Introduction

Hello Readers!!
Covid-19 has affected our lives very much in very accepts it could be economical, mentally, etc. In this blog, we are going to explore how the vaccination drive is going around the world. For the past 1 year, we have been hoping for vaccines so that we can enjoy our life as we were doing before.
Hope this vaccination drive will help millions of people and save them. We are going to first read the dataset, then clean and draw some beautiful visuals.

IMPORT LIBRARIES

For analyzing data, we need some libraries. In this section, we are importing all the required libraries like pandas, NumPy, matplotlib, plotly, seaborn, and word cloud that are required for data analysis. Check the below code to import all the required libraries.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
import matplotlib.patches as mpatches
from plotly.subplots import make_subplots
from wordcloud import WordCloud
import seaborn as sns
sns.set(color_codes = True)
sns.set(style="whitegrid")
import plotly.figure_factory as ff
from plotly.colors import n_colors
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw
0	Albania	ALB	2021-01-10	0.0	0.0	NaN	NaN
1	Albania	ALB	2021-01-11	NaN	NaN	NaN	NaN
2	Albania	ALB	2021-01-12	128.0	128.0	NaN	NaN
3	Albania	ALB	2021-01-13	188.0	188.0	NaN	60.0
4	Albania	ALB	2021-01-14	266.0	266.0	NaN	78.0

Observation:

Dataset has columns like country, iso_code, date, total_vaccinations, people_vaccinated, people_fully vaccinated, etc. An initial look at the above table shows that data has null values too. We will deal with null values later.

info() function is used to get the overview of data like data type of feature, a number of null values in each column, and many more.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4568 entries, 0 to 4567
Data columns (total 15 columns):
 #   Column                               Non-Null Count  Dtype  
---  ------                               --------------  -----  
 0   country                              4568 non-null   object 
 1   iso_code                             4260 non-null   object 
 2   date                                 4568 non-null   object 
 3   total_vaccinations                   2988 non-null   float64
 4   people_vaccinated                    2541 non-null   float64
 5   people_fully_vaccinated              1702 non-null   float64
 6   daily_vaccinations_raw               2523 non-null   float64
 7   daily_vaccinations                   4409 non-null   float64
 8   total_vaccinations_per_hundred       2988 non-null   float64
 9   people_vaccinated_per_hundred        2541 non-null   float64
 10  people_fully_vaccinated_per_hundred  1702 non-null   float64
 11  daily_vaccinations_per_million       4409 non-null   float64
 12  vaccines                             4568 non-null   object 
 13  source_name                          4568 non-null   object 
 14  source_website                       4568 non-null   object 
dtypes: float64(9), object(6)
memory usage: 535.4+ KB
```

Observation:

The above picture shows that there are many null values in our dataset. We will deal with these null values later in this blog. There are two data types as seen from the table object means string and float.

The below function is used to get the total count of null values in each feature.

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

```
country                                0
iso_code                             308
date                                   0
total_vaccinations                  1580
people_vaccinated                   2027
people_fully_vaccinated             2866
daily_vaccinations_raw              2045
daily_vaccinations                   159
total_vaccinations_per_hundred      1580
people_vaccinated_per_hundred       2027
people_fully_vaccinated_per_hundred 2866
daily_vaccinations_per_million       159
vaccines                               0
source_name                            0
source_website                         0
dtype: int64
```

DATA CLEANING

Dataset has many null values as we have seen before. To get rid of it we need to clean the data first, After cleaning we will perform our further analysis. For cleaning the dataset we will perform many steps. Some of these steps are shown below

- Handling and Filling null values
- Change the data type of features
- Handling strings like splitting.

Check the below code for all the data cleaning that we are performing here:

```
df.fillna(value = 0, inplace = True)
df.total_vaccinations = df.total_vaccinations.astype(int)
df.people_vaccinated = df.people_vaccinated.astype(int)
df.people_fully_vaccinated = df.people_fully_vaccinated.astype(int)
df.daily_vaccinations_raw = df.daily_vaccinations_raw.astype(int)
df.daily_vaccinations = df.daily_vaccinations.astype(int)
df.total_vaccinations_per_hundred =
df.total_vaccinations_per_hundred.astype(int)
df.people_fully_vaccinated_per_hundred =
df.people_fully_vaccinated_per_hundred.astype(int)
df.daily_vaccinations_per_million = df.daily_vaccinations_per_million.astype(int)

df.people_vaccinated_per_hundred =
df.people_vaccinated_per_hundred.astype(int)
date = df.date.str.split('-', expand =True)
date
```

	0	1	2
0	2021	01	10
1	2021	01	11
2	2021	01	12
3	2021	01	13
4	2021	01	14
...	...	...	...
4563	2021	02	24
4564	2021	02	25
4565	2021	02	26
4566	2021	02	27
4567	2021	02	28

DATA VISUALIZATION

In this section, we are going to draw some visuals to get insights from our dataset. So let's started.

describe() function in pandas used to get the statistics of each feature present in our dataset. Some of the information we get include count, max, min, standard deviation, median, etc.

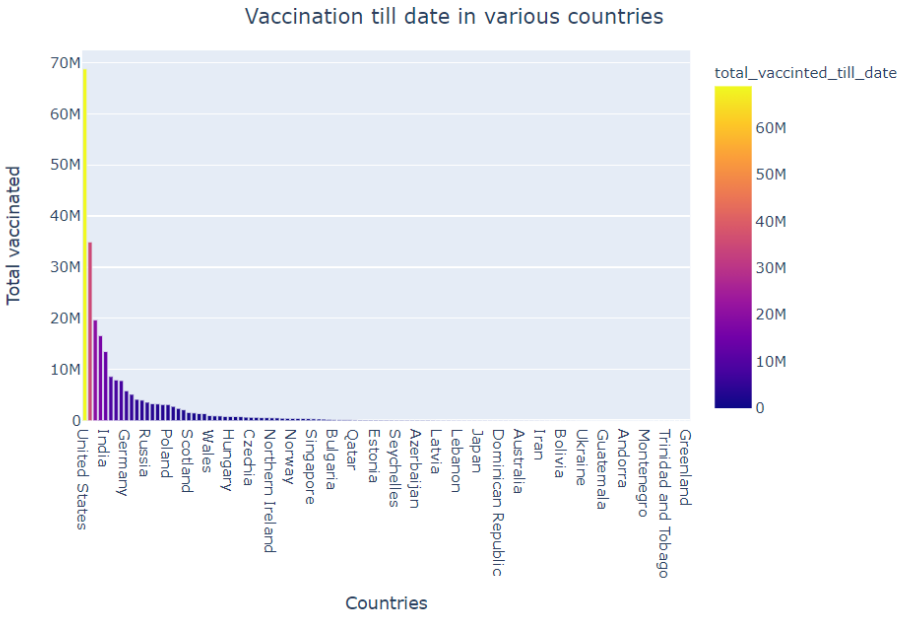
```
df.describe()
```

	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	t
count	4.568000e+03	4.568000e+03	4.568000e+03	4.568000e+03	4.568000e+03	4
mean	1.138244e+06	8.391915e+05	1.887296e+05	4.163373e+04	5.302986e+04	4
std	4.836662e+06	3.604334e+06	1.227922e+06	1.636365e+05	1.714283e+05	1
min	0.000000e+00	0.000000e+00	0.000000e+00	-5.001200e+04	0.000000e+00	0
25%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	9.940000e+02	0
50%	2.776050e+04	8.216500e+03	0.000000e+00	5.250000e+02	5.144000e+03	0
75%	4.094480e+05	2.412348e+05	1.901900e+04	1.515300e+04	2.443450e+04	4
max	7.523600e+07	4.977218e+07	2.477992e+07	2.429823e+06	1.916190e+06	1

Total Vaccinated Till Date

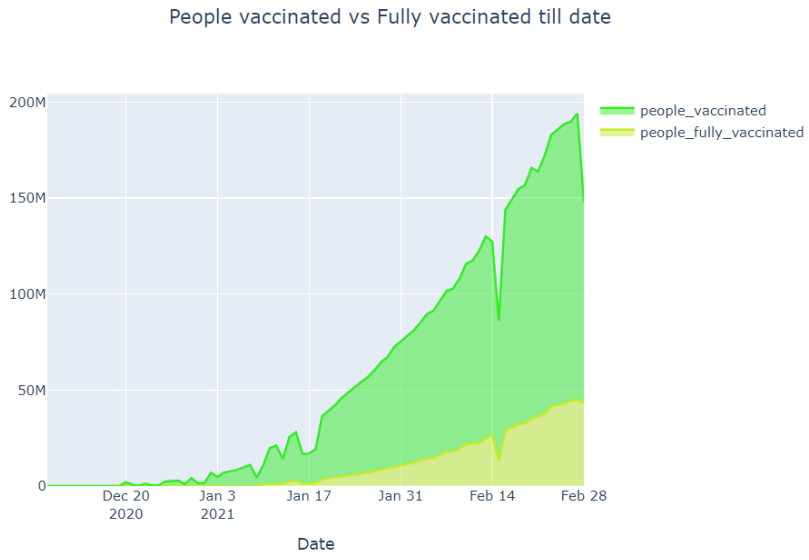
In this section, we are going to see how many total vaccines have been used in each country. Check the below code for more information. The data shows the United States has administrated most vaccines in the world followed by China, United Kingdom, England, India and at the last some countries includes Saint Helena, San Marino has 0 vaccination

	total_vaccinted_till_date
United States	68767620
China	34922496
United Kingdom	19660299
England	16602591
India	13483116
...	...
Trinidad and Tobago	441
Venezuela	155
Saint Helena	0
San Marino	0
Greenland	0



People vaccinated vs people fully vaccinated in the world :

In this section, let's analyze how many people vaccinated vs the people which are fully vaccinated in the world. We are drawing a kind of curve where the x-axis is Date and the y-axis is the count of people that are fully vaccinated in the world



End Notes: So in this article, we had a detailed discussion on Covid Vaccination Progress.