**PROJECT TITLE**: COVID VACCINE ANALYSIS

**PHASE 3**:DATA PREPROCESSING

**1. Download the dataset:**

Download the dataset from the provided Kaggle link: [COVID-19 World Vaccination Progress](https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress).

**2. Import the necessary libraries:**

Import the necessary libraries in your programming environment, typically using Python and libraries like Pandas, NumPy, and Matplotlib for data analysis and visualization.

**3. Load the dataset:**

Load the dataset into your environment using Pandas. You can use the `read\_csv` function to read the CSV file. For example:

python

import pandas as pd

data = pd.read\_csv("your\_file\_path.csv")

**4. Explore the data** :

Explore the data to understand its structure. You can use commands like `data.head()` to see the first few rows, `data.info()` to get information about data types, and `data.describe()` to get basic statistics.

**5. Preprocess the data:**

Preprocess the data as needed for your analysis. This may include handling missing values, converting data types, and filtering the dataset to focus on relevant columns. For example, you can use `data.dropna()` to remove rows with missing values and `data['column\_name'].astype(type)` to change data types.

**6. Perform any specific data:**

Perform any specific data transformations and cleaning required for your analysis. This could include aggregating data, calculating new variables, or merging datasets.

**7. Visualize the data:**

Visualize the data using libraries like Matplotlib or Seaborn to gain insights and better understand the trends in COVID-19 vaccination progress.

**8. ANALYZE DATA:**

Once your data is prepared and analyzed, you can proceed to the next phase of your project, which might involve building models or conducting statistical analyses.



































