**Project 7: COVID-19 using Cognos**

**Phase 1: Problem Definition and Design Thinking**

**By Badri N**

**Project Definition:** The project involves analysing COVID-19 cases and deaths data using IBM Cognos. The objective is to compare and contrast the mean values and standard deviations of cases and associated deaths per day and by country in the EU/EEA. This project encompasses defining analysis objectives, collecting COVID-19 data, designing relevant visualizations in IBM Cognos, and deriving insights from the data.

**Design Thinking:**

* 1. **Analysis Objectives**: Define the specific objectives of analysing COVID-19 cases and deaths data, such as comparing mean values and standard deviations.
  2. **Data Collection**: Obtain the provided data file containing COVID-19 cases and deaths information per day and by country in the EU/EEA.
  3. **Visualization Strategy**: Plan how to visualize the mean values and standard deviations using IBM Cognos to create informative charts and graphs.
  4. **Insights Generation**: Identify potential insights from the comparison of mean values and standard deviations of cases and deaths.

**Analysis Objectives:**

* To identify and understand the different trends in COVID-19 cases and deaths.
* Calculate the Mean and Standard Deviation of Daily, Monthly and Yearly cases and deaths per Country/Territory.
* Compare the Mean and Standard Deviation values of EU region to calculate the mortality rate.
* To visualize the data in a graphical representation.

**Data Collection:**

* The dataset used here is the COVID 19 dataset.
* It contains information about the cases and deaths per country affected by COVID-19.
* The dataset contains 7 fields: DateReported, day, month, year, cases, deaths, countries and 2730 records but we choose a sample data of 31 records only.

**Visualization Strategy:**

* To first calculate the Mean and Standard Deviation for both Cases and Death individually, we modify the dataset by adding calculation to existing table using “Calculations” option and calculate Standard Deviation, Average for each attribute deaths and cases individually.
* Then, we create a new Report and select COVID-19 data module as Source
* Now to create the graph, we select the “Visualization” option and select Clustered Column graph and substitute Average cum Standard Deviation values of deaths and cases on X and Y comparatively to produce 2 graphs.

**Insights Generation:**

|  |  |  |
| --- | --- | --- |
|  | Cases | Deaths |
| Mean | 896.7742 | 13 |
| Standard Deviation | 475.9697 | 6.801961 |

* Description/Comparison of Means:

The mean values of the cases are far higher than that of the deaths. The mean value of Cases is 896.7742 and the Deaths having 13. This explains that people were infected in large numbers, but only smallest number of people lost their lives.

* Description/Comparison of Standard Deviation:

The stand deviation values of the Cases are also far higher than that of Deaths. The Cases have an S.D. value of 475.9697 and the Deaths have an S.D. value of 6.801961.

This shows that the average number of cases are more spread out than the death’s average data.

**Conclusion:**

Throughout the EU region, the infectivity rate of COVID-19 was far higher than its lethality and severity rate. Thus, the number of cases is at a drastic amount compared to the lesser values in number of average deaths in the month of May as taken as the Sample Data for this analysis.