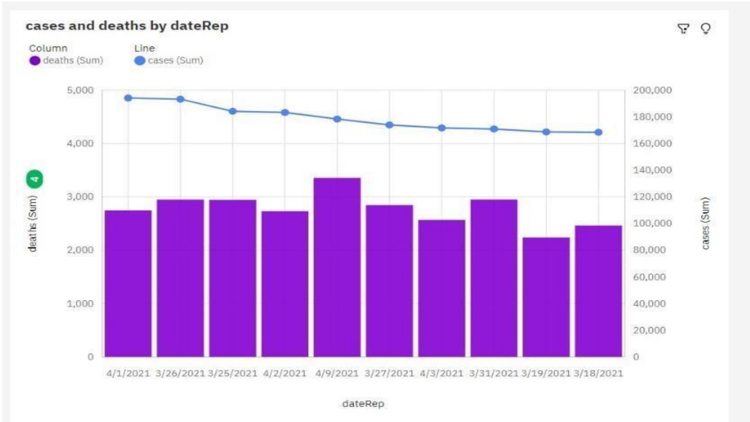
### Data Cleaning:

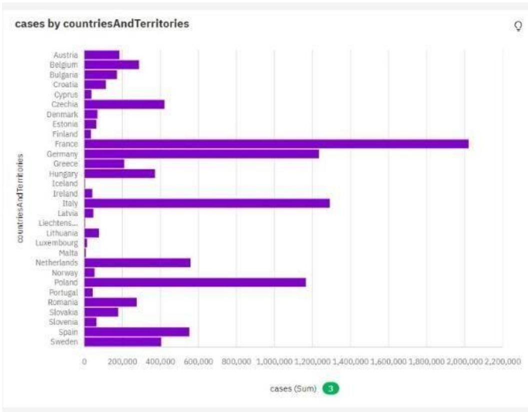
- Inspect your data for inconsistencies, outliers, or irregularities. Ensure that the data is clean and standardized. This may include dealing with irregular units, correcting typos, or removing duplicates.

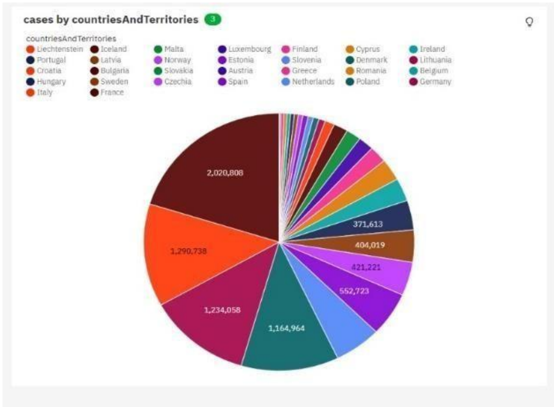
## 4.6. Predictive Model training:

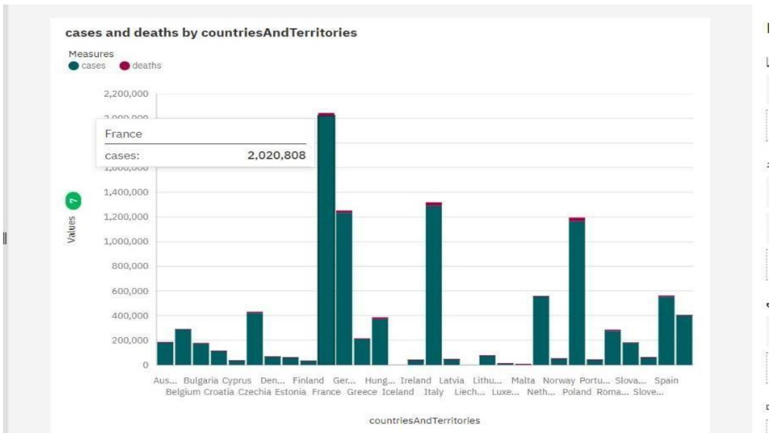
- Choose Support Vector Machine (SVM) for regression and classification tasks, handling complex data relationships.
- Split the data into training and testing sets.
- Train the model using preprocessed dataset and target variables.
- Train the model on the training data and evaluate its performance on the test data using relevant metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

# 5. Visualization using IBM Cognos:



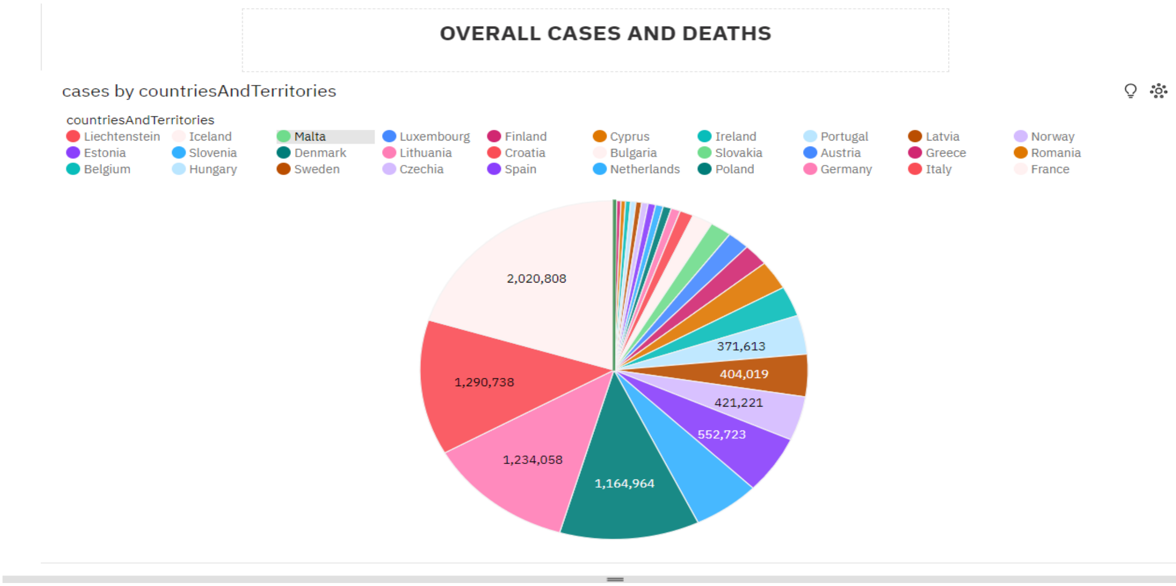



# Visualisation using IBM Cognos and insights:

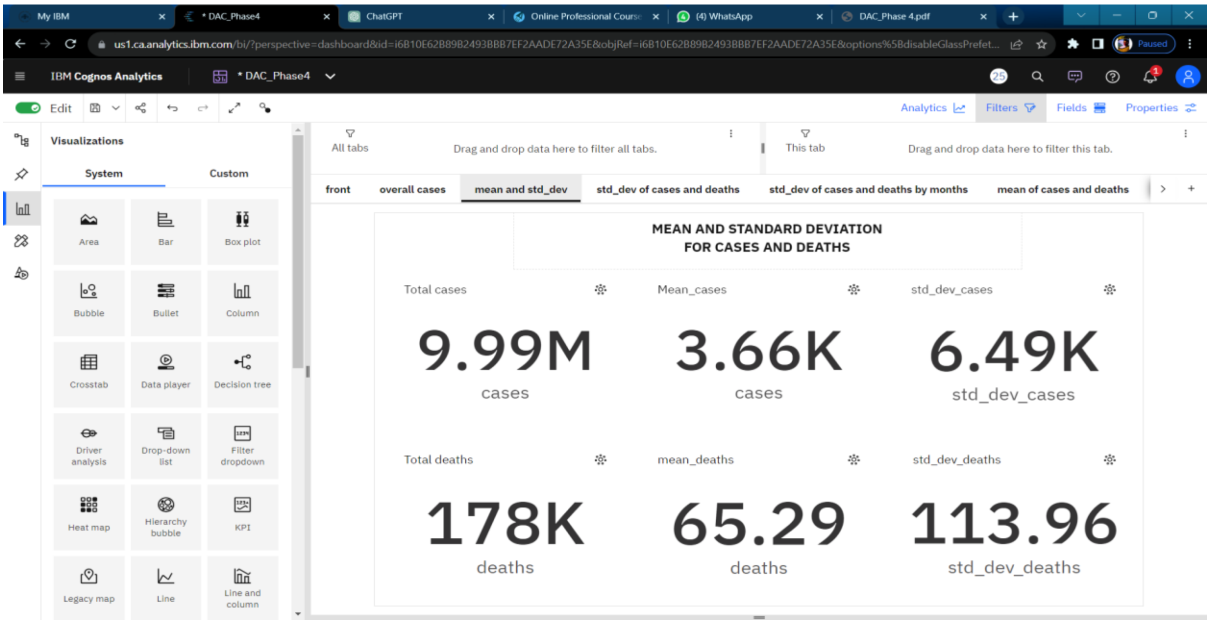
### 5.1.Home


### 5.2.Visualisation for overall cases and analyses:




In this pie-chart, various countries and territories, the total number of cases of COVID-19 is significantly high, with France having the highest number of cases, exceeding 2.0 million, while Liechtenstein has the lowest, with 437 cases. Notably, on 2021-03-29 to 2021-03-30, France experienced a staggering 937% increase in cases, indicating a significant surge in infections within a short timeframe. It is projected that by 2021-06-19, France will surpass Germany in cases by more than 14 thousand, highlighting the severity of the situation in France. Overall, the cumulative cases across all countries and territories are nearing 10.0 million, underscoring the global impact of the pandemic.

### 5.3. Visualisation for total Mean and Standard Deviation to analysis cases and deaths:




**For Cases:**

For cases, there is a noticeable strong weekly trend, with the highest values typically occurring on Thursdays and the lowest on Mondays. Additionally, there is a moderate downward trend in the number of cases. Notably, on 2021-04-06 and 2021-04-07, there were unusual spikes in cases, with a 69% increase in just one day. The lowest average cases were reported on 2021-05-25 at 953.87 and 2021-05-26 at 989.0, while the highest average cases were observed on 2021-04-01 at 6467.87 and 2021-03-26 at 6438.93. According to current forecasts, cases are expected to reach almost 1500 by 2021-06-19, and the dataset contains over 2500 results for cases.

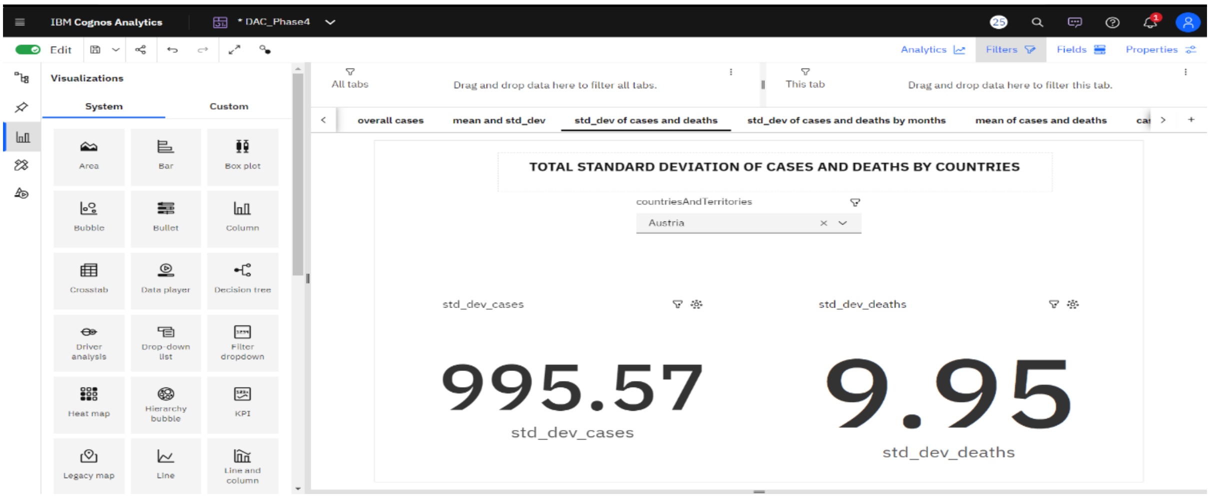
**For Deaths:**

As for deaths, there is also a strong weekly trend, with the highest values tending to occur on Wednesdays and the lowest on Mondays. However, there is a weak downward trend in the number of deaths. On 2021-04-08, an unusually high value was reported. The lowest average deaths were recorded on 2021-05-31 at 11.87 and 2021-05-30 at 18.07, while the highest average deaths were observed on 2021-04-09 at 111.83 and 2021-04-08 at 109.77. Current forecasts suggest that deaths may reach 13.27 by 2021-06-19, and the dataset contains over 2500 results for deaths.

**For Standard Deviation:**

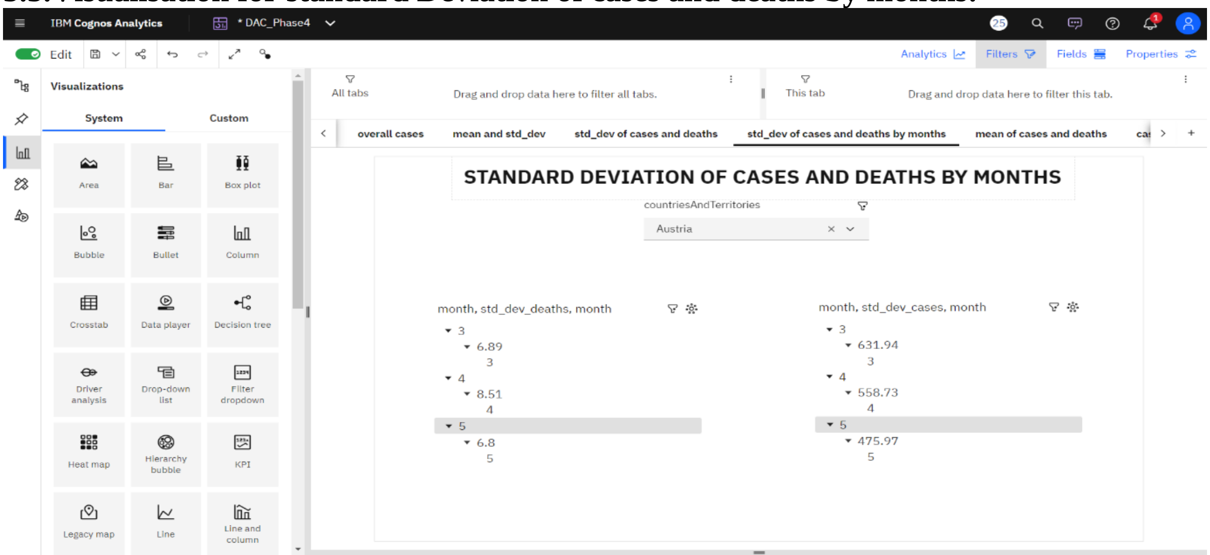
In the context of standard deviation, the number of cases has a deviation of 6.49k from the mean value, which is 3.66k. This indicates a relatively high variability in the number of cases. Similarly, the number of deaths shows a deviation of 113.96 from the mean of 65.29, suggesting a significant spread in the data. These values suggest that there is considerable variation in both cases and deaths, which may have implications for analyzing and managing the situation they represent.

**5.4. Visualisation for Standard Deviation of cases and deaths by months:**



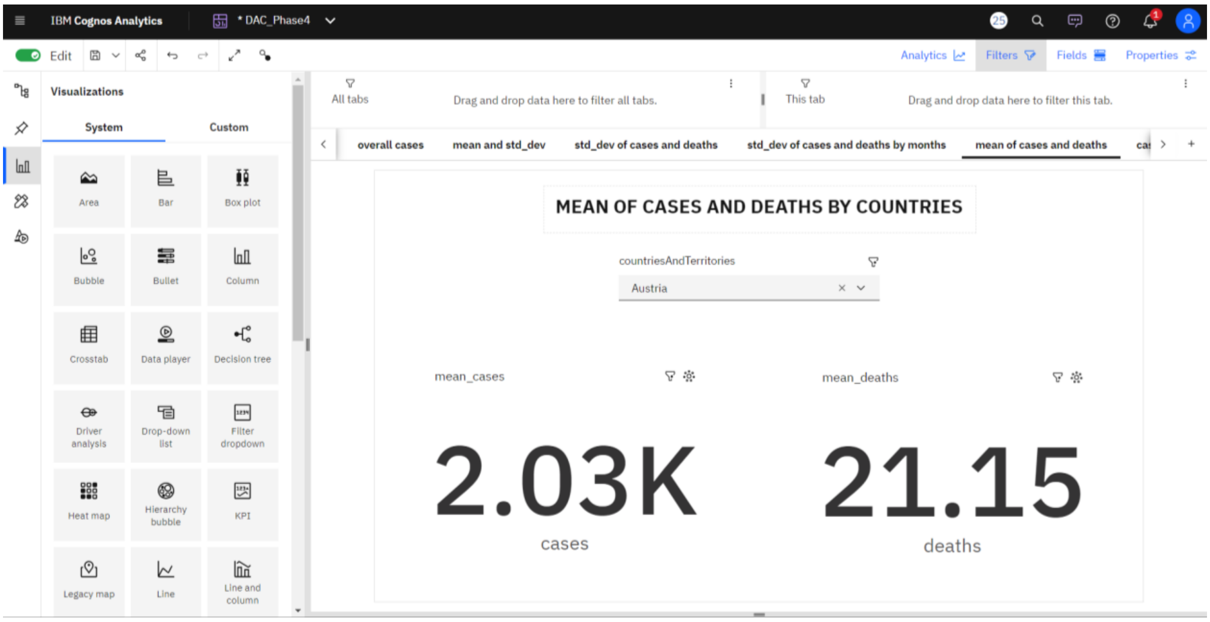

In Australia, the standard deviation for COVID-19 cases is 995.97, indicating a relatively high degree of variability in case numbers. In contrast, the standard deviation for deaths in Australia is much lower at 9.95, suggesting less variability in mortality figures. Similarly, in France, there were 13.1k cases with a standard deviation of 995.97 and 122.02 deaths with a standard deviation of 9.95, reflecting differing patterns in the spread of the virus and its impact on these two countries. Utilizing IBM Cognos, we can create a comprehensive visualization to explore and compare these statistics across various countries and territories, offering valuable insights into the COVID-19 situation worldwide.

### 5.5.Visualisation for standard Deviation of cases and deaths by months:



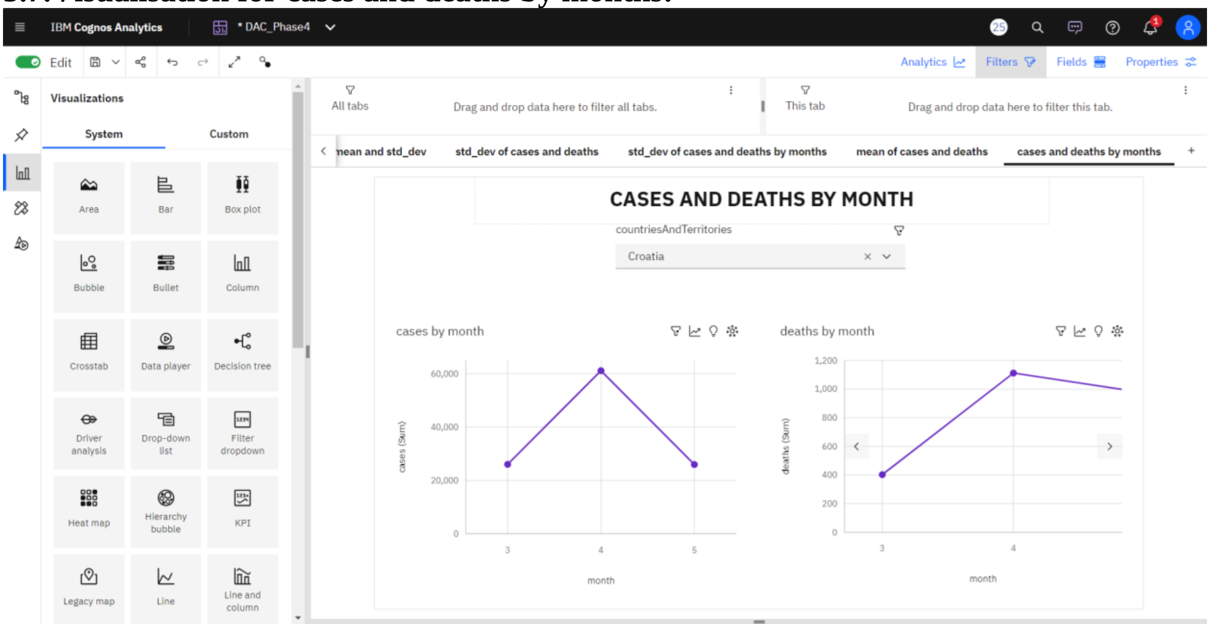

In this visualization, we can leverage IBM Cognos to conduct a detailed analysis of the standard deviation for both COVID-19 cases and deaths on a monthly basis for every country and territory. By breaking down the data into monthly increments, we gain a more granular understanding of the fluctuations in case numbers and mortality rates over time. This approach allows for the identification of trends, anomalies, and patterns that might not be apparent when examining aggregate data. Such insights can be invaluable for policymakers, healthcare professionals, and researchers in tailoring strategies and responses to the evolving dynamics of the pandemic worldwide.

### 5.6. Visualisation for Mean of cases and deaths by countries:



In the IBM Cognos visualization, a comprehensive analysis of COVID-19 cases and deaths reveals intriguing insights. Cases exhibit a strong weekly trend, with the highest values typically observed on Tuesdays and the lowest on Mondays. Additionally, there's a weak downward trend in cases. Notably, the date Rep for May 25, 2021, records the lowest average cases at just over 2,000, while April 7, 2021, has the highest average cases at almost 54,000.
For deaths, a moderate weekly trend is observed, with the largest values occurring on Wednesdays and the smallest on Mondays. Similar to cases, deaths also exhibit a weak downward trend. Remarkably, March 27, 2021, stands out with the highest average deaths at 897, and April 20, 2021, follows with an average of 447.
The forecasting data suggests that both cases and deaths may increase by June 19, 2021, with cases reaching over 18,000 and deaths potentially rising to 68.47. Several unusual data points are noted, particularly on March 27, 2021, for deaths and May 25, 2021, for cases. Moreover, during the period from March 27 to March 28, deaths dropped by 79%.
Overall, the IBM Cognos visualization provides comprehensive insights based on 91 data points for both cases and deaths, offering a valuable tool for understanding the dynamics of the COVID-19 pandemic.

### 5.7.Visualisation for cases and deaths by months:



In the IBM Cognos analysis, intriguing patterns emerge when comparing monthly data for COVID-19 cases and deaths. Month 4 holds the highest total deaths but is ranked second in total cases, indicating a discrepancy between case severity and mortality. Similarly, month 3 boasts the highest total cases but is ranked second in total deaths, highlighting variations in the impact of the virus.

Across all months, the cumulative sum of cases exceeds 288,000, underscoring the significance of the pandemic's reach. Cases fluctuate, with the lowest occurring in month 5 at over 69,000, and the highest in month 3 at over 114,000. Months 3 and 4 stand out significantly in terms of cases, contributing to nearly 76% of the total.

For deaths, the collective sum across all months surpasses 2,500. Deaths range from 708 in month 5 to over a thousand in month 4, revealing variations in mortality trends. This comprehensive analysis provides valuable insights into the dynamics of COVID-19 cases and deaths across different months.

# Conclusion:

Designing a COVID-19 test case model for data analysis is essential for understanding the patterns and trends of the disease. By analyzing various parameters such as demographics, geographic locations, testing methods, and outcomes, researchers can gain valuable insights into the spread and impact of the virus. A well-constructed COVID-19 test case model facilitates informed decision-making for healthcare professionals, policymakers, researchers. It enables them to identify high-risk areas, allocate resources efficiently, and develop targeted interventions.