**COVID -19 ANALAYSIS**

**PHASE 5**

|  |  |
| --- | --- |
| **Date** | **31-10-2023** |
| **Team ID** | **720** |
| **Project Name** | **Covid-19 Cases Analysis** |

|  |  |  |
| --- | --- | --- |
| **SNO** | **TABLE OF CENTENTS** | **PAGE NO** |
| 1 | INTRODUCTION | 2 |
| 2 | PROBLEM STATEMENT | 2 |
| 3 | PROBLEM IDENTIFIED | 3 |
| 4 | LITERATURE SURVEY | 3 |
| 5 | DESIGN THINKING | 5 |
| 6 | DEVELOPMENT PHASE | 12 |
| 7 | ANALYSIS OBJECTIVE | 16 |
| 8 | DATA COLLECTION | 16 |
| 9 | DATA VISUALISATION | 18 |
| 10 | CONCLUSION | 36 |

1. **INTRODUCTION**

The year 2021 marked a pivotal period in the global response to the COVID-19 pandemic. During the months of March, April, and May, the world continued its relentless battle against the virus, characterized by a complex interplay of rising cases, vaccination campaigns, and evolving public health measures. The data from this critical period holds the key to understanding the dynamics of the pandemic and shaping effective strategies for containment and mitigation.

In response to the pressing need highlighted by our problem statement, this study embarks on a comprehensive analysis of COVID-19 cases, deaths, and associated trends within the specified timeframe. Our aim is to provide a deeper insight into the impact of the virus during these months, examining variations across countries and regions, identifying potential hotspots, and assessing the effectiveness of public health responses.

As we navigate the intricate landscape of COVID-19 data for March, April, and May 2021, our focus extends beyond mere statistics. It encompasses the stories of resilience, sacrifice, and scientific innovation that unfolded during this period. We seek to illuminate the challenges and milestones, capturing the experiences of different nations and regions as they grappled with the pandemic's waves.

Through this analysis, we hope to empower decision-makers, healthcare professionals, and the public with valuable insights that can guide evidence-based responses, resource allocation, and the ongoing global effort to combat the pandemic. The data from these three crucial months is a treasure trove of information, and by exploring it in-depth, we aim to contribute to the collective knowledge that shapes our fight against COVID-19.

1. **PROBLEM STATEMENT:**

The COVID-19 pandemic represents an unprecedented global challenge characterized by rapidly evolving data. The core issue at hand is the pressing need to establish a robust data analysis framework capable of delivering timely and precise insights. This framework should enable us to comprehensively comprehend the spread, impact, and management of the virus. In an environment where information is continually changing, accurate data analysis becomes paramount for healthcare professionals, policymakers, and researchers alike. The challenge lies in developing methods and tools that can adapt to the dynamic nature of the pandemic, providing actionable insights to guide effective response efforts and mitigate the pandemic's far-reaching consequences on public health and society.

.

1. **PROBLEM IDENTIFIED:**

The global COVID-19 pandemic has created an urgent need for comprehensive and data-driven analysis of cases to inform public health policies and responses. As the virus's impact evolves, accurate and timely analysis of infection rates, transmission patterns, and vaccination effectiveness is paramount. However, the vast and dynamic nature of COVID-19 data, coupled with the need to integrate diverse sources of information, presents a complex challenge. Developing an accessible, robust, and up-to-date system for COVID-19 cases analysis is crucial to facilitate evidence-based decision-making and resource allocation.

**4. LITERATURE SURVEY**

* **1. “Detection of COVID-19 from Medical Images and/ or Symptoms of Patient using Machine Learning Approaches”, Akshay Kumar Siddhu1 , Dr. Ashok Kumar2 and Dr. Shakti Kundu3**

This article focuses on detecting COVID-19 using X-ray and CT-Scan images, emphasizing the urgency of efficient detection due to the rapid global spread of the virus. It provides background information on COVID-19, its origin, and respiratory symptoms, stressing the need for early detection and care. The article discusses diagnostic methods, including CT-Scans and Nucleic Acid Tests. It also highlights the development of smart imaging programs and data collection techniques, such as a smart CT imaging system and smartphone data collection. Deep learning techniques, like ResNet50 and InceptionV3, are discussed for COVID-19 detection using chest X-ray images. In conclusion, the article underscores the importance of early detection and efficient medical facilities, advocating for further development of deep learning models and public databases to enhance accuracy in COVID-19 diagnosis and prediction.

* **2. “Covid-19 impact on students”,** **Andria Pragholapati**

Globally, the COVID-19 epidemic has had a major effect on businesses, tourism, health, and students. As the virus spreads among people, social and physical barriers are imposed, impacting not only educational institutions but also the commercial, tourist, and health sectors. The Ministry of Education and Culture in Indonesia has released a circular directing colleges to offer remote learning opportunities and recommending that student's study from home. Millions of pupils have been impacted by this, particularly those from marginalized and vulnerable areas. Healthy individuals, persons with monitoring (ODP), people without symptoms (OTG), patients with supervision (PDP), Covid-19 patients, and vulnerable groups are all subject to the policies of the Indonesian Ministry of Health. UNESCO assists nations in their attempts to lessen the immediate effects of closing schools and enable universal access to remote education.

* **3.“Covid-19 impacts on society”,** **Jaspreet Singh, Jagandeep Singh**

Since its start in March 2020, the COVID-19 pandemic has had a significant negative influence on social interactions and relationships, which has resulted in stress, anxiety, sadness, mental disorders, and health risks. Over 420 million children and youth have been impacted by the virus, which has forced the closure of schools, colleges, universities, and other institutions. The virus can spread through inhalation or contact with sick individuals. People are being compelled to work from home and avoid social gatherings due to social distancing measures, which is resulting in a new phase of social suffering. Sociologist Eric Kleinberg of New York University cautions that social suffering associated with social separation and isolation will occur to an extent that is not yet widely discussed. Stressful states of anxiety, loneliness, anxiety drives, despair, and so on can result from the lack of deeply meaningful connections.

* **4. “ Investigation of the Covid-19 Reasearch”, Yong xu**

This study analyzes over 16,000 COVID-19 publications from Web of Science, emphasizing the lack of big data research on the pandemic. It identifies key findings, including a focus on vulnerable groups, primarily lung involvement in COVID-19, and a reliance on clinical and modeling methods. Notably, respiratory symptoms are highlighted as crucial for diagnosis. The study identifies frequently cited journals, such as the Journal of Medical Virology, and reveals commonly mentioned countries, keywords, and research methods in COVID-19 research. It provides valuable insights for pandemic control strategies, emphasizing the importance of respiratory symptoms in diagnosis and the need to address undeveloped countries and vulnerable populations.

* **5.“** **Prediction of COVID-19 Confirmed, Death, and Cured Cases in India Using Random Forest Model”,V.K.Guptha**

This paper investigates COVID-19 case prediction in India using the Random Forest model, aiming to forecast confirmed, death, and cured cases for public health management. It highlights the model's adaptability and predictive power, especially relevant in addressing the dynamic challenges of the pandemic in India.The research discusses data sources, emphasizing the importance of time-based features and preprocessing steps, including cases, deaths, recoveries, meteorological data, and population statistics.The Random Forest model shows promising performance, outperforming other machine learning models. Feature selection reveals the significance of factors like the previous day's cases and population density. The study also emphasizes feature engineering and visualization for a deeper understanding of the pandemic's dynamics.

**5.Design Thinking Approach:**

**5.1 Empathize:**

We will start by understanding the experiences and concerns of people affected by COVID-19, acknowledging their fears and needs.

**5.2 Define:**

This phase involves clearly defining the specific problems we aim to address, such as identifying high-risk communities or assessing the impact of vaccination campaigns.

**5.3 Objectives:**

Develop data cleaning processes in IBM Cognos that ensure data accuracy and completeness.

Create IBM Cognos dashboard and reports that provide real-time COVID-19 insights.

Produce data-driven narratives and visualization within IBM Cognos that inform policy decisions.

**5.4 Ideate:**

Innovative solutions will be generated, such as predictive models for case trends, targeted public health campaigns, or vaccine distribution strategies.

**5.5 Actions:**

Brainstorm creative solutions, such as predictive modeling for case trends or novel communication strategies. Foster an open and collaborative environment to encourage diverse idea generation. Prioritize ideas based on feasibility, impact, and alignment with the defined problems.

**5.6 Prototype:**

We will develop data-driven prototypes, leveraging IBM Cognos for visualizations and dashboards, to communicate insights effectively.

**5.7 Actions:**

Create visualizations and dashboards using tools like IBM Cognos to convey data-driven insights. Develop mock-up interventions or communication materials for testing. Ensure prototypes effectively communicate complex information to diverse audiences.

**5.8 Test:**

Proposed solutions will be rigorously tested against real-world data to ensure their feasibility and impact.

**5.9 Actions:**

Apply proposed solutions to real-world data to assess their performance. Solicit feedback from end-users and stakeholders on the effectiveness and usability of interventions. Continuously monitor the impact of strategies and refine them based on outcomes.

**5.10 Implement:**

Insights will be translated into actionable policies and interventions, aligning with the needs of different stakeholders.

**5.12 Actions:**

Translate insights into actionable policies, guidelines, and interventions. Collaborate with relevant stakeholders to ensure the seamless execution of strategies. Communicate and educate the public, healthcare professionals, and policymakers about the implemented actions.

**5.13 Iterate:**

Continuous monitoring and adaptation of strategies based on outcomes will be central to our approach.

**5.14 Actions:**

Continuously monitor the evolving COVID-19 situation and adapt strategies accordingly .Collect and analyze feedback from stakeholders to drive ongoing improvements. Stay agile and flexible in responding to new challenges and emerging data trends.

1. **DEVELOPMENT PHASE**
   1. **Data discovery**
2. **Data Import and Initial Inspection:**

Begin by importing the COVID-19 dataset into your data analysis environment .Inspect the dataset's structure, including the number of rows and columns, data types, and the first few rows to get a sense of the data.

1. **Data Exploration:**

Calculate summary statistics for cases and deaths to understand the overall scale of the pandemic. Explore the distribution of cases and deaths by creating histograms, box plots, or density plots.

1. **Temporal Analysis:**

Analyse how COVID-19 cases and deaths have evolved over time. Consider the day, month, and year columns. Calculate daily, monthly, and yearly totals for cases and deaths. Visualize trends using line plots or time series charts to identify peaks and patterns.

1. **Geographical Analysis:**

Explore how COVID-19 cases and deaths are distributed across different countries and territories. Calculate country-wise totals for cases and deaths. Create maps, bar charts, or heatmaps to visualize the geographical spread.

* 1. **Data preparation :**

Data preparation for COVID-19 analysis involves cleaning, structuring, and organizing the dataset to ensure that it is ready for analysis. Below are the key steps involved in data preparation for COVID-19 data:

1. **Data Collection:**

Obtain COVID-19 data from reliable sources, such as health agencies, government repositories, or research organizations. Ensure that the data is up-to-date and comprehensive.

**2. Data Import:**

Import the raw data into your chosen data analysis tool or programming environment (e.g., Python, R, Excel).

1. **Data Inspection:**

Examine the dataset's structure by checking the number of rows and columns, data types, and the first few rows of data to gain an initial understanding.

1. **Handling Missing Values:**

Identify and handle missing values appropriately. Options include imputation (replacing missing values with estimated values), removal of rows or columns with missing data, or considering the missingness pattern.

1. **Data Cleaning:**

Address data inconsistencies and errors. This may involve correcting data entry mistakes, resolving duplicates, and ensuring uniform formats for dates and locations.

1. **Data Transformation:**

Convert data types as needed (e.g., date columns to date objects) and create new variables if necessary (e.g., calculating daily cases from cumulative data).

1. **Outlier Handling:**

Identify and handle outliers, which are extreme values that can skew analysis results. Options include capping values, transforming data, or conducting separate analyses with and without outliers.

**6.2 Data modelling:**

Modelling COVID-19 is a complex task that requires collaboration among epidemiologists, data scientists, and domain experts. It plays a critical role in understanding the pandemic's dynamics, guiding public health measures, and making informed decisions. However, models are simplifications of reality and should be used in conjunction with real-world data and expert judgment to inform decision-making.

* 1. **Dashboarding Visualizing Insights**

1. **Define Dashboard Objectives:**

Determine the main objectives of your COVID-19 dashboard. Are you aiming to track cases, deaths, vaccinations, or other key metrics? What insights do you want to provide to your audience?

1. **Choose Visualization Tools**:

Select a suitable visualization tool for creating your dashboard. Common choices include Tableau, Power BI, Python libraries like Matplotlib and Plotly, R with Shiny, or even custom web development using JavaScript libraries like D3.js.

1. **Data Integration**:

Integrate clean and up-to-date COVID-19 data into your chosen visualization tool. Ensure that data sources are connected and regularly updated.

1. **Design Layout:**

Create a user-friendly and visually appealing layout for your dashboard. Consider the needs of different users, such as public health officials, policymakers, or the general public.

1. **Interactive Elements:**

Incorporate interactive features, such as dropdown menus, filters, or sliders, to allow users to explore the data and customize their views.

1. **Choose Visualization Types:**

Select appropriate chart types and visualizations for presenting COVID-19 insights. Examples include line charts for time series, bar charts for comparisons, maps for geographical data, and heatmaps for intensity visualization.

1. **Time Series Analysis:**

Use time series charts to illustrate trends over time, such as daily or weekly changes in COVID-19 cases, deaths, or vaccination rates.

1. **Geographical Mapping:**

If relevant, include maps to display the geographic distribution of COVID-19 cases, vaccination coverage, or other location-based data. Use color coding or symbols to represent data points.

1. **Key Metrics and KPIs:**

Highlight key metrics and key performance indicators (KPIs) relevant to the COVID-19 situation. Display daily case counts, test positivity rates, vaccination percentages, or other critical data.

* 1. **Communication of Results**

1. **Finalize Report Layout:**

Review and finalize the layout of your report. Ensure that all elements, such as tables, charts, and text, are properly arranged and formatted.

1. **Formatting:**

Apply formatting options to enhance the report's appearance. You can format fonts, colors, borders, and backgrounds to make the report visually appealing and easy to read.

1. **Report Styling:**

Apply a consistent styling theme to your report to maintain a unified look and feel. Cognos Analytics provides predefined styles that you can choose from or customize.

1. **Testing and Validation:**

Test your report to ensure that all data elements are displayed correctly. Verify that data calculations, aggregations, and filters are working as expected.

1. **Save and Publish:**

Save your report in Cognos Analytics and publish it to the appropriate location, such as a shared folder, report server, or email distribution list.

1. **Security and Permissions:**

Set up security and permissions for the report to control who can access, view, or modify it. Define user roles and access levels as needed.

1. **Documentation:**

Document the report creation process, including data sources, calculations, and any customizations.

1. **ANALYSIS OBJECTIVE**

The analysis objective appears to be centered on calculating the mean (average) and the corresponding standard deviation for a dataset. The standard deviation measures the spread or dispersion of the data points around the mean. In the context of this analysis, it's important to determine the level of variability in the data.

**Calculate the Mean:** This involves adding up all the data points and dividing by the total number of data points. The mean provides a measure of the central tendency of the data.

**Calculate the Standard Deviation:** The standard deviation quantifies how individual data points deviate from the mean. A higher standard deviation indicates greater data variability, while a lower standard deviation suggests data points are closer to the mean.

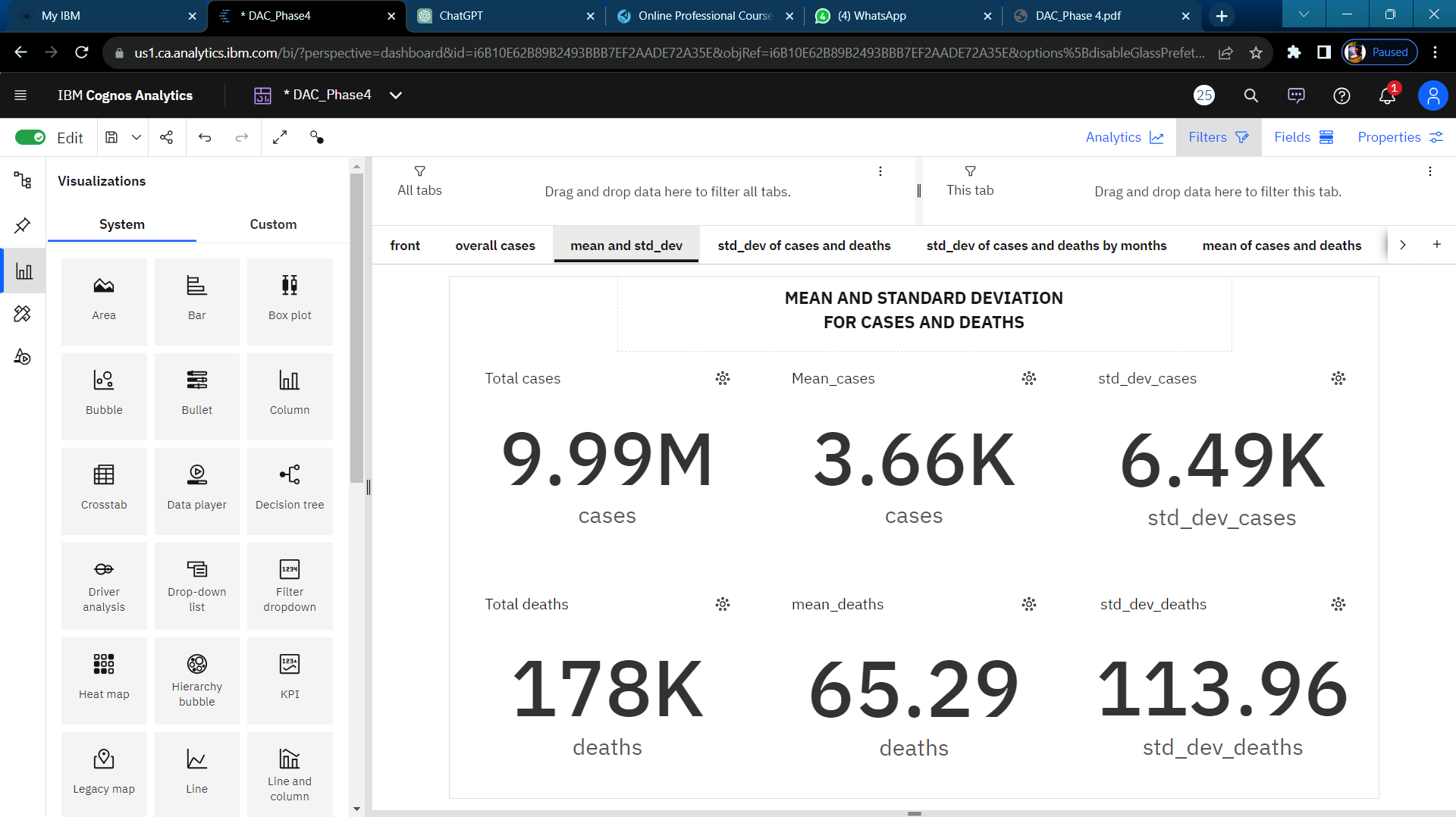
**Interpret the Results:** Once you have the mean and standard deviation, you can analyze the data distribution. If the data follows a normal distribution, the mean and standard deviation provide essential information about the data's characteristics.

In addition to the mean and standard deviation, you might also want to consider other statistics, depending on the specific objectives of your analysis, such as median, quartiles, skewness, and kurtosis. These additional statistics provide a more comprehensive understanding of the data's distribution and shape.

1. **DATA COLLECTION**

* The data collection process for COVID-19 typically involves the following steps:
* Source Identification: Identify sources such as health authorities, hospitals, and laboratories that report COVID-19 data.
* Data Variables: Define the specific data variables to collect, including cases, deaths, recoveries, testing, and demographics.
* Data Points: Collect daily or periodic data points, including location, date, and case status.
* Reporting Methods: Establish reporting methods, which can be manual, electronic, or through online portals.
* Standardization: Ensure data is reported in a standardized format, following national or international guidelines.
* Privacy Considerations: Protect individuals' privacy and adhere to data protection regulations.
* Quality Control: Implement data validation and quality control measures to minimize errors.
* Data Aggregation: Aggregate data at various levels (local, regional, national) for analysis.
* Data Transparency: Publish data in a transparent and accessible manner for the public and researchers.
* Data Sharing: Collaborate with other organizations to share data for a comprehensive picture.
* Historical Data: Maintain historical data for trend analysis and research.
* Data Security: Secure data to prevent breaches and maintain public trust.
* Data Analytics: Employ data analytics tools to identify patterns, hotspots, and potential outbreaks.
* Public Communication: Disseminate findings and updates to the public through official channels.
* Research Use: Encourage researchers to utilize the data for studies and modelling.
* Feedback Loops: Establish feedback mechanisms with data providers to improve data collection.
* International Collaboration: Collaborate with other countries to monitor global trends and travel-related cases.
* Data Storage: Store data securely for long-term research and policymaking.
* Policy Adaptation: Use data insights to adapt public health policies and interventions.
* Continuous Improvement: Continuously refine the data collection process to respond to evolving needs and challenges.
* This process helps in tracking and understanding the COVID-19 pandemic's dynamics and informs public health responses.

1. **DATA VISUALISATION**
2. MEAN AND STANDARD DEVIATION OF THE CASES AND ASSOCIATED 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 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.
* .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.

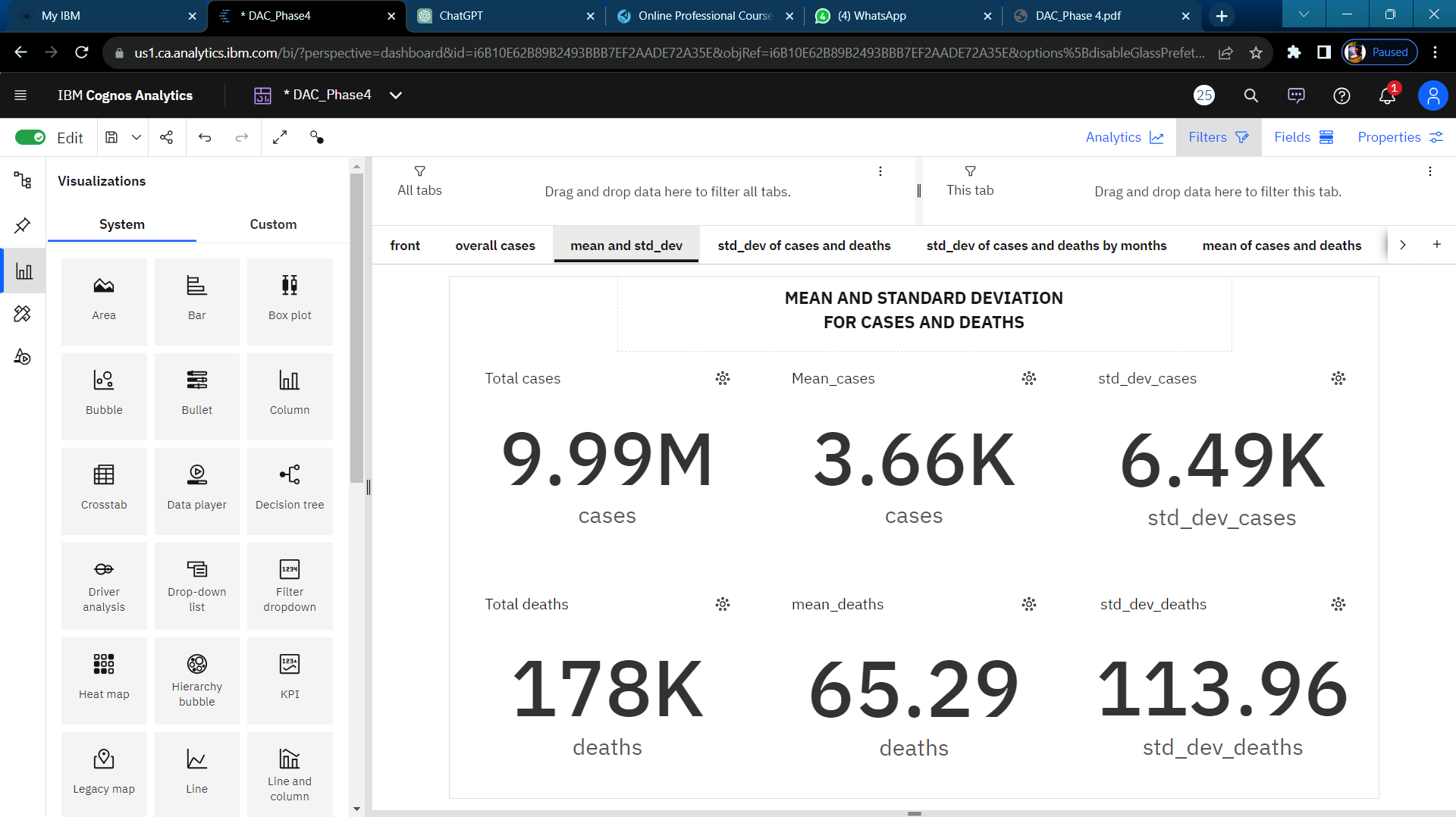
**For Deaths:**

**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.

1. **PROVIDING THE OVERVIEW OF CASES AND DEATHS**

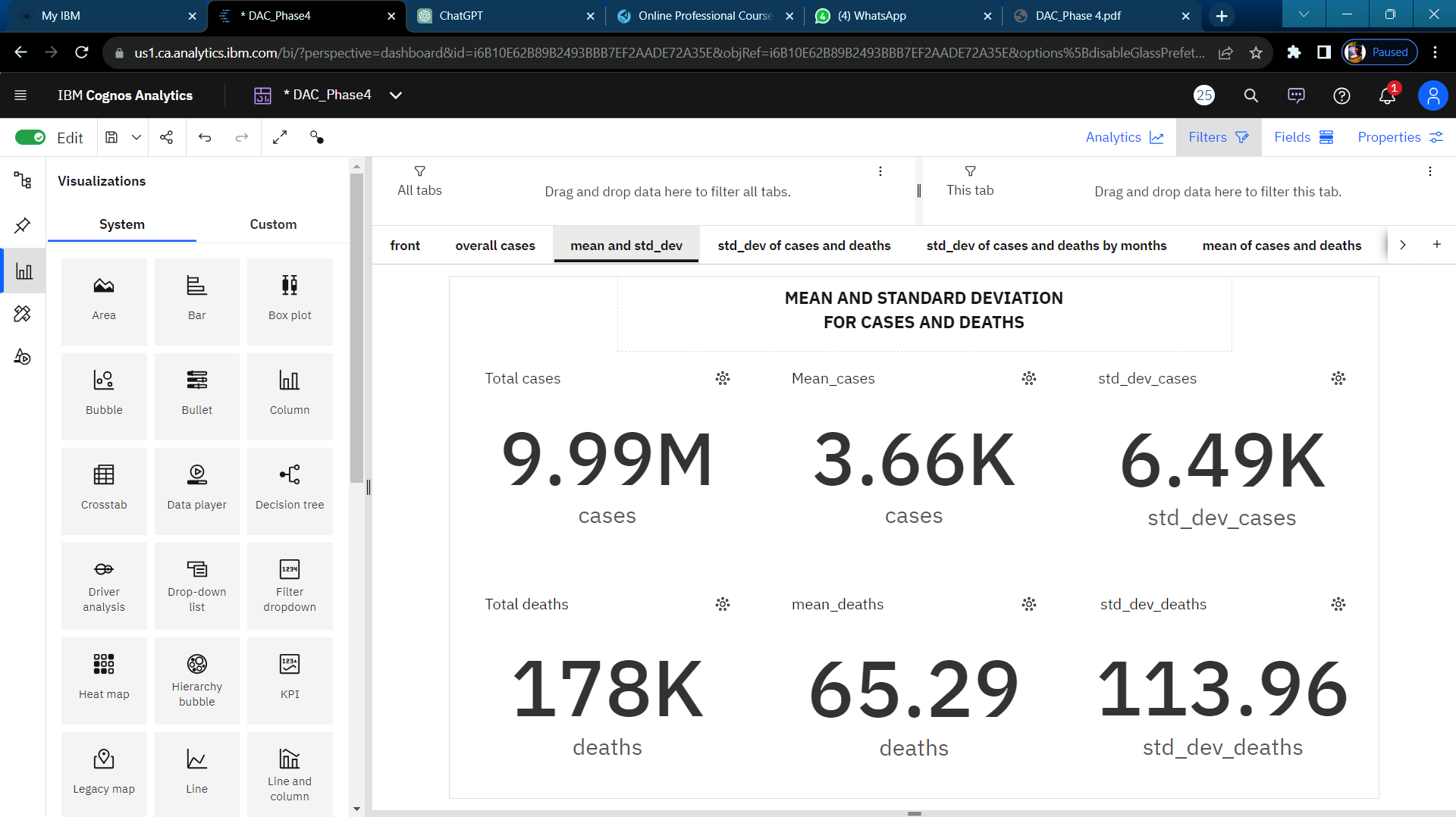
**TOTAL CASES**



**INSIGHTS**

* The insight from the dataset reveals that during the months of **March, April, and May, 2021**, the cumulative number of COVID-19 cases in the dataset reached approximately **9.99 million.**
* This indicates a substantial level of infections within this three-month period, underlining the significant impact of the virus using the **summary visual**.

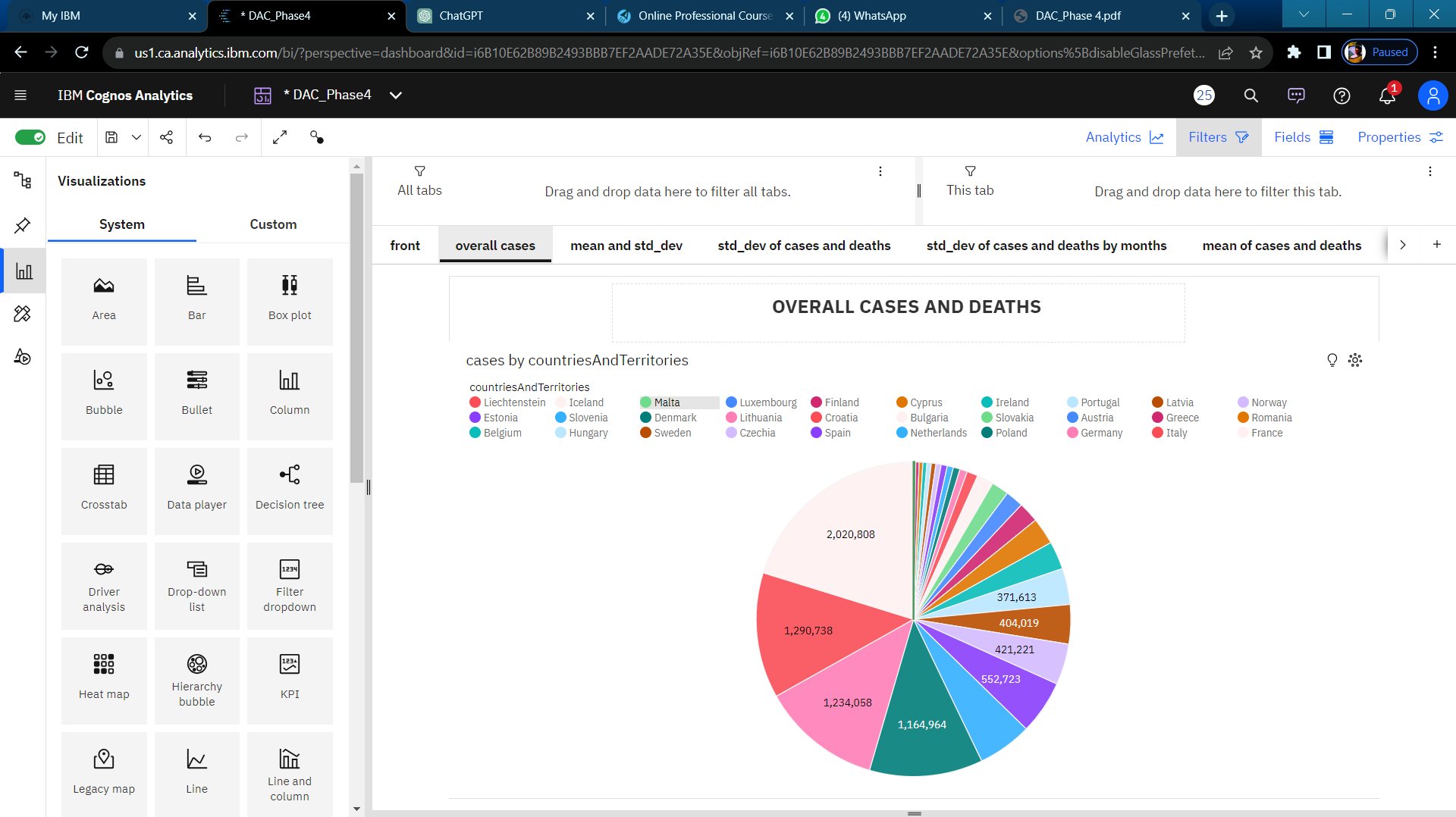
**TOTAL DEATHS**



**INSIGHTS**

* The dataset indicates that during the combined months of **March, April, and May, 2021,** there were a total of **178,000** reported COVID-19 cases.
* This data suggests a notable level of infections within this three-month period using the **summary visual**.

1. **WHICH COUNTRY HAS THE HIGHEST NUMBER OF CASES**

****

**INSIGHTS**

* Cases ranges from 437 in Liechtenstein, to over 2.0 million in France.
* From 2021-03-29 to 2021-03-30, France's cases increased by 937%.
* Overall sum of cases is nearly 10.0 million.
* It is projected that by 2021-06-19, France will exceed Germany in cases by over 14 thousand.

1. **TOP 5 COUNTRIES WITH HIGHEST NUMBER OF COVID 19 CASES**

**INSIGHTS**

* France (2,020,808 cases): France has faced a substantial number of COVID-19 cases, reflecting the challenges in containing the virus. This high case count underscores the need for ongoing public health measures and vaccination campaigns to mitigate the spread.
* Italy (1,290,738 cases): Italy, one of the early epicenters of the pandemic, continues to grapple with a significant number of cases. It highlights the importance of sustained vigilance and healthcare preparedness.
* Germany (1,234,758 cases): Germany, known for its robust healthcare system, still faces a considerable case load. The data emphasizes the persistence of COVID-19 and the need for adaptive strategies.
* Poland (1,164,964 cases): Poland's notable case count suggests ongoing challenges in controlling the virus. This underscores the importance of vaccination and adhering to public health guidelines.
* Netherlands (557,983 cases): The Netherlands has experienced a substantial number of COVID-19 cases. This data reflects the global nature of the pandemic and the need for international cooperation in controlling its spread.
* These insights highlight that COVID-19 remains a global challenge, with countries continuing to grapple with significant case numbers. It underscores the importance of vaccination, public health measures, and data-driven decision-making to curb the spread of the virus.

1. **TOP 5 COUNTRIES WITH LOWER NUMBER OF COVID 19 CASES.**

* Liechtenstein (437 cases): Liechtenstein's low case count reflects effective containment measures and a small population. It underscores the advantages of early intervention in limiting virus spread.
* Iceland (527 cases): Iceland's success in maintaining a low case count is attributed to a robust healthcare system and early testing and contact tracing efforts.
* Malta (7,586 cases): Malta has managed to keep its case count comparatively low. Effective public health measures and vaccination campaigns likely play a role in this achievement.
* Luxembourg (14,464 cases): Luxembourg's case count, while higher than the previous countries, is still relatively low. Stringent measures and a responsive healthcare system may contribute to this outcome.
* Finland (34,760 cases): Finland has maintained a modest case count compared to larger European countries. The data suggests a combination of strict public health measures and a health-conscious population.
* These insights highlight the effectiveness of timely public health measures, testing, contact tracing, and vaccination strategies in managing and limiting COVID-19 cases in these countries.

1. **TOP 5 COUNTRIES WITH HIGHEST NUMBER OF DEATHS.**

**INSIGHTS**

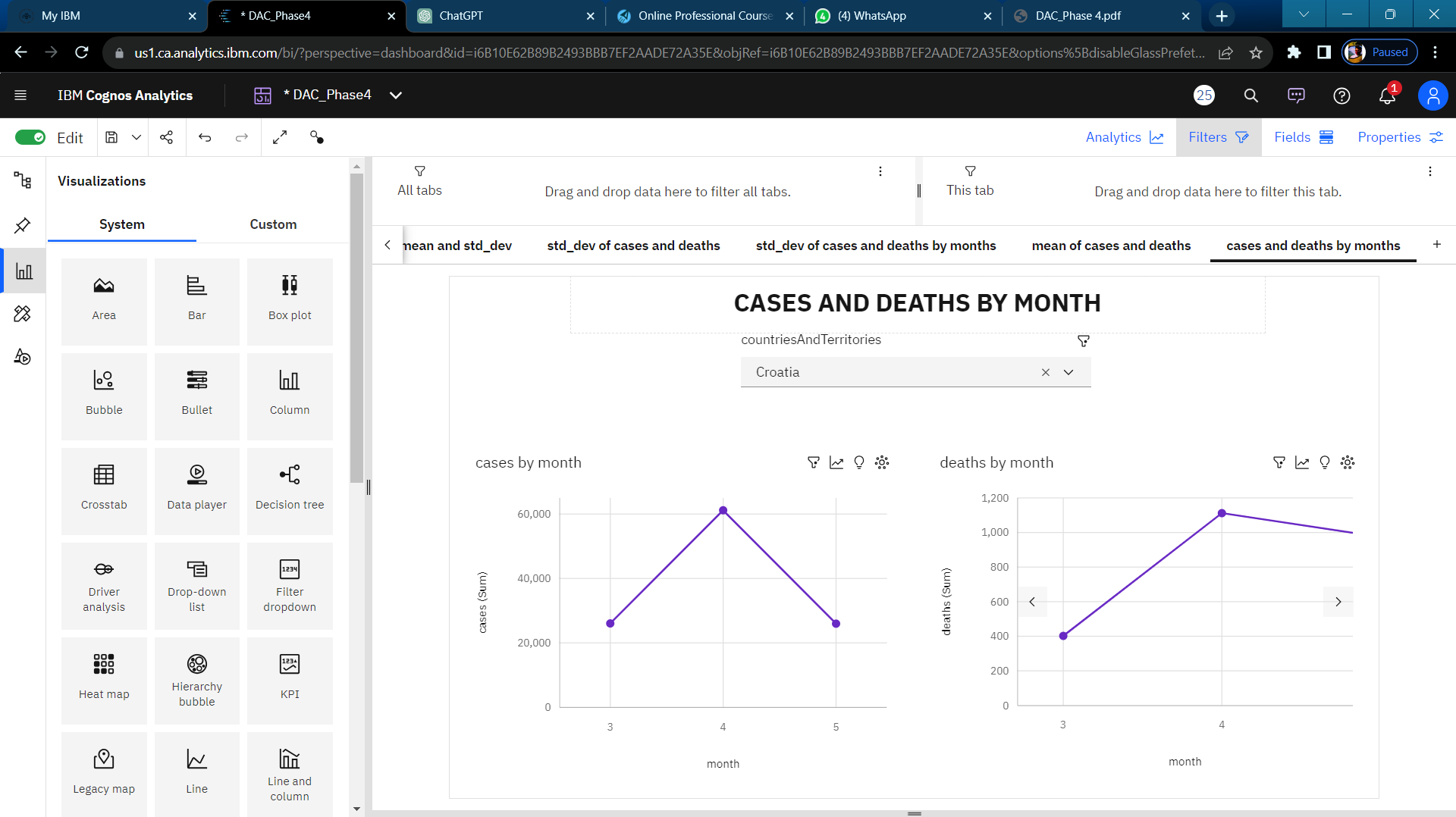
* Poland (29,969 deaths): Poland has experienced a significant number of COVID-19 fatalities, highlighting the importance of continued efforts to manage the pandemic effectively.
* Italy (28,347 deaths): Italy, which faced a severe outbreak early in the pandemic, continues to grapple with a substantial death toll, emphasizing the need for ongoing vigilance.
* France (22,977 deaths): France has also faced a considerable number of COVID-19 fatalities, underscoring the importance of healthcare preparedness and public health measures.
* Germany (18,337 deaths): Germany, known for its robust healthcare system, has experienced a notable number of deaths, indicating the global challenges in managing the virus.
* Hungary (14,675 deaths): Hungary's death count, while lower than the previously mentioned countries, is still significant. It highlights the need for vaccination campaigns and adherence to public health guidelines.
* These insights illustrate the impact of the COVID-19 pandemic on these European countries and emphasize the importance of ongoing measures to reduce mortality and prevent further infections.

1. **TOP 5 COUNTRIES WITH LOWEST NUMBER OF DEATHS.**

**INSIGHTS**

* Iceland (1 death): Iceland's exceptionally low death count underscores the effectiveness of its response and healthcare system in minimizing fatalities.
* Liechtenstein (4 deaths): Liechtenstein's minimal number of deaths reflects successful efforts in containment and healthcare management.
* Malta (104 deaths): Malta's relatively low death count suggests effective public health measures and healthcare responses in the face of the pandemic.
* Cyprus (129 deaths): Cyprus has managed to keep its death count relatively low through responsive measures and effective healthcare provision.
* Denmark (155 deaths): Denmark's modest death count, compared to larger European countries, highlights the efficacy of its public health policies and healthcare system.
* These insights illustrate how these countries have been successful in keeping COVID-19 death numbers relatively low, primarily through proactive public health measures and efficient healthcare systems.

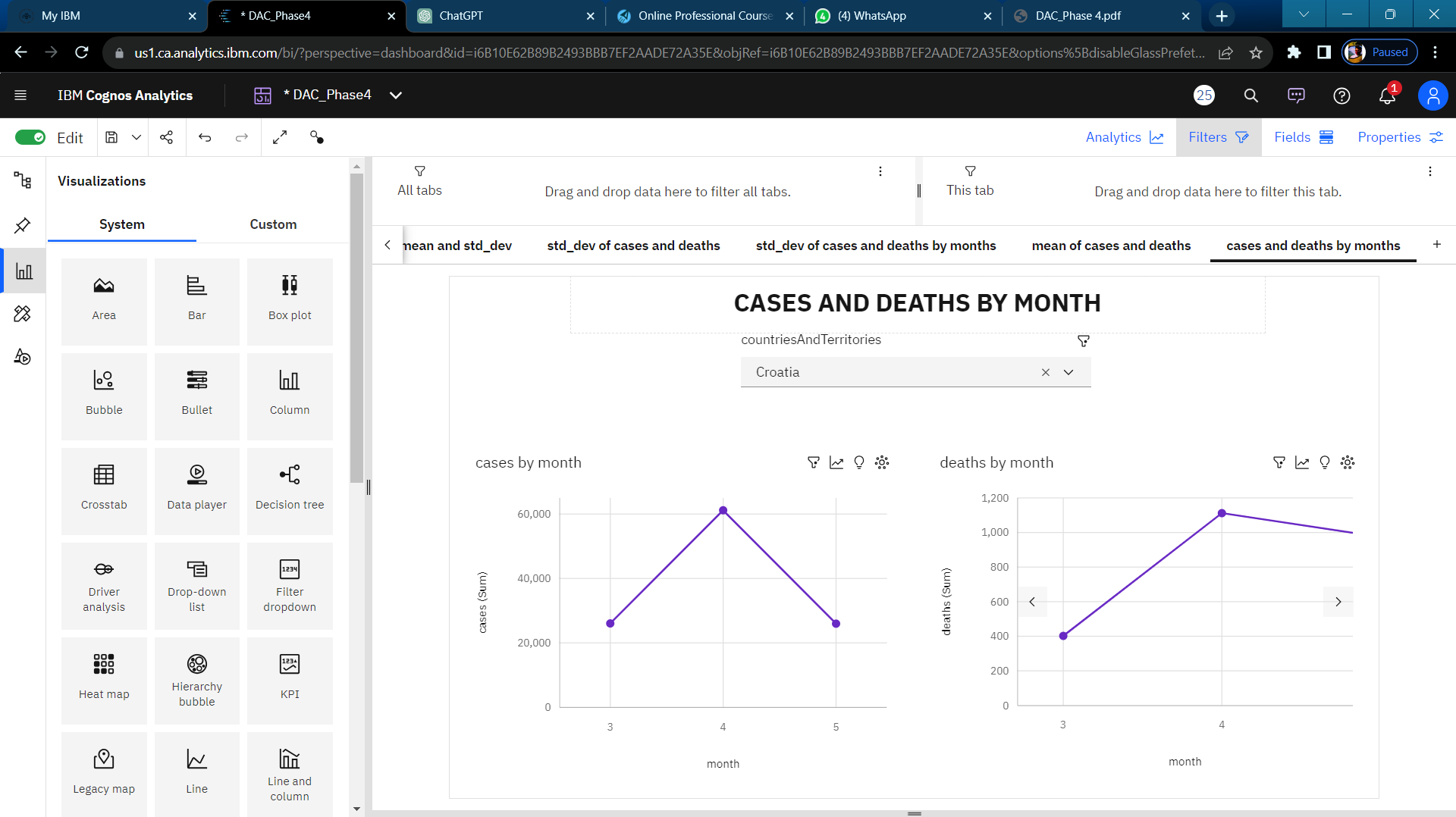
1. **MONTHLY CASES CHART**



**INSIGHTS**

* Cases is unusually low in the month of May.
* April month has the highest Total deaths but is ranked #2 in Total cases.
* March month has the highest Total cases but is ranked #2 in Total deaths.
* Over all months, the sum of cases is nearly 10.0 million.
* Cases ranges from over 1.8 million, when month is 5, to over 4.2 million, in the month of march.
* For cases, the most significant values of month are march and april, whose respective cases values add up to almost 8.2 million, or 81.7 % of the total.

1. **MONTHLY DEATHS CHART**



**INSIGHTS**

* Deaths is unusually low when month of May.
* March has the highest Total cases but is ranked #2 in Total deaths.
* April has the highest Total deaths but is ranked #2 in Total cases.
* Across all months, the sum of deaths is over 178 thousand.
* Deaths ranges from nearly 38 thousand, when month is May, to over 72 thousand, when month is April.
* For deaths, the most significant values of month are March and April, whose respective deaths values add up to over 140 thousand, or 78.8 % of the total.

**Conclusion:**

The COVID-19 data analysis project aims to provide actionable insights and predictions for managing the pandemic. By employing innovative strategies such as real-time data collection, automated data cleaning, advanced modeling, sentiment analysis, we seek to contribute to informed decision-making and crisis management. This comprehensive approach combines data science, epidemiology, and technology to address the challenges posed by the ongoing pandemic.