Part C
Mini-Project

Name: Rupali Sunil Khairnar

Roll No: 13219

Title: Use the following covid_vaccine_statewise.csv dataset and perform following analytics on the given dataset https:

//www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv

a. Describe the dataset

b. Number of persons state wise vaccinated for first dose in India

c. Number of persons state wise vaccinated for second dose in India

d. Number of Males vaccinated d. Number of females vaccinated

Introduction:

The COVID-19 pandemic has brought unprecedented challenges globally, leading to the urgent development and distribution of vaccines. The vaccination process involves complex logistics and varies from region to region. This mini-project aims to analyze the statewise distribution of COVID-19 vaccines in the United States, employing data science and big data analytics techniques.

a. Description of the dataset:

- The dataset provides information on COVID-19 vaccination in India, categorized by states and union territories.
- It contains the following columns:
    - State_Code: Code representing the state or union territory.
    - State: Name of the state or union territory.
    - First_Dose_Administered: Number of individuals vaccinated with the first dose.
    - Second_Dose_Administered: Number of individuals vaccinated with the second dose.
    - Male(Individuals Vaccinated): Number of male individuals vaccinated.
    - Female(Individuals Vaccinated): Number of female individuals vaccinated.
    - Additional columns might include total doses administered, total beneficiaries vaccinated, etc.

b. Number of persons state-wise vaccinated for the first dose in India:

- To find this, we sum the values in the First_Dose_Administered column for each state.

- This gives us the total number of individuals vaccinated with the first dose in each state.

c. Number of persons state-wise vaccinated for the second dose in India:

- Similarly, we sum the values in the Second_Dose_Administered column for each state.
- This provides the total number of individuals vaccinated with the second dose in each state.

d. Number of Males vaccinated:

- We sum the values in the Male(Individuals Vaccinated) column across all states.
- This yields the total number of male individuals vaccinated in India.

e. Number of Females vaccinated:

- Likewise, we sum the values in the Female(Individuals Vaccinated) column across all states.
- This gives us the total number of female individuals vaccinated in India.

Objective:

The primary objective is to gain insights into the distribution patterns of COVID-19 vaccines across different states in the US. Specifically, the project aims to:

1. Analyze the distribution rates of vaccines across states.
2. Identify any disparities or discrepancies in vaccine distribution.
3. Explore factors influencing vaccine distribution, such as population density, demographics, and healthcare infrastructure.
4. Visualize the data to provide intuitive insights for stakeholders and policymakers.

Tools and Technologies:

- Programming Language: Python (utilizing libraries such as Pandas, NumPy, Matplotlib, Seaborn)
- Big Data Technologies: Apache Spark for handling large-scale datasets efficiently.
- Machine Learning Libraries: Scikit-learn, TensorFlow (if applicable)
- Data Visualization: Matplotlib, Seaborn, Plotly
- Data Storage: PostgreSQL, MongoDB (if necessary for storing and retrieving large datasets)

Expected Outcome:

1. Insights into the distribution patterns of COVID-19 vaccines across different states.
2. Identification of factors influencing vaccine distribution.
3. Visualizations and reports summarizing key findings for stakeholders and policymakers.
4. Recommendations for optimizing vaccine distribution strategies based on data-driven insights.

Conclusion:

This mini-project aims to leverage data science and big data analytics techniques to analyze the statewise distribution of COVID-19 vaccines in the US. By uncovering distribution trends, identifying influential factors, and providing actionable insights, this analysis contributes to the optimization of vaccine distribution strategies, ultimately aiding in the effective management of the pandemic.