

Practical C4

Use the following covid_vaccine_statewise.csv dataset and perform following analytics on the given dataset

https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv

- a. Describe the dataset
- b. Number of persons state wise vaccinated for first dose in India
- c. Number of persons state wise vaccinated for second dose in India
- d. Number of Males vaccinated
- e. Number of females vaccinated

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('covid_vaccine_statewise.csv')
```

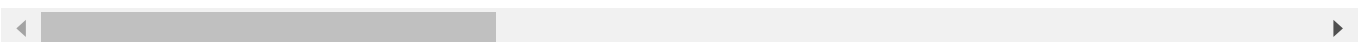
a. Describe the dataset

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

Out []:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	NaN	NaN
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	NaN	NaN
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	NaN	NaN
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	NaN	NaN
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	NaN	NaN

5 rows × 10 columns

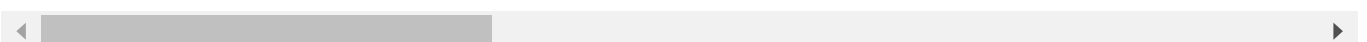


```
In [ ]: df.tail()
```

Out []:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)
7840	11/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7841	12/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7842	13/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7843	14/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7844	15/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 10 columns



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

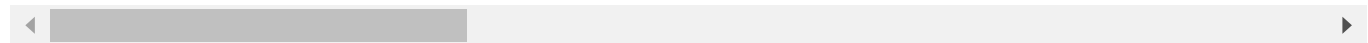
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7845 entries, 0 to 7844
Data columns (total 24 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Updated On                               7845 non-null   object
1   State                                    7845 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 [ ]: df.describe()
```

Out []:

	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)
count	7.621000e+03	7.621000e+03	7621.000000	7.621000e+03	7.621000e+03	7.461000e+03	7.461000e+03
mean	9.188171e+06	4.792358e+05	2282.872064	7.414415e+06	1.773755e+06	3.620156e+06	3.168416e+06
std	3.746180e+07	1.911511e+06	7275.973730	2.995209e+07	7.570382e+06	1.737938e+07	1.515310e+07
min	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00	2.000000e+00
25%	1.356570e+05	6.004000e+03	69.000000	1.166320e+05	1.283100e+04	5.655500e+04	5.210700e+04
50%	8.182020e+05	4.547000e+04	597.000000	6.614590e+05	1.388180e+05	3.897850e+05	3.342380e+05
75%	6.625243e+06	3.428690e+05	1708.000000	5.387805e+06	1.166434e+06	2.735777e+06	2.561513e+06
max	5.132284e+08	3.501031e+07	73933.000000	4.001504e+08	1.130780e+08	2.701636e+08	2.395186e+08

8 rows × 22 columns



Data Cleaning

```
In [ ]: # convert First Dose Administered to int
df['First Dose Administered'].isnull().sum()
df['First Dose Administered'].fillna(0,inplace=True)

df['First Dose Administered']=df['First Dose Administered'].astype(int)
```

```
In [ ]: # for Second Dose Administered
df['Second Dose Administered'].isnull().sum()
df['Second Dose Administered'].fillna(0,inplace=True)

df['Second Dose Administered']=df['Second Dose Administered'].astype(int)
```

```
In [ ]: # For Male(Individuals Vaccinated)
df['Male(Individuals Vaccinated)'].isnull().sum()
df['Male(Individuals Vaccinated)'].fillna(0,inplace=True)

df['Male(Individuals Vaccinated)']=df['Male(Individuals Vaccinated)'].astype(int)
```

```
In [ ]: # For Female(Individuals Vaccinated)
df['Female(Individuals Vaccinated)'].isnull().sum()
df['Female(Individuals Vaccinated)'].fillna(0,inplace=True)

df['Female(Individuals Vaccinated)']=df['Female(Individuals Vaccinated)'].astype(int)
```

a. Number of persons state wise vaccinated for first dose in India

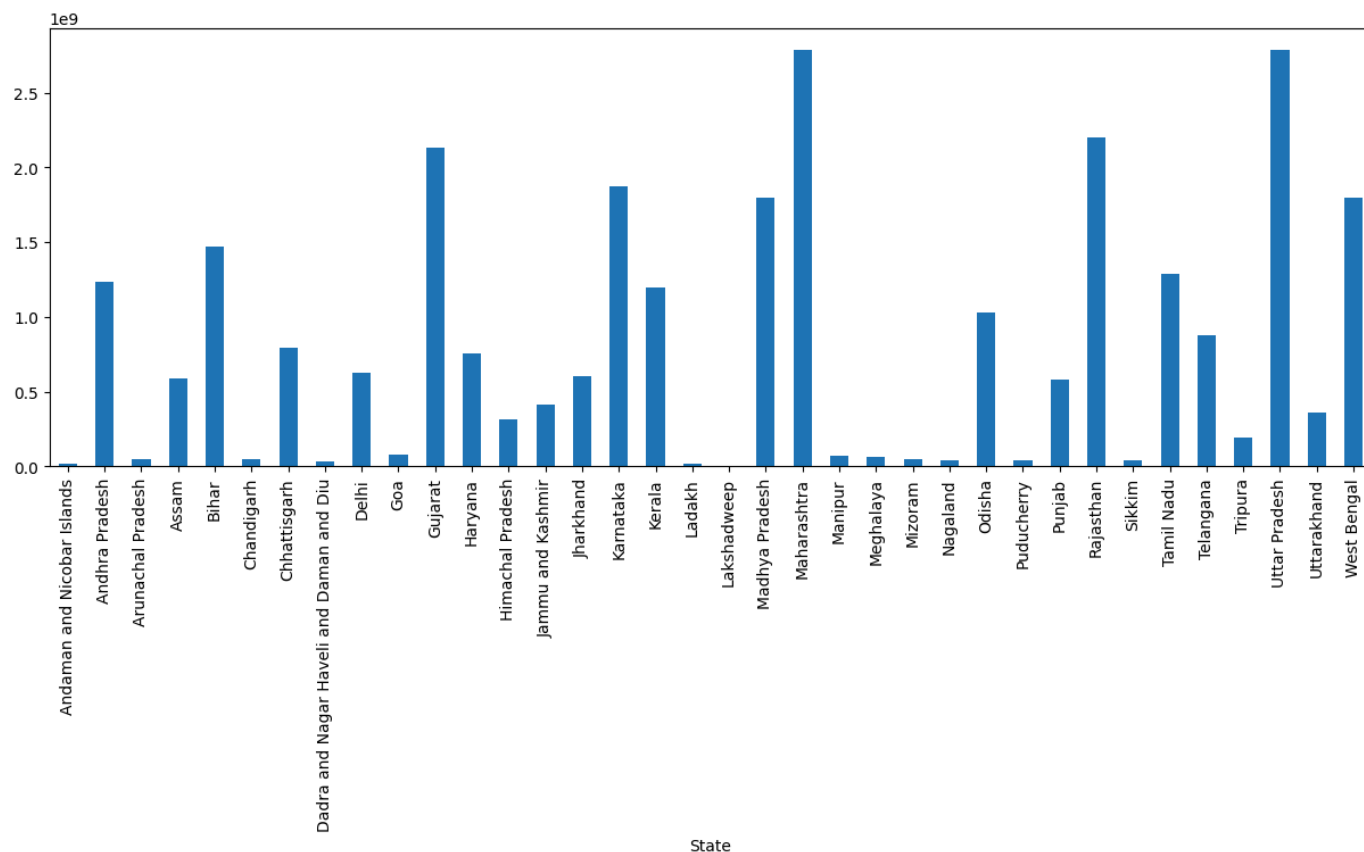
```
In [ ]: # get sum of First Dose Administered for group by state
first_dose_vaccinated = df.groupby('State')['First Dose Administered'].sum()
# remove india from the list
first_dose_vaccinated = first_dose_vaccinated.drop('India')
first_dose_vaccinated
```

```
Out[ ]: State
Andaman and Nicobar Islands    16425854
Andhra Pradesh                 1232860845
Arunachal Pradesh              49004980
Assam                          585600226
Bihar                          1470502878
Chandigarh                     44703105
Chhattisgarh                   796002902
Dadra and Nagar Haveli and Daman and Diu  33595063
Delhi                           624339473
Goa                             75991368
Gujarat                        2131646009
Haryana                         755798352
Himachal Pradesh               316294004
Jammu and Kashmir               410101777
Jharkhand                       603673726
Karnataka                       1873329968
Kerala                          1193845072
Ladakh                          17809249
Lakshadweep                     4363655
Madhya Pradesh                 1796604591
Maharashtra                     2784364331
Manipur                         67409568
Meghalaya                       62615974
Mizoram                         47873077
Nagaland                        42410766
Odisha                          1032633168
Puducherry                      41346858
Punjab                          584346582
Rajasthan                       2201044187
Sikkim                          36980929
Tamil Nadu                      1288532512
Telangana                       880320645
Tripura                         192689726
Uttar Pradesh                   2788411358
Uttarakhand                     363191446
West Bengal                     1796449989
Name: First Dose Administered, dtype: int64
```

```
In [ ]: # Plot graph of states vs first done vaccinated
```

```
first_dose_vaccinated.plot(kind='bar',figsize=(15,5))
```

Out[]: <Axes: xlabel='State'>



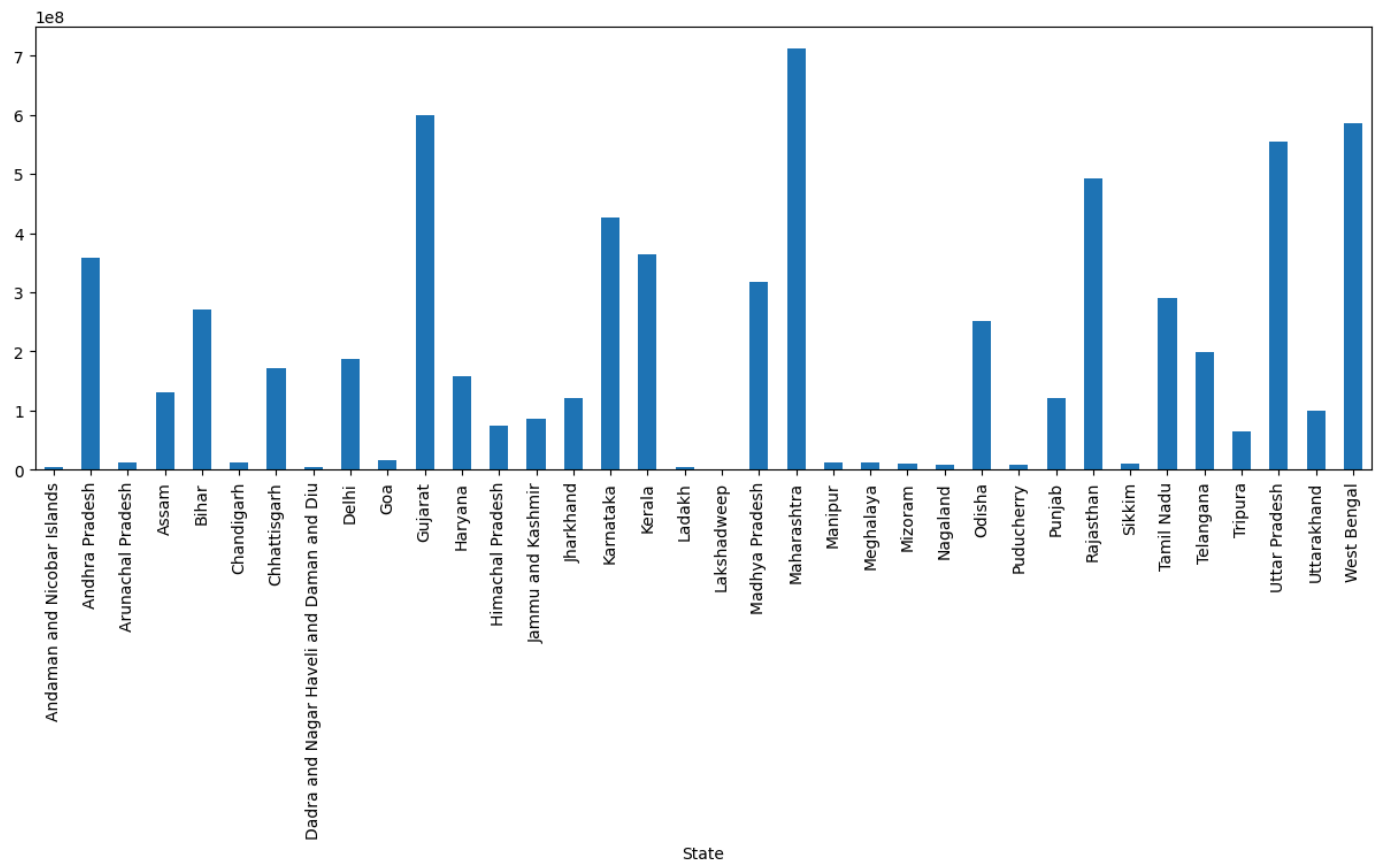
b. Number of persons state wise vaccinated for Second dose in India

```
In [ ]: # get sum of Second Dose Administered for group by state
second_dose_vaccinated = df.groupby('State')['Second Dose Administered'].sum()
# remove india from the list
second_dose_vaccinated = second_dose_vaccinated.drop('India')
second_dose_vaccinated
```

```
Out[ ]: State
Andaman and Nicobar Islands      4118554
Andhra Pradesh                    358817595
Arunachal Pradesh                 11932317
Assam                             130788792
Bihar                             270790571
Chandigarh                       11593735
Chhattisgarh                     172120400
Dadra and Nagar Haveli and Daman and Diu 4594416
Delhi                             188218946
Goa                               16198174
Gujarat                          600418376
Haryana                          158656058
Himachal Pradesh                  73838582
Jammu and Kashmir                 85951651
Jharkhand                        122121060
Karnataka                        427187178
Kerala                           364048753
Ladakh                           5453762
Lakshadweep                      1056446
Madhya Pradesh                   316932957
Maharashtra                      712881086
Manipur                          11858150
Meghalaya                        12166633
Mizoram                          9998418
Nagaland                         9204637
Odisha                           251302794
Puducherry                       8608859
Punjab                           121120995
Rajasthan                        491702988
Sikkim                           9723640
Tamil Nadu                       290670622
Telangana                        198152946
Tripura                          65270138
Uttar Pradesh                    554435112
Uttarakhand                      100085040
West Bengal                      586146880
Name: Second Dose Administered, dtype: int64
```

```
In [ ]: # Plot graph of states vs second done vaccinated
second_dose_vaccinated.plot(kind='bar',figsize=(15,5))
```

```
Out[ ]: <Axes: xlabel='State'>
```



d. Number of Males vaccinated

```
In [ ]: # get sum of Male(Individuals Vaccinated)
male_vaccinated = df['Male(Individuals Vaccinated)'].sum()
print(f"Total Male Vaccinated are {male_vaccinated}")
```

Total Male Vaccinated are 7138698858

e. Number of Females vaccinated

```
In [ ]: # get sum of Female(Individuals Vaccinated)
Female_vaccinated = df['Female(Individuals Vaccinated)'].sum()
print(f"Total Female Vaccinated are {Female_vaccinated}")
```

Total Female Vaccinated are 6321628736