Covid - 19 vaccine analysis project

Aim:-

This project aims to analyze data in detail on vaccinated people and fully vaccinated by year and country.

Introduction:-

The main objective of this project is to analyze the data on, COVID-19 Vaccinations. Through the Analysis of data, we can find out some important insights.

Step 1: Data cleaning

- Load the dataset into a data frame using Pandas
- Explore the number of rows & columns, ranges of values etc.
- Handle missing, incorrect and invalid data

In [ ]:
```
# importing libraries
import pandas as pd
import scipy
import numpy as np
from sklearn.preprocessing import MinMaxScaler
import seaborn as sns
import matplotlib.pyplot as plt
```

In [ ]:
```
# Load the dataset
df = pd.read_csv('D:\country_vaccinations.csv')
print(df.head())
```

In [3]:
```
df.info()
```

Covid - 19 vaccine analysis project

```
In [3]: 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  int64
 4   people_vaccinated                  86512 non-null  int64
 5   people_fully_vaccinated            86512 non-null  int64
 6   daily_vaccinations_raw             86512 non-null  int64
 7   daily_vaccinations                 86512 non-null  int64
 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  int64
 12  vaccines                           86512 non-null  object
 13  source_name                        86512 non-null  object
 14  source_website                     86512 non-null  object
dtypes: float64(3), int64(6), object(6)
memory usage: 7.9+ MB
```

```
In [4]: df.describe()
```

Out[4]:

	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinat
count	8.651200e+04	8.651200e+04	8.651200e+04	8.651200e+04	8.651200e+04	86512.000000	
mean	2.315117e+07	8.451007e+06	6.341251e+06	1.106083e+05	1.308517e+05	40.419616	
std	1.611037e+08	4.969867e+07	3.890729e+07	7.864756e+05	7.669487e+05	62.707869	
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	
25%	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	8.770000e+02	0.000000	
50%	1.008000e+03	0.000000e+00	0.000000e+00	0.000000e+00	7.245000e+03	0.010000	
75%	3.697554e+06	1.843103e+06	1.137869e+06	1.280625e+04	4.370450e+04	68.750000	
max	3.263129e+09	1.275541e+09	1.240777e+09	2.474100e+07	2.242429e+07	345.370000	

Covid - 19 vaccine analysis project

```
In [9]: corr['total_vaccinations'].sort_values(ascending = False)
```

```
Out[9]: total_vaccinations                  1.000000
        daily_vaccinations                  0.688296
        daily_vaccinations_raw              0.662729
        people_fully_vaccinated             0.571087
        people_vaccinated                   0.535036
        total_vaccinations_per_hundred      0.222264
        people_vaccinated_per_hundred       0.106979
        people_fully_vaccinated_per_hundred 0.104074
        daily_vaccinations_per_million      0.050911
        Name: total_vaccinations, dtype: float64
```

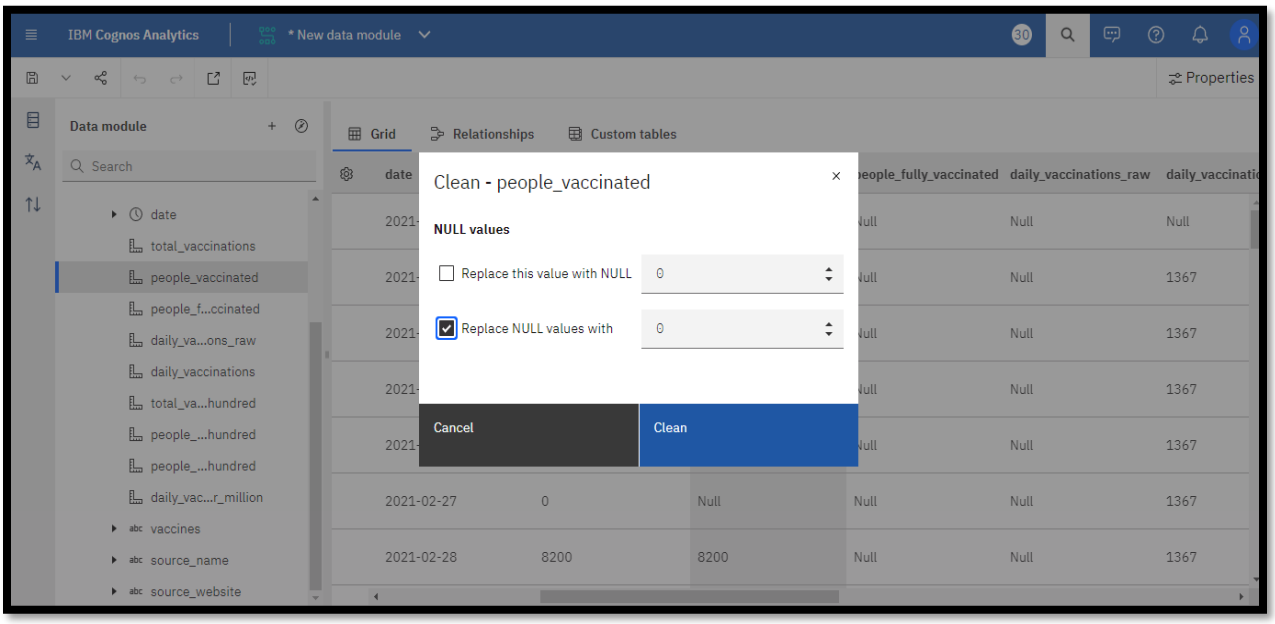
```
In [8]: vaccinations_df.isnull().sum()
```

```
Out[8]: 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
```

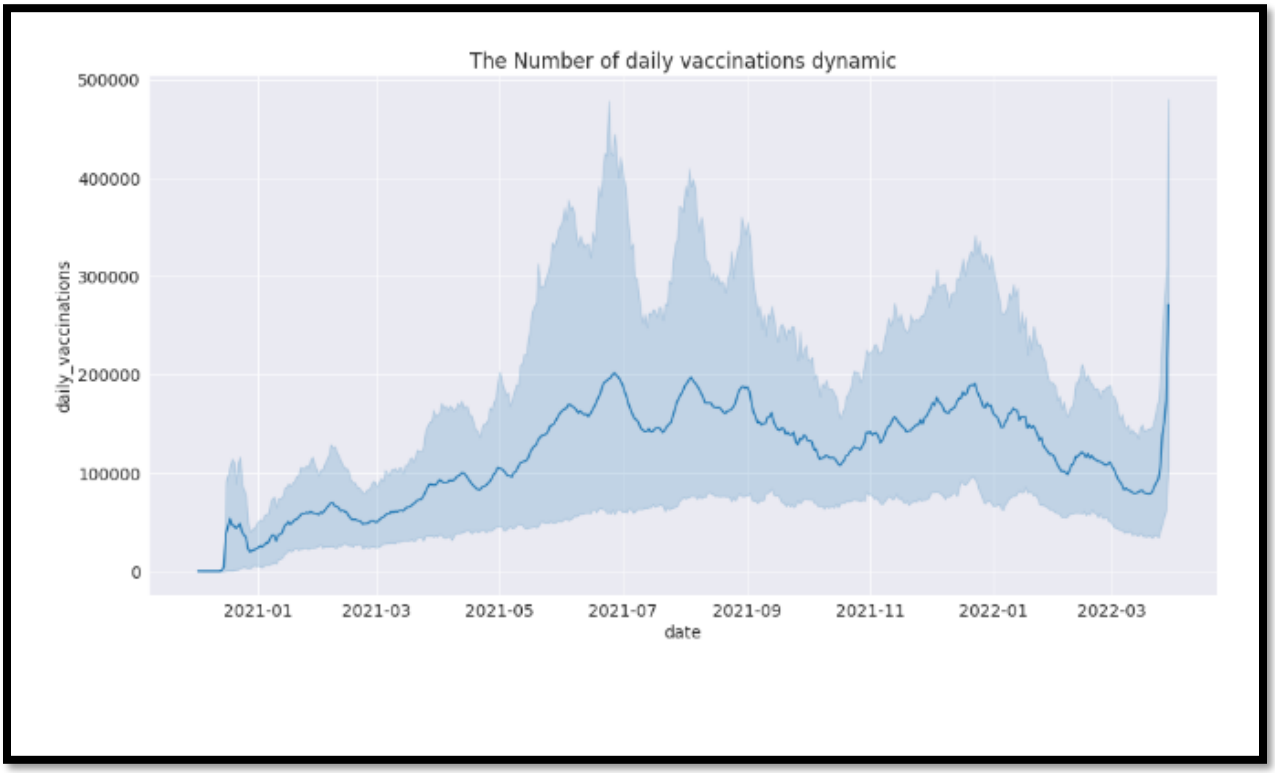
Step 2: Data interpretation

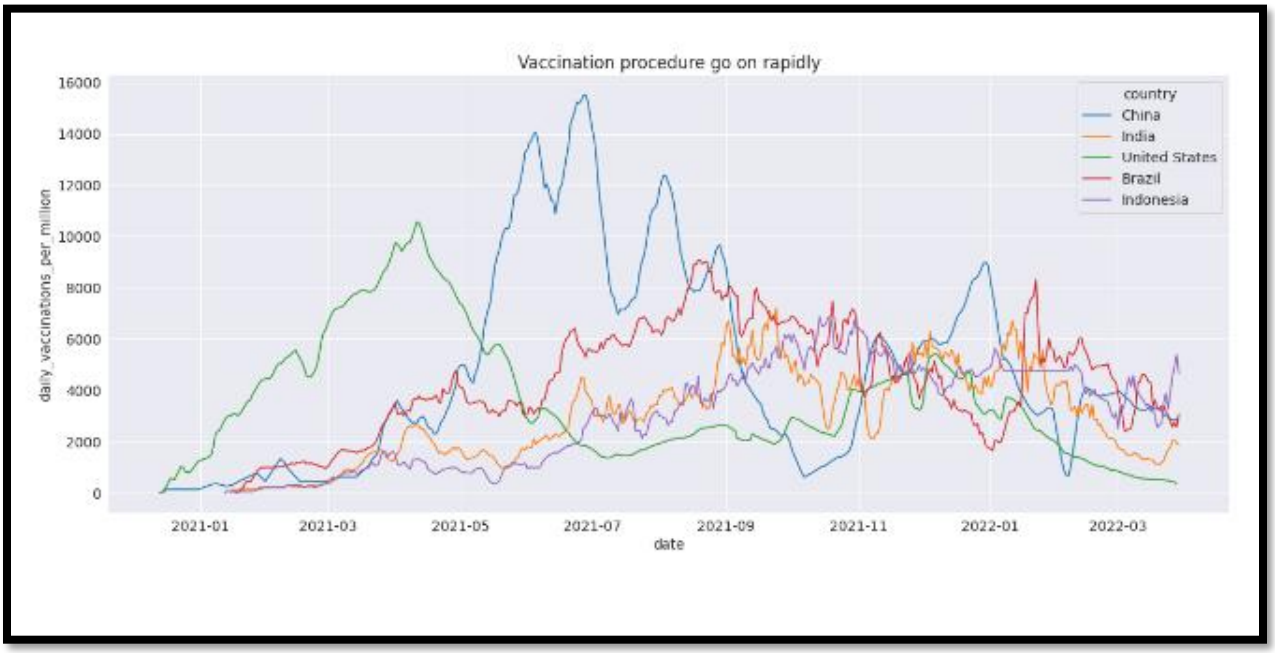
In this data interpretation step, I found out some important information about the data set.

- The data set was almost clear and completed.
- I replaced the null values with 0.
- There is total of 86512 rows and 15 columns.



Step 3: Visualization





Insights:-

• In conclusion, we can take a look at the final Report for further analysis.
• We can see the top 10 sources.
• Countries by fully vaccination etc.

Recommendations:-

• We can collect good datasets for analysis.
• We can analyze the data of the overall world.
• Like this dataset we can perform operations with various categories, city-wise or region-wise.
• We can work on large datasets and analyze them with the proper format of the chart.

Conclusion:-

In conclusion, we can take a look at the final Report for further analysis.
As we can see in the Report the top countries vaccination-wise, Top Country by the vaccination in daily and best top 10 countries with our vaccines. Vaccination peeks in 2022. The Top 2 vaccine sources are India, and China.