

# **COVID 19 VACCINE ANALYSIS**

## **Phase 5: Project Documentation & Submission**

### **Aim:**

To conduct an in-depth analysis of Covid-19 vaccine data, including efficacy, distribution, and adverse effects, to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies.

### **INTRODUCTION:**

The COVID 19 pandemic caused due to the virus named corona. This virus was originated in Wuhan, China in 2019 and was later spread throughout the world. The disease showed symptoms such as mild fever and cold but also caused life threatening symptoms like breathing problems caused by damage to the lungs. As this virus was new to the world and there was no vaccine or cure to it at the earlier period there were several deaths around the world. So the countries around the world announced lockdowns to prevent the spread of the disease. An effort was made by several health organizations to discover the cure or a vaccine to fight against the virus.

In later stages of 2020 several experimental vaccines were developed and was tested to humans. The efforts were successful as the vaccines were helpful in reducing the affects the virus and even if people were infected, they were not in any life threatening situation and escaped the illness having only minor symptoms. Many countries later developed their own vaccines and also helped other countries without the resources by providing them with vaccines developed.

### **Problem Statement:**

The problem is to conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. This project involves data collection, data preprocessing, exploratory data analysis, statistical analysis, and visualization.

### **Data collection:**

The dataset that I will be using here for the task of covid-19 vaccines analysis is taken from the online resource called kaggle and the collected file was imported to the desktop.

## Data Processing:

After loading the data and after analyzing the data | 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.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing the libraries

```
7 data_reading = pandas.read_csv("country_vaccinations_by_manufacturer.csv")
8 print(data_reading)
```

Reading the data

```
C:\Users\Dell\AppData\Local\Programs\Python\Python39\python.exe "C:\Users\Dell\Desktop\covid 19 vaccine analysis IBM project\ADS sept 20.py"
   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        600
35620  European Union  2022-03-29  Sinopharm/Beijing      2501516
35621  European Union  2022-03-29      Sinovac           1809
35622  European Union  2022-03-29      Sputnik V      1845103

[35623 rows x 4 columns]

Process finished with exit code 0
```

Output

```

1 import pandas as pd
2 import numpy as np
3
4 data_reading = pd.read_csv("country_vaccinations_by_manufacturer.csv")
5 print(data_reading)
6 data_reading = pd.series(np.arange(1,51))
7 print(data_reading.head(6))

```

```

1 import pandas as pd
2 import numpy as np
3
4 data_reading = pd.read_csv("country_vaccinations_by_manufacturer.csv")
5 print(data_reading)
6 data_reading = pd.series(np.arange(1,51))
7 print(data_reading.tail(6))

```

```

C:\Users\Dell\AppData\Local\Programs\Python\Python39\python.exe "C:\Users\Dell\Desktop\covid 19 vaccine analysis IBM project\ADS sept 20.py"

```

	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

```

1 import pandas as pd
2
3 data_reading = pd.read_csv("country_vaccinations_by_manufacturer.csv")
4
5
6 print(data_reading.describe())
7

```

```
C:\Users\Dell\AppData\Local\Programs\Python\Python39\python.exe "C:\Users\Dell\Desktop\covid 19 vaccine analysis IBM project\ADS sept 20.py"
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

Process finished with exit code 0
```

output

```
print(data_reading.value_counts())
```

```
location      date      vaccine      total_vaccinations
Argentina    2020-12-29  Moderna              2              1
Netherlands  2021-12-03  Pfizer/BioNTech    19383352         1
              2021-11-26  Oxford/AstraZeneca 2800809          1
                           Pfizer/BioNTech    19055501         1
              2021-12-03  Johnson&Johnson   865506           1
              ..
European Union 2021-11-21  Johnson&Johnson   17156871         1
                           Covaxin           11              1
              2021-11-20  Sputnik V          1845101          1
                           Sinovac           1788             1
Uruguay       2022-03-29  Sinovac            3247932          1
Name: count, Length: 35623, dtype: int64

Process finished with exit code 0
```

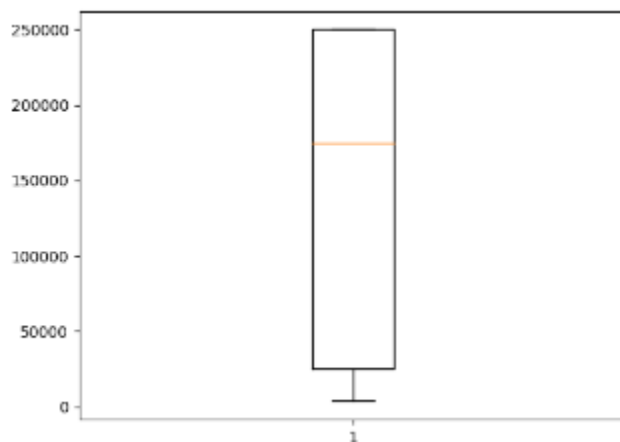
output

**Boxplot:**

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data_reading = pd.read_csv("country_vaccinations_by_manufacturer.csv")
4
5
6 y = list(data_reading.population)
7 plt.boxplot(y)
8 plt.show()

```



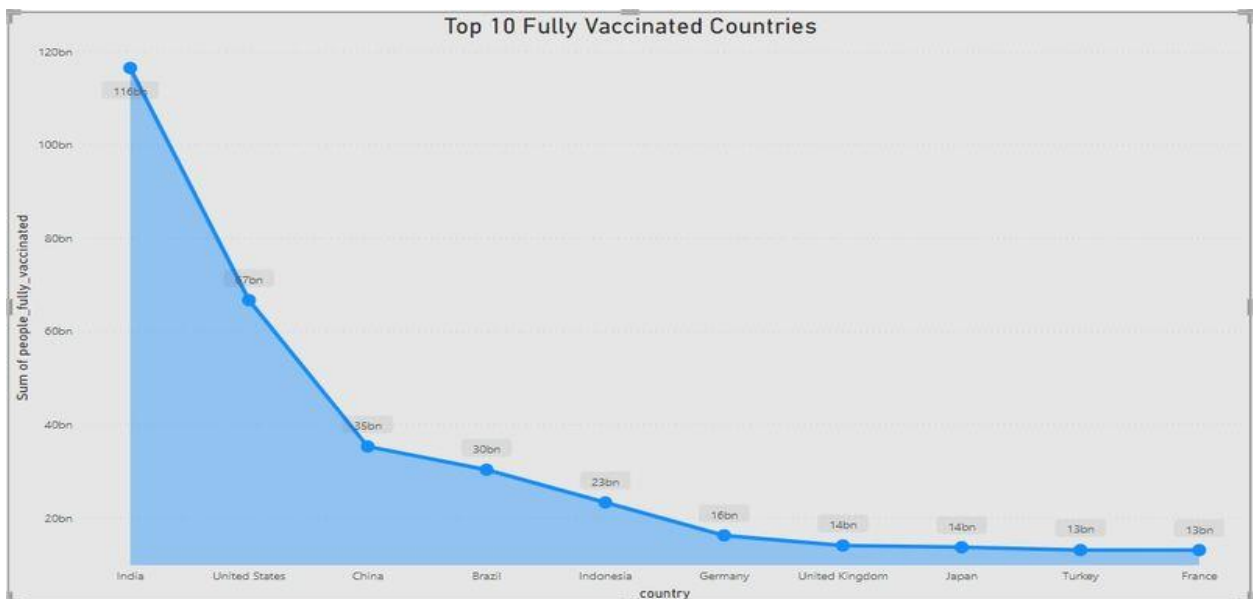
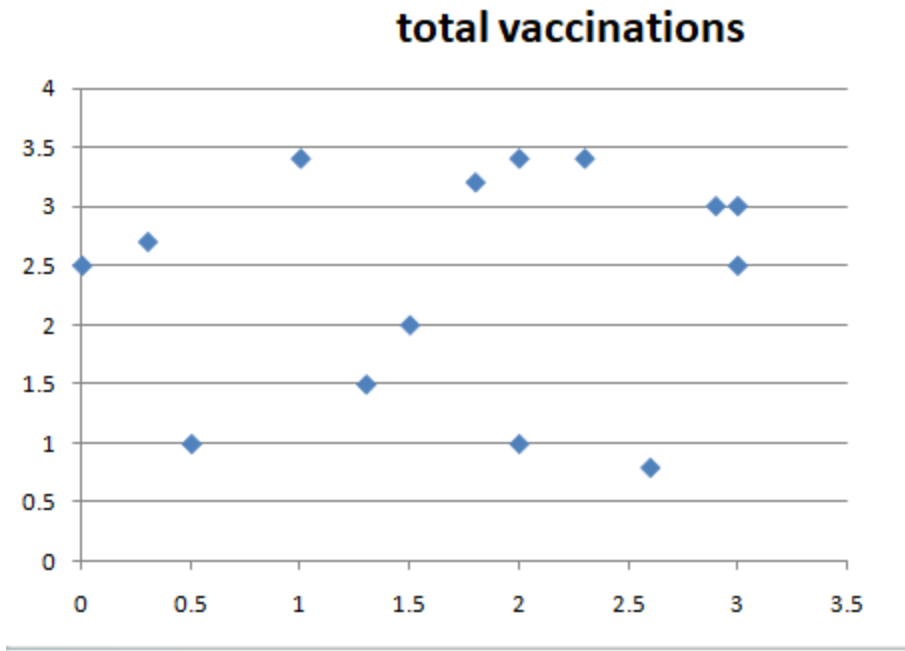
Output

## VISUALIZATION:

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3 data_reading = pd.read_csv("country_vaccinations_by_manufacturer.csv")
4
5
6 plt.scatter(data_reading["location"], data_reading["total_vaccinations"])
7 plt.title("scatter plot")
8 plt.xlabel('location')
9 plt.ylabel('total_vaccinations')
10 plt.show()
11
12

```



Output

## Explanation:

The data set named country vaccinations by manufacturer was downloaded from the online resource called [kaggle.com](https://www.kaggle.com) and it was imported to the local desktop. A python file was created to get started with the analysis. Using the command pip install the

required libraries are installed, now the installed libraries has to be imported to the python file where we are going to do the analysis, For importing those libraries into the python file the import command is used after importing the required libraries we have to read the data. By using

```
data_reading =pandas.read(country_vaccinations_by_manufacturer.csv)
```

this code we can read the data after reading the data we can get the clear vision of the data

Now we are going to process the data in our hands for the sake of better understanding of the data by using the head and tail functions in the pandas library we can process our data and gain valuable insights.

EDA is a phenomenon under data analysis used for gaining a better understanding of data aspects.

Descriptive statistics is a helpful way to understand characteristics of your data and to get a quick summary of it. Pandas in python provide an interesting method **describe()**. The describe function applies basic statistical computations on the dataset like extreme values, count of data points standard deviation etc. Any missing value or NaN value is automatically skipped. describe() function gives a good picture of distribution of data. Another useful method is value\_counts() which can get count of each category in a categorical attributed series of values. For an instance suppose you are dealing with a dataset of customers who are divided as youth, medium and old categories under column name age and your dataframe is "DF". You can run this statement to know how many people fall in respective categories. In our data set example education column can be used One more useful tool is boxplot which you can use through matplotlib module. Boxplot is a pictorial representation of distribution of data which shows extreme values, median and quartiles. We can easily figure out outliers by using boxplots. Now consider the dataset we've been dealing with again and lets draw a boxplot on attribute *population* Scatter plots are used to observe relationships between variables and uses dots to represent the relationship between them. The **scatter()** method in the matplotlib library is used to draw a scatter plot. Bar Chart can be of two types horizontal bars and vertical bars. Each can be created using the hbar() and vbar() functions of the plotting interface respectively This is how we can effectively use available libraries to make various visualizations using python which will be very helpful in analyzing a dataset effectively.

## **Conclusion:**

From the inferences and conjectures, it can be concluded that people from all parts of the world are educating themselves and willingly taking the vaccines under the governments' free vaccination program. Also, these vaccines have been proved effective against COVID-19 till now. If the rate of people taking the vaccine continues to grow then all the countries can vaccinate their people before the end of this year and that concludes our journey through the COVID-19 Vaccine Analysis Project! We hope you found this exploration informative and insightful. Remember, knowledge is power, and by understanding the data, we can make more informed decisions.