

Covid 19 India Dataset Analysis by Python

Tanmoy Banerjee

Dataset :- <https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india> (<https://www.kaggle.com/datasets/sudalairajkumar/covid19-in-india>).

```
In [6]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
```

```
In [7]: covid_df=pd.read_csv('H:/covid.csv') #dataset calling
```

```
In [55]: covid_df.style.background_gradient(cmap='cubebehelix') #dataset presenting
```

Out[55]:

	Date	State/UnionTerritory	Cured	Deaths	Confirmed	Active_cases
0	2020-01-30	Kerala	0	0	1	1
1	2020-01-31	Kerala	0	0	1	1
2	2020-02-01	Kerala	0	0	2	2
3	2020-02-02	Kerala	0	0	3	3
4	2020-02-03	Kerala	0	0	3	3
5	2020-02-04	Kerala	0	0	3	3
6	2020-02-05	Kerala	0	0	3	3

In [56]: `covid_df.head(10)`

Out[56]:

	Date	State/UnionTerritory	Cured	Deaths	Confirmed	Active_cases
0	2020-01-30	Kerala	0	0	1	1
1	2020-01-31	Kerala	0	0	1	1
2	2020-02-01	Kerala	0	0	2	2
3	2020-02-02	Kerala	0	0	3	3
4	2020-02-03	Kerala	0	0	3	3
5	2020-02-04	Kerala	0	0	3	3
6	2020-02-05	Kerala	0	0	3	3
7	2020-02-06	Kerala	0	0	3	3
8	2020-02-07	Kerala	0	0	3	3
9	2020-02-08	Kerala	0	0	3	3

In [11]: `covid_df.info() #checking basic information about covid19 india data`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18110 entries, 0 to 18109
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Sno              18110 non-null   int64  
 1   Date             18110 non-null   object  
 2   Time             18110 non-null   object  
 3   State/UnionTerritory  18110 non-null   object  
 4   ConfirmedIndianNational  18110 non-null   object  
 5   ConfirmedForeignNational  18110 non-null   object  
 6   Cured            18110 non-null   int64  
 7   Deaths           18110 non-null   int64  
 8   Confirmed        18110 non-null   int64  
dtypes: int64(4), object(5)
memory usage: 1.2+ MB
```

In [13]: `covid_df.describe().round(2) #statistical description of the data`

Out[13]:

	Sno	Cured	Deaths	Confirmed
count	18110.00	18110.00	18110.00	18110.00
mean	9055.50	278637.52	4052.40	301031.40
std	5228.05	614890.89	10919.08	656148.87
min	1.00	0.00	0.00	0.00
25%	4528.25	3360.25	32.00	4376.75
50%	9055.50	33364.00	588.00	39773.50
75%	13582.75	278869.75	3643.75	300149.75
max	18110.00	6159676.00	134201.00	6363442.00

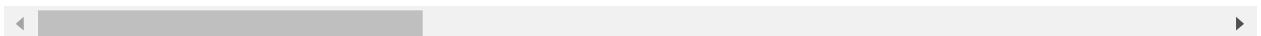
In [14]: `vaccine_df=pd.read_csv('H:/vaccine.csv') #vaccination dataset calling`

In [15]: `vaccine_df #checking vaccination dataset`

Out[15]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Dose Administered)
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	N
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	N
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	N
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	N
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	N
...
7840	11/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	N
7841	12/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	N
7842	13/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	N
7843	14/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	N
7844	15/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	N

7845 rows × 24 columns



In [17]: `vaccine_df.head(8) #displaying top 8 rows`

Out[17]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	NaN
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	NaN
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	NaN
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	NaN
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	NaN
5	21/01/2021	India	365965.0	32226.0	12600.0	365965.0	0.0	NaN
6	22/01/2021	India	549381.0	36988.0	14115.0	549381.0	0.0	NaN
7	23/01/2021	India	759008.0	43076.0	15605.0	759008.0	0.0	NaN

8 rows × 24 columns

In [18]: `vaccine_df.tail(7)`

Out[18]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)
7838	09/08/2021	West Bengal	32390378.0	866173.0	2086.0	23257417.0	9132961.0	17234284
7839	10/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN
7840	11/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN
7841	12/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN
7842	13/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN
7843	14/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN
7844	15/08/2021	West Bengal		NaN	NaN	NaN	NaN	NaN

7 rows × 24 columns

```
In [21]: vaccine_df.describe().round(4)      #statistical description of vaccine dataset
```

Out[21]:

	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Admini
count	7.621000e+03	7.621000e+03	7621.0000	7.621000e+03	7.621000e+03	7.461000e+03	7.4610
mean	9.188171e+06	4.792358e+05	2282.8721	7.414415e+06	1.773755e+06	3.620156e+06	3.1684
std	3.746180e+07	1.911511e+06	7275.9737	2.995209e+07	7.570382e+06	1.737938e+07	1.5153
min	7.000000e+00	0.000000e+00	0.0000	7.000000e+00	0.000000e+00	0.000000e+00	2.0000
25%	1.356570e+05	6.004000e+03	69.0000	1.166320e+05	1.283100e+04	5.655500e+04	5.2107
50%	8.182020e+05	4.547000e+04	597.0000	6.614590e+05	1.388180e+05	3.897850e+05	3.3423
75%	6.625243e+06	3.428690e+05	1708.0000	5.387805e+06	1.166434e+06	2.735777e+06	2.5615
max	5.132284e+08	3.501031e+07	73933.0000	4.001504e+08	1.130780e+08	2.701636e+08	2.3951

8 rows × 22 columns



In [20]: `vaccine_df.info() #basic info checking of vaccination dataset`

```
<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 [22]: `covid_df.columns #checking columns of covid data`

```
Out[22]: Index(['Sno', 'Date', 'Time', 'State/UnionTerritory',
               'ConfirmedIndianNational', 'ConfirmedForeignNational', 'Cured',
               'Deaths', 'Confirmed'],
               dtype='object')
```

In [23]: `#drop specific blank columns which are useless
covid_df`

Out[23]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational
	0	1	2020-01-30	6:00 PM	Kerala	1
	1	2	2020-01-31	6:00 PM	Kerala	1
	2	3	2020-02-01	6:00 PM	Kerala	2
	3	4	2020-02-02	6:00 PM	Kerala	3
	4	5	2020-02-03	6:00 PM	Kerala	3

18105	18106	18106	2021-08-11	8:00 AM	Telangana	-
18106	18107	18107	2021-08-11	8:00 AM	Tripura	-
18107	18108	18108	2021-08-11	8:00 AM	Uttarakhand	-
18108	18109	18109	2021-08-11	8:00 AM	Uttar Pradesh	-
18109	18110	18110	2021-08-11	8:00 AM	West Bengal	-

18110 rows × 9 columns

In [24]: `covid_df.drop(['Sno','Time','ConfirmedIndianNational','ConfirmedForeignNational'])`

In [25]: covid_df

Out[25]:

	Date	State/UnionTerritory	Cured	Deaths	Confirmed
0	2020-01-30	Kerala	0	0	1
1	2020-01-31	Kerala	0	0	1
2	2020-02-01	Kerala	0	0	2
3	2020-02-02	Kerala	0	0	3
4	2020-02-03	Kerala	0	0	3
...
18105	2021-08-11	Telangana	638410	3831	650353
18106	2021-08-11	Tripura	77811	773	80660
18107	2021-08-11	Uttarakhand	334650	7368	342462
18108	2021-08-11	Uttar Pradesh	1685492	22775	1708812
18109	2021-08-11	West Bengal	1506532	18252	1534999

18110 rows × 5 columns

In [26]: *#new column added of active cases in covid_df dataset*
`covid_df['Active_cases']=covid_df['Confirmed']-(covid_df['Cured']+covid_df['Deaths'])`

In [28]: covid_df.tail()

Out[28]:

	Date	State/UnionTerritory	Cured	Deaths	Confirmed	Active_cases
18105	2021-08-11	Telangana	638410	3831	650353	8112
18106	2021-08-11	Tripura	77811	773	80660	2076
18107	2021-08-11	Uttarakhand	334650	7368	342462	444
18108	2021-08-11	Uttar Pradesh	1685492	22775	1708812	545
18109	2021-08-11	West Bengal	1506532	18252	1534999	10215

In [33]: *#creating a pivot table statewise for covid data*
`statewise=pd.pivot_table(covid_df,values=['Confirmed','Deaths','Cured'],index='St`

In [34]: `statewise['Recoveryrate']=statewise['Cured']*100/statewise['Confirmed']`

In [35]: `statewise['Mortalityrate']=statewise['Deaths']*100/statewise['Confirmed']`

In [39]: `statewise=statewise.sort_values(by='Confirmed',ascending=False)`

In [40]: `statewise.style.background_gradient(cmap='cubehelix') #by confirmed cases`

Out[40]:

State/UnionTerritory	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
Maharashtra	6363442	6159676	134201	96.797865	2.108937
Maharashtra***	6229596	6000911	130753	96.329056	2.098900
Kerala	3586693	3396184	18004	94.688450	0.501967
Karnataka	2921049	2861499	36848	97.961349	1.261465
Karanataka	2885238	2821491	36197	97.790581	1.254559
Tamil Nadu	2579130	2524400	34367	97.877967	1.332504
Andhra Pradesh	1985182	1952736	13564	98.365591	0.683262
Uttar Pradesh	1708812	1685492	22775	98.635309	1.332797
West Bengal	1534999	1506532	18252	98.145471	1.189056
Delhi	1436852	1411280	25068	98.220276	1.744647
Chhattisgarh	1003356	988189	13544	98.488373	1.349870
Odisha	988997	972710	6565	98.353180	0.663804
Rajasthan	953851	944700	8954	99.040626	0.938721
Gujarat	825085	814802	10077	98.753704	1.221329
Madhya Pradesh	791980	781330	10514	98.655269	1.327559
Madhya Pradesh***	791656	780735	10506	98.620487	1.327092
Haryana	770114	759790	9652	98.659419	1.253321
Bihar	725279	715352	9646	98.631285	1.329971
Bihar****	715730	701234	9452	97.974655	1.320610
Telangana	650353	638410	3831	98.163613	0.589065
Punjab	599573	582791	16322	97.201008	2.722271
Assam	576149	559684	5420	97.142232	0.940729
Telengana	443360	362160	2312	81.685312	0.521472
Jharkhand	347440	342102	5130	98.463620	1.476514
Uttarakhand	342462	334650	7368	97.718871	2.151480
Jammu and Kashmir	322771	317081	4392	98.237140	1.360717
Himachal Pradesh	208616	202761	3537	97.193408	1.695460
Himanchal Pradesh	204516	200040	3507	97.811418	1.714780
Goa	172085	167978	3164	97.613389	1.838626
Puducherry	121766	119115	1800	97.822873	1.478245
Manipur	105424	96776	1664	91.796934	1.578388
Tripura	80660	77811	773	96.467890	0.958344
Meghalaya	69769	64157	1185	91.956313	1.698462

State/UnionTerritory	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
Chandigarh	61992	61150	811	98.641760	1.308233
Arunachal Pradesh	50605	47821	248	94.498567	0.490070
Mizoram	46320	33722	171	72.802245	0.369171
Nagaland	28811	26852	585	93.200514	2.030474
Sikkim	28018	25095	356	89.567421	1.270612
Ladakh	20411	20130	207	98.623291	1.014159
Dadra and Nagar Haveli and Daman and Diu	10654	10646	4	99.924911	0.037545
Dadra and Nagar Haveli	10377	10261	4	98.882143	0.038547
Lakshadweep	10263	10165	51	99.045114	0.496931
Cases being reassigned to states	9265	0	0	0.000000	0.000000
Andaman and Nicobar Islands	7548	7412	129	98.198198	1.709062
Unassigned	77	0	0	0.000000	0.000000
Daman & Diu	2	0	0	0.000000	0.000000

```
In [41]: statewise=statewise.sort_values(by='Mortalityrate',ascending=False)
```

In [42]: `statewise.style.background_gradient(cmap='cubehelix') #by mortality rate`

Out[42]:

State/UnionTerritory	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
Punjab	599573	582791	16322	97.201008	2.722271
Uttarakhand	342462	334650	7368	97.718871	2.151480
Maharashtra	6363442	6159676	134201	96.797865	2.108937
Maharashtra***	6229596	6000911	130753	96.329056	2.098900
Nagaland	28811	26852	585	93.200514	2.030474
Goa	172085	167978	3164	97.613389	1.838626
Delhi	1436852	1411280	25068	98.220276	1.744647
Himachal Pradesh	204516	200040	3507	97.811418	1.714780
Andaman and Nicobar Islands	7548	7412	129	98.198198	1.709062
Meghalaya	69769	64157	1185	91.956313	1.698462
Himachal Pradesh	208616	202761	3537	97.193408	1.695460
Manipur	105424	96776	1664	91.796934	1.578388
Puducherry	121766	119115	1800	97.822873	1.478245
Jharkhand	347440	342102	5130	98.463620	1.476514
Jammu and Kashmir	322771	317081	4392	98.237140	1.360717
Chhattisgarh	1003356	988189	13544	98.488373	1.349870
Uttar Pradesh	1708812	1685492	22775	98.635309	1.332797
Tamil Nadu	2579130	2524400	34367	97.877967	1.332504
Bihar	725279	715352	9646	98.631285	1.329971
Madhya Pradesh	791980	781330	10514	98.655269	1.327559
Madhya Pradesh***	791656	780735	10506	98.620487	1.327092
Bihar****	715730	701234	9452	97.974655	1.320610
Chandigarh	61992	61150	811	98.641760	1.308233
Sikkim	28018	25095	356	89.567421	1.270612
Karnataka	2921049	2861499	36848	97.961349	1.261465
Karanataka	2885238	2821491	36197	97.790581	1.254559
Haryana	770114	759790	9652	98.659419	1.253321
Gujarat	825085	814802	10077	98.753704	1.221329
West Bengal	1534999	1506532	18252	98.145471	1.189056
Ladakh	20411	20130	207	98.623291	1.014159
Tripura	80660	77811	773	96.467890	0.958344
Assam	576149	559684	5420	97.142232	0.940729
Rajasthan	953851	944700	8954	99.040626	0.938721

State/UnionTerritory	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
Andhra Pradesh	1985182	1952736	13564	98.365591	0.683262
Odisha	988997	972710	6565	98.353180	0.663804
Telangana	650353	638410	3831	98.163613	0.589065
Telengana	443360	362160	2312	81.685312	0.521472
Kerala	3586693	3396184	18004	94.688450	0.501967
Lakshadweep	10263	10165	51	99.045114	0.496931
Arunachal Pradesh	50605	47821	248	94.498567	0.490070
Mizoram	46320	33722	171	72.802245	0.369171
Dadra and Nagar Haveli	10377	10261	4	98.882143	0.038547
Dadra and Nagar Haveli and Daman and Diu	10654	10646	4	99.924911	0.037545
Cases being reassigned to states	9265	0	0	0.000000	0.000000
Unassigned	77	0	0	0.000000	0.000000
Daman & Diu	2	0	0	0.000000	0.000000

In [45]: statewise=statewise.sort_values(by='Recoveryrate', ascending=False)

In [46]: `statewise.style.background_gradient(cmap='cubehelix') #by recovery rate`

Out[46]:

State/UnionTerritory	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
Dadra and Nagar Haveli and Daman and Diu	10654	10646	4	99.924911	0.037545
Lakshadweep	10263	10165	51	99.045114	0.496931
Rajasthan	953851	944700	8954	99.040626	0.938721
Dadra and Nagar Haveli	10377	10261	4	98.882143	0.038547
Gujarat	825085	814802	10077	98.753704	1.221329
Haryana	770114	759790	9652	98.659419	1.253321
Madhya Pradesh	791980	781330	10514	98.655269	1.327559
Chandigarh	61992	61150	811	98.641760	1.308233
Uttar Pradesh	1708812	1685492	22775	98.635309	1.332797
Bihar	725279	715352	9646	98.631285	1.329971
Ladakh	20411	20130	207	98.623291	1.014159
Madhya Pradesh***	791656	780735	10506	98.620487	1.327092
Chhattisgarh	1003356	988189	13544	98.488373	1.349870
Jharkhand	347440	342102	5130	98.463620	1.476514
Andhra Pradesh	1985182	1952736	13564	98.365591	0.683262
Odisha	988997	972710	6565	98.353180	0.663804
Jammu and Kashmir	322771	317081	4392	98.237140	1.360717
Delhi	1436852	1411280	25068	98.220276	1.744647
Andaman and Nicobar Islands	7548	7412	129	98.198198	1.709062
Telangana	650353	638410	3831	98.163613	0.589065
West Bengal	1534999	1506532	18252	98.145471	1.189056
Bihar****	715730	701234	9452	97.974655	1.320610
Karnataka	2921049	2861499	36848	97.961349	1.261465
Tamil Nadu	2579130	2524400	34367	97.877967	1.332504
Puducherry	121766	119115	1800	97.822873	1.478245
Himachal Pradesh	204516	200040	3507	97.811418	1.714780
Karanataka	2885238	2821491	36197	97.790581	1.254559
Uttarakhand	342462	334650	7368	97.718871	2.151480
Goa	172085	167978	3164	97.613389	1.838626
Punjab	599573	582791	16322	97.201008	2.722271
Himachal Pradesh	208616	202761	3537	97.193408	1.695460
Assam	576149	559684	5420	97.142232	0.940729
Maharashtra	6363442	6159676	134201	96.797865	2.108937

	Confirmed	Cured	Deaths	Recoveryrate	Mortalityrate
State/UnionTerritory					
Tripura	80660	77811	773	96.467890	0.958344
Maharashtra***	6229596	6000911	130753	96.329056	2.098900
Kerala	3586693	3396184	18004	94.688450	0.501967
Arunachal Pradesh	50605	47821	248	94.498567	0.490070
Nagaland	28811	26852	585	93.200514	2.030474
Meghalaya	69769	64157	1185	91.956313	1.698462
Manipur	105424	96776	1664	91.796934	1.578388
Sikkim	28018	25095	356	89.567421	1.270612
Telengana	443360	362160	2312	81.685312	0.521472
Mizoram	46320	33722	171	72.802245	0.369171
Cases being reassigned to states	9265	0	0	0.000000	0.000000
Unassigned	77	0	0	0.000000	0.000000
Daman & Diu	2	0	0	0.000000	0.000000

In [64]:

```
.groupby(by='State/UnionTerritory').max()[['Active_cases','Date']].sort_values(by=
```

The syntax to find out most cases , deaths, confirmed cases etc max,min cases across india

```
top_1_active_cases=covid_df.groupby(by='State/UnionTerritory').max()
[['Active_cases','Date']].sort_values(by=
['Active_cases'],ascending=False).reset_index()
```

In [62]:

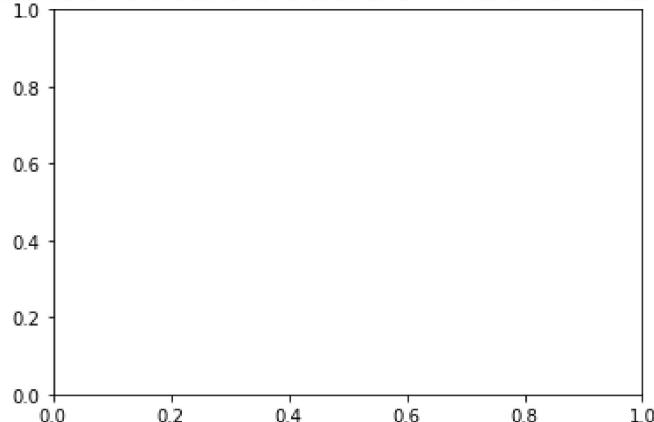
```
fig=plt.figure(figsize=(16,9))
```

<Figure size 1152x648 with 0 Axes>

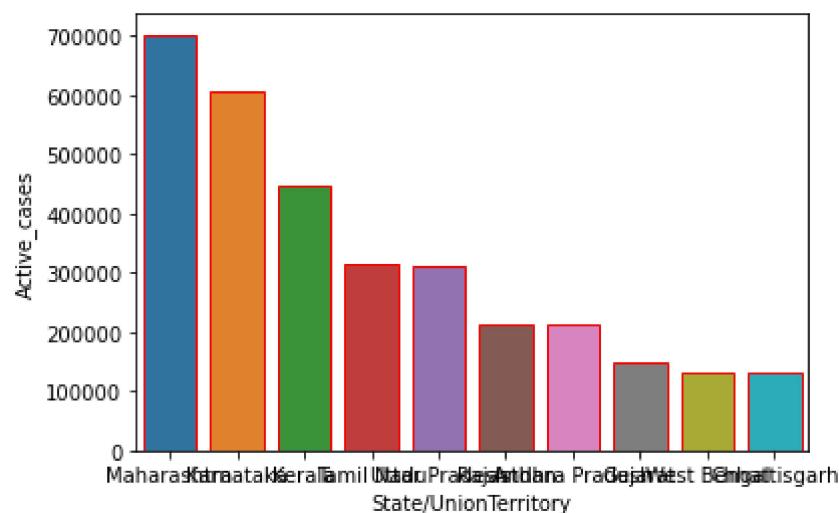
```
In [63]: plt.title("Top 10 States With Most Active Cases In India",size=30)
```

```
Out[63]: Text(0.5, 1.0, 'Top 10 States With Most Active Cases In India')
```

Top 10 States With Most Active Cases In India

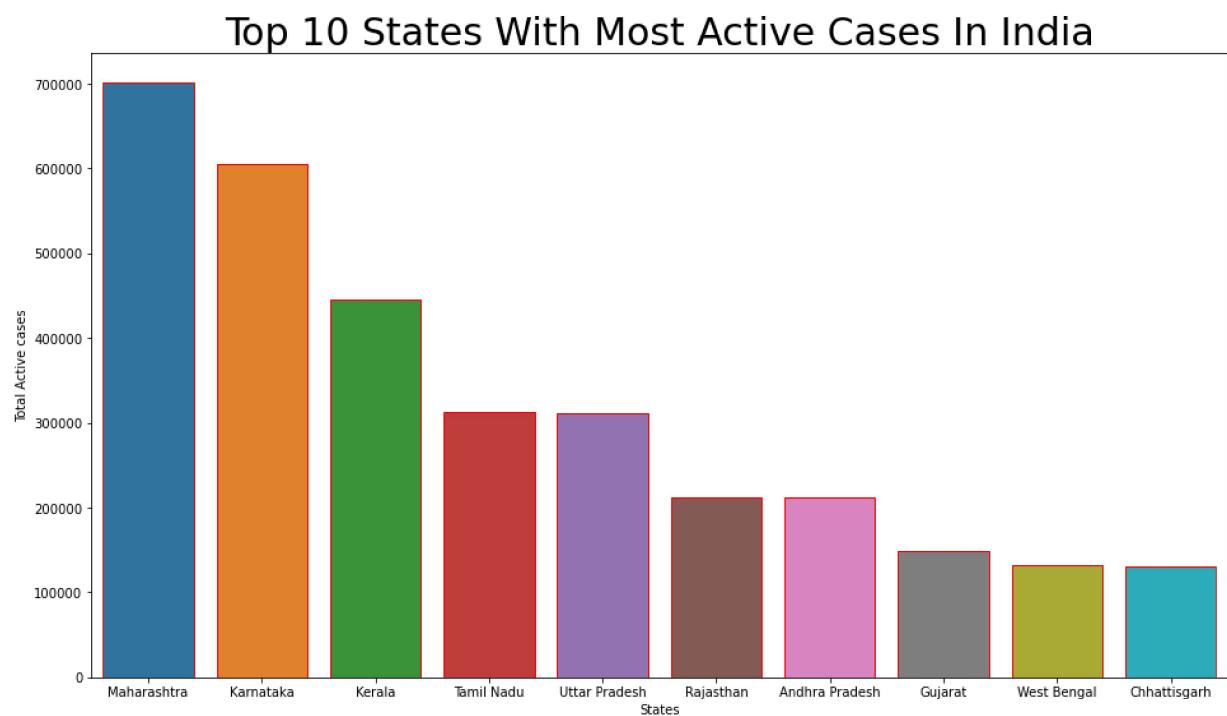


```
In [68]: ax=sns.barplot(data=top_10_active_cases.iloc[:10],y='Active_cases',x='State/UnionTerritory')
```



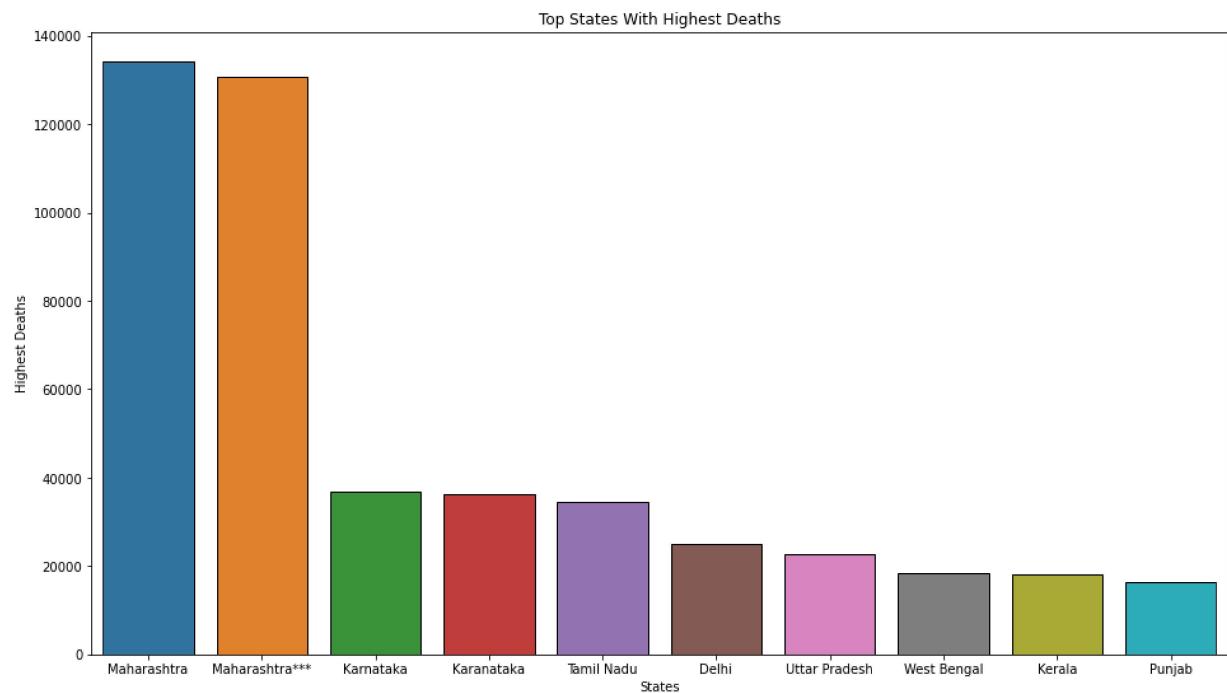
all those above line we took it one place to re_assemble

```
In [69]: top_10_active_cases=covid_df.groupby(by='State/UnionTerritory').max()[['Active_cases']]
fig=plt.figure(figsize=(16,9))
plt.title("Top 10 States With Most Active Cases In India",size=30)
ax=sns.barplot(data=top_10_active_cases.iloc[:10],y='Active_cases',x='State/UnionTerritory')
plt.xlabel('States')
plt.ylabel('Total Active cases')
plt.show()
```



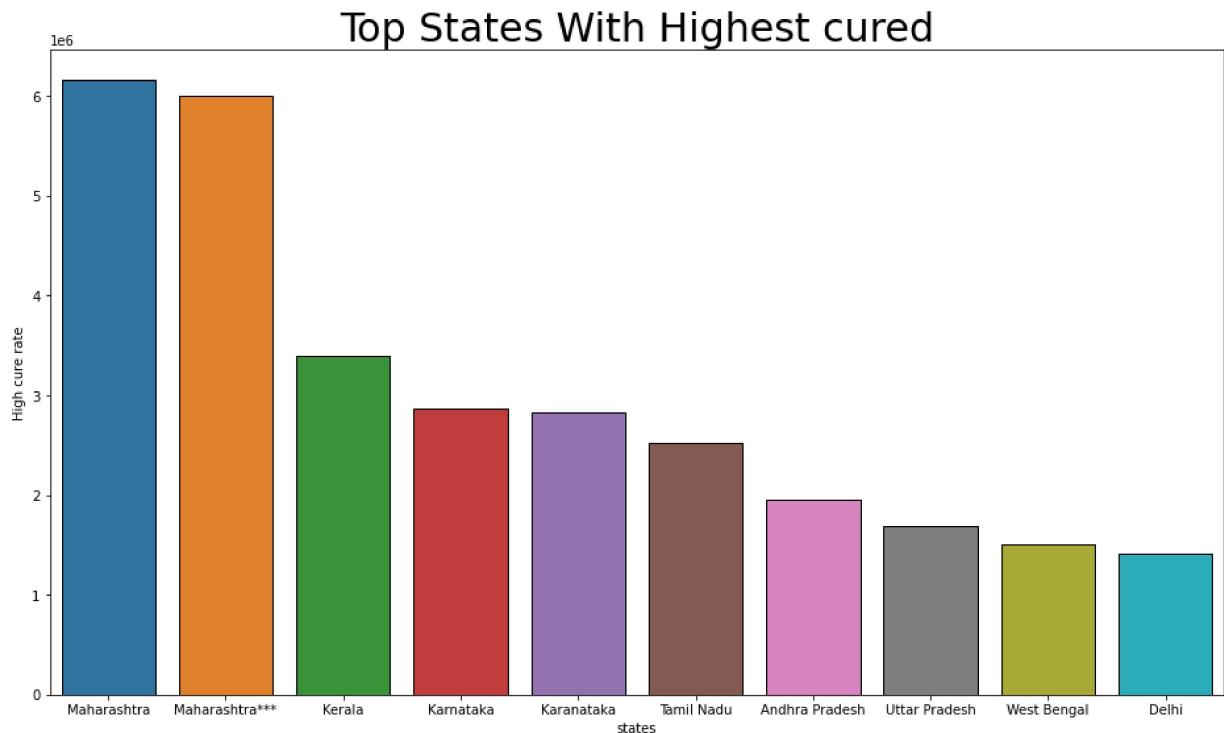
In [72]: #top states with highest deaths

```
top_states_with_highest_deaths=covid_df.groupby(by='State/UnionTerritory').max()
fig=plt.figure(figsize=(16,9))
plt.title('Top States With Highest Deaths')
ax=sns.barplot(data=top_states_with_highest_deaths.iloc[:10],y='Deaths',x='State',
plt.xlabel('States')
plt.ylabel('Highest Deaths')
plt.show()
```



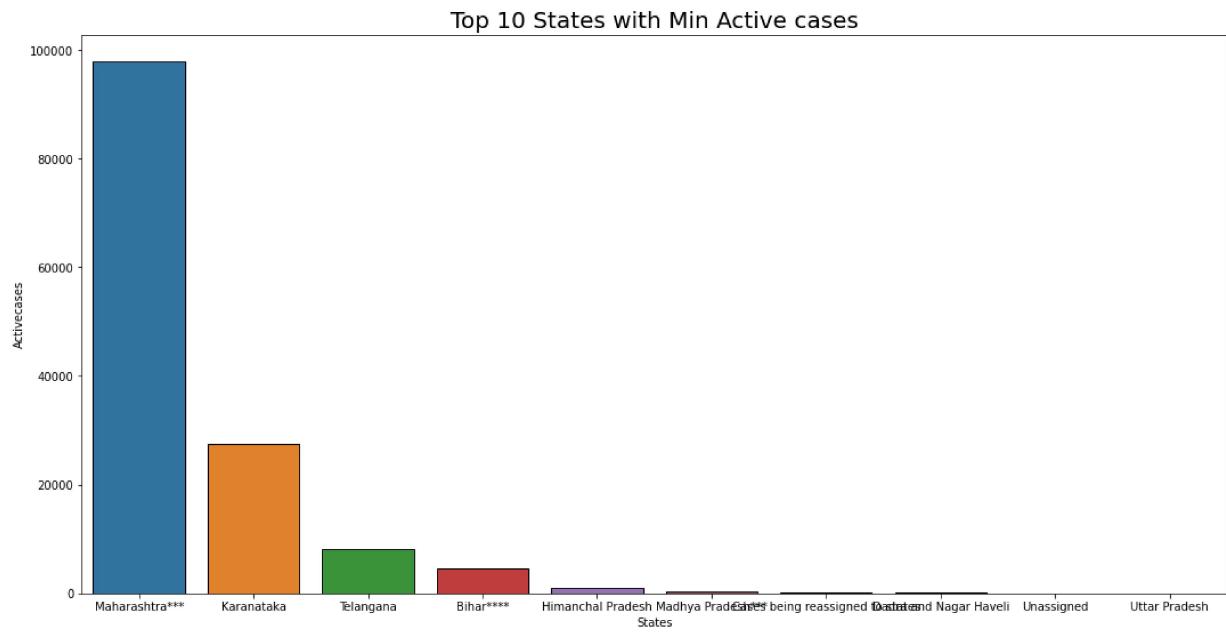
In [77]: # top states with highest curing rate

```
High_cure_rate=covid_df.groupby(by='State/UnionTerritory').max()[['Cured','Date']]
fig=plt.figure(figsize=(16,9))
plt.title('Top States With Highest cured',size=30)
ax=sns.barplot(data=High_cure_rate.iloc[:10],y='Cured',x='State/UnionTerritory',)
plt.xlabel('states')
plt.ylabel('High cure rate')
plt.show()
```



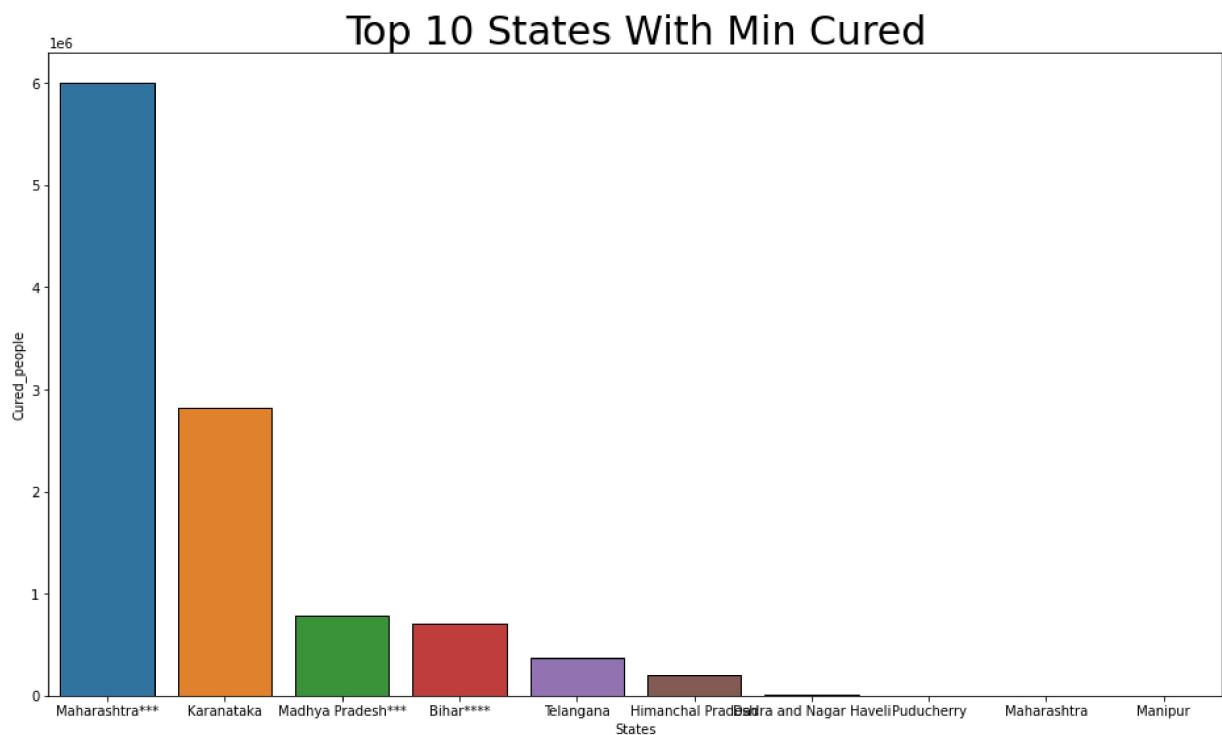
In [80]: #top states with min active cases

```
min_active_cases=covid_df.groupby(by=['State/UnionTerritory']).min()[['Active_cases']]
fig=plt.figure(figsize=(18,9))
plt.title('Top 10 States with Min Active cases',size=20)
ax=sns.barplot(data=min_active_cases.iloc[:10],y='Active_cases',x='State/UnionTerritory')
plt.xlabel('States')
plt.ylabel('Active cases')
plt.show()
```



In [82]: #top 10 min cured states in india

```
min_cured=covid_df.groupby(by=['State/UnionTerritory']).min()[['Cured','Date']].reset_index()
fig=plt.figure(figsize=(16,9))
plt.title('Top 10 States With Min Cured',size=30)
ax=sns.barplot(data=min_cured.iloc[:10],y='Cured',x='State/UnionTerritory',linewidth=2)
plt.xlabel('States')
plt.ylabel('Cured_people')
plt.show()
```



syntax:- fig=plt.figure(figsize=(12,6))

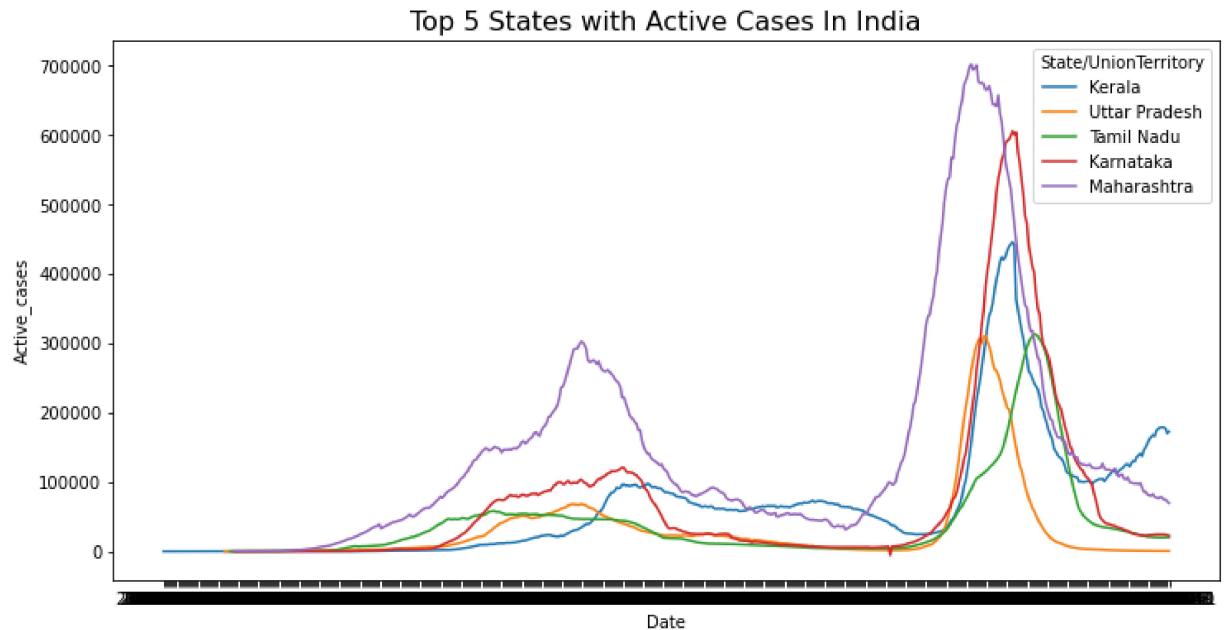
```
ax=sns.lineplot(data=covid_df[covid_df['State/UnionTerritory'].isin(['Maharashtra','Karnataka','Tamil Nadu','Uttar Pradesh'])],x='Date',y='Active_cases',hue='State/UnionTerritory')
```

```
ax.set_title('Top 5 States with Active Cases In India',size=16)
```

In [95]: #Growth Trend Analysis of active_cases

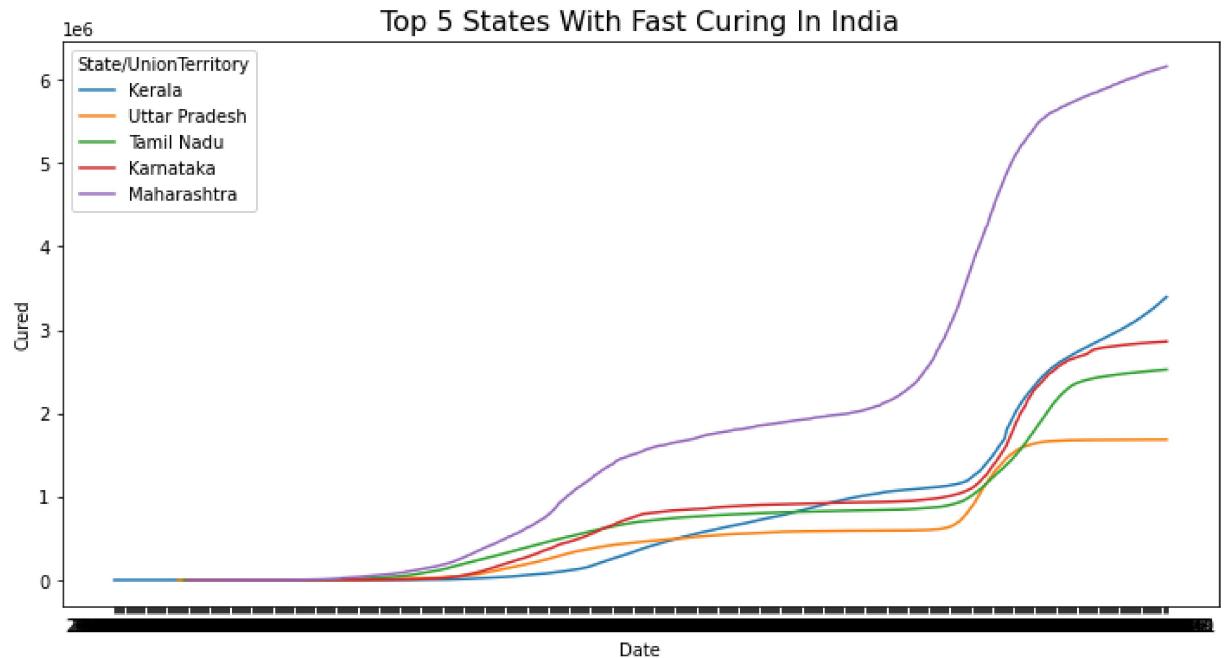
```
fig=plt.figure(figsize=(12,6))
ax=sns.lineplot(data=covid_df[covid_df['State/UnionTerritory'].isin(['Maharashtra','Karnataka','Tamil Nadu','Uttar Pradesh','Kerala'])].groupby('Date').sum().reset_index(),x='Date',y='Active_Cases',hue='State/UnionTerritory')
ax.set_title('Top 5 States with Active Cases In India',size=16)
```

Out[95]: Text(0.5, 1.0, 'Top 5 States with Active Cases In India')



```
In [93]: #Growth Trend analysis of cured cases
fig=plt.figure(figsize=(12,6))
ax=sns.lineplot(data=covid_df[covid_df['State/UnionTerritory'].isin(['Maharashtra','Karnataka','Tamil Nadu','Uttar Pradesh','Kerala'])].sort_values('Date').tail(100),x='Date',y='Cured',hue='State/UnionTerritory')
ax.set_title('Top 5 States With Fast Curing In India',size=16)
```

```
Out[93]: Text(0.5, 1.0, 'Top 5 States With Fast Curing In India')
```

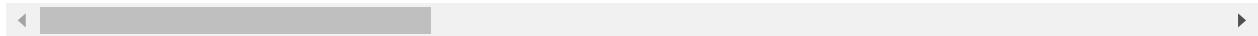


In [97]: `vaccine_df.head()`

Out[97]:

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

5 rows × 24 columns



In [99]: `vaccine_df.rename(columns={'Updated On':'Vaccine_Date'},inplace=True)`

In [100]: `vaccine_df.head()` #name changed from updated on to vaccine_date

Out[100]:

Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	18-44 Years (Doses Administered)	45-60 Years (Doses Administered)	...
48276.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN
58604.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN
99449.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN
195525.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN
251280.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN



```
In [101]: #total count of null values in each columns
vaccine_df.isnull().sum()
```

```
Out[101]: Vaccine_Date          0
State                  0
Total Doses Administered 224
Sessions               224
Sites                  224
First Dose Administered 224
Second Dose Administered 224
Male (Doses Administered) 384
Female (Doses Administered) 384
Transgender (Doses Administered) 384
Covaxin (Doses Administered) 224
CoviShield (Doses Administered) 224
Sputnik V (Doses Administered) 4850
AEFI                  2407
18-44 Years (Doses Administered) 6143
45-60 Years (Doses Administered) 6143
60+ Years (Doses Administered) 6143
18-44 Years(Individuals Vaccinated) 4112
45-60 Years(Individuals Vaccinated) 4111
60+ Years(Individuals Vaccinated) 4111
Male(Individuals Vaccinated) 7685
Female(Individuals Vaccinated) 7685
Transgender(Individuals Vaccinated) 7685
Total Individuals Vaccinated 1926
dtype: int64
```

```
In [107]: vaccine_df.rename(columns={'Sputnik V (Doses Administered)':'abc'},inplace=True)
```

```
In [108]: vaccine_df.rename(columns={'18-44 Years (Doses Administered)':'ab'},inplace=True)
```

```
In [109]: vaccine_df.rename(columns={'45-60 Years (Doses Administered)':'abd'},inplace=True)
```

```
In [110]: vaccine_df.rename(columns={'60+ Years (Doses Administered)':'xyz'},inplace=True)
```

```
In [112]: #blank columns drop from vaccine table
```

```
vaccination=vaccine_df.drop(columns=['AEFI','abc','abd','ab','xyz'],axis=1)
```

In [115]: `vaccination.head()`

Out[115]:

	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	Covaxin (Doses Administered)	CoviShield (Doses Administered)	18-44 Years(Individuals Vaccinated)	Years(
)	NaN	NaN	NaN	579.0	47697.0		NaN
)	NaN	NaN	NaN	635.0	57969.0		NaN
)	NaN	NaN	NaN	1299.0	98150.0		NaN
)	NaN	NaN	NaN	3017.0	192508.0		NaN
)	NaN	NaN	NaN	3946.0	247334.0		NaN

first install plotly then import it

In [130]: `!pip3 install plotly`

Requirement already satisfied: plotly in c:\programdata\anaconda3\lib\site-packages (5.9.0)

Requirement already satisfied: tenacity>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from plotly) (8.0.1)

In [131]: `import plotly`

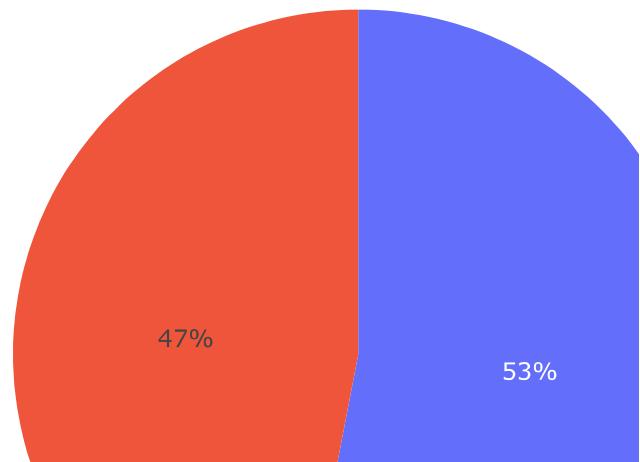
In [132]: `plotly.__version__`

Out[132]: '5.9.0'

In [133]: `import plotly.express as px
from plotly.subplots import make_subplots`

```
In [138]: male=vaccination['Male(Individuals Vaccinated)'].sum()  
female=vaccination['Female(Individuals Vaccinated)'].sum()  
px.pie(names=['Male','Female'],values=[male,female],title='Male and Female vaccination')
```

Male and Female vaccination



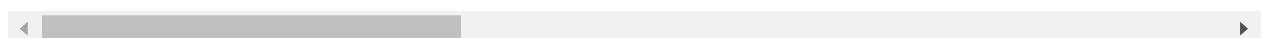
In [157]: `#remove rows where state = india`

```
vaccine=vaccine_df[vaccine_df.State != 'India']
vaccine
```

Out[157]:

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (I Adminis
212	16/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	
213	17/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	
214	18/01/2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	
215	19/01/2021	Andaman and Nicobar Islands	89.0	12.0	2.0	89.0	0.0	
216	20/01/2021	Andaman and Nicobar Islands	124.0	16.0	3.0	124.0	0.0	
...
7840	11/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN
7841	12/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN
7842	13/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN
7843	14/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN
7844	15/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN

7633 rows × 24 columns

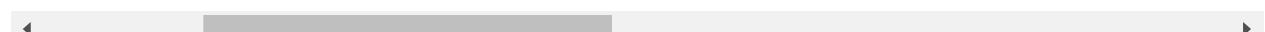


```
In [159]: vaccine=vaccine_df.drop(columns=['AEFI','abc','abd','ab','xyz'],axis=1)

vaccine.rename(columns={'Total Individuals Vaccinated':'Total'},inplace=True)
vaccine
```

Out[159]:

Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	A
176.0	3455.0	2957.0	48276.0	0.0	NaN	NaN	NaN	NaN
104.0	8532.0	4954.0	58604.0	0.0	NaN	NaN	NaN	NaN
49.0	13611.0	6583.0	99449.0	0.0	NaN	NaN	NaN	NaN
25.0	17855.0	7951.0	195525.0	0.0	NaN	NaN	NaN	NaN
80.0	25472.0	10504.0	251280.0	0.0	NaN	NaN	NaN	NaN
...
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN



```
In [160]: # most vaccinated state
```

```
max_vacc=vaccine.groupby('State')['Total'].sum().to_frame('Total')
max_vacc=max_vacc.sort_values('Total',ascending=False)[:5]
max_vacc.style.background_gradient(cmap='winter')
```

Out[160]:

Total	
State	
India	13462306895.000000
Maharashtra	1403075494.000000
Uttar Pradesh	1200575482.000000
Rajasthan	1141162807.000000
Gujarat	1078260544.000000

```
In [166]: fig=plt.figure(figsize=(10,5))
x=sns.barplot(data=max_vacc.iloc[:10],y=max_vacc.Total,x=max_vacc.index,linewidth=2)
plt.title('Top 5 Vaccinated States In India',size=20)
plt.xlabel('States')
plt.ylabel('Vaccination')
plt.show()
```

