**PHASE 5: DOCUMENTATION**

**TECHNOLOGY NAME:** Data analytics with Cognos

**PROJECT NAME:** COVID-19 Cases Analysis

**Analysis Objectives:**

To analyze COVID-19 cases and deaths data, such as comparing mean values and standard deviations.

**DOCUMENTATION:**

In this part,start to build the COVID-19 cases analysis using IBM Cognos for visualization.

Defining the analysis objectives and obtain the COVID-19 cases and deaths data file.

**Data Collection:**

Obtain the provided data file containing COVID-19 cases and deaths information per day and by country in the EU/EEA.

**DOCUMENTATION:**

The data that can be taken from the given dataset and processing it. The data can be cleansed to ensure its accuracy and reliability. In this , begin to build a project by loading and preprocessing the dataset.

**SUBMISSION:**

The code to cleanse and preprocess a data can be submitted in Github repository.

**Visualization Strategy:**

Plan how to visualize the mean values and standard deviations using IBM Cognos to create informative charts and graphs.

**DOCUMENTATION:**

In this,build the analysis by creating visualizations using IBM Cognos and deriving insights from the data.

Create charts and graphs in IBM Cognos to visualize and compare the mean values and standard deviations of COVID-19 cases and associated deaths.

Analyze the visualizations to identify trends, variations, and potential correlations between cases and deaths.

**SUBMISSION:**

The code that can be used to create a charts and graphs in IBM Cognos to visualize and compare the mean values and standard deviations of COVID-19 cases and deaths.

**Insights Generation:**

Identify potential insights from the comparison of mean values and standard deviations of cases and deaths.

**DOCUMENTATION:**

The result identifying the potential insights from the comparison of mean values and standard deviations of cases and deaths can be documented.

**SUBMISSION:**

The code that can ce identify the insights from the comparison of mean and standard deviations of cases and deaths can be submitted in Github repository.

**DOCUMENTATION OF CHARTS AND GRAPHS:**

**VISUALIZATIONS BETWEEN DAY AND CASES:**

**￼**

**VISUALIZATIONS BETWEEN DAY AND DEATHS:**

**￼**

**VISUALIZATION BETWEEN THE MEAN VALUES AND STANDARD DEVIATION OF CASES AND**

**ASSOCIATED DEATHS:**

**SCATTER PLOT:**

**￼**

**SIMPLE GRAPH:**

**￼**

**SIMPLE GRAPH AND SCATTER PLOT:**

**￼**

**DOCUMENTATION OF MEAN AND SD VALUES:**

Mean value for cases: 3661.010989010989

Mean value for deaths: 65.29194139194139

Variance for CASES: 2344.631646785771

Standard deviation for CASES: 48.421396580290526

Variance for DEATHS: 0.7849611127600501

Standard deviation for DEATHS: 0.8859803117225857

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

#extracting data from a given dataset

df=pd.read\_csv("Covid\_19\_cases4.csv")

#extracting an associated columns "cases" and "deaths"

x=df['cases']

y=df['deaths']

z=df['day']

#Creating visualizations using bar graph between cases and day

plt.bar(z,x)

plt.xlabel("DAY")

plt.ylabel("CASES")

plt.title("DAY AND CASES")

#Creating visualizations using bar graph between deaths and day

plt.bar(z,y)

plt.xlabel("DAY")

plt.ylabel("DEATHS")

plt.title("DAYS AND DEATHS")

#Printing the values of x and y

print("Cases:",x)

print("Deaths:",y)

#Finding the mean value for cases

x\_mean=np.mean(x)

print("Mean value for cases:",x\_mean)

#Finding the mean value for deaths

y\_mean=np.mean(y)

print("Mean value for deaths:",y\_mean)

#Finding the Variance for CASES

for i in range(0,2730):

variance\_x=((x[i]-x\_mean)\*\*2)/2730

print("Variance for CASES:",variance\_x)

#Finding SD for CASES

sd\_x=np.sqrt(variance\_x)

print("Standard deviation for CASES:",sd\_x)

#Finding the Variance for DEATHS

for i in range(0,2730):

variance\_y=((y[i]-y\_mean)\*\*2)/2730

print("Variance for DEATHS:",variance\_y)

#Finding SD for DEATHS

sd\_y=np.sqrt(variance\_y)

print("Standard deviation for DEATHS:",sd\_y)

#Creating a visualizations between mean values and associated deaths

a=[x\_mean,y\_mean]

b=[sd\_x,sd\_y]

#Visualize a scatter plot

plt.scatter(a,b)

plt.xlabel("Mean of deaths and cases")

plt.ylabel("Standard deviation of deaths and cases")

plt.title("Scatter")

#visualize a simple graph plot

plt.plot(a,b)

plt.xlabel("Mean of deaths and cases")

plt.ylabel("Standard deviation of deaths and cases")

plt.title("Simple Graph")