PROJECT: COVID-19 VACCINES ANALYSIS

DEVELOPMENT PART-1

1.COLLECT THE DATA.

There are many different sources of COVID-19 vaccine data, including:

- Government agencies such as the World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC) publish data on COVID-19 vaccination rates, side effects, and effectiveness.
- Academic institutions often conduct research on COVID-19 vaccines and publish their findings in scientific journals.
- Social media platforms such as Twitter and Facebook can be used to collect data on public sentiment towards COVID-19 vaccines.
- Once we have identified a source of data, we need to decide how to collect it. For example, we may be able to download a dataset from a government website or use a web scraping tool to collect data from social media.

In this project we have collected data from the **World Health Organization (WHO) COVID-19** Dashboard.

The data includes the following information:

- Country
- Iso code
- Date
- Total vaccinations
- Daily vaccinations
- Total vaccinations per hundred
- People fully vaccinated
- People vaccinated per hundred
- People fully vaccinated per hundred
- Daily vaccinations per million
- vaccines

2.PREPROCESS THE DATA.

Once we have collected the data, we need to preprocess it to make it suitable for analysis. This may involve the following steps:

- **Cleaning the data**. This involves removing any errors or inconsistencies in the data. For example, we may need to remove duplicate rows, correct spelling errors, or convert data to a consistent format.
- **Transforming the data**. This involves converting the data into a format that is suitable for our analysis. For example, we may need to calculate new variables, such as the percentage of people who are vaccinated or the risk of hospitalization for vaccinated and unvaccinated people.
- **Feature engineering.** This involves creating new features from the existing data that may be more informative for your analysis. For example, you could create a feature that indicates whether a person is at high risk of severe COVID-19 disease based on their age and underlying health conditions.

3.ANALYZE THE DATA.

- Once we have pre-processed the data, we can use a variety of statistical and machine learning techniques to analyse it and answer our research questions.
- For example, we could use a linear regression model to predict the number of doses administered per day in a country based on its population size and other factors.

Here are some specific examples of COVID-19 vaccine analysis questions that we could answer using the dataset we have provided:

1. What is the global trend of COVID-19 vaccine administration?
2. Which countries have the highest and lowest vaccination rates?
3. How does the vaccination rate in a country compare to its population size?
4. How does the vaccination rate in a country correlate with its COVID-19 case and death rates?
5. Which factors are associated with higher COVID-19 vaccination rates?
6. What is the impact of COVID-19 vaccine administration on COVID-19 transmission and disease severity?

We can use the answers to these questions to inform public policy, improve vaccine distribution strategies, and better understand the impact of COVID-19 vaccines on global health.

```
// importing the required python libraries//
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
%matplotlib inline
```

```
//using the read_cv method to read the dataset//
df=pd.read_csv("/content/drive/MyDrive/DATASET/country_vaccinations.csv")
```

```
// head() will display the columns of the dataset//
df.head()
```

```
df.head()
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	to
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	

```
// tail() will display the last data values of the dataset//
df.tail()
```

```
[ ] df.tail()
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	tot
86507	Zimbabwe	ZWE	2022-03-25	8691642.0	4814582.0	3473523.0	139213.0	69579.0	
86508	Zimbabwe	ZWE	2022-03-26	8791728.0	4886242.0	3487962.0	100086.0	83429.0	
86509	Zimbabwe	ZWE	2022-03-27	8845039.0	4918147.0	3493763.0	53311.0	90629.0	
86510	Zimbabwe	ZWE	2022-03-28	8934360.0	4975433.0	3501493.0	89321.0	100614.0	
86511	Zimbabwe	ZWE	2022-03-29	9039729.0	5053114.0	3510256.0	105369.0	103751.0	

```
// isnull() will display whether there is any null data values of the dataset//
df.isnull()
```

```
df.isnull()
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations
0	False	False	False	False	False	True	True	True
1	False	False	False	True	True	True	True	False
2	False	False	False	True	True	True	True	False
3	False	False	False	True	True	True	True	False
4	False	False	False	True	True	True	True	False
...	...	...	...	...	...	...	...	...
86507	False	False	False	False	False	False	False	False
86508	False	False	False	False	False	False	False	False
86509	False	False	False	False	False	False	False	False
86510	False	False	False	False	False	False	False	False

```
//fillna() will fill the  null values with 0//
df.fillna(0,inplace=True)
```

```
df.isnull().sum()
```

```
country                               0
iso_code                              0
date                                  0
total_vaccinations                    0
people_vaccinated                     0
people_fully_vaccinated               0
daily_vaccinations_raw                0
daily_vaccinations                    0
total_vaccinations_per_hundred        0
people_vaccinated_per_hundred         0
people_fully_vaccinated_per_hundred   0
daily_vaccinations_per_million        0
vaccines                              0
source_name                           0
source_website                        0
dtype: int64
```

```
[ ] df.info()
```

```
<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                   86512 non-null  float64
 4   people_vaccinated                    86512 non-null  float64
 5   people_fully_vaccinated              86512 non-null  float64
 6   daily_vaccinations_raw               86512 non-null  float64
 7   daily_vaccinations                   86512 non-null  float64
 8   total_vaccinations_per_hundred       86512 non-null  float64
 9   people_vaccinated_per_hundred        86512 non-null  float64
 10  people_fully_vaccinated_per_hundred  86512 non-null  float64
 11  daily_vaccinations_per_million       86512 non-null  float64
 12  vaccines                             86512 non-null  object
 13  source_name                          86512 non-null  object
 14  source website                       86512 non-null  object
```

```
//the date given in the dataset is converted into specified format//
df['date']=pd.to_datetime(df['date'],format='%Y-%m-%d')
```

// Now we are performing the covid-19 vaccine analysis on the country USA //

USA

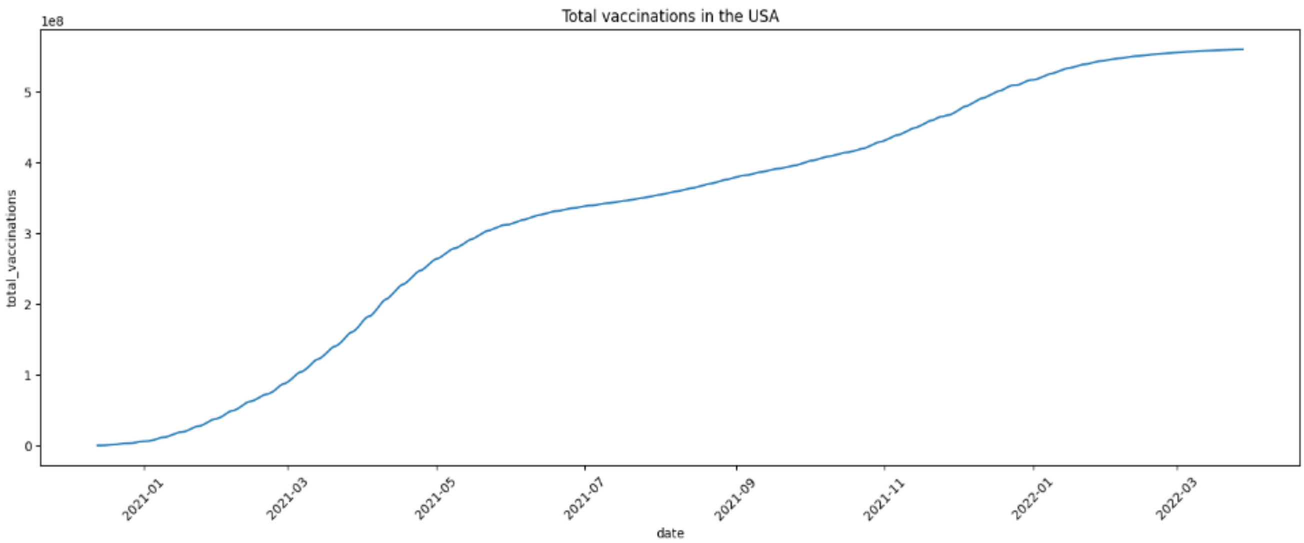
```
[ ] df_USA=df[df["iso_code"]=='USA'].copy()
    df_USA
```

	country	iso_code	date	total_vaccinations	people_vaccinated	daily_vaccinations	total_vaccinations_per_hundred	vaccines
82360	United States	USA	2020-12-13	30288.0	25125.0	0.0	0.01	Johnson&Johnson, Moderna, Pfizer/BioNTech
82361	United States	USA	2020-12-14	34867.0	29543.0	4579.0	0.01	Johnson&Johnson, Moderna, Pfizer/BioNTech
82362	United States	USA	2020-12-15	84638.0	76984.0	27175.0	0.03	Johnson&Johnson, Moderna, Pfizer/BioNTech
82363	United States	USA	2020-12-16	244549.0	231496.0	71420.0	0.07	Johnson&Johnson, Moderna, Pfizer/BioNTech
82364	United States	USA	2020-12-17	517161.0	496980.0	121718.0	0.16	Johnson&Johnson, Moderna, Pfizer/BioNTech
...	...	...	...	...	...	...	...	...
82826	United States	USA	2022-03-24	559861103.0	255273235.0	156889.0	168.63	Johnson&Johnson, Moderna, Pfizer/BioNTech
82827	United States	USA	2022-03-25	560045501.0	255322519.0	149541.0	168.68	Johnson&Johnson, Moderna, Pfizer/BioNTech
82828	United States	USA	2022-03-26	560137738.0	255348742.0	143396.0	168.71	Johnson&Johnson, Moderna, Pfizer/BioNTech

```
df_USA.drop(df_USA.index[df_USA['total_vaccinations']==0],inplace=True)
```

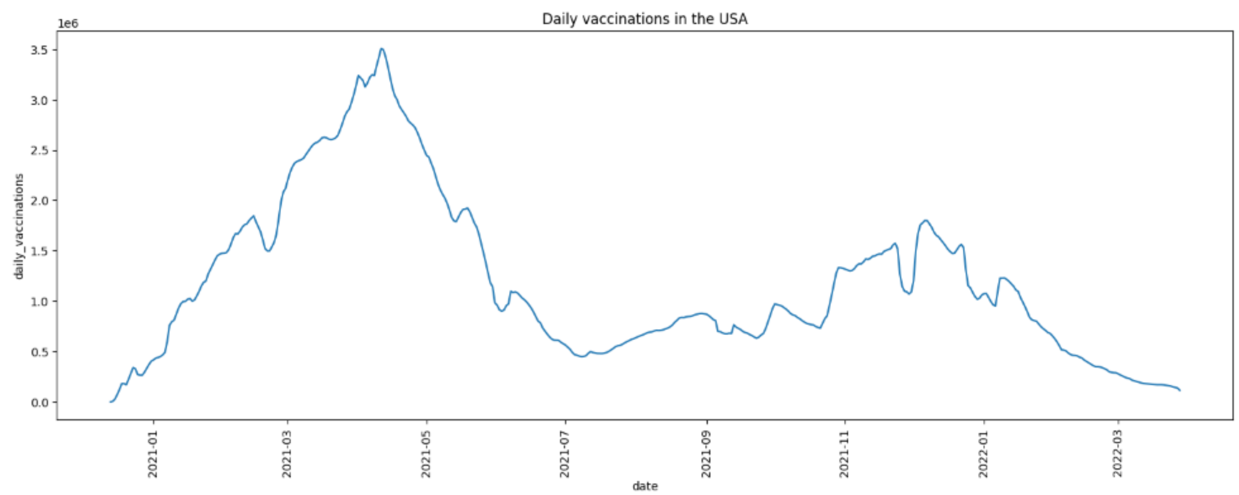
// Virtualization analysis of Total Vaccinations in the USA //

```
plt.figure(figsize=(18,6))
sns.lineplot(data=df_USA,x="date",y="total_vaccinations")
plt.title("Total vaccinations in the USA")
plt.xticks(rotation=45)
plt.show()
```



//// **Virtualization analysis of Total Vaccinations in the USA //**

```
plt.figure(figsize=(18,6))
sns.lineplot(data=df_USA,x="date",y="daily_vaccinations")
plt.title("Daily vaccinations in the USA")
plt.xticks(rotation=90)
plt.show()
```



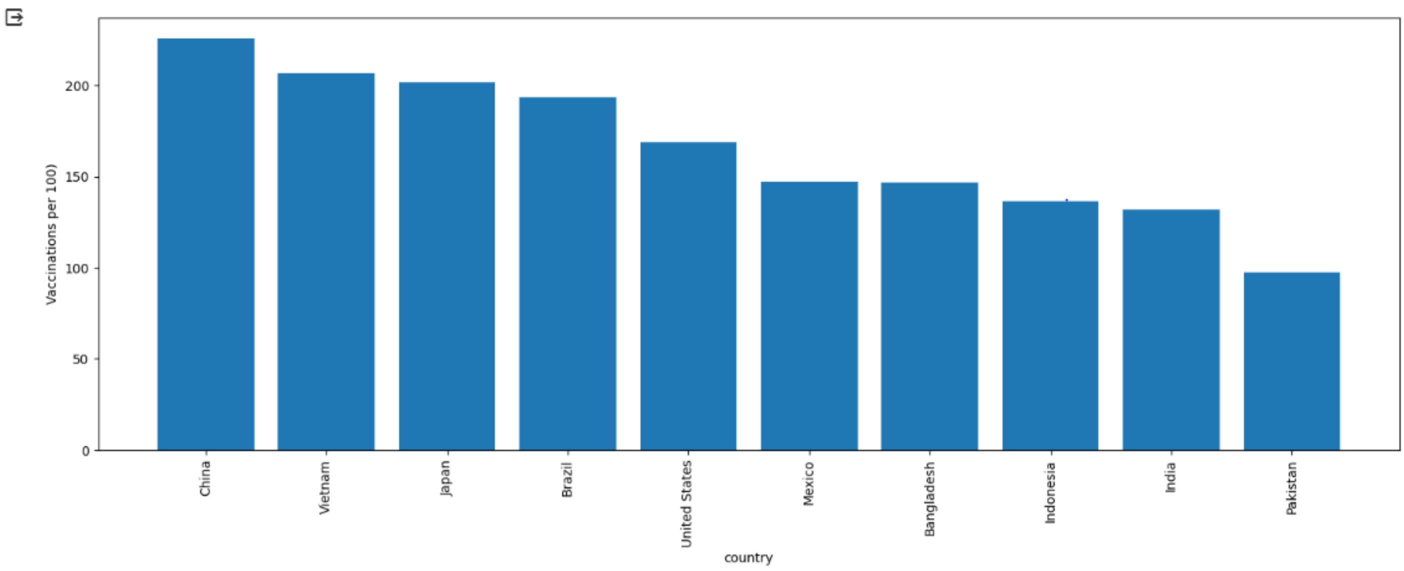
// TOP 10 COUNTRIES WITH HIGHEST AMOUNT OF VACCINATED PEOPLE PER 100 OF THE POPULATION//

```
[ ] vacc_by_country=vacc_by_country.sort_values('total_vaccinations_per_hundred',ascending=False)
    vacc_by_country
```

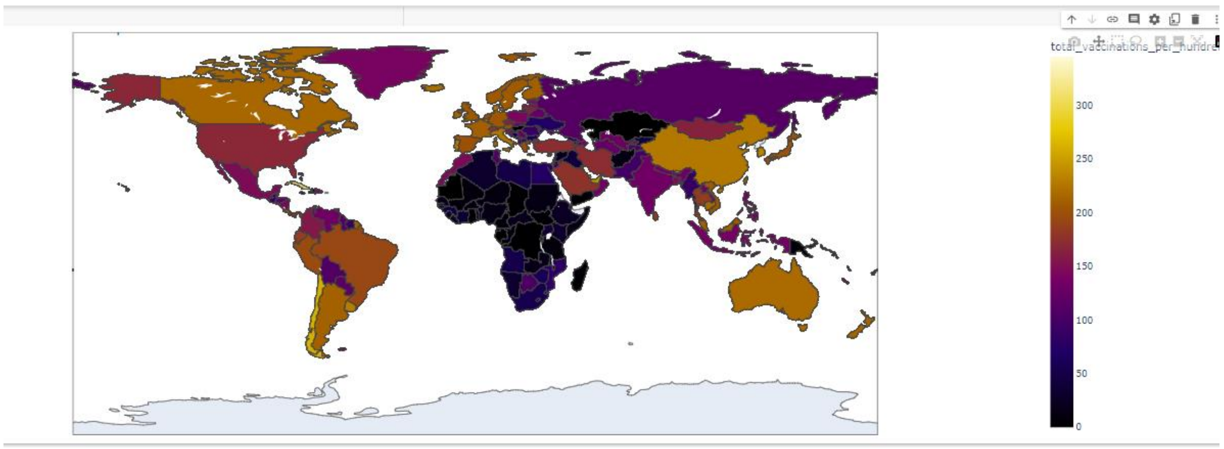
country	iso_code	date	total_vaccinations	people_vaccinated	daily_vaccinations	total_vaccinations_per_hundred
China	CHN	2022-03-29	3.263129e+09	1.275541e+09	22424286.0	225.94
Vietnam	VNM	2022-03-22	2.031444e+08	7.994719e+07	1675471.0	206.93
Japan	JPN	2022-03-29	2.543456e+08	1.024675e+08	1997542.0	201.78
Brazil	BRA	2022-03-29	4.135596e+08	1.810781e+08	1941268.0	193.26
United States	USA	2022-03-28	5.601818e+08	2.553624e+08	3506960.0	168.72
Mexico	MEX	2022-03-29	1.919079e+08	8.558029e+07	1648223.0	147.32
Bangladesh	BGD	2022-03-29	2.436427e+08	1.275441e+08	3758404.0	146.50
Indonesia	IDN	2022-03-29	3.771089e+08	1.962409e+08	1897011.0	136.45
India	IND	2022-03-29	1.834501e+09	9.848381e+08	10037995.0	131.66
Pakistan	PAK	2022-03-10	2.193686e+08	1.280741e+08	2175773.0	97.41

//BAR DIAGRAM VACCINATIONS PER HUNDRED WITHH COUNTRY//

```
plt.figure(figsize=(18,6))
plt.bar(vacc_by_country.index,vacc_by_country.total_vaccinations_per_hundred)
plt.xticks(rotation=90)
plt.ylabel("Vaccinations per 100)")
plt.xlabel("country")
plt.show()
```



`// GEOGRAPHICAL REPRESENTATION OF COUNTRY WITH TOTAL VACCINATIONS PER 100 //`



4.INTERPRET THE RESULTS:

Once we have analyzed the data, we need to interpret the results and draw conclusions. For example, the logistic regression model shows that vaccinated people are less likely to be hospitalized with COVID-19 than unvaccinated people, we can conclude that the COVID-19 vaccine is effective at preventing severe illness.

Here are some additional tips for conducting COVID-19 vaccine analysis:

- Use multiple sources of data. This will help you to get a more complete picture of the situation. For example, you could use government data to track vaccination rates and social media data to track public sentiment.
- Be aware of the limitations of your data. No dataset is perfect. There may be errors or inconsistencies in the data, or the data may not be representative of the entire population. It is important to be aware of these limitations when interpreting your results.
- Use appropriate statistical and machine learning techniques. There are a variety of statistical and machine learning techniques that can be used to analyze COVID-19 vaccine data. It is important to choose the techniques that are appropriate for your research questions and the type of data you have.
- Interpret your results carefully. It is important to avoid making overgeneralizations from your results. For example, if your study shows that the COVID-19 vaccine is effective in preventing severe illness in young adults, you should not conclude that it is effective in all age groups.