COVID-19 Vaccination Data Analysis

This project is about "**COVID-19 World Vaccination Progress**" Data Analysis with Python. Collected this Dataset from "Kaggle" which is the world's largest data science community with powerful tools and resources.

COVID-19 Vaccination image

This dataset contains 35310 rows and 15 columns which is really informaive to analysis. In this project,an attempt has been made to analyze various information of COVID-19 World Vaccination Progress such as country, total_Vaccinations, people_vaccinated, daily_vaccinations total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, vaccines and many more.

Library Used:

- pandas

- matplotlib

- seaborn

Data Preparation and 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 [1]:

```python
import pandas as pd
```

```python
vaccinations_df = pd.read_csv('../inp
ut/covid-world-vaccination-progress/c
ountry_vaccinations.csv')
```

```python
vaccinations_df
```

	country	iso_code	date	total_vaccinat
0	Afghanistan	AFG	2021-02-22	0.0
1	Afghanistan	AFG	2021-02-23	NaN
2	Afghanistan	AFG	2021-02-24	NaN
3	Afghanistan	AFG	2021-02-25	NaN
4	Afghanistan	AFG	2021-02-26	NaN
...	...	...	...	...

86507	Zimbabwe	ZWE	2022-03-25	8691642.0
86508	Zimbabwe	ZWE	2022-03-26	8791728.0
86509	Zimbabwe	ZWE	2022-03-27	8845039.0
86510	Zimbabwe	ZWE	2022-03-28	8934360.0
86511	Zimbabwe	ZWE	2022-03-29	9039729.0

86512 rows × 15 columns

In [4]:

```
vaccinations_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
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_hund
red  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
```

In [5]:

```
vaccinations_df.columns
```

Out[5]:

```
Index(['country', 'iso_code', 'date',
'total_vaccinations',
        'people_vaccinated', 'people_f
ully_vaccinated',
        'daily_vaccinations_raw', 'dai
ly_vaccinations',
        'total_vaccinations_per_hundre
d', 'people_vaccinated_per_hundred',
        'people_fully_vaccinated_per_h
```

```
ion',
       'vaccines', 'source_name', 'so
urce_website'],
      dtype='object')
```

In [6]:

```
vaccinations_df.shape
```

Out[6]:

```
(86512, 15)
```

In [7]:

```
vaccinations_df.describe()
```

Out[7]:

	total_vaccinations	people_vaccinated	people_f
count	4.360700e+04	4.129400e+04	3.88020
mean	4.592964e+07	1.770508e+07	1.41383
std	2.246004e+08	7.078731e+07	5.71392
min	0.000000e+00	0.000000e+00	1.00
25%	5.264100e+05	3.494642e+05	2.43962

50%	3.590096e+06	2.187310e+06	1.72214
75%	1.701230e+07	9.152520e+06	7.55987
max	3.263129e+09	1.275541e+09	1.24077

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
dtype: int64
```

In [9]:

```python
vaccinations_df.fillna(value=0, inplace=True)
date = vaccinations_df.date.str.split('-', expand=True)
```

```
('-', expand=True)
date
```

	0	1	2
0	2021	02	22
1	2021	02	23
2	2021	02	24
3	2021	02	25
4	2021	02	26
...	...	...	...
86507	2022	03	25
86508	2022	03	26
86509	2022	03	27
86510	2022	03	28
86511	2022	03	29

86512 rows × 3 columns

```
vaccinations_df['year'] = date[0]
vaccinations df['month'] = date[1]
```

```python
vaccinations_df['month'] = date[1]
vaccinations_df['day'] = date[2]


vaccinations_df.year = pd.to_numeric
(vaccinations_df.year)
vaccinations_df.month = pd.to_numeric
(vaccinations_df.month)
vaccinations_df.day = pd.to_numeric(v
accinations_df.day)


vaccinations_df.date = pd.to_datetime
(vaccinations_df.date)


vaccinations_df.head()
```

Out[10]:

	country	iso_code	date	total_vaccinations
0	Afghanistan	AFG	2021-02-22	0.0
1	Afghanistan	AFG	2021-02-23	0.0
2	Afghanistan	AFG	2021-02-24	0.0

			2021-02-25	0.0
3	Afghanistan	AFG		
4	Afghanistan	AFG	2021-02-26	0.0

In [11]:

```
vaccinations_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 18 columns):
 #    Column
Non-Null Count  Dtype
---   ------

---------------  -----
 0    country
86512 non-null  object
 1    iso_code
86512 non-null  object
 2    date
86512 non-null  datetime64[ns]
```