**COURSE NAME: DATA ANALYTICS**

**PROJECT TITLE: COVID-19 USING COGNOS**

**PHASE-1**

**Problem Definition and Design Thinking**

**NAME:M A MANISHA**

**NM ID: au211521243098**

**PROJECT OVERVIEW:**

"In the wake of the global COVID-19 pandemic, our data analytics project using IBM Cognos aims to provide valuable insights into the spread and impact of the virus. Leveraging a diverse dataset that includes infection rates, vaccination data, hospitalizations, and demographic information, we seek to uncover trends, patterns, and correlations that can inform public health decisions and policy recommendations. Our project will involve data cleaning, visualization, and advanced analytics techniques to produce meaningful reports and dashboards that not only track the progression of the pandemic but also facilitate data-driven decision-making for healthcare authorities and policymakers. Through this project, we aspire to contribute to a better understanding of the COVID-19 crisis and support efforts to mitigate its effects on society."

**DESIGN THINKING:**

**ANALYSIS OBJECTIVES:**

**1. Comparing Mean Values and Standard Deviations:**

* Utilize IBM Cognos to calculate the mean (average) number of COVID-19 cases and deaths over a specified time period. This provides a central measure of the data distribution.
* Calculate the standard deviation using Cognos to measure the spread or dispersion of the data points. A higher standard deviation indicates greater variability in cases and deaths.

**2. Identifying Trends and Outliers:**

* Create line charts or time series visualizations in Cognos to identify trends and fluctuations in COVID-19 cases and deaths over time. This can help in spotting surges or declines in specific periods.
* Use Cognos to identify outliers or anomalies in the data, such as unusually high or low numbers of cases and deaths. This can be crucial for detecting potential data reporting errors or significant events.

**3. Comparing Regional or Demographic Data:**

* Break down the analysis by regions or demographics using Cognos' filtering and grouping capabilities. Compare mean values and standard deviations across different geographic areas or demographic groups to identify disparities.
* Create bar charts or heat maps in Cognos to visualize these variations, making it easier to understand and communicate the differences in COVID-19 statistics.

**4. Forecasting and Predictive Modelling:**

* Utilize Cognos for time series forecasting and predictive modelling to estimate future COVID-19 cases and deaths based on historical data. This can help in resource planning and decision-making.
* Evaluate the accuracy of your forecasts by comparing them with actual data over time, allowing for adjustments to response strategies.

**DATA COLLECTION:**

**1. Data Import and Collection:**

* Use IBM Cognos to import the provided data file containing COVID-19 cases and deaths information. Cognos allows you to connect to various data sources, such as CSV files or databases. You can set up a data connection to regularly update your dataset with the latest information.
* Create a data source connection in Cognos to ensure that your data is up-to date and accessible for analysis.

**2. Data Preparation and Cleansing:**

* Use Cognos Data Preparation capabilities to clean and transform the data as needed. This may include removing duplicates, handling missing values, and formatting data columns appropriately.
* Ensure that the data is structured correctly with relevant fields for date, country, COVID-19 cases, and deaths. Cognos provides tools for data cleansing and transformation to streamline this process.

**3. Data Modelling and Aggregation:**

* Create a data model in Cognos that reflects the structure of your COVID-19 dataset, defining relationships between different data elements. This allows you to efficiently aggregate data by day and country.
* Use Cognos Query Studio or Data Modules to build queries and aggregations that summarize the COVID-19 cases and deaths per day and by country. You can calculate daily totals, averages, and other relevant statistics.

**4. Data Visualization and Reporting:**

* Leverage Cognos' reporting and visualization capabilities to create interactive dashboards and reports. You can design visualizations such as line charts, heat maps, and tables to display COVID-19 trends, comparisons between EU/EEA countries, and more.
* Utilize Cognos' scheduling and distribution features to automatically generate and share reports with stakeholders on a regular basis.

**VISUALIZATION STRATEGY:**

**1. Bar Charts for Mean Values:**

* Utilize Cognos to create bar charts that display the mean values of COVID-19 cases and deaths. Group the data by a relevant dimension, such as time (e.g., by month or week) or geography (e.g., by country or region).
* Label the y-axis as "Mean Value" and the x-axis with the corresponding dimension (e.g., time or geography).
* Use colour coding to differentiate between COVID-19 cases and deaths. This allows for a clear visual comparison of the mean values over the selected dimension.
* Add data labels to the bars to display the exact mean values, making it easy for viewers to interpret the chart.

**2. Line Charts for Standard Deviations:**

* Create line charts in Cognos to represent standard deviations of COVID-19 cases and deaths over time or across regions.
* Plot two lines on the same chart, one for cases and one for deaths, with time on the x-axis (if analysing trends over time) or regions (if comparing regions).
* Apply shading or bands around each line to indicate the range of standard deviations, effectively highlighting the degree of variability.
* Include a legend to clarify which line represents cases and which represents deaths.

**3. Combo Charts for Comprehensive Comparison:**

* For a comprehensive comparison, consider using combo charts in Cognos. These charts allow you to combine different chart types, such as bars and lines, on the same graph.
* Use bars to represent mean values (e.g., mean cases and mean deaths) and lines to represent standard deviations (e.g., cases' standard deviation and deaths' standard deviation).
* This approach provides a clear visual correlation between the mean values and their variability, enhancing the audience's understanding of the data.

**4. Interactive Tooltips and Filters:**

* Enhance the user experience by enabling interactive tooltips in Cognos. When viewers hover over data points or bars, display additional information, including the exact mean and standard deviation values.
* Implement interactive filters to allow users to focus on specific time periods, countries, or regions. This empowers users to explore the data and draw insights relevant to their interests.

**INSIGHTS GENERATION:**

**1. Identification of Hotspots and Variability:**

* By comparing the mean values and standard deviations of cases and deaths across different regions or countries in Cognos, you can identify potential COVID-19 hotspots.
* High mean values coupled with high standard deviations may indicate regions with both a high number of cases and significant variability in daily or weekly reports. These areas might require targeted interventions and resources.

**2. Trend Analysis and Anomalies Detection:**

* Analysing mean values and standard deviations over time in Cognos can help in spotting trends and anomalies.
* A sudden spike in mean cases with a large standard deviation might signify an outbreak or a change in testing and reporting protocols. Conversely, a decreasing mean with a decreasing standard deviation could indicate a successful containment strategy.

**3. Resource Allocation and Preparedness:**

* Insights from mean values and standard deviations can guide resource allocation and preparedness efforts.
* Regions with consistently high mean values and low standard deviations might require a steady supply of medical resources, while areas with fluctuating mean values and high standard deviations may need flexible and adaptive responses.

**4. Effectiveness of Interventions:**

* Comparing mean values and standard deviations before and after the implementation of public health interventions, like lockdowns or vaccination campaigns, can assess their effectiveness.
* A decrease in mean cases with a decrease in standard deviation postintervention may indicate successful control measures, whereas a sustained high mean with increasing standard deviation could suggest the need for further action.

**5. Data Quality Assessment:**

* Unexpectedly high standard deviations in certain regions or time periods might also indicate data quality issues or reporting inconsistencies.
* Use Cognos to drill down into these anomalies and investigate potential data collection or reporting problems, ensuring that the data used for decision making is accurate and reliable.