

COVID-19 VACCINES ANALYSIS

Phase 1: Problem Definition and Design Thinking

PROBLEM DEFINITION:

Analyzing COVID-19 vaccines entails accessing their safety, efficacy and performance against emerging variants. Evaluating vaccine distribution, strategies, addressing vaccine hesitancy and understanding their long-term effects are pivotal. The goal is to inform evidence-based policies, optimize vaccine deployment and ensure equitable global access to combat the COVID-19 pandemic effectively.

DESIGN THINKING:

1.Data collection:

Globally there have been 770,875,433 confirmed cases of COVID-19, including 6,959,316 deaths, reported to WHO. As of 19 September 2023, a total of 13,505,262,477 vaccine doses have been administered.

2.Data Preprocessing:

Data Importing- In desktop with the help of the get data option import the CSV data which is named as `country_vaccination` and clicked load option.

```
import pandas as pd
import numpy as np
```

Data Cleaning – After loading the data and after analyzing the data I understood that there are 86512 rows and 15 columns. And in that some of the columns contained null values I have replaced the null values by 0 with the use of replace functions and started working on the data.

```
df = pd.read_csv('country_vaccination.csv')
new_df = df.dropna()
print(new_df.to_string())
```

3.Exploratory data Analysis:

Descriptive Statistics, which is a way of giving a brief overview of the dataset we are dealing with, including some measures and features of the sample.

```
import pandas as pd
Df = pd.read_csv ("country_vaccinations.csv")
pf.head()
df.info()
df.describe()
df.duplicated.sum()
```

3.Statistical analysis:

statistical hypothesis testing, apply estimation statistics and interpret the results. We will also validate this with the findings from part one. We will apply both parametric and non-parametric tests.

```
import numpy as np
import pandas as pd
import scipy.stats as stats

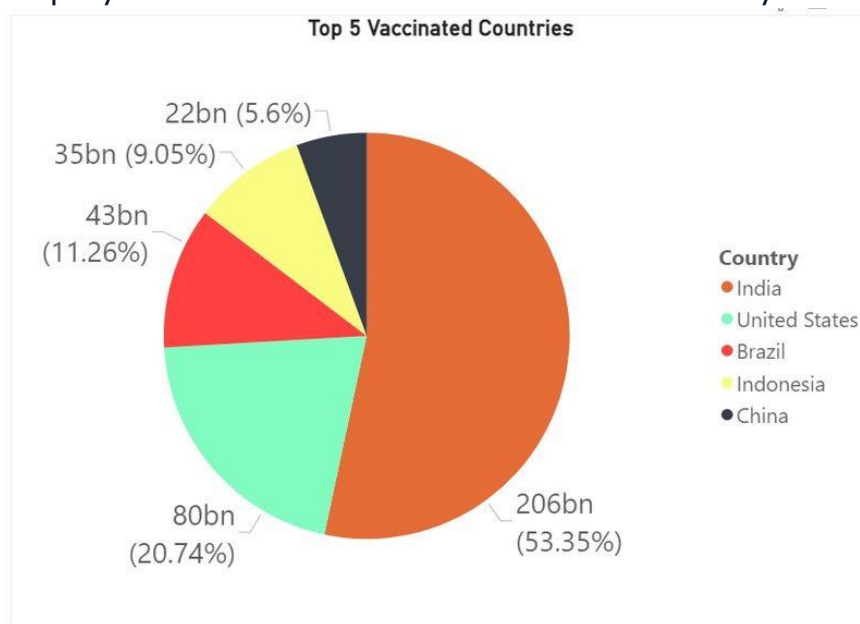
from sklearn.preprocessing import LabelEncoder
from sklearn import feature_selection

import seaborn as sns
from matplotlib import pyplot
from statsmodels.graphics.gofplots import qqplot

sns.set()
```

5. Data Virtualization:

In visualization part with the help of power BI desktop software I have used different kinds of charts, graphs, cards and table to display the data in the format which will be easy to understand.



```
import seaborn
import pandas
data = pandas.read_csv("country_vaccinated.csv")
seaborn.scatterplot(data['deaths'], data['affected'])
```

6. Insights and Recommendations:

Real-world vaccine effectiveness varies by factors such as age, variants, and duration of immunity.

Vaccine hesitancy must be addressed through targeted, transparent, and empathetic communication.

Global coordination is essential for equitable vaccine distributions.

Continuously monitor and adapt vaccination strategies based on real-world data.

Invest in public education campaigns to combat vaccine hesitancy.

Foster international collaboration to ensure widespread vaccine access.