**COVID-19 VACCINE ANALYSIS**

**PHASE-3**

**TEAM MEMBERS:**

1.NISHANTH.S. D 2021115070

2.NITHYASRI INDHUMATHI.S 2021115072

3.NITHYASRI.S. R 2021115073

4.PRABHAVATHI.R 2021115074

5.YAZHINIYAN.P 2021115306

**INTRODUCTION:**

In the ongoing battle against the COVID-19 pandemic, the collection and preprocessing of COVID vaccine-related data have emerged as indispensable elements in shaping effective vaccination strategies and public health policies. Diverse sources, including government health agencies, the World Health Organization, research institutions, and pharmaceutical companies, contribute to the wealth of data that informs our understanding of vaccine distribution, coverage, and efficacy. Once gathered, data preprocessing becomes imperative to ensure the data's quality and suitability for analysis. This includes addressing missing values, standardizing formats, and transforming variables. Clean and well-structured data empowers analysts and researchers to extract meaningful insights, aiding in the global effort to combat the pandemic and manage the COVID-19 vaccine deployment with greater precision and impact.

**DATA PREPROCESSING OBJECTIVE:**

The primary objectives of data preprocessing for COVID vaccine analysis include ensuring data accuracy and consistency by handling missing values and outliers, standardizing data formats and units, transforming and encoding variables for analysis, and organizing data into a structured format that facilitates meaningful insights. This process sets the foundation for effective vaccine efficacy assessments, coverage rate evaluations, and informed decision-making to support global vaccination efforts in the fight against the COVID-19 pandemic.

**CODE :**

import pandas as pd

import numpy as np

'''Reading the data set using Pandas'''

df = pd.read\_csv(r"../country\_vaccinations.csv")

'''Now extracting some basic information from the dataset'''

print("The first 5 entries in the set")

print(df.head())

print("\n\nOverall information of the file")

print(df.info())

print("\n\nNull information in the file")

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

'''Now removing the rows that have null values in it'''

print("\n\nRemoving rows with null values")

df.dropna(inplace=True)

df = df.reset\_index(drop=True)

print("Removed")

print("Saving the cleaned CSV in a new folder named /cleaned\_csv")

df.to\_csv(r"../cleaned\_csv/preprocessed\_country\_vaccination.csv",index=False)

'''Now creating a new CSV to store the Vaccines used in a country'''

columns\_of\_interest = ['iso\_code','vaccines']

selected\_columns\_df = df[columns\_of\_interest]

unique\_values\_df = selected\_columns\_df.drop\_duplicates()

unique\_values\_df = unique\_values\_df.reset\_index(drop=True)

print("Created a helping dataset to know about the vaccines used in each county...")

unique\_values\_df.to\_csv(r"../cleaned\_csv/vacc.csv")

**OUTPUT**

The first 5 entries in the set

country iso\_code ... source\_name source\_website

0 Afghanistan AFG ... World Health Organization https://covid19.who.int/

1 Afghanistan AFG ... World Health Organization https://covid19.who.int/

2 Afghanistan AFG ... World Health Organization https://covid19.who.int/

3 Afghanistan AFG ... World Health Organization https://covid19.who.int/

4 Afghanistan AFG ... World Health Organization https://covid19.who.int/

[5 rows x 15 columns]

* Overall information of the file

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 86512 entries, 0 to 86511

Data columns (total 15 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 country 86512 non-null object

1 iso\_code 86512 non-null object

2 date 86512 non-null object

3 total\_vaccinations 43607 non-null float64

4 people\_vaccinated 41294 non-null float64

5 people\_fully\_vaccinated 38802 non-null float64

6 daily\_vaccinations\_raw 35362 non-null float64

7 daily\_vaccinations 86213 non-null float64

8 total\_vaccinations\_per\_hundred 43607 non-null float64

9 people\_vaccinated\_per\_hundred 41294 non-null float64

10 people\_fully\_vaccinated\_per\_hundred 38802 non-null float64

11 daily\_vaccinations\_per\_million 86213 non-null float64

12 vaccines 86512 non-null object

13 source\_name 86512 non-null object

14 source\_website 86512 non-null object

dtypes: float64(9), object(6)

memory usage: 9.9+ MB

None

* Null information in the file

country 0

iso\_code 0

date 0

total\_vaccinations 42905

people\_vaccinated 45218

people\_fully\_vaccinated 47710

daily\_vaccinations\_raw 51150

daily\_vaccinations 299

total\_vaccinations\_per\_hundred 42905

people\_vaccinated\_per\_hundred 45218

people\_fully\_vaccinated\_per\_hundred 47710

daily\_vaccinations\_per\_million 299

vaccines 0

source\_name 0

source\_website 0

dtype: int64

Removing rows with null values

Removed

**EXPLANATION:**

The provided Python code is designed to prepare a COVID vaccine dataset for analysis. It starts by loading the dataset from a CSV file and then conducts an initial exploration, displaying the first five entries and providing an overview of the dataset's structure. Missing values are identified and subsequently removed, ensuring that only complete and reliable data remains. The cleaned dataset is then saved to a new file for future analysis. Additionally, the code creates a separate dataset to track the unique combinations of 'iso\_code' and 'vaccines' for each country, shedding light on the vaccines used across different regions. By conducting these data preprocessing steps, the code paves the way for meaningful insights into vaccine efficacy, distribution, and coverage, contributing to more effective global efforts to combat the COVID-19 pandemic.

**CONCLUSION:**

The data preprocessing steps play a vital role in preparing the dataset for COVID vaccine analysis. By removing missing data and creating a dataset to understand vaccine usage in different countries, the code ensures that the analysis can be conducted with high-quality, complete data. These preprocessed datasets are essential for obtaining valuable insights into vaccine efficacy, distribution, and coverage, aiding global efforts to combat the COVID-19 pandemic more effectively.