COVID 19 VACCINE ANALYSIS

Phase 2

2.1 Short explanation:

Covid vaccine analysis informs vaccine distribution strategies,addressing logistical challenges and promoting equitable access. It plays a cruical role in managing vaccine hesitancy by providing data on a vaccine safety and efficacy. It guides decisions on potential booster doses and adaptation to combat emerging variants of the virus.

2.2 Data set link:

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

The data (country vaccinations) contains the following information:

- **Country**- this is the country for which the vaccination information is provided;
- **Country ISO Code** - ISO code for the country;
- **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
- **Total number of vaccinations** - this is the absolute number of total immunizations in the country;
- **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccinations might be larger than the number of people;
- **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;
- **Daily vaccinations (raw)** - for a certain data entry, the number of vaccinations for that date/country;
- **Daily vaccinations** - for a certain data entry, the number of vaccinations for that date/country;

- **Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country;
- **Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country;
- **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;
- **Number of vaccinations per day** - number of daily vaccinations for that day and country;
- **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;
- **Vaccines used in the country** - total number of vaccines used in the country (up to date);
- **Source name** - source of the information (national authority, international organization, local organization etc.);

2.3 Columns to be used :

- Total vaccination
- People vaccinated
- People fully vaccinated
- Country

2.4 Libraries :

NUMPY:

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

- To download : Pip install numpy
- To import: import numpy as np

PANDAS:

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive.

- To download : Pip install pandas
- To import: import pandas as pd

MATPLOTLIB:

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

- To download : Pip install matplotlib
- To import: import matplotlib.pyplot as plt

SEABORN:

Seaborn library is a widely popular data visualization library that is commonly used for data science and machine learning tasks

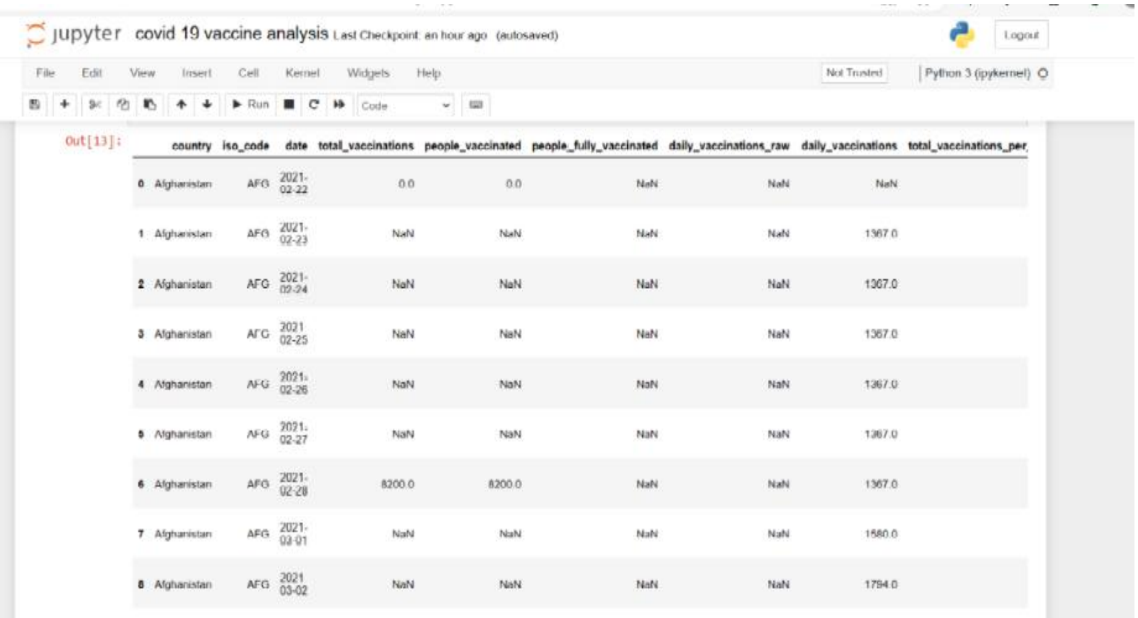
- To download : Pip install seaborn
- To import: import seaborn as sns

2.5 TEST AND TRAIN:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
df=pd.read_csv("country_vaccinations.csv")
print(df.info()) #TO FIND BASIC INFORMATIION ABOUT DATASET
```

```
<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                   43607 non-null  float64
 4   people_vaccinated                   41294 non-null  float64
 5   people_fully_vaccinated             38802 non-null  float64
 6   daily_vaccinations_raw              35362 non-null  float64
 7   daily_vaccinations                  86213 non-null  float64
 8   total_vaccinations_per_hundred      43607 non-null  float64
 9   people_vaccinated_per_hundred       41294 non-null  float64
 10  people_fully_vaccinated_per_hundred  38802 non-null  float64
 11  daily_vaccinations_per_million       86213 non-null  float64
 12  vaccines                             86512 non-null  object
 13  source_name                          86512 non-null  object
 14  source_website                       86512 non-null  object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None
```

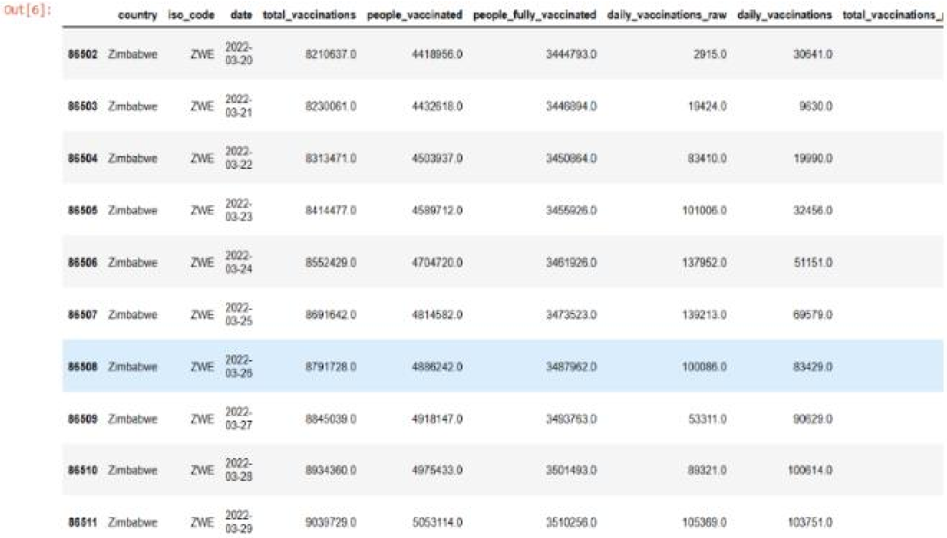
```
df.head(10) #DISPLAY FIRST 10 DATA IN DATASET
```



Out[13]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per
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	
5	Afghanistan	AFG	2021-02-27	NaN	NaN	NaN	NaN	1367.0	
6	Afghanistan	AFG	2021-02-28	8200.0	8200.0	NaN	NaN	1367.0	
7	Afghanistan	AFG	2021-03-01	NaN	NaN	NaN	NaN	1580.0	
8	Afghanistan	AFG	2021-03-02	NaN	NaN	NaN	NaN	1794.0	

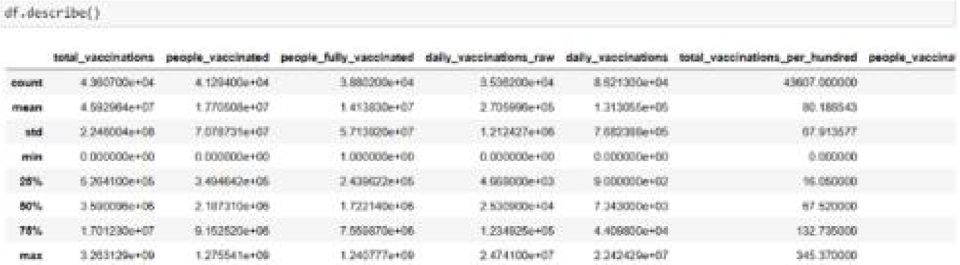
df.tail(10)



Out[6]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_
86502	Zimbabwe	ZWE	2022-03-20	8210637.0	4418956.0	3444793.0	2915.0	30641.0	
86503	Zimbabwe	ZWE	2022-03-21	8230061.0	4432618.0	3446894.0	19424.0	9630.0	
86504	Zimbabwe	ZWE	2022-03-22	8313471.0	4503937.0	3450864.0	83410.0	19990.0	
86505	Zimbabwe	ZWE	2022-03-23	8414477.0	4589712.0	3455926.0	101006.0	32456.0	
86506	Zimbabwe	ZWE	2022-03-24	8552429.0	4704720.0	3461926.0	137952.0	51151.0	
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	

df.describe()



```
df.describe()
```

	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccina
count	4.360700e+04	4.129400e+04	3.880200e+04	3.536200e+04	8.621300e+04	43607.000000	
mean	4.592984e+07	1.770508e+07	1.413830e+07	2.705996e+05	1.313055e+05	80.188543	
std	2.248004e+08	7.078731e+07	5.713020e+07	1.212427e+06	7.682388e+05	67.913577	
min	0.000000e+00	0.000000e+00	1.000000e+00	0.000000e+00	0.000000e+00	0.000000	
25%	5.204100e+05	3.494642e+05	2.439622e+05	4.668000e+03	9.000000e+02	16.050000	
50%	3.590096e+06	2.187310e+06	1.722140e+06	2.530900e+04	7.343000e+03	67.520000	
75%	1.701230e+07	9.152520e+06	7.559870e+06	1.234925e+05	4.409800e+04	132.735000	
max	3.263129e+09	1.275541e+09	1.240777e+09	2.474100e+07	2.242429e+07	345.370000	

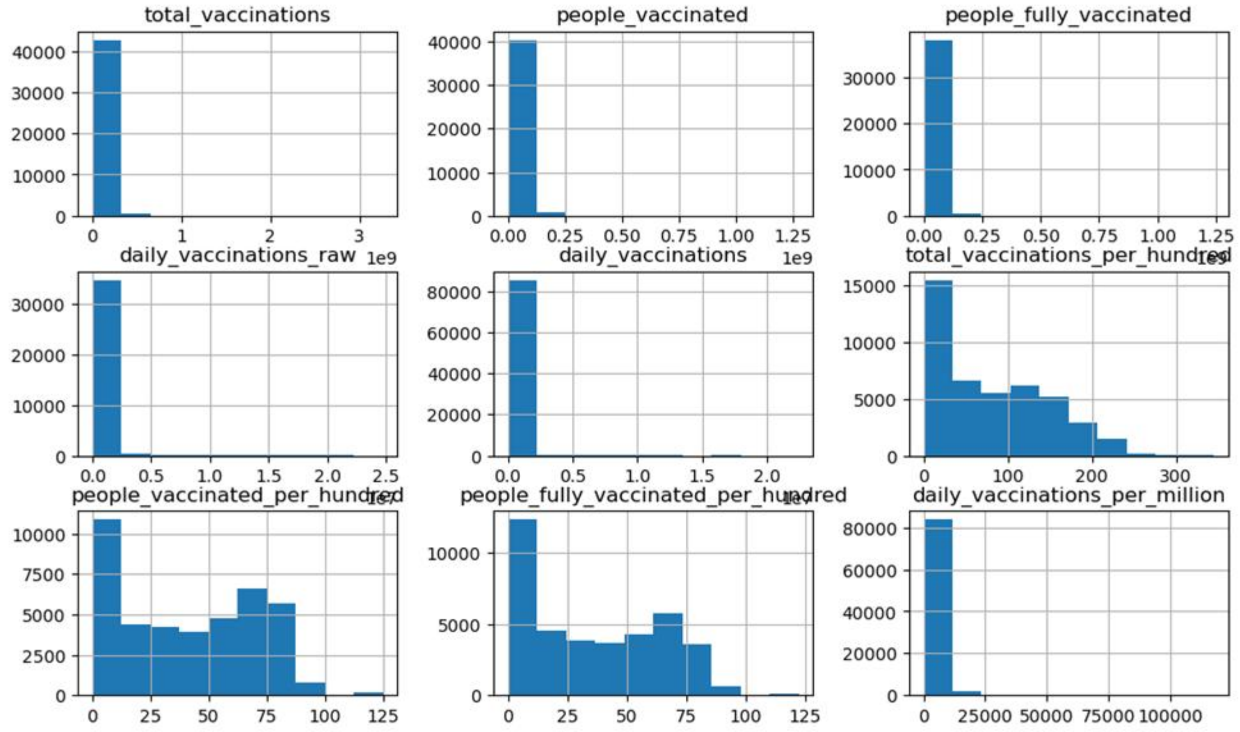
df.isnull().sum()

```
country                              0
iso_code                              0
date                               0
total_vaccinations                42905
people_vaccinated                  45218
people_fully_vaccinated             47710
daily_vaccinations_raw              51150
daily_vaccinations                  299
total_vaccinations_per_hundred        42905
people_vaccinated_per_hundred         45218
people_fully_vaccinated_per_hundred    47710
daily_vaccinations_per_million          299
vaccines                              0
source_name                           0
source_website                         0
dtype: int64
```

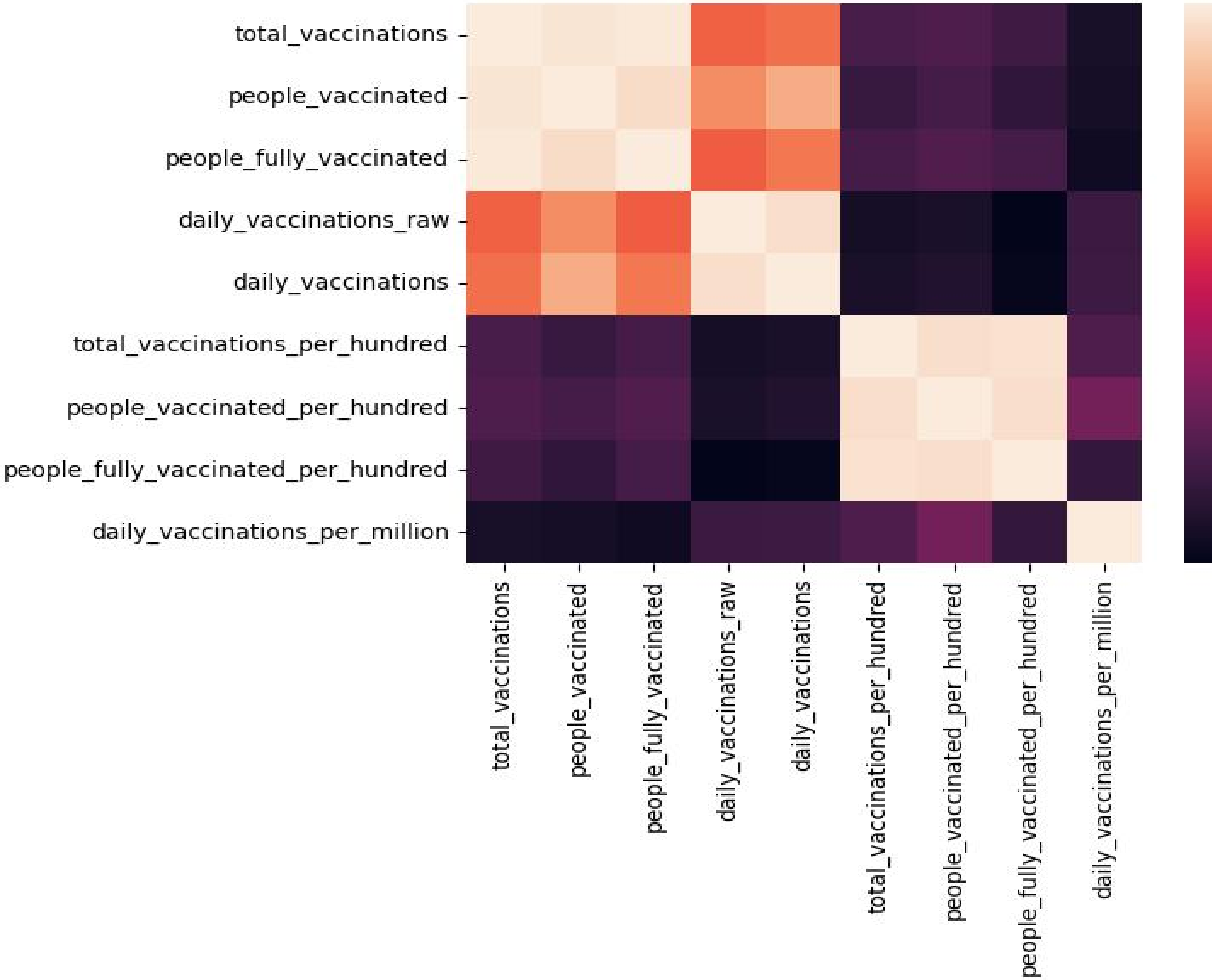
df.value_counts("vaccines")

```
vaccines
Johnson vaccines
&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
 7608
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
        6263
Oxford/AstraZeneca
        6022
Oxford/AstraZeneca, Pfizer/BioNTech
        4629
Johnson&Johnson, Moderna, Novavax, Oxford/AstraZeneca, Pfizer/BioNTech
        3564
        ...
Johnson&Johnson, Oxford/AstraZeneca, Sinovac
        312
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V
        311
Johnson&Johnson, Moderna
        251
Johnson&Johnson, Pfizer/BioNTech, Sinopharm/Beijing
        228
EpiVacCorona, Oxford/AstraZeneca, QazVac, Sinopharm/Beijing, Sputnik V, ZF
2001     190
Length: 84, dtype: int64
```

```
df.hist(figsize=(12,12),layout=(5,3))
array([[<AxesSubplot:title={'center':'total_vaccinations'}>,
    <AxesSubplot:title={'center':'people_vaccinated'}>,
    <AxesSubplot:title={'center':'people_fully_vaccinated'}>],
   [<AxesSubplot:title={'center':'daily_vaccinations_raw'}>,
    <AxesSubplot:title={'center':'daily_vaccinations'}>,
    <AxesSubplot:title={'center':'total_vaccinations_per_hundred'}>],
   [<AxesSubplot:title={'center':'people_vaccinated_per_hundred'}>,
    <AxesSubplot:title={'center':'people_fully_vaccinated_per_hundred'}>,
    <AxesSubplot:title={'center':'daily_vaccinations_per_million'}>],
   [<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>],
   [<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>]], dtype=object)
```



```
import seaborn as sns
sns.heatmap(df.corr())
<AxesSubplot:>
```



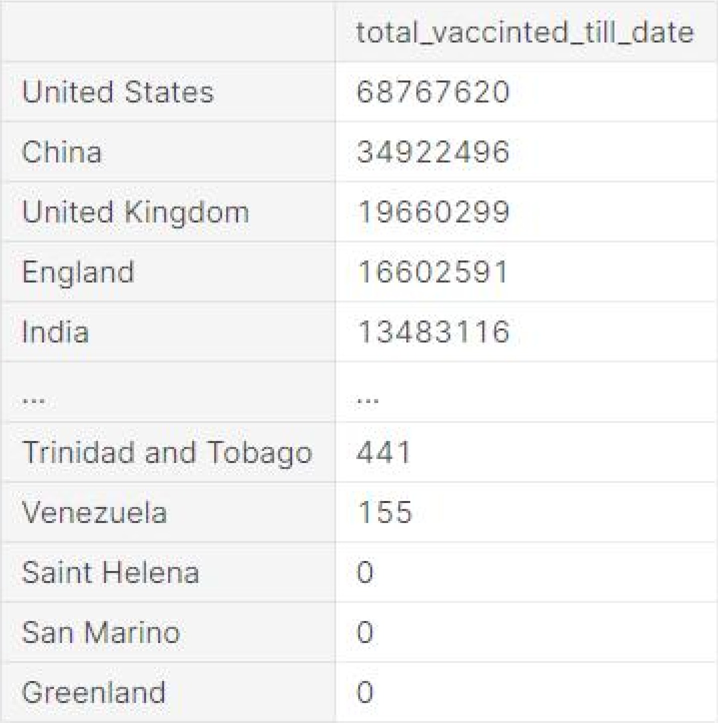
In [21]:

2.6 EXPLANATION:

Total Vaccinated till Date

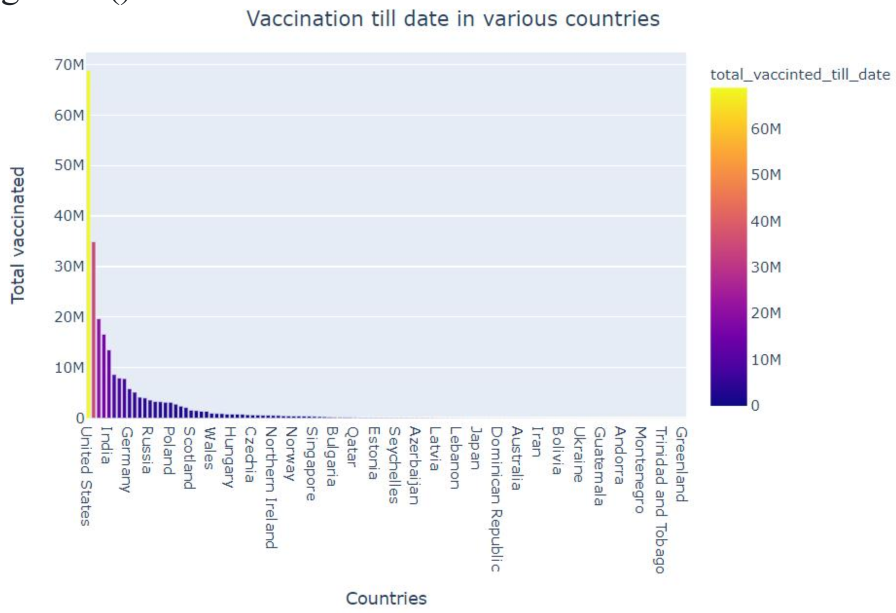
In this section, we are going to see how many total vaccines have been used in each country. Check the code below for more information. The data shows the United States has administrated most vaccines in the world followed by China, United Kingdom, England, India and at the last some countries include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

```
country_wise_total_vaccinated = {}
for country in df.country.unique() :
    vaccinated = 0
    for i in range(len(df)) :
        if df.country[i] == country :
            vaccinated += df.daily_vaccinations[i]
    country_wise_total_vaccinated[country] = vaccinated
#     made a seperate dict from the df
    country_wise_total_vaccinated_df =
pd.DataFrame.from_dict(country_wise_total_vaccinated,
                                                        orient='index',
                                                        columns = ['total_vaccinted_till_date'])
#     converted dict to df
country_wise_total_vaccinated_df.sort_values(by = 'total_vaccinted_till_date',
ascending = False, inplace = True)
country_wise_total_vaccinated_df
```



	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

```
fig = px.bar(country_wise_total_vaccinated_df,
             y = 'total_vaccinted_till_date',
             x = country_wise_total_vaccinated_df.index,
             color = 'total_vaccinted_till_date',
             color_discrete_sequence= px.colors.sequential.Viridis_r
            )
fig.update_layout(
    title={
            'text' : "Vaccination till date in various countries",
            'y':0.95,
            'x':0.5
        },
    xaxis_title="Countries",
    yaxis_title="Total vaccinated",
    legend_title="Total vaccinated"
)
fig.show()
```

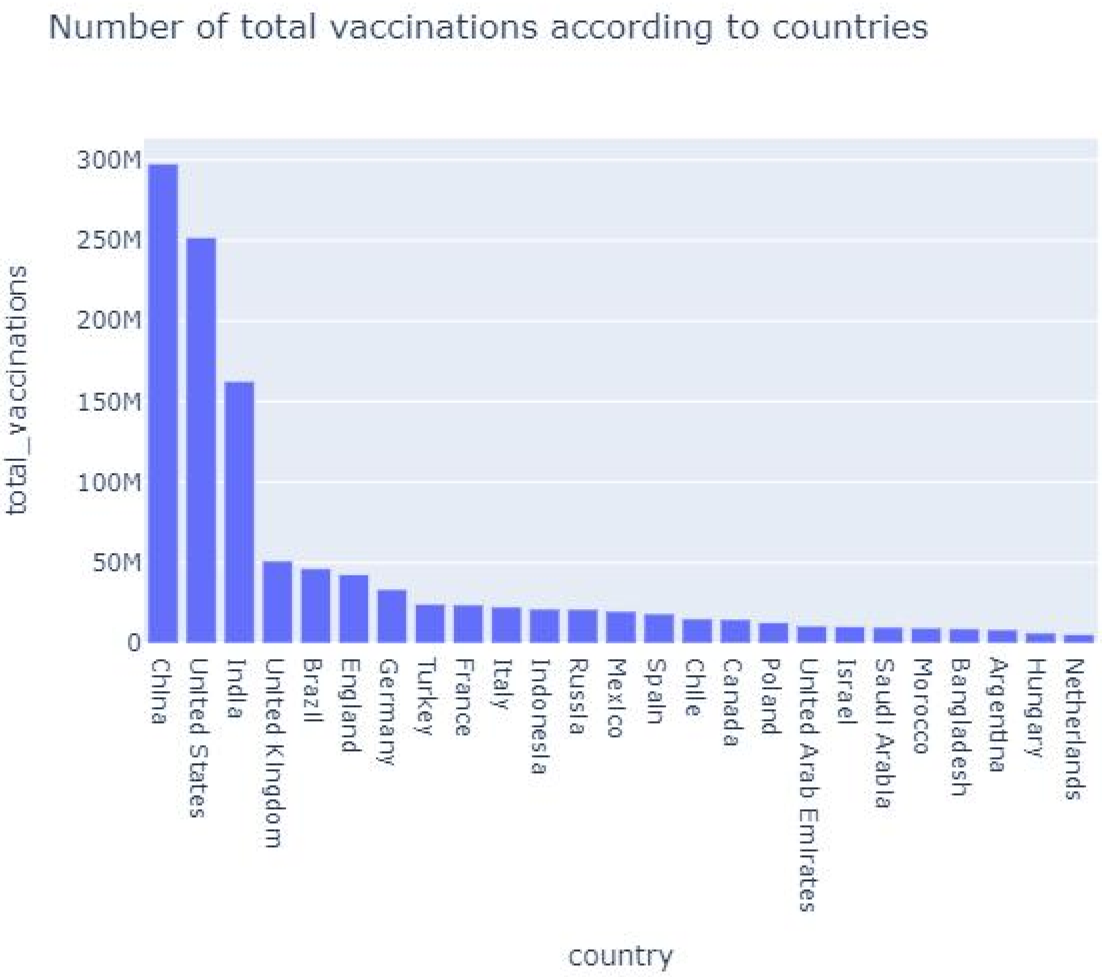


Country Wise Daily Vaccination

To check what is the vaccination trend in each country, check the below code. We are drawing the line plot where the x-axis is the date and the y-axis is the count of daily vaccination, Colours Is set to be the **country**.

```
data = new_df[['country','total_vaccinations']].nlargest(25,'total_vaccinations')
```

```
fig = px.bar(data, x = 'country',y = 'total_vaccinations',title="Number of total
vaccinations according to countries",)
fig.show()
```



2.7 METRICS USED FOR ACCURACY:

Precision is used for accuracy checks. Precision is a measure of a model's performance that tells you how many of the positive predictions made by the model are actually correct. It is calculated as the number of true positive predictions divided by the number of true positive and false positive predictions.