

```
In [28]: import numpy as np
import pandas as pd
import os
```

```
In [29]: df=pd.read_csv("covid_vaccine_statewise.csv")
```

```
In [30]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7845 entries, 0 to 7844
Data columns (total 24 columns):
 #   Column                                     Non-Null Count  Dtype
---  -
 0   Updated On                               7845 non-null   object
 1   State                                    7845 non-null   object
 2   Total Doses Administered                 7621 non-null   float64
 3   Sessions                                7621 non-null   float64
 4   Sites                                   7621 non-null   float64
 5   First Dose Administered                  7621 non-null   float64
 6   Second Dose Administered                 7621 non-null   float64
 7   Male (Doses Administered)               7461 non-null   float64
 8   Female (Doses Administered)             7461 non-null   float64
 9   Transgender (Doses Administered)        7461 non-null   float64
10   Covaxin (Doses Administered)            7621 non-null   float64
11   CoviShield (Doses Administered)         7621 non-null   float64
12   Sputnik V (Doses Administered)          2995 non-null   float64
13   AEFI                                     5438 non-null   float64
14   18-44 Years (Doses Administered)        1702 non-null   float64
15   45-60 Years (Doses Administered)        1702 non-null   float64
16   60+ Years (Doses Administered)          1702 non-null   float64
17   18-44 Years(Individuals Vaccinated)     3733 non-null   float64
18   45-60 Years(Individuals Vaccinated)     3734 non-null   float64
19   60+ Years(Individuals Vaccinated)       3734 non-null   float64
20   Male(Individuals Vaccinated)            160 non-null    float64
21   Female(Individuals Vaccinated)          160 non-null    float64
22   Transgender(Individuals Vaccinated)     160 non-null    float64
23   Total Individuals Vaccinated            5919 non-null   float64
dtypes: float64(22), object(2)
memory usage: 1.4+ MB
```

```
In [31]: df.describe()
```

Out[31]:

	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Admi
count	7.621000e+03	7.621000e+03	7621.000000	7.621000e+03	7.621000e+03	7.461000e+03	7.46
mean	9.188171e+06	4.792358e+05	2282.872064	7.414415e+06	1.773755e+06	3.620156e+06	3.16
std	3.746180e+07	1.911511e+06	7275.973730	2.995209e+07	7.570382e+06	1.737938e+07	1.51
min	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00	2.00
25%	1.356570e+05	6.004000e+03	69.000000	1.166320e+05	1.283100e+04	5.655500e+04	5.21
50%	8.182020e+05	4.547000e+04	597.000000	6.614590e+05	1.388180e+05	3.897850e+05	3.34
75%	6.625243e+06	3.428690e+05	1708.000000	5.387805e+06	1.166434e+06	2.735777e+06	2.56
max	5.132284e+08	3.501031e+07	73933.000000	4.001504e+08	1.130780e+08	2.701636e+08	2.39

8 rows × 22 columns



```
In [32]: df.columns
```

```
Out[32]: Index(['Updated On', 'State', 'Total Doses Administered', 'Sessions',
               'Sites ', 'First Dose Administered', 'Second Dose Administered',
               'Male (Doses Administered)', 'Female (Doses Administered)',
               'Transgender (Doses Administered)', 'Covaxin (Doses Administered)',
               'CoviShield (Doses Administered)', 'Sputnik V (Doses Administered)',
               'AEFI', '18-44 Years (Doses Administered)',
               '45-60 Years (Doses Administered)', '60+ Years (Doses Administered)',
               '18-44 Years(Individuals Vaccinated)',
               '45-60 Years(Individuals Vaccinated)',
               '60+ Years(Individuals Vaccinated)', 'Male(Individuals Vaccinated)',
               'Female(Individuals Vaccinated)', 'Transgender(Individuals Vaccinated)',
               'Total Individuals Vaccinated'],
              dtype='object')
```

## Data cleaning

```
In [33]: c=['State','Total Doses Administered','First Dose Administered','Second Dose Adminis
for i in df.columns:
    if str(i) not in c :
        df.drop(i,axis=1,inplace=True)
```

```
In [34]: df.head(100)
```

```
Out[34]:
```

	State	Total Doses Administered	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)
0	India	48276.0	48276.0	0.0	NaN	NaN
1	India	58604.0	58604.0	0.0	NaN	NaN
2	India	99449.0	99449.0	0.0	NaN	NaN
3	India	195525.0	195525.0	0.0	NaN	NaN
4	India	251280.0	251280.0	0.0	NaN	NaN
...	...	...	...	...	...	...
95	India	129646105.0	111673361.0	17972744.0	NaN	NaN
96	India	132754608.0	113584269.0	19170339.0	NaN	NaN
97	India	135658324.0	115461202.0	20197122.0	NaN	NaN
98	India	138185956.0	117105343.0	21080613.0	NaN	NaN
99	India	139185173.0	117795008.0	21390165.0	NaN	NaN

100 rows × 6 columns

```
In [35]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7845 entries, 0 to 7844
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   State                                7845 non-null   object
1   Total Doses Administered             7621 non-null   float64
2   First Dose Administered              7621 non-null   float64
3   Second Dose Administered             7621 non-null   float64
4   Male (Doses Administered)            7461 non-null   float64
5   Female (Doses Administered)          7461 non-null   float64
```

```
dtypes: float64(5), object(1)
memory usage: 367.9+ KB
```

```
df.describe()
```

```
In [36]: df.isnull().sum()
```

```
Out[36]: State                                0
Total Doses Administered                    224
First Dose Administered                     224
Second Dose Administered                    224
Male (Doses Administered)                   384
Female (Doses Administered)                  384
dtype: int64
```

```
In [37]: df.dropna(inplace=True )
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7461 entries, 160 to 7838
Data columns (total 6 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   State                                7461 non-null   object
 1   Total Doses Administered             7461 non-null   float64
 2   First Dose Administered               7461 non-null   float64
 3   Second Dose Administered              7461 non-null   float64
 4   Male (Doses Administered)            7461 non-null   float64
 5   Female (Doses Administered)          7461 non-null   float64
dtypes: float64(5), object(1)
memory usage: 408.0+ KB
```

## Number of person state wise vaccinated for first dose in india

```
In [38]: df["State"].unique()
```

```
Out[38]: array(['India', 'Andaman and Nicobar Islands', 'Andhra Pradesh',
                'Arunachal Pradesh', 'Assam', 'Bihar', 'Chandigarh',
                'Chhattisgarh', 'Dadra and Nagar Haveli and Daman and Diu',
                'Delhi', 'Goa', 'Gujarat', 'Haryana', 'Himachal Pradesh',
                'Jammu and Kashmir', 'Jharkhand', 'Karnataka', 'Kerala', 'Ladakh',
                'Lakshadweep', 'Madhya Pradesh', 'Maharashtra', 'Manipur',
                'Meghalaya', 'Mizoram', 'Nagaland', 'Odisha', 'Puducherry',
                'Punjab', 'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana',
                'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal'],
              dtype=object)
```

```
In [39]: state_wise_first_dose={}
for state in df.State.unique():
    vaccinated = 0
    for i in range(len(df)):
        if df.State.iloc[i] == state:
            vaccinated=df["First Dose Administered"].iloc[i]
    state_wise_first_dose[state]=vaccinated
state_wise_first_dose_df=pd.DataFrame.from_dict(state_wise_first_dose,orient="index")
state_wise_first_dose_df.sort_values(by='state_wise_first_dose',ascending = False, inplace=True)
```

```
In [40]: state_wise_first_dose_df
```

```
Out[40]:
```

	state_wise_first_dose
India	400150406.0

state_wise_first_dose	
Uttar Pradesh	45932488.0
Maharashtra	35040812.0
Madhya Pradesh	29723036.0
Gujarat	28101222.0
Rajasthan	27008606.0
Karnataka	25847691.0
Bihar	23350171.0
West Bengal	23257417.0
Tamil Nadu	20836674.0
Andhra Pradesh	17628583.0
Kerala	15670747.0
Odisha	13954592.0
Telangana	11649268.0
Assam	10495293.0
Haryana	10086831.0
Chhattisgarh	9181482.0
Jharkhand	8382280.0
Punjab	8005636.0
Delhi	7835546.0
Jammu and Kashmir	5318516.0
Uttarakhand	5070544.0
Himachal Pradesh	4249849.0
Tripura	2411195.0
Manipur	1159424.0
Goa	1094392.0
Meghalaya	938572.0
Chandigarh	700285.0
Arunachal Pradesh	692475.0
Mizoram	654946.0
Nagaland	632120.0
Puducherry	601591.0
Dadra and Nagar Haveli and Daman and Diu	584370.0
Sikkim	497851.0
Andaman and Nicobar Islands	216046.0
Ladakh	188699.0
Lakshadweep	51156.0

# Number of person state wise vaccinated for second dose in india

```
In [41]: state_wise_second_dose={}
for state in df.State.unique():
    vaccinated = 0
    for i in range(len(df)):
        if df.State.iloc[i] == state:
            vaccinated=df["Second Dose Administered"].iloc[i]
    state_wise_second_dose[state]=vaccinated
    state_wise_second_dose_df=pd.DataFrame.from_dict(state_wise_second_dose,orient="state_wise_second_dose_df.sort_values(by='state_wise_second_dose',ascending = False,
```

```
In [42]: state_wise_second_dose_df
```

```
Out[42]:
```

	state_wise_second_dose
India	113077994.0
Maharashtra	12112554.0
West Bengal	9132961.0
Gujarat	9051153.0
Uttar Pradesh	8515236.0
Rajasthan	8375056.0
Karnataka	7432852.0
Kerala	6426984.0
Andhra Pradesh	6214312.0
Madhya Pradesh	5733640.0
Tamil Nadu	4686034.0
Bihar	4484768.0
Odisha	4200094.0
Telangana	3965624.0
Delhi	3000536.0
Haryana	2923550.0
Chhattisgarh	2587695.0
Punjab	2285629.0
Assam	2208577.0
Jharkhand	1996014.0
Uttarakhand	1596572.0
Jammu and Kashmir	1489826.0
Himachal Pradesh	1382592.0
Tripura	804099.0
Goa	302519.0

	state_wise_second_dose
Manipur	246694.0
Meghalaya	231982.0
Chandigarh	223534.0
Mizoram	206773.0
Arunachal Pradesh	186619.0
Nagaland	159388.0
Puducherry	151771.0
Sikkim	151538.0
Andaman and Nicobar Islands	94597.0
Dadra and Nagar Haveli and Daman and Diu	80851.0
Ladakh	70337.0
Lakshadweep	17139.0

## Total Number of Male state wise vaccinated for first dose in india

```
In [43]: male_dose={}
for state in df.State.unique():
    vaccinated = 0
    for i in range(len(df)):
        if df.State.iloc[i] == state:
            vaccinated=df["Male (Doses Administered)"].iloc[i]
    male_dose[state]=vaccinated
male_dose_df=pd.DataFrame.from_dict(male_dose,orient="index",columns=['male_dose'])
male_dose_df.sort_values(by='male_dose',ascending = False, inplace = True)
```

```
In [44]: male_dose_df
```

```
Out[44]:
```

	male_dose
India	270163622.0
Uttar Pradesh	30643444.0
Maharashtra	25563569.0
Gujarat	20266401.0
Madhya Pradesh	19265405.0
Rajasthan	18618674.0
West Bengal	17234284.0
Karnataka	16724770.0
Bihar	14926420.0
Tamil Nadu	12952604.0
Andhra Pradesh	10852932.0
Kerala	10623457.0

	male_dose
Odisha	9617376.0
Telangana	8068394.0
Haryana	7206601.0
Assam	6739027.0
Delhi	6228216.0
Chhattisgarh	5916437.0
Punjab	5734736.0
Jharkhand	5699723.0
Jammu and Kashmir	3919555.0
Uttarakhand	3405375.0
Himachal Pradesh	2772475.0
Tripura	1646249.0
Manipur	759282.0
Goa	748770.0
Meghalaya	618575.0
Chandigarh	524263.0
Arunachal Pradesh	486874.0
Nagaland	451752.0
Dadra and Nagar Haveli and Daman and Diu	444656.0
Mizoram	433324.0
Puducherry	367047.0
Sikkim	354056.0
Andaman and Nicobar Islands	165554.0
Ladakh	146068.0
Lakshadweep	37277.0

```
In [45]: male_df=male_dose_df.drop('India')
```

```
In [46]: male_df.sum()
```

```
Out[46]: male_dose    270163622.0
dtype: float64
```

## Total Number of Female state wise vaccinated for first dose in india

```
In [47]: female_dose={}
for state in df.State.unique():
    vaccinated = 0
    for i in range(len(df)):
        if df.State.iloc[i] == state:
```

```
vaccinated=df["Female (Doses Administered)"].iloc[i]
female_dose[state]=vaccinated
female_dose_df=pd.DataFrame.from_dict(female_dose,orient="index",columns=['female_dose'])
female_dose_df.sort_values(by='female_dose',ascending = False, inplace = True)
```

In [48]: female\_dose\_df

Out[48]:

	female_dose
India	239518609.0
Uttar Pradesh	23785865.0
Maharashtra	21582082.0
Gujarat	16880326.0
Rajasthan	16758710.0
Karnataka	16550206.0
Madhya Pradesh	16184296.0
West Bengal	15151152.0
Andhra Pradesh	12986129.0
Bihar	12902990.0
Tamil Nadu	12565628.0
Kerala	11470142.0
Odisha	8533720.0
Telangana	7543612.0
Assam	5962985.0
Chhattisgarh	5851349.0
Haryana	5801370.0
Jharkhand	4676520.0
Delhi	4605508.0
Punjab	4555264.0
Uttarakhand	3259384.0
Jammu and Kashmir	2887608.0
Himachal Pradesh	2858812.0
Tripura	1568576.0
Goa	648014.0
Manipur	646498.0
Meghalaya	551829.0
Mizoram	428257.0
Chandigarh	399424.0
Arunachal Pradesh	391993.0
Puducherry	386171.0
Nagaland	339640.0



	female_dose
Sikkim	295116.0
Dadra and Nagar Haveli and Daman and Diu	220429.0
Andaman and Nicobar Islands	145049.0
Ladakh	112942.0
Lakshadweep	31013.0

```
In [49]: female_dose=female_dose_df.drop("India")
female_dose
```

	female_dose
Uttar Pradesh	23785865.0
Maharashtra	21582082.0
Gujarat	16880326.0
Rajasthan	16758710.0
Karnataka	16550206.0
Madhya Pradesh	16184296.0
West Bengal	15151152.0
Andhra Pradesh	12986129.0
Bihar	12902990.0
Tamil Nadu	12565628.0
Kerala	11470142.0
Odisha	8533720.0
Telangana	7543612.0
Assam	5962985.0
Chhattisgarh	5851349.0
Haryana	5801370.0
Jharkhand	4676520.0
Delhi	4605508.0
Punjab	4555264.0
Uttarakhand	3259384.0
Jammu and Kashmir	2887608.0
Himachal Pradesh	2858812.0
Tripura	1568576.0
Goa	648014.0
Manipur	646498.0
Meghalaya	551829.0
Mizoram	428257.0

	female_dose
Chandigarh	399424.0
Arunachal Pradesh	391993.0
Puducherry	386171.0
Nagaland	339640.0
Sikkim	295116.0
Dadra and Nagar Haveli and Daman and Diu	220429.0
Andaman and Nicobar Islands	145049.0
Ladakh	112942.0
Lakshadweep	31013.0

```
In [50]: female_dose.sum()

Out[50]: female_dose    239518609.0
dtype: float64

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:
```