COVID Vaccines Analysis

NAME: BHARATHI.S
NAANMUDHALVAN ID:au920821104011
PROJECT 5

Problem definition:

The problem of COVID-19 vaccine analysis involves examining various aspects of COVID-19 vaccines to assess their safety,efficacy,distribution and impact. This analysis can encompass several specific objectives.

To deal with COVID-19 various countries have made many efforts, including the research and development of vaccines.

1.Safety Assessment:

Evaluate the safety profile of COVID-19 vaccines by analyzing adverse events reported post vaccination.

2.Efficacy Assessment:

Measure the effectiveness of vaccines in preventing COVID-19 infection,severe illness and transmission.

3.Distribution and Access:

Analyze the distribution and accessibility of vaccines globally,including issues related to vaccine distribution equity.

4.Vaccine Variants:

Assess how vaccines perform against emerging COVID-19 variants.

5.Vaccine Hesitancy:

Study factors contributing to vaccine hesitancy and develop strategies to address it.

6. Impact on Public Health:

Analyze the overall imapact of vaccination on reducing COVID-19 cases, hospitalization and deaths.

7.Long –Term Effects:

Investigate potential long-term effects of COVID-19 vaccination.

8.Economic Impact:

Assess the economic implications of vaccination, including its role in reopening economies.

9.Methods:

This article reviewed the existing literature to see development of the COVID-19 vaccine.

10.Result:

We found the different type of vaccines had their own advantages and disadvantages. At the same time, the side effects of vaccines, the dose of vaccination , the evaluation of efficacy and the application of the vaccine were all things are studying.

Design and Thinking:

Designing and analyzing a COVID-19 vaccine involves a complex, multi-stage process that requires careful planning, scientific expertise and rigorous testing.

1.Research and Development Phase:

Researchers must identify specific antigens (usually proteins) on the virus that can be targeted by the immune system.

2.Clinical Trails:

Testing the vaccine in a small group of healthy volubteers to assess safety and dosage.

Expanding the trail to a large group to evaluate safety, immunogenicity and optimal dosing.

3.Regulatory Approval:

Submitting comprehensive data to regulatory agencies(eg. FDA, EMA) for approval.

4.Manufacturing and Distribution:

Establishing large-scale manufacturing facilities to produce the vaccine at the required volume.

Ensuring the cold chain for vaccines that require specific storage conditions.

5.Public Engagement:

Communicating transparently with the public, addressing concerns and providing accurate information about the vaccine.

6.Ethical and Legal Considerations:

Ensuring that research and distribution are conducted ethically and in compliance with local and international laws.

Throughtout this process, collaboration between scientists, healthcare professionals, governments and phaemaceutical companies is crucial.

7.Adaptation and Response:

Being prepared to adapt the vaccine or develop new ones if new variants of the virus emerge.

Collaborating with international health organization to respond to global health crises.

COVID Vaccine Innovation Analysis

Introduction:

- The COVID-19 pandemic has underscored the critical importance of developing and distributing effective vaccines to combat the spread of the virus. This analysis focuses on a novel vaccine innovation project aimed at addressing the ongoing challenges presented by COVID-19.

Project Scope and Objectives:

Scope:

- This project aims to develop a next-generation COVID-19 vaccine that is highly effective, easily distributable, and adaptable to potential variants.

Objectives:

- Develop a vaccine that provides robust and lasting immunity against COVID-19.
- Ensure the vaccine is suitable for all age groups.
- Create a distribution plan to ensure equitable access on a global scale.
- Develop a mechanism for rapid response to emerging variants.
- Establish safety and efficacy benchmarks in line with regulatory agencies.

Detailed Project Plan:

Research and Development (R&D) Phase:

- Conduct comprehensive literature review on existing COVID-19 vaccines.
- Formulate and test vaccine candidates.
- Identify the most promising candidate based on safety and efficacy.

Clinical Trials:

- Phase I: Small-scale human trials to assess safety.
- Phase II: Expanded trials to determine efficacy.
- Phase III: Large-scale trials involving diverse populations.

Regulatory Approval:

- Compile trial data and submit for regulatory review.
- Collaborate with regulatory agencies for expedited approvals.

Manufacturing and Distribution:

- Establish manufacturing facilities.
- Secure distribution agreements with global partners.
- Develop distribution infrastructure for global reach.

Surveillance and Variant Monitoring:

- Implement surveillance programs to detect new variants.
- Adapt the vaccine as necessary to address emerging variants.

Public Awareness and Education:

- Launch public awareness campaigns to promote vaccine uptake.
- Educate healthcare professionals about the vaccine's benefits.

➢ Scaling and Continuous Improvement:

- Increase production capacity to meet global demand.
- Continue monitoring vaccine effectiveness and safety.

❖ **Conclusion**:

➢ In conclusion, this innovative COVID vaccine project seeks to address the COVID-19 pandemic comprehensively. By focusing on cutting-edge research, rigorous testing, equitable distribution, and adaptability to variants, this project aims to provide a powerful tool in the fight against the virus. Successful execution of this project will not only mitigate the immediate crisis but also establish a framework for addressing future global health challenges.

Collect And Preprocess The COVID Vaccine Analysis

Steps for data Analysis

1.Data collection

2. Data Exploration

3.Data Preprocessing

4.Descriptive statistics

5.Save processed data

6.Data Analysis

➢ Data Collection:

- ➢ Find a reliable source for COVID-19 vaccine data.Common sources include government health agencies, reputable research institutions, or datasets on platforms like Kaggle.
- ➢ Download or access the dataset in a format that's compatible with your analysis tools (e.g., CSV, Excel, JSON).

```
import pandas as pd
data_path = "C:/Users/My pc/Desktop/COVID.csv"
df = pd.read_csv(data_path)
print(df)
```

```
Type "copyright", "credits" or "license" for more information.

IPython 8.2.0 -- An enhanced Interactive Python.

In [1]: runfile('F:/data/untitled0.py', wdir='F:/data')
            location        date             vaccine  total_vaccinations
0          Argentina  2020-12-29             Moderna                   2
1          Argentina  2020-12-29  Oxford/AstraZeneca                   3
2          Argentina  2020-12-29   Sinopharm/Beijing                   1
3          Argentina  2020-12-29           Sputnik V               20481
4          Argentina  2020-12-30             Moderna                   2
...              ...         ...                 ...                 ...
35618  European Union  2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union  2022-03-29     Pfizer/BioNTech           600519998
35620  European Union  2022-03-29   Sinopharm/Beijing             2301516
35621  European Union  2022-03-29             Sinovac                1809
35622  European Union  2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
```

➢ Data Exploration:

- ❖ Load the dataset using a data manipulation library such as Pandas for Python or a tool that fits your preference.Examine the dataset's structure, column names, and the type of information it contains.

```
#step2:Data Exploration
print(df.head())
print(df.info())
```

```
35622  European Union  2022-03-29          Sputnik V            1845103

[35623 rows x 4 columns]
    location        date            vaccine  total_vaccinations
0  Argentina  2020-12-29            Moderna                   2
1  Argentina  2020-12-29  Oxford/AstraZeneca                   3
2  Argentina  2020-12-29   Sinopharm/Beijing                   1
3  Argentina  2020-12-29           Sputnik V               20481
4  Argentina  2020-12-30            Moderna                   2
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35623 entries, 0 to 35622
Data columns (total 4 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   location            35623 non-null  object
 1   date                35623 non-null  object
 2   vaccine             35623 non-null  object
 3   total_vaccinations  35623 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
```

➢ Data Preprocessing:

- ❖ Handle missing data: Check for missing values and decide on an appropriate strategy, like imputation or removal of incomplete rows.

```
# Step 3: Data Preprocessing
df = df.dropna()
df['date'] = pd.to_datetime(df['date'])
print(df)
```

```
 0   location            35623 non-null  object
 1   date                35623 non-null  object
 2   vaccine             35623 non-null  object
 3   total_vaccinations  35623 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
            location       date             vaccine  total_vaccinations
0          Argentina 2020-12-29             Moderna                   2
1          Argentina 2020-12-29  Oxford/AstraZeneca                   3
2          Argentina 2020-12-29   Sinopharm/Beijing                   1
3          Argentina 2020-12-29           Sputnik V               20481
4          Argentina 2020-12-30             Moderna                   2
...              ...        ...                 ...                 ...
35618  European Union 2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union 2022-03-29     Pfizer/BioNTech           600519998
35620  European Union 2022-03-29   Sinopharm/Beijing             2301516
35621  European Union 2022-03-29             Sinovac                1809
35622  European Union 2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
```

➢ Descriptive Statistics:

- ❖ Calculate basic statistics like mean, median, and standard deviation to understand the central tendencies and variability of the data.

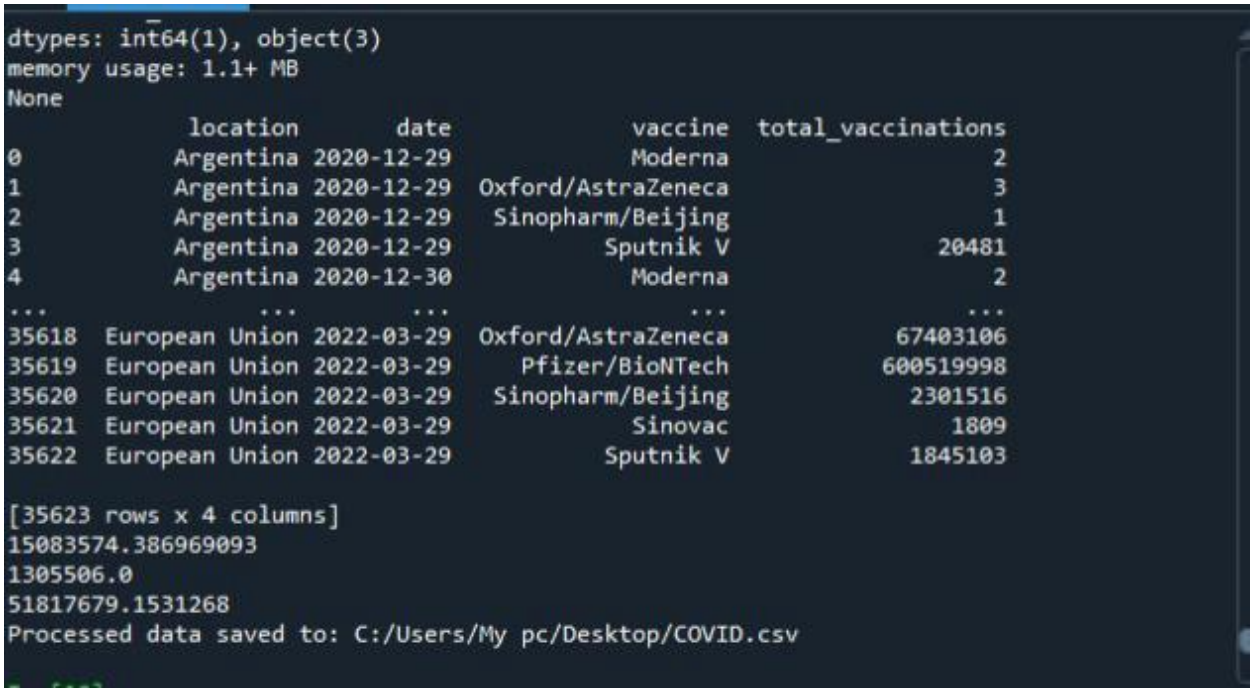
```python
# Step 4: Descriptive Statistics
mean = df['total_vaccinations'].mean()
median = df['total_vaccinations'].median()
std_dev = df['total_vaccinations'].std()
print(mean)
print(median)
print(std_dev)
```

```
 3   total_vaccinations  35623 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
            location       date             vaccine  total_vaccinations
0          Argentina 2020-12-29             Moderna                   2
1          Argentina 2020-12-29  Oxford/AstraZeneca                   3
2          Argentina 2020-12-29   Sinopharm/Beijing                   1
3          Argentina 2020-12-29           Sputnik V               20481
4          Argentina 2020-12-30             Moderna                   2
...              ...        ...                 ...                 ...
35618  European Union 2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union 2022-03-29     Pfizer/BioNTech           600519998
35620  European Union 2022-03-29   Sinopharm/Beijing             2301516
35621  European Union 2022-03-29             Sinovac                1809
35622  European Union 2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
15083574.386969093
1305506.0
51817679.1531268
```

➢ Save Processed Data:

- ❖ After preprocessing, save the clean dataset to ensure you can work with it in future analysis without repeating these steps.

```
# Step 5: Save Processed Data
processed_data_path = "C:/Users/My pc/Desktop/COVID.csv"
df.to_csv(processed_data_path, index=False)
print("Processed data saved to:", processed_data_path)
```



```
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
            location       date             vaccine  total_vaccinations
0          Argentina 2020-12-29             Moderna                   2
1          Argentina 2020-12-29  Oxford/AstraZeneca                   3
2          Argentina 2020-12-29   Sinopharm/Beijing                   1
3          Argentina 2020-12-29           Sputnik V               20481
4          Argentina 2020-12-30             Moderna                   2
...              ...        ...                 ...                 ...
35618  European Union 2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union 2022-03-29     Pfizer/BioNTech           600519998
35620  European Union 2022-03-29   Sinopharm/Beijing             2301516
35621  European Union 2022-03-29             Sinovac                1809
35622  European Union 2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
15083574.386969093
1305506.0
51817679.1531268
Processed data saved to: C:/Users/My pc/Desktop/COVID.csv
```

➢ Data Analysis:

- ❖ Once your data is preprocessed, you can start your analysis, which could include trends, correlations, and more, depending on your specific research questions.

```python
import pandas as pd
data_path = "C:/Users/My pc/Desktop/COVID.csv"
df = pd.read_csv(data_path)
print(df)


#step2:Data Exploration
print(df.head())
print(df.info())


# Step 3: Data Preprocessing
df = df.dropna()
df['date'] = pd.to_datetime(df['date'])
print(df)


# Step 4: Descriptive Statistics
mean = df['total_vaccinations'].mean()
median = df['total_vaccinations'].median()
std_dev = df['total_vaccinations'].std()
print(mean)
print(median)
print(std_dev)


# Step 5: Save Processed Data
processed_data_path = "C:/Users/My pc/Desktop/COVID.csv"
df.to_csv(processed_data_path, index=False)
print("Processed data saved to:", processed_data_path)


# step 6:Data analysis
total_vaccinations = df['total_vaccinations'].sum()
print("Total Vaccinations Administered:", total_vaccinations)
```

```
memory usage: 1.1+ MB
None
            location       date             vaccine  total_vaccinations
0          Argentina 2020-12-29            Moderna                   2
1          Argentina 2020-12-29  Oxford/AstraZeneca                   3
2          Argentina 2020-12-29   Sinopharm/Beijing                   1
3          Argentina 2020-12-29           Sputnik V               20481
4          Argentina 2020-12-30            Moderna                   2
...              ...        ...                 ...                 ...
35618  European Union 2022-03-29  Oxford/AstraZeneca            67403106
35619  European Union 2022-03-29     Pfizer/BioNTech           600519998
35620  European Union 2022-03-29   Sinopharm/Beijing             2301516
35621  European Union 2022-03-29             Sinovac                1809
35622  European Union 2022-03-29           Sputnik V             1845103

[35623 rows x 4 columns]
15083574.386969093
1305506.0
51817679.1531268
Processed data saved to: C:/Users/My pc/Desktop/COVID.csv
Total Vaccinations Administered: 537322170387
```

Exploratory data analysis, Statistical analysis, Visualization

Abstraction:

- The goal of this research is to analyze data on vaccinations, vaccination administration, and forecasting vaccination rates on a country-by-country basis for the general public, policymakers, vaccine manufacturers, national governments, and international governments to better understand the current state of COVID-19 vaccination.
- Summarize your findings in a clear and concise manner.
- Highlight key insights and trends discovered during EDA and statistical analysis.
- Provide actionable recommendations based on your analysis, such as targeting vaccination campaigns in regions with low vaccination rates.

Data set link:

https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress

Output:

```
location,date,vaccine,total_vaccinations
Argentina,29-12-2020,Moderna,2
Argentina,29-12-2020,Oxford/AstraZeneca,3
Argentina,29-12-2020,Sinopharm/Beijing,1
Argentina,29-12-2020,Sputnik V,20481
Argentina,30-12-2020,Moderna,2
Argentina,30-12-2020,Oxford/AstraZeneca,3
Argentina,30-12-2020,Sinopharm/Beijing,1
Argentina,30-12-2020,Sputnik V,40583
Argentina,31-12-2020,Moderna,2
Argentina,31-12-2020,Oxford/AstraZeneca,3
Argentina,31-12-2020,Sinopharm/Beijing,1
Argentina,31-12-2020,Sputnik V,43388
Argentina,01-01-2021,Moderna,2
Argentina,01-01-2021,Oxford/AstraZeneca,5
Argentina,01-01-2021,Sinopharm/Beijing,1
Argentina,01-01-2021,Sputnik V,43513
Argentina,02-01-2021,Moderna,2
Argentina,02-01-2021,Oxford/AstraZeneca,6
Argentina,02-01-2021,Sinopharm/Beijing,1
Argentina,02-01-2021,Sputnik V,46824
Argentina,03-01-2021,Moderna,2
Argentina,03-01-2021,Oxford/AstraZeneca,6
Argentina,03-01-2021,Sinopharm/Beijing,1
Argentina,03-01-2021,Sputnik V,47266
Argentina,04-01-2021,Moderna,2
Argentina,04-01-2021,Oxford/AstraZeneca,6
Argentina,04-01-2021,Sinopharm/Beijing,1
Argentina,04-01-2021,Sputnik V,57726
Argentina,05-01-2021,Moderna,2
Argentina,05-01-2021,Oxford/AstraZeneca,6
Argentina,05-01-2021,Sinopharm/Beijing,5
Argentina,05-01-2021,Sputnik V,68445
Argentina,06-01-2021,Moderna,2
Argentina,06-01-2021,Oxford/AstraZeneca,6
Argentina,06-01-2021,Sinopharm/Beijing,8
```

- **Exploratory data analysis :**
  - Conduct descriptive statistics like mean, median, and standard deviation.

- Explore the distribution of vaccination rates across different regions or demographics. Identify trends and patterns in the data through plots, such as histograms, box plots, and scatter plots. Perform correlation analysis to understand relationships between variables (e.g., vaccination rates and infection rates).

```
import pandas as pd
data_path="C:/Users/Lenovo/Desktop/covid/ex1.csv"
df=pd.read_csv(data_path)
print(df)
```

```
            location        date             vaccine  total_vaccinations
0          Argentina  29-12-2020             Moderna                   2
1          Argentina  29-12-2020  Oxford/AstraZeneca                   3
2          Argentina  29-12-2020   Sinopharm/Beijing                   1
3          Argentina  29-12-2020           Sputnik V               20481
4          Argentina  30-12-2020             Moderna                   2
...              ...         ...                 ...                 ...
35618  European Union  29-03-2022  Oxford/AstraZeneca            67403106
35619  European Union  29-03-2022     Pfizer/BioNTech           600519998
35620  European Union  29-03-2022   Sinopharm/Beijing             2301516
35621  European Union  29-03-2022             Sinovac                1809
35622  European Union  29-03-2022           Sputnik V             1845103

[35623 rows x 4 columns]
```

```
import pandas as pd
data_path="C:/Users/Lenovo/Desktop/covid/ex1.csv"
df=pd.read_csv(data_path)
print(df)
#step2:Exploratory data analysis

#head() will display the top 5 observations of the dataset
print(df.head())

#tail() will display the last 5 observations of the dataset
print(df.tail())

#info() helps to information about the data
print(df.info())

#describe() method return description of the data
print(df.describe())

#Missing values calculation
print(df.isnull().sum())

#check for Duplication
print(df.nunique())
```

```
 0   location            35623 non-null  object
 1   date                35623 non-null  object
 2   vaccine             35623 non-null  object
 3   total_vaccinations  35623 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
       total_vaccinations
count        3.562300e+04
mean         1.508357e+07
std          5.181768e+07
min          0.000000e+00
25%          9.777600e+04
50%          1.305506e+06
75%          7.932423e+06
max          6.005200e+08
location              0
date                  0
vaccine               0
total_vaccinations    0
dtype: int64
location                43
date                   473
vaccine                 10
total_vaccinations    29210
dtype: int64
```

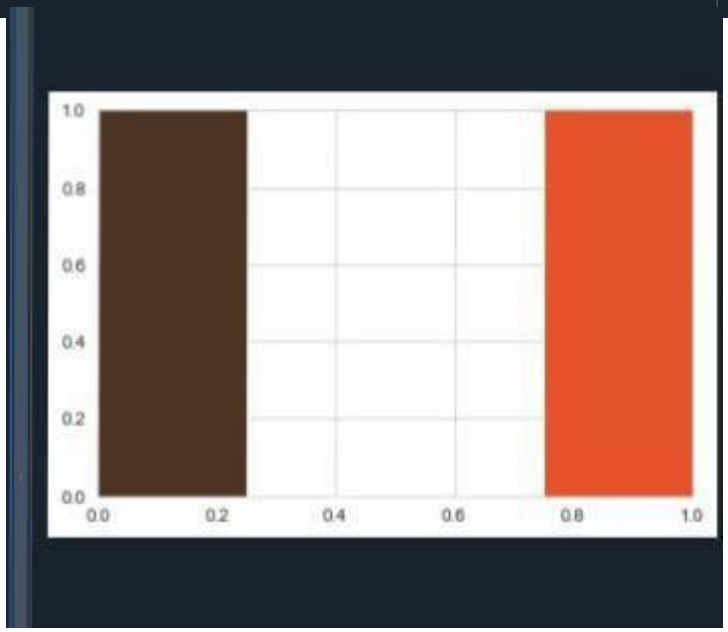
```
[35623 rows x 4 columns]
    location        date            vaccine  total_vaccinations
0  Argentina  29-12-2020            Moderna                   2
1  Argentina  29-12-2020  Oxford/AstraZeneca                   3
2  Argentina  29-12-2020   Sinopharm/Beijing                   1
3  Argentina  29-12-2020          Sputnik V                20481
4  Argentina  30-12-2020            Moderna                   2
            location        date            vaccine  total_vaccinations
35618  European Union  29-03-2022  Oxford/AstraZeneca            67403106
35619  European Union  29-03-2022     Pfizer/BioNTech           600519998
35620  European Union  29-03-2022   Sinopharm/Beijing             2301516
35621  European Union  29-03-2022             Sinovac                1809
35622  European Union  29-03-2022          Sputnik V             1845103
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35623 entries, 0 to 35622
Data columns (total 4 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   location            35623 non-null  object
 1   date                35623 non-null  object
 2   vaccine             35623 non-null  object
 3   total_vaccinations  35623 non-null  int64
dtypes: int64(1), object(3)
memory usage: 1.1+ MB
None
```

```
            location        date             vaccine  total_vaccinations
0          Argentina  29-12-2020             Moderna                   2
1          Argentina  29-12-2020  Oxford/AstraZeneca                   3
2          Argentina  29-12-2020   Sinopharm/Beijing                   1
3          Argentina  29-12-2020           Sputnik V               20481
4          Argentina  30-12-2020             Moderna                   2
...              ...         ...                 ...                 ...
35618  European Union  29-03-2022  Oxford/AstraZeneca            67403106
35619  European Union  29-03-2022     Pfizer/BioNTech           600519998
35620  European Union  29-03-2022   Sinopharm/Beijing             2301516
35621  European Union  29-03-2022             Sinovac                1809
35622  European Union  29-03-2022           Sputnik V             1845103

[35623 rows x 4 columns]
    location        date             vaccine  total_vaccinations
0  Argentina  29-12-2020             Moderna                   2
1  Argentina  29-12-2020  Oxford/AstraZeneca                   3
2  Argentina  29-12-2020   Sinopharm/Beijing                   1
3  Argentina  29-12-2020           Sputnik V               20481
4  Argentina  30-12-2020             Moderna                   2
            location        date             vaccine  total_vaccinations
35618  European Union  29-03-2022  Oxford/AstraZeneca            67403106
```

❖ STATISTICAL ANALYSIS:

- Conduct hypothesis tests to assess the significance of differences in vaccination rates between groups (e.g., age groups or geographical regions).
- Apply regression analysis to understand the impact of various factors on vaccination rates.
- Calculate confidence intervals to estimate the precision of your findings.

```
29
30  colors = ['#4D3425', '#E4512B']
31  ax = (df['total_vaccinations'].value_counts() * 100.0 / len(df)).plot(kind='bar', stacked=True, rot=0, color=col
32  ax.set_ylabel('% date')
33  ax.set_xlabel('total_vaccinations')
34  ax.set_ylabel('% vaccine')
35  ax.set_title('covid 19 vaccine analysis')
36
```



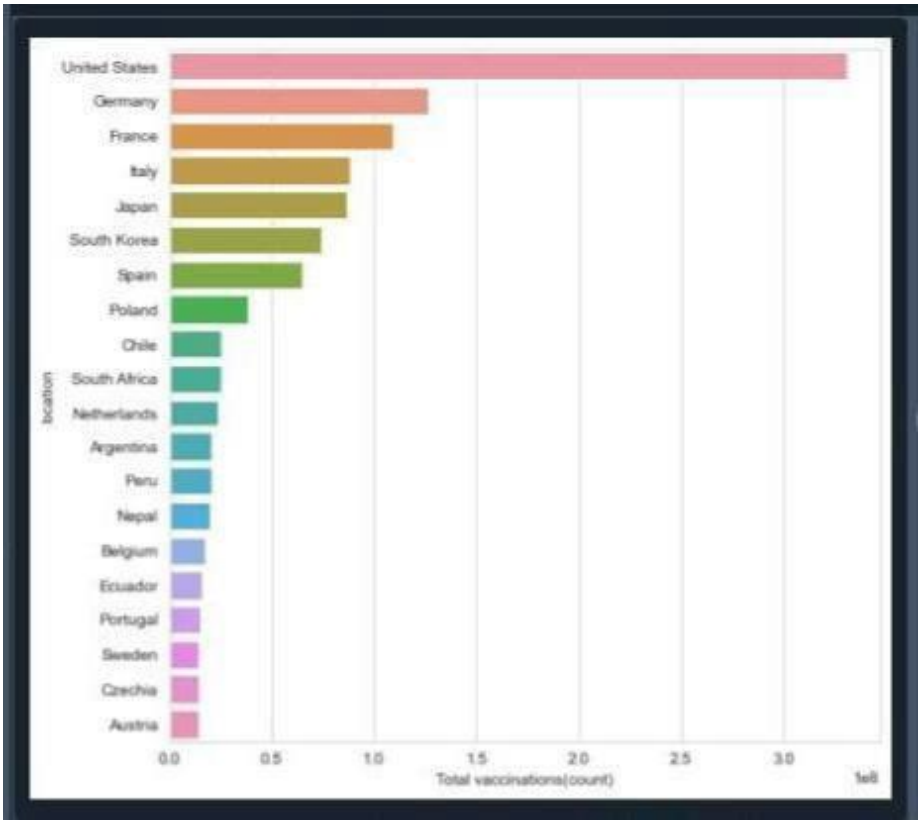
❖Visualization

- Create visually engaging plots to communicate your findings effectively:
- Bar charts or stacked bar charts to compare vaccination rates by region or age group.
- Time series plots to track vaccination progress over time.
- Heatmaps to visualize correlations between variables.
- Utilize interactive dashboards for dynamic exploration of data.

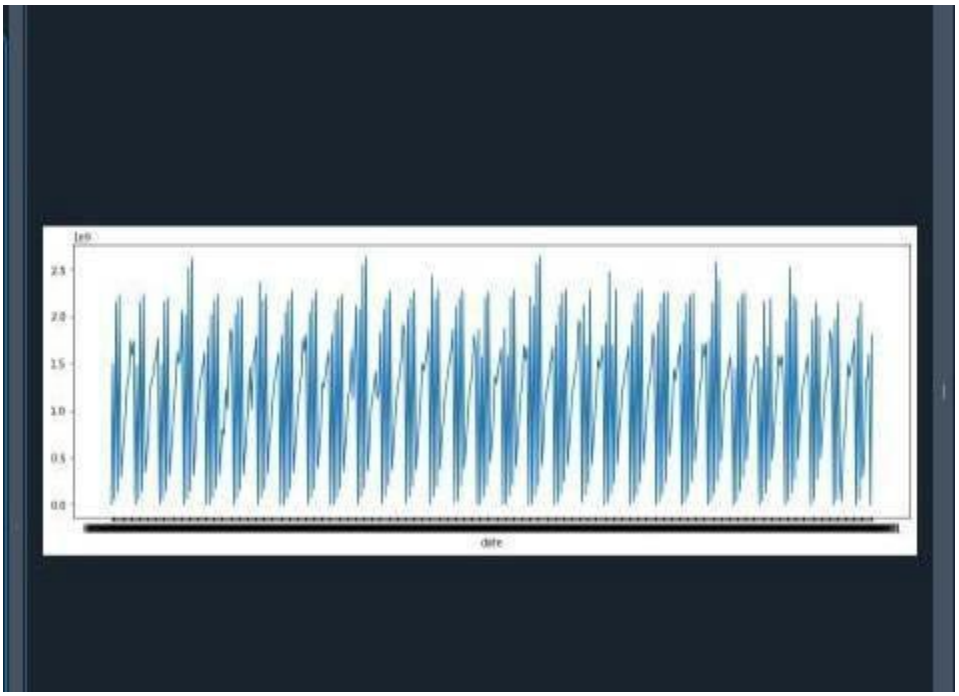
```
29
30  df["date"]= pd.to_datetime(df.date)
31  df["Total_vaccinations(count)"]= df.groupby("location").total_vaccinations.tail(1)
32  df.groupby("location")["Total_vaccinations(count)"].mean().sort_values(ascending= False).head(20)
33
34
```

```
Out[14]:
location
United States    330748673.0
Germany          126041243.0
France           109187212.0
Italy             88556058.0
Japan             86647668.0
South Korea       74173406.0
Spain             65141278.0
Poland            38515444.0
Chile             25596927.0
South Africa      25379326.0
Netherlands       23997608.0
Argentina         20405678.0
Peru              20304277.0
Nepal             19580285.0
Belgium           17451842.0
Ecuador           15812935.0
Portugal          14981060.0
Sweden            14709074.0
Czechia           14604323.0
Austria           14584985.0
Name: Total_vaccinations(count), dtype: float64

In [15]:
```

```
34
35  #step3:visualization
36  #barplot visualization of top countries with most vaccinations
37  x= df.groupby("location")["Total_vaccinations(count)"].mean().sort_values(ascending= False).head(26
38  sns.set_style("whitegrid")
39  plt.figure(figsize= (8,8))
40  ax= sns.barplot(x.values,x.index)
41  ax.set_xlabel("Total vaccinations(count)")
42  plt.show()
43
```



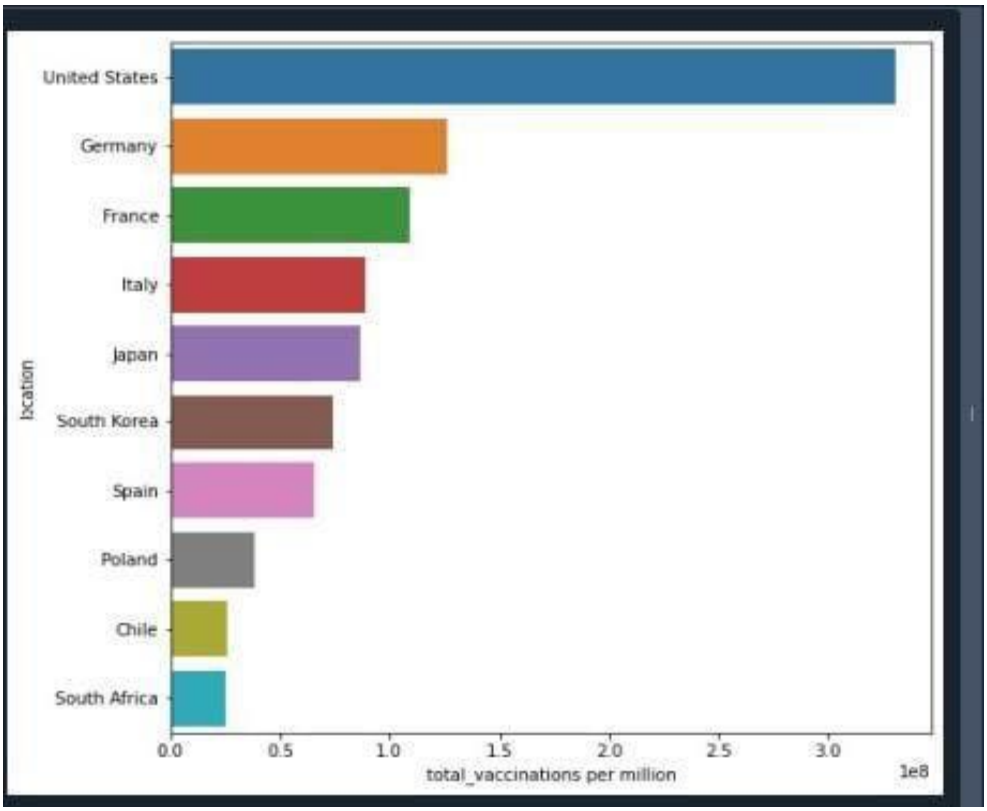
```
import numpy as np
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
plt.figure(figsize= (20,20))
words= "".join(df["vaccines"])
final = WordCloud(width = 2000, height = 800, background_color ="black",min_font_size = 10).generate(words)
plt.imshow(final)
plt.axis("off")
plt.show()

plt.figure(figsize=(15,15))
sns.countplot(y= "vaccine",data= df)
plt.show()

#total_vaccinations
x= df.groupby("date").total_vaccinations.sum()
plt.figure(figsize= (15,5))
sns.lineplot(x.index,x.values)
plt.show()
```



```
36
37
38    #location with best total_vaccinations
39    x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
40    x
41
```

```
Out[38]:
location
United States     330748673.0
Germany           126041243.0
France            109187212.0
Italy              88556058.0
Japan              86647668.0
South Korea        74173406.0
Spain              65141278.0
Poland             38515444.0
Chile              25596927.0
South Africa       25379326.0
Netherlands        23997608.0
Argentina          20405678.0
Peru               20304277.0
Nepal              19580285.0
Belgium            17451842.0
Ecuador            15812935.0
Portugal           14981060.0
Sweden             14709074.0
Czechia            14604323.0
Austria            14584985.0
Name: total_vaccinations, dtype: float64
```

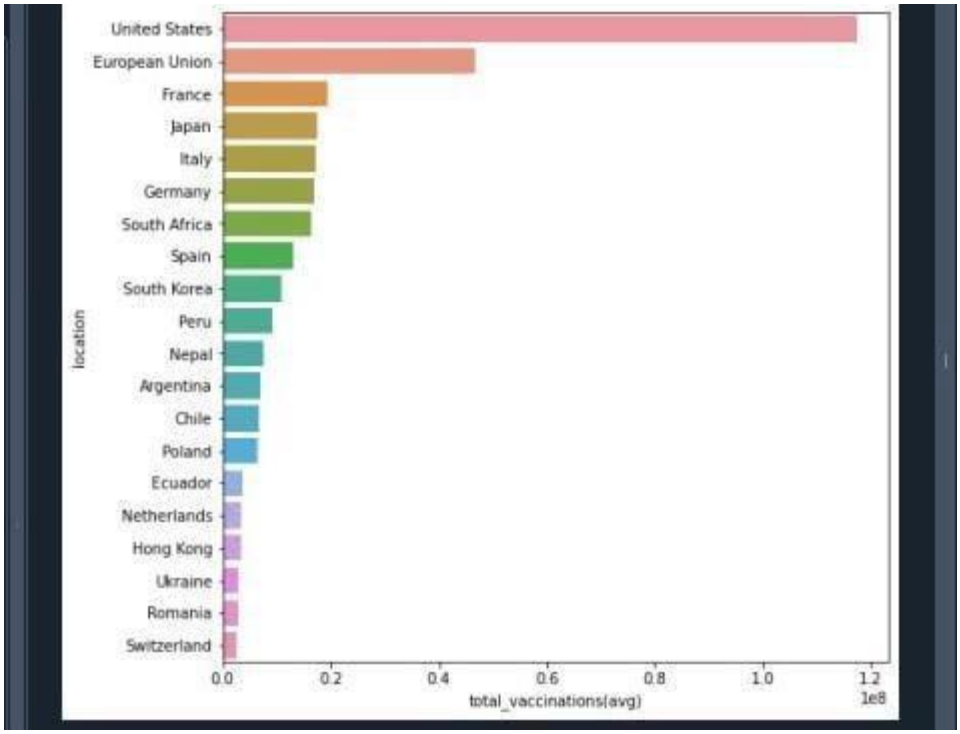
```
8
9     #total_vaccinations per million top location
0     df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(20)
1     #total_vaccination per million
2     plt.figure(figsize= (15,5))
3     sns.lineplot(x= "date",y= "total_vaccinations per million",data= df)
4     plt.show()
```




```
36
37
38  #location with best total_vaccinations
39  x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
40  x
41
```

```
.... x
Out[33]:
location
United States      1.174027e+08
European Union     4.682682e+07
France             1.947763e+07
Japan              1.737966e+07
Italy              1.727065e+07
Germany            1.702307e+07
South Africa       1.641744e+07
Spain              1.294769e+07
South Korea        1.097557e+07
Peru               9.228180e+06
Nepal              7.675604e+06
Argentina          7.104964e+06
Chile              6.652215e+06
Poland             6.491354e+06
Ecuador            3.598674e+06
Netherlands        3.489179e+06
Hong Kong          3.424323e+06
Ukraine            2.922647e+06
Romania            2.786433e+06
Switzerland        2.663743e+06
Name: total_vaccinations, dtype: float64
```

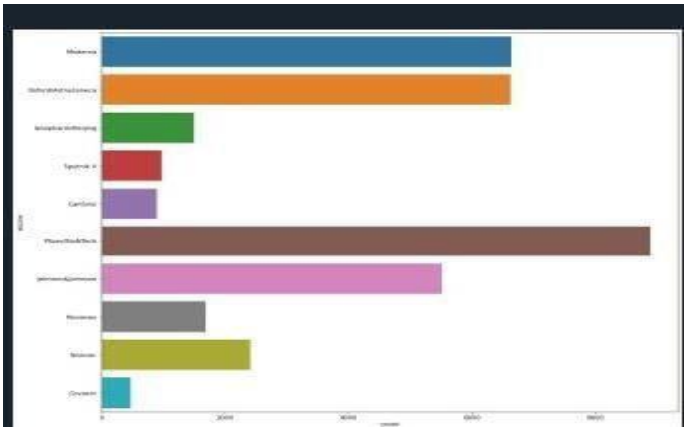
```
#location with best total_vaccinations
x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
x
#total_vaccinations barplot
plt.figure(figsize= (8,8))
ax= sns.barplot(x.values,x.index)

ax.set_xlabel("total_vaccinations(avg)")
plt.show()
```




```
47
48
49    #total_vaccination per hundred top location
50    df["total_vaccinations"]= df.groupby("location").total_vaccinations.tail(1)
51
```

```
ascending=False).head(20)
Out[38]:
location
United States     330748673.0
Germany           126041243.0
France            109187212.0
Italy              88556058.0
Japan              86647668.0
South Korea        74173406.0
Spain              65141278.0
Poland             38515444.0
Chile              25596927.0
South Africa       25379326.0
Netherlands        23997608.0
Argentina          20405678.0
Peru               20304277.0
Nepal              19580285.0
Belgium            17451842.0
Ecuador            15812935.0
Portugal           14981060.0
Sweden             14709074.0
Czechia            14604323.0
Austria            14584985.0
Name: total_vaccinations, dtype: float64
```

```
print(df)

import numpy as np
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
plt.figure(figsize= (20,20))
words= "".join(df["vaccines"])
final = WordCloud(width = 2000, height = 800, background_color ="black",min_font_size = 10).generate(words)
plt.imshow(final)
plt.axis("off")
plt.show()

plt.figure(figsize=(15,15))
sns.countplot(y= "vaccine",data= df)
plt.show()
```



❖ Conclusion

- Based on the data analysis, addressing questions or hypotheses posed earlier. Discuss the practical implications of your findings in the context of COVID-19 vaccination efforts.