

```
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
<b>count</b>	7.621000e+03	7.621000e+03	7621.000000	7.621000e+03	7.621000e+03	7.461000e+03	7.46
<b>mean</b>	9.188171e+06	4.792358e+05	2282.872064	7.414415e+06	1.773755e+06	3.620156e+06	3.16e
<b>std</b>	3.746180e+07	1.911511e+06	7275.973730	2.995209e+07	7.570382e+06	1.737938e+07	1.51!
<b>min</b>	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00	2.00!
<b>25%</b>	1.356570e+05	6.004000e+03	69.000000	1.166320e+05	1.283100e+04	5.655500e+04	5.21!
<b>50%</b>	8.182020e+05	4.547000e+04	597.000000	6.614590e+05	1.388180e+05	3.897850e+05	3.34!
<b>75%</b>	6.625243e+06	3.428690e+05	1708.000000	5.387805e+06	1.166434e+06	2.735777e+06	2.56
<b>max</b>	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)

	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="in
state_wise_first_dose_df.sort_values(by='state_wise_first_dose',ascending = False, i
```

In [40]: state\_wise\_first\_dose\_df

	state_wise_first_dose
India	400150406.0

<b>state_wise_first_dose</b>	
<b>Uttar Pradesh</b>	45932488.0
<b>Maharashtra</b>	35040812.0
<b>Madhya Pradesh</b>	29723036.0
<b>Gujarat</b>	28101222.0
<b>Rajasthan</b>	27008606.0
<b>Karnataka</b>	25847691.0
<b>Bihar</b>	23350171.0
<b>West Bengal</b>	23257417.0
<b>Tamil Nadu</b>	20836674.0
<b>Andhra Pradesh</b>	17628583.0
<b>Kerala</b>	15670747.0
<b>Odisha</b>	13954592.0
<b>Telangana</b>	11649268.0
<b>Assam</b>	10495293.0
<b>Haryana</b>	10086831.0
<b>Chhattisgarh</b>	9181482.0
<b>Jharkhand</b>	8382280.0
<b>Punjab</b>	8005636.0
<b>Delhi</b>	7835546.0
<b>Jammu and Kashmir</b>	5318516.0
<b>Uttarakhand</b>	5070544.0
<b>Himachal Pradesh</b>	4249849.0
<b>Tripura</b>	2411195.0
<b>Manipur</b>	1159424.0
<b>Goa</b>	1094392.0
<b>Meghalaya</b>	938572.0
<b>Chandigarh</b>	700285.0
<b>Arunachal Pradesh</b>	692475.0
<b>Mizoram</b>	654946.0
<b>Nagaland</b>	632120.0
<b>Puducherry</b>	601591.0
<b>Dadra and Nagar Haveli and Daman and Diu</b>	584370.0
<b>Sikkim</b>	497851.0
<b>Andaman and Nicobar Islands</b>	216046.0
<b>Ladakh</b>	188699.0
<b>Lakshadweep</b>	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="index")
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
<b>India</b>	113077994.0
<b>Maharashtra</b>	12112554.0
<b>West Bengal</b>	9132961.0
<b>Gujarat</b>	9051153.0
<b>Uttar Pradesh</b>	8515236.0
<b>Rajasthan</b>	8375056.0
<b>Karnataka</b>	7432852.0
<b>Kerala</b>	6426984.0
<b>Andhra Pradesh</b>	6214312.0
<b>Madhya Pradesh</b>	5733640.0
<b>Tamil Nadu</b>	4686034.0
<b>Bihar</b>	4484768.0
<b>Odisha</b>	4200094.0
<b>Telangana</b>	3965624.0
<b>Delhi</b>	3000536.0
<b>Haryana</b>	2923550.0
<b>Chhattisgarh</b>	2587695.0
<b>Punjab</b>	2285629.0
<b>Assam</b>	2208577.0
<b>Jharkhand</b>	1996014.0
<b>Uttarakhand</b>	1596572.0
<b>Jammu and Kashmir</b>	1489826.0
<b>Himachal Pradesh</b>	1382592.0
<b>Tripura</b>	804099.0
<b>Goa</b>	302519.0

state_wise_second_dose	
<b>Manipur</b>	246694.0
<b>Meghalaya</b>	231982.0
<b>Chandigarh</b>	223534.0
<b>Mizoram</b>	206773.0
<b>Arunachal Pradesh</b>	186619.0
<b>Nagaland</b>	159388.0
<b>Puducherry</b>	151771.0
<b>Sikkim</b>	151538.0
<b>Andaman and Nicobar Islands</b>	94597.0
<b>Dadra and Nagar Haveli and Daman and Diu</b>	80851.0
<b>Ladakh</b>	70337.0
<b>Lakshadweep</b>	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
<b>India</b>	270163622.0
<b>Uttar Pradesh</b>	30643444.0
<b>Maharashtra</b>	25563569.0
<b>Gujarat</b>	20266401.0
<b>Madhya Pradesh</b>	19265405.0
<b>Rajasthan</b>	18618674.0
<b>West Bengal</b>	17234284.0
<b>Karnataka</b>	16724770.0
<b>Bihar</b>	14926420.0
<b>Tamil Nadu</b>	12952604.0
<b>Andhra Pradesh</b>	10852932.0
<b>Kerala</b>	10623457.0

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

In [45]: male\_df=males\_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=[ 'femal
female_dose_df.sort_values(by='female_dose',ascending = False, inplace = True)
```

In [48]: female\_dose\_df

Out[48]:

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

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

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

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

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

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

```
Out[50]: female_dose    239518609.0
dtype: float64
```

```
In [ ]:
```