### COVID VACCIENS ANALYSIS

**TEAM MEMBER**

**412621243020: Gokul P**

# Phase-4 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 4:*Development Part 2:***

**Title:** Data **Exploratory data analysis, Statistical analysis, Visualization**  For Covid-19 Vaccine Analysis

*1. Exploratory Data Analysis (EDA):*

*EDA is the initial step in understanding your data. It helps identify patterns, relationships, and outliers in the dataset. Here's what you can do:*

*.*

* *Data Cleaning: Start by cleaning the data to handle missing values, duplicates, and inconsistencies.*

* *Summary Statistics: Calculate basic statistics like mean, median, standard deviation, and quartiles for numerical variables. For categorical variables, count the frequency of each category.*
* *Data Visualization:Create visualizations like histograms, box plots, and bar charts to understand the distribution of data. Scatter plots can help you find relationships between variables. Pandas and libraries like Matplotlib or Seaborn can be very helpful.*
* *Correlation Analysis:Determine the correlation between variables to identify potential relationships. You can use tools like correlation matrices or heatmaps.*

*2. Statistical Analysis:*

*Statistical analysis helps you draw meaningful insights from your data. Depending on your dataset and research questions, you may perform various statistical tests, such as:*

* *Hypothesis Testing:Use tests like t-tests, chi-squared tests, or ANOVA to compare groups or conditions. For example, you might want to test if there's a significant difference in vaccination rates between different regions.*
* *Regression Analysis:Conduct linear or logistic regression to model relationships between variables. This could help you predict vaccination rates based on other factors.*
* *Time Series Analysis: If your dataset includes time data, analyze trends and patterns over time. You can use techniques like moving averages, autocorrelation, or seasonal decomposition.*

*3. Visualization:*

*Data visualization is essential for presenting your findings effectively. Consider the following visualization techniques:*

* *Time Series Plots:Create line charts to show how vaccination rates change over time.*
* *Geospatial Maps:If you have geographic data, plot vaccination rates on maps to visualize regional disparities.*
* *Bar Charts and Pie Charts:Use these to compare vaccination rates in different categories or regions.*
* *Box Plots:Visualize the distribution of vaccination rates among different groups.*
* *Heatmaps:Show correlations between variables in a visually intuitive way.*
* *Interactive Dashboards:Tools like Tableau or Power BI can be used to create interactive dashboards to explore the data and share insights.*

***DataSet Link:***

[**https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress**](https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress)

***program:***

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv('covid\_vaccine\_data.csv')

print(df.head())

print(df.describe())

print(df.isnull().sum())

correlation\_matrix = df.corr()

print(correlation\_matrix)

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

plt.plot(df['Date'], df['Vaccinations'], marker='o', linestyle='-', color='b')

plt.xlabel('Date')

plt.ylabel('Number of Vaccinations')

plt.title('Vaccination Trends Over Time')

plt.xticks(rotation=45)

plt.grid(True)

plt.show()

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

sns.barplot(x='Region', y='VaccinationRate', data=df, palette='viridis')

plt.xlabel('Region')

plt.ylabel('Vaccination Rate (%)')

plt.title('Vaccination Rates by Region')

plt.xticks(rotation=45)

plt.show()

plt.figure(figsize=(8, 6))

sns.boxplot(x='Manufacturer', y='Efficacy', data=df, palette='Set3')

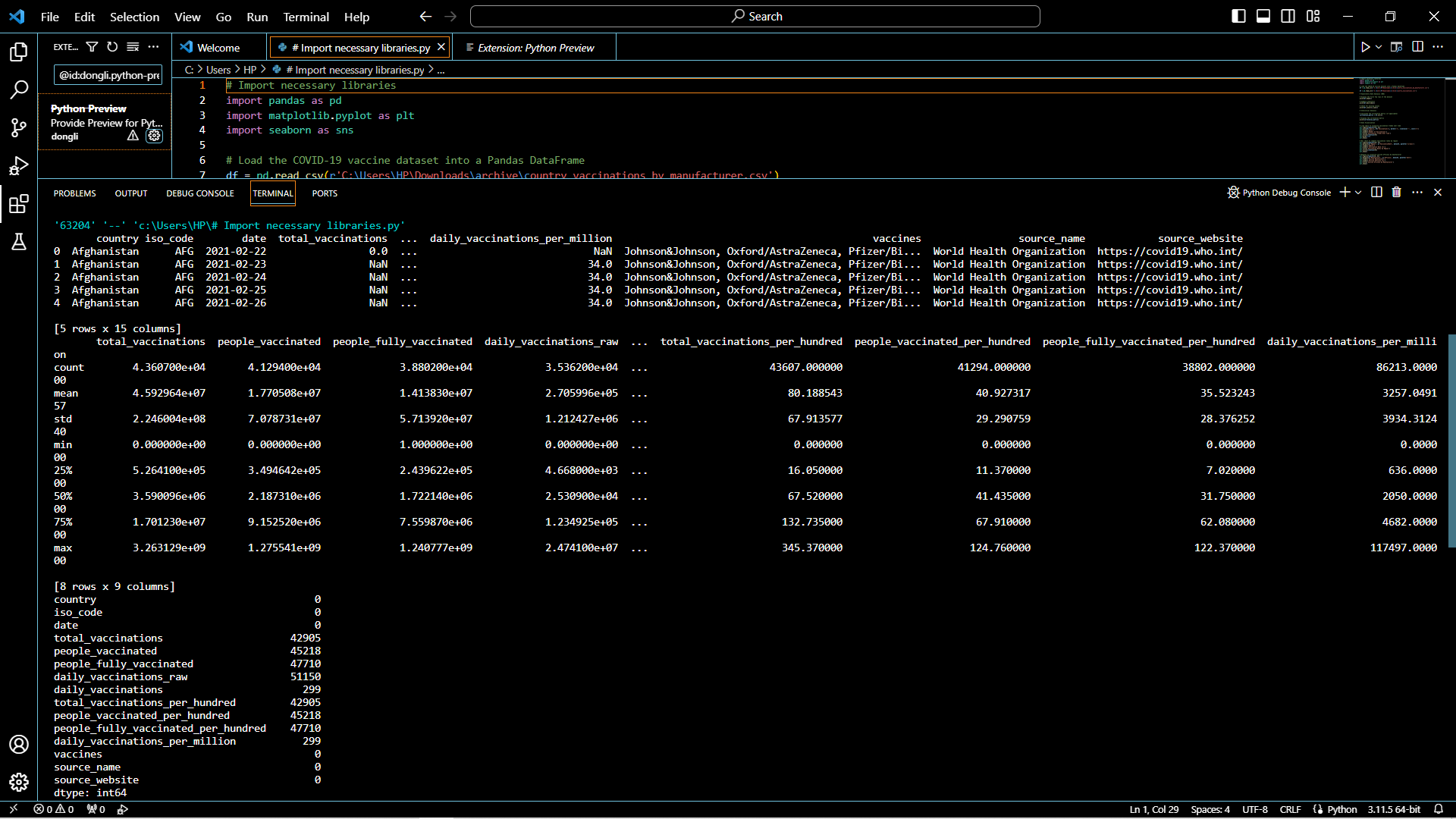
plt.xlabel('Vaccine Manufacturer')

plt.ylabel('Vaccine Efficacy (%)')

plt.title('Vaccine Efficacy by Manufacturer')

plt.show()

***OutPut :***

******

