COVID-19 VACCINE ANALYSIS

- **PROBLEM DEFINITION:**
- **Vaccine Effectiveness: Assess the real-world effectiveness of different COVID-19 vaccines, considering their ability to prevent infections, reduce severe illness and mortality, and protect against emerging variants**
- **Vaccine Distribution and Equity: Evaluate the strategies and mechanisms used for the distribution of COVID-19 vaccines, examining issues of fairness, equity, and access across diverse populations and regions**

- **DATA SET LINK**

https://drive.google.com/file/d/1DFirmNno2-hmVkTYdHk0-pDcPKRyA8WZ/view?usp=drivesdk

The dataset That We have For this tash data contains:

Country name

Vaccine name

Vaccinated people

Unvaccinated people

- **DESIGN THINKING:**
- **Set objective**
- **Collect record data**
- **Measure analyze data**
- **Budget accordingly**

SET OBJECTIVE:

Assess Vaccine Effectiveness:

Evaluate the real-world effectiveness of different COVID-19 vaccines in preventing infections, reducing severe illness and mortality, and protecting against emerging variants.

Understand Vaccine Hesitancy and Public Perception:

Analyze the factors contributing to vaccine hesitancy and skepticism, and provide insights to address public perceptions, trust, and misinformation.

Analyze Vaccine Distribution Equity:

Investigate the fairness and equity in the distribution of COVID-19 vaccines across diverse populations and regions, identifying disparities and proposing solutions.

COLLECT AND RECORD DATA

Vaccine Effectiveness Data:

Clinical trial data from vaccine manufacturers.

Real-world data from healthcare systems and public health agencies.

Global Data Sources:

Data from international organizations like the World Bank, WHO, and UNICEF for a global perspective.

Research publications and studies related to COVID-19 vaccines.

TOOL USED:

JUPYTER NOTEBOOK

COLAB

GITHUB

EXCEL(.CSV FILE)

- **CODE FOR VISUALIZATION**
- **Import matplotlib.pyplot as plt**
-
- **# Sample data (replace with your own data)**
- **vaccination_dates = [“2021-01-01”, “2021-02-01”, “2021-03-01”, “2021-04-01”]**
- **vaccination_counts = [1000, 5000, 10000, 15000]**
-
- **# Create a line plot**
- **plt.figure(figsize=(10, 5))**
- **plt.plot(vaccination_dates, vaccination_counts, marker=‘o’, linestyle=‘-’, color=‘b’)**
-
- **# Customize the plot**
- **plt.title(‘COVID-19 Vaccination Progress’)**
- **plt.xlabel(‘Date’)**
- **plt.ylabel(‘Total Vaccinations’)**
- **plt.xticks(rotation=45)**
- **plt.grid(True)**
-
- **# Show the plot**
- **plt.tight_layout()**
- **plt.show()**

- **INNOVATIVE IDEAS:**
- **Spatial Visualization of Vaccine Distribution: Create interactive maps that show the spatial distribution of COVID-19 vaccine distribution. This can help identify underserved areas and track the efficiency of vaccine delivery.**
- **Genomic Analysis: Analyze genomic data to understand the genetic factors influencing vaccine responses. This can lead to personalized vaccine recommendations.**
- **Behavioral Economics Analysis: Apply behavioral economics principles to understand the incentives and nudges that can increase vaccine acceptance and compliance.**

- **CONCLUSION**
- **The analysis not only aids in achieving herd immunity and reducing the impact of COVID-19 but also serves as a model for future vaccine distribution and public health responses to global health crises. It underscores the importance of interdisciplinary collaboration, data transparency, and scientific innovation in addressing complex public health challenges. As we continue to adapt and refine our strategies, vaccine analysis remains a critical tool in our fight against COVID-19.**

- **DATA PREPROCESSING**

Steps in Data Preprocessing

Step 1: Import the necessary libraries

```
# importing libraries
import pandas as pd
import scipy
import numpy as np
from sklearn.preprocessing import MinMaxScaler
import seaborn as sns
import matplotlib.pyplot as plt
```

Step 2: Load the dataset

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
# Load the dataset
df = pd.read_csv('/content/gdrive/MyDrive/covid-vaccine-willingness-and-people-vaccinated-by-country.csv
print(df.head())
```

```
       Entity Code         Day  people_vaccinated_per_hundred  \
0   Australia  AUS  2021-02-28                           0.13
1      Canada  CAN  2021-01-31                           2.24
2      Canada  CAN  2021-02-28                           3.61
3      Canada  CAN  2021-03-31                          13.26
4      Canada  CAN  2021-04-30                          32.67

   willingness_covid_vaccinate_this_week_pct_pop  \
0                                          52.91
1                                          54.26
2                                          53.56
3                                          51.96
4                                          36.12

   uncertain_covid_vaccinate_this_week_pct_pop  \
0                                        19.03
1                                        15.56
2                                        15.65
3                                        12.21
4                                        10.09

   unwillingness_covid_vaccinate_this_week_pct_pop
0                                            27.93
1                                            27.94
2                                            27.18
3                                            22.57
4                                            21.12
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 7 columns):
 #   Column                                         Non-Null Count  Dtype
---  ------                                         --------------  -----
 0   Entity                                         49 non-null     object
 1   Code                                           49 non-null     object
 2   Day                                            49 non-null     object
 3   people_vaccinated_per_hundred                  49 non-null     float64
 4   willingness_covid_vaccinate_this_week_pct_pop  49 non-null     float64
 5   uncertain_covid_vaccinate_this_week_pct_pop    49 non-null     float64
 6   unwillingness_covid_vaccinate_this_week_pct_pop  49 non-null   float64
dtypes: float64(4), object(3)
memory usage: 2.8+ KB
```

```
df.head()
```

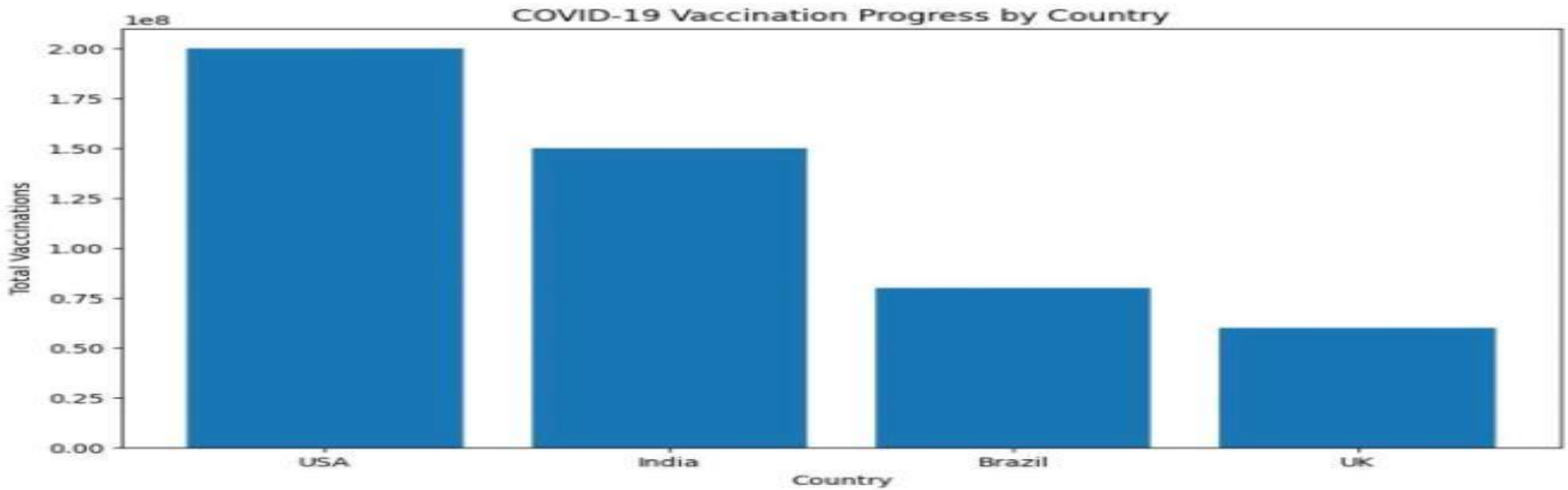
performing exploratory data analysis

```
from google.colab import drive
drive.mount('/content/drive')


import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load your dataset
data = pd.read_csv('/content/drive/MyDrive/covid-vaccine-willingness-and-people-vaccinated-by-country.csv')

# Display the first few rows of the dataset
print(data.head())

# Summary statistics
print(data.describe())

# Data distribution visualization
# Example: Histogram of a numeric column
plt.figure(figsize=(8, 6))
plt.title("Histogram of a Numeric Column")

plt.xlabel('Value')
plt.ylabel('Frequency')
plt.show()

# Example: Count plot for a categorical column
plt.figure(figsize=(8, 6))
plt.title("Count Plot of a Categorical Column")

plt.xticks(rotation=45)
plt.show()

# Correlation heatmap (for numeric columns)
correlation_matrix = data.corr()
plt.figure(figsize=(10, 8))
plt.title("Correlation Heatmap")
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", linewidths=.5)
plt.show()
```

```
      Entity Code         Day  people_vaccinated_per_hundred  \
0  Australia  AUS  2021-02-28                           0.13
1     Canada  CAN  2021-01-31                           2.24
2     Canada  CAN  2021-02-28                           3.61
3     Canada  CAN  2021-03-31                          13.26
4     Canada  CAN  2021-04-30                          32.67

   willingness_covid_vaccinate_this_week_pct_pop  \
0                                          52.91
1                                          54.26
2                                          53.56
3                                          51.96
4                                          36.12

   uncertain_covid_vaccinate_this_week_pct_pop  \
0                                        19.03
1                                        15.56
2                                        15.65
3                                        12.21
4                                        10.09

   unwillingness_covid_vaccinate_this_week_pct_pop
0                                            27.93
1                                            27.94
2                                            27.18
3                                            22.57
4                                            21.12
       people_vaccinated_per_hundred  \
count                      49.000000
mean                       11.295306
std                        12.917389
min                         0.000000
25%                         2.280000
50%                         5.490000
75%                        13.690000
max                        50.620000

       willingness_covid_vaccinate_this_week_pct_pop  \
count                                      49.000000
mean                                       47.548980
std                                        10.784062
min                                        19.570000
25%                                        40.160000
50%                                        51.270000
```

visulaization

```
import matplotlib.pyplot as plt
import pandas as pd

# Sample data
data = {
    'Country': ['USA', 'India', 'Brazil', 'UK'],
    'Total Vaccinations': [200000000, 150000000, 80000000, 60000000],
}

df = pd.DataFrame(data)

plt.figure(figsize=(10, 6))
plt.bar(df['Country'], df['Total Vaccinations'])
plt.title('COVID-19 Vaccination Progress by Country')
plt.xlabel('Country')
plt.ylabel('Total Vaccinations')
plt.show()
```

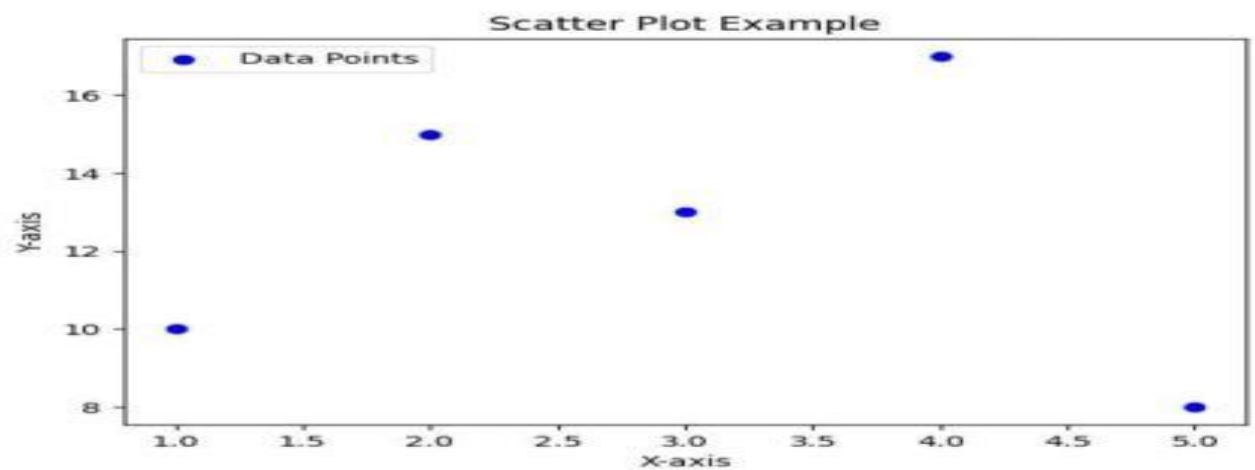


visualization for scatter plots

```
import matplotlib.pyplot as plt

# Sample data
x = [1, 2, 3, 4, 5]
y = [10, 15, 13, 17, 8]

# Create a scatter plot
plt.scatter(x, y, label='Data Points', color='blue', marker='o')

# Add labels and a title
plt.xlabel('X-axis')
```

https://colab.research.google.com/drive/1cKRKLksG6oYOqKJOCR3B_gdSSgoUm4GI#scrollTo=WmGwOv0mjcrx&printMode=true


```
plt.ylabel('Y-axis')
plt.title('Scatter Plot Example')

# Add a legend
plt.legend()

# Show the plot
plt.show()
```



statistical analysis

Double-click (or enter) to edit

step 1:To find mean ,median,standard deviation?

```
import numpy as np

data = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])

# Mean
mean_value = np.mean(data)

# Median
median_value = np.median(data)

# Standard Deviation
std_deviation = np.std(data)

print(f"Mean: {mean_value}, Median: {median_value}, Standard Deviation: {std_deviation}")
```

```
Mean: 5.5, Median: 5.5, Standard Deviation: 2.8722813232690143
```

-
- **THANK YOU**