COVID 19 Cases Analysis Using Data Analytics Tool

DATE	27-10-2023
TEAM ID	**720**
PROJECT NAME	**COVID 19 CASE ANALYSIS**

TABLE OF CONTENTS

1. Introduction

The COVID-19 pandemic has had a profound impact on societies, healthcare systems, and economies worldwide. As the pandemic continues to evolve, it is crucial to conduct a comprehensive case analysis to better understand, manage, and mitigate its effects. This problem statement outlines the key aspects of the analysis required to address the on going challenges posed by COVID-19

The task at hand is to design and implement a comprehensive COVID-19 case analysis system for tracking, visualizing, and deriving insights from pandemic data

2. Problem Statement:

Designing a project to analyze COVID-19 cases and deaths using IBM cognos, The objective is to compare and contrast the mean and standard deviation of cases and deaths, which is a valuable undertaking. This project will involve data analysis, visualization, and deriving insights from the data.

3. Project Objectives:

➢ **Data Collection**: Gather covid19 data from from reliable sources across the globe. This data will include parameters such as cases, deaths, countries/Territories, and other relevant variables.

➢ **Data Analysis:** Perform exploratory data analysis (EDA) to understand the distribution of covid cases, detect outliers and identify trends and patterns.

➢ **Visualization:** Utilize data visualization techniques to represent covid19 data geospatially and temporally. This will help identify hotspots areas and understand trends over time.

➢ **Identification of Highly-Affected Areas:** Determine areas with consistently high case levels and investigate the factors contributing to this covid trend.

➢ **Predictive Model:** Develop a predictive model, likely using machine learning techniques, to estimate case and death rate . This model will be valuable for forecasting covid19 spreads, fatalities and identifying hotspot areas that require immediate attention.

4. Steps Involved in Model Evaluation:

4.1. Data Collection:

First, ensure you have access to COVID 19 case data ,Gather data from reliable sources such as government health agencies, the World Health Organization (WHO), and reputable research institutions. Collect data on the number of cases, deaths, recoveries, vaccination rates, and other relevant variables .

4.2. import Libraries:

Start by importing the necessary libraries such as numpy, pandas for data manipulations, matplotlib and seaborn for visualisations etc

IMPORT LIBRARIES

```python
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import plotly as py
import plotly.graph_objects as go
import plotly.io as pio
import plotly.express as px
```

4.3. Load the Dataset:

➢ This step involves loading your COVID 19 dataset into your Python environment. The dataset should be in a format that Pandas can easily handle, such as a CSV file.

LOADING DATASET

```
covid_data=pd.read_csv('Covid_19_cases4 (1).csv')
```

➤ The read_csv () function is used to load a CSV (Comma-Separated Values) file into a Pandas DataFrame. You specify the file path within the parentheses.

➤ The result of this operation is a DataFrame, which is a tabular data structure that's similar to a spreadsheet. It allows you to work with your data in a structured and flexible way.

4.4. Explore the Dataset

Before diving into data preprocessing, it's important to understand your dataset. You can use various Pandas functions to explore it:

data.head ():

➤ This function displays the first few rows of your dataset, giving you a glimpse of its structure.

```
covid_data.head()
```

Data.describe():

➤ It provides basic statistical information about your data, including measures like mean, standard deviation, and quartiles for numerical columns.

```
covid_data.describe()
```

data.columns():

➤ This helps you see the names of all the columns in your dataset

```
covid_data.columns
```

Data.info():

➤ This method prints information about a DataFrame including the index dtype and columns, non-null values and memory usage.

```
covid_data.info()
```

4.5. Data Pre-proccessing:

> ➤ Data preprocessing is crucial for ensuring the quality and usability of your data:

Handle Missing Values:

- Check for missing values in your dataset and decide on an appropriate strategy to handle them. You can fill missing values using methods like forward-fill, backward-fill, mean, median, or simply remove rows with missing values.

```
covid_data.isnull().sum()
```

```
covid_data.isnull().any()
```

Data Transformation:

- If your dataset contains date or time columns, convert them to the datetime data type for time-based analysis.

```
"# Example: Convert a date column to datetime
data['Date'] = pd.to_datetime (data['Date'])"
```

Data Cleaning:

- Inspect your data for inconsistencies, outliers, or irregularities. Ensure that the data is clean and standardized. This may include dealing with irregular units, correcting typos, or removing duplicates.

4.6. Predictive Model training:

> ➤ Choose Support Vector Machine (SVM) for regression and classification tasks, handling complex data relationships.
> ➤ Split the data into training and testing sets.
> ➤ Train the model using preprocessed dataset and target variables.
> ➤ Train the model on the training data and evaluate its performance on the test data using relevant metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

5. Visualization using IBM Cognos:
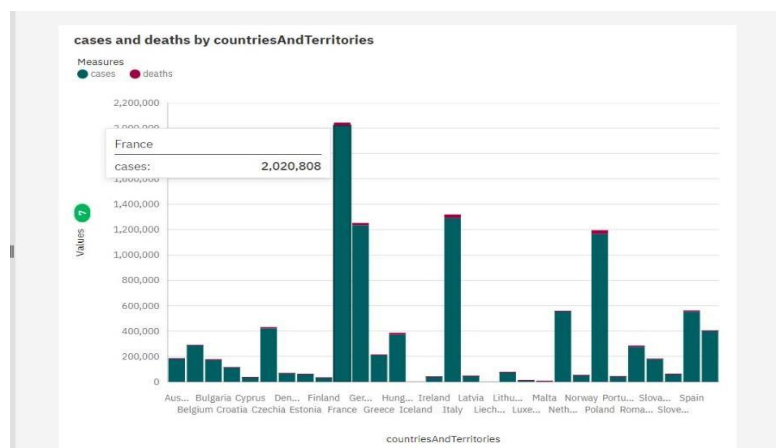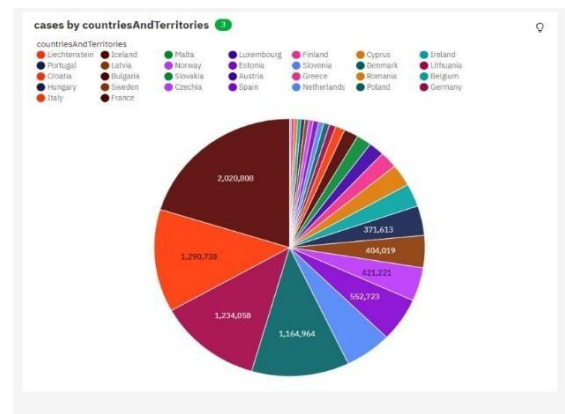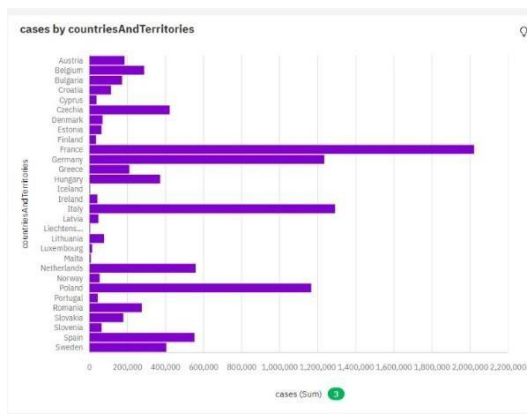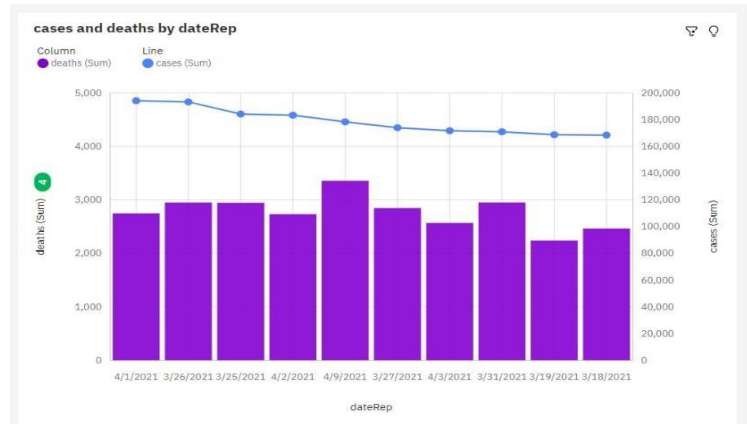
6. Conclusion:

Designing a COVID-19 test case model for data analysis is essential for understanding the patterns and trends of the disease. By analyzing various parameters such as demographics, geographic locations, testing methods, and outcomes, researchers can gain valuable insights into the spread and impact of the virus. A well-constructed COVID-19 test case model facilitates informed decision-making for healthcare professionals, policymakers, researchers. It enables them to identify high-risk areas, allocate resources efficiently, and develop targeted interventions.