PHASE 2

INNOVATION- Analyzing COVID-19 Cases and Deaths Data using IBM Cognos

DETAILED EXPLANATION OF THE PROJECT

Introduction:

The primary objective of this project is to conduct a comprehensive analysis of COVID-19 cases and associated deaths within the European Union and the European Economic Area (EU/EEA) using the powerful analytical and visualization capabilities offered by IBM Cognos. In light of the global pandemic's ongoing impact, it is imperative to gain deeper insights into the trends and variations of COVID-19 data, specifically in this region.

The project aims to delve into a nuanced understanding of the dynamics of COVID-19 by focusing on daily cases and fatalities. By harnessing the capabilities of IBM Cognos, we will not only gather, process, and visualize data but also uncover hidden patterns and trends that can provide valuable insights for policymakers, healthcare professionals, and the public at large.

Project Objectives:

The objectives are clearly defined: comparing mean values,

analyzing standard deviations, and examining variations in cases and deaths by country.

Data Collection:

Tools and Technologies: APIs (such as WHO API or ECDC API), Python for data fetching (using libraries like requests).

Algorithm/Methodology: Use APIs to fetch daily records of COVID-19 cases and deaths for each country within the EU/EEA.

Data Preparation:

Tools and Technologies: Python (pandas, numpy), Excel for data cleaning and preprocessing.

Algorithm/Methodology:

1. Cleaning the data: Remove inconsistencies and handle missing values using techniques like interpolation or dropping null values.

2. Aggregating the data: Group data by date to calculate daily mean and standard deviation values.

3. Grouping the data by country: Aggregate data by country for country-level analysis.

Analysis Approach:

Tools and Technologies: Python (pandas, numpy), statistical libraries for mean and standard deviation calculations.

Algorithm/Methodology:

1. Calculate Mean and Standard Deviation:

- Use pandas to calculate the mean and standard deviation of daily COVID-19 cases and deaths for the entire EU/EEA region.

- Calculate the mean and standard deviation for each individual country within the EU/EEA.

2. Comparison and Contrast:

- Use statistical tests (t-tests, ANOVA) to compare means and standard deviations for the EU/EEA as a whole and across different countries.

Data Visualization Design:

Tools and Technologies: IBM Cognos, Python (matplotlib, seaborn) for pre-processing if needed.

Visualization Techniques:

1.Line Charts: To display trends in daily cases and deaths over time.

- Python Example:

```
python
import matplotlib.pyplot as plt
plt.plot(dates, cases_data, label='Cases')
plt.plot(dates, deaths_data, label='Deaths')
plt.xlabel('Date')
plt.ylabel('Count')
plt.legend()
plt.show()
```

2. Bar Charts: To compare mean values between countries or regions.

3. Box Plots: To visualize the spread of data and identify outliers.

Maps: To geographically represent case and death data by country.

- Tools: IBM Cognos has built-in mapping capabilities.

Insights Derivation:

Tools and Technologies: Statistical analysis tools (Python, R), IBM Cognos for visualization.

Algorithm/Methodology:

1. Identifying countries with higher/lower mean cases and deaths: Use sorting and filtering techniques on the processed data.

2. Detecting regions with significant variations: Statistical methods can be used to identify countries with higher standard deviations.

3. Understanding the evolution of COVID-19: Time series analysis techniques to observe trends over time.

TEAM MEMBERS: -

AATHITYAN MARIRAJ (2021115002)

AJAY RAJ AGARWAL S (2021115009)

M.AHAMED IJAZ MOHAMED THAHEER (2021115008)

M. ANISH (2021115328)

ARAVINDHAN S (2021115327)