### COVID VACCIENS ANALYSIS

**TEAM MEMBER**

**412621243020 : GOKUL P**

# Phase-3 Document Submission

**Project: Covid Vaccines Analysis**



### OBJECTIVE:

The objective of COVID-19 vaccine analysis is to assess the safety, efficacy, and overall performance of COVID-19 vaccines in preventing COVID-19 infections, reducing the severity of the disease, and minimizing its impact on public health. This analysis involves rigorous testing, clinical trials, and ongoing monitoring to ensure that vaccines are effective and safe for widespread use. Additionally, vaccine analysis aims to identify any potential adverse effects and to provide data to guide vaccination strategies and public health policies.

**Phase 3:*Development Part 1:***

**Title:** Data Preprocessing For Covid-19 Vaccine Analysis

**1.Data Collection:**

Identify reliable sources for COVID-19 vaccine data. These sources can include government health agencies, research institutions, or reputable datasets from organizations like the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC).

2. **Data Retrieval:**

Ensure that the data includes relevant attributes such as date, location (country/region), vaccine types, doses administered, and population statistics.

3. **Data Cleaning:**

Clean the dataset to ensure it is in a usable format. This may involve:

* Handling missing data by imputation or removal.
* Removing duplicates and irrelevant columns.
* Standardizing date formats and column names.
* Correcting any inconsistencies or errors in the data.

4. **Data Transformation:**

Depending on the research questions and analysis you plan to perform, you may need to create new variables or aggregate data at different levels (e.g., daily, weekly, or by country/region).

5. **Data Exploration:**

Conduct initial exploratory data analysis (EDA) to understand the characteristics of the dataset. You can create summary statistics, visualizations, and plots to gain insights into the data.

6. **Data Integration:**

If you have data from multiple sources, you may need to integrate them into a single dataset for a comprehensive analysis.

7. **Data Validation:**

Check the data for any anomalies, outliers, or inconsistencies. It's essential to ensure the data is accurate and reliable.

8. **Data Preprocessing:**

Depending on your specific analysis, you may need to perform additional preprocessing steps such as normalization, scaling, or encoding categorical variables.

9. **Data Splitting:**

If you plan to perform predictive modeling or machine learning, split the data into training and testing sets.

10. **Data Documentation:**

Document the data collection and preprocessing steps. This documentation is essential for transparency and reproducibility.

***DataSet Link:* [https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress](https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress" \t "_blank)**

***PROGRAM:***

import pandas as pd

data = pd.read\_csv(“country\_vaccinations\_by\_manufacturer.csv")

data = pd.read\_csv("country\_vaccinations.csv")

print(data.head())

print(data.info())

data = data[['country', 'date', 'total\_vaccinations', 'people\_vaccinated']]

data['date'] = pd.to\_datetime(data['date'])

data.fillna(0, inplace=True)

grouped\_data = data.groupby('country')

total\_vaccinations\_per\_country = grouped\_data['total\_vaccinations'].max()

people\_vaccinated\_per\_country = grouped\_data['people\_vaccinated'].max()

import matplotlib.pyplot as plt

us\_data = grouped\_data.get\_group('United States')

plt.figure(figsize=(10, 6))

plt.plot(us\_data['date'], us\_data['total\_vaccinations'])

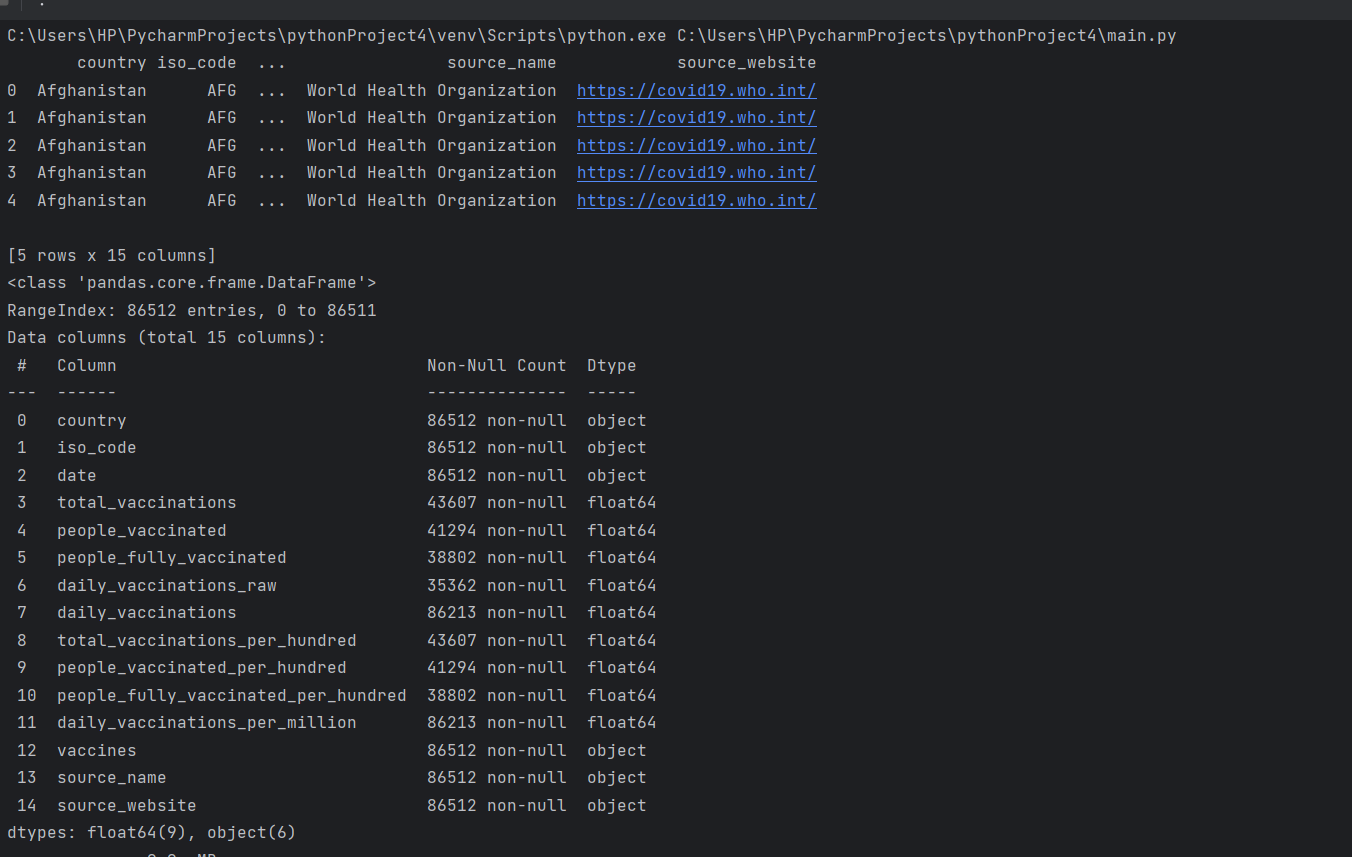
plt.title('Total Vaccinations in the United States')

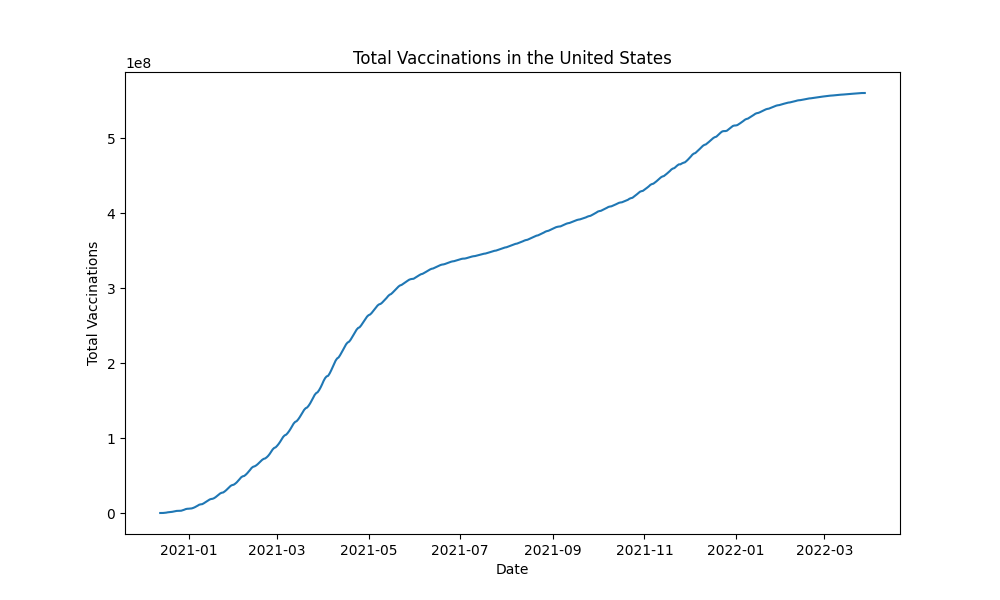
plt.xlabel('Date')

plt.ylabel('Total Vaccinations')

plt.show()

*Output:*

**

**