Covid-19 Data Analysis

In [1]:

```
#importing necessary libraries

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from plotly.subplots import make_subplots
from datetime import datetime
```

In [2]:

```
#creating dataframe 1

covid_df = pd.read_csv('covid_19_india.csv')
covid_df.head(10)
```

Out[2]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Dea
0	1	2020-01-30	6:00 PM	Kerala	1	0	0	
1	2	2020-01-31	6:00 PM	Kerala	1	0	0	
2	3	2020-02-01	6:00 PM	Kerala	2	0	0	
3	4	2020-02-02	6:00 PM	Kerala	3	0	0	
4	5	2020-02-03	6:00 PM	Kerala	3	0	0	
5	6	2020-02-04	6:00 PM	Kerala	3	0	0	
6	7	2020-02-05	6:00 PM	Kerala	3	0	0	
7	8	2020-02-06	6:00 PM	Kerala	3	0	0	
8	9	2020-02-07	6:00 PM	Kerala	3	0	0	
9	10	2020-02-08	6:00 PM	Kerala	3	0	0	

In [3]:

```
covid_df.describe()
```

Out[3]:

	Sno	Cured	Deaths	Confirmed
count	18110.000000	1.811000e+04	18110.000000	1.811000e+04
mean	9055.500000	2.786375e+05	4052.402264	3.010314e+05
std	5228.051023	6.148909e+05	10919.076411	6.561489e+05
min	1.000000	0.000000e+00	0.000000	0.000000e+00
25%	4528.250000	3.360250e+03	32.000000	4.376750e+03
50%	9055.500000	3.336400e+04	588.000000	3.977350e+04
75%	13582.750000	2.788698e+05	3643.750000	3.001498e+05

max	18110.000000	6.159676e+06	134201.000000	6.363442e+06

In [4]:

```
covid_df.info()
```

```
<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 [5]:

```
#creating dataframe 2
vaccine_df = pd.read_csv('covid_vaccine_statewise.csv')
vaccine_df.head(10)
```

Out[5]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	I Adminis
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	24/01/2021	India	835058.0	49851.0	18111.0	835058.0	0.0	NaN	
9	25/01/2021	India	1277104.0	55151.0	19682.0	1277104.0	0.0	NaN	

10 rows × 24 columns

In [6]:

```
vaccine_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 [7]:

```
vaccine_df.describe()
```

Out[7]:

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

8 rows × 22 columns

In [8]:

```
covid_df = covid_df.copy()
vaccine_df = vaccine_df.copy()
```

Dropping Unwanted columns

In [9]:

```
covid_df.drop(['Sno','Time','ConfirmedIndianNational','ConfirmedForeignNational'], axis
covid_df.head()
```

Out[9]:

	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

In [10]:

```
#convert date from object to datetime64 data type using datetime method
```

```
covid_df['Date'] = pd.to_datetime(covid_df['Date'], format = '%Y-%m-%d')
covid_df.head()
```

Out[10]:

	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

Adding Active Cases Column

In [11]:

```
#Active Cases
covid_df['Active_Cases'] = covid_df['Confirmed'] - (covid_df['Cured'] + covid_df['Deaths
covid_df.tail()
```

Out[11]:

	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 [12]:

```
#Pivot Table
statewise = pd.pivot_table(covid_df, values = ['Cured','Deaths','Confirmed'], index = ['
```

In [13]:

```
statewise['Recovery_Rate'] = statewise['Cured'] * 100 / statewise['Confirmed']
```

In [14]:

```
statewise['Mortality_Rate'] = statewise['Deaths'] * 100 / statewise['Confirmed']
```

In [15]:

```
statewise = statewise.sort_values(by = 'Confirmed', ascending = False)
```

PIVOT TABLE

In [16]:

```
statewise.style.background_gradient(cmap = 'ocean')
```

Out[16]:

	Confirmed	Cured	Deaths	Recovery_Rate	Mortality_Rate
State/UnionTerritory					
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
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 [17]:
```
#top 10 active cases

top_10_active_cases = covid_df.groupby(by = 'State/UnionTerritory').max()[['Active_Cases
```

In [18]:
```
top_10_active_cases.head()
```

Out[18]:

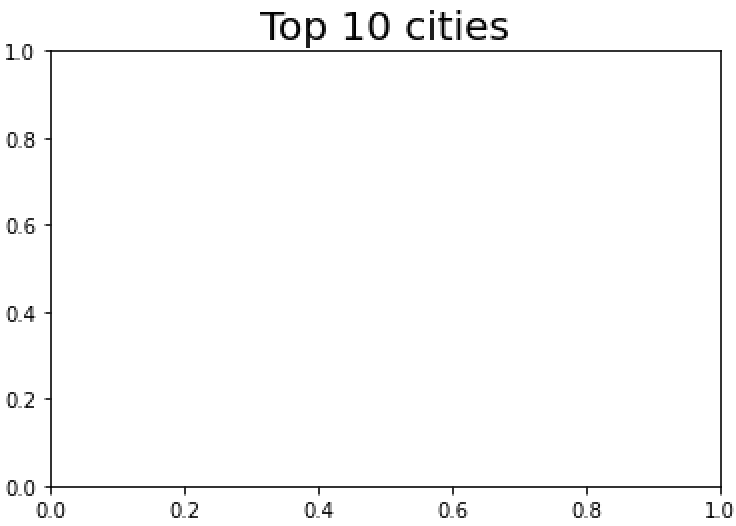
	State/UnionTerritory	Active_Cases	Date
0	Maharashtra	701614	2021-08-11
1	Karnataka	605515	2021-08-11
2	Kerala	445692	2021-08-11
3	Tamil Nadu	313048	2021-08-11
4	Uttar Pradesh	310783	2021-08-11

In [19]:
```
fig = plt.figure(figsize = (16,9))
```

```
<Figure size 1152x648 with 0 Axes>
```

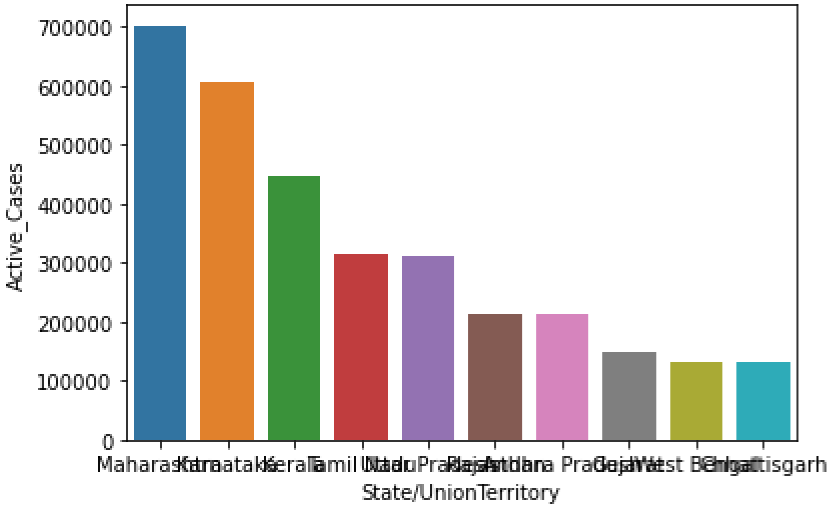
In [20]:
```
plt.title('Top 10 cities', size = 20)
```

Out[20]:
```
Text(0.5, 1.0, 'Top 10 cities')
```



TOP 10 Active Cases by State

In [21]:
```
ax = sns.barplot(x = 'State/UnionTerritory', y = 'Active_Cases', data = top_10_active_ca
```



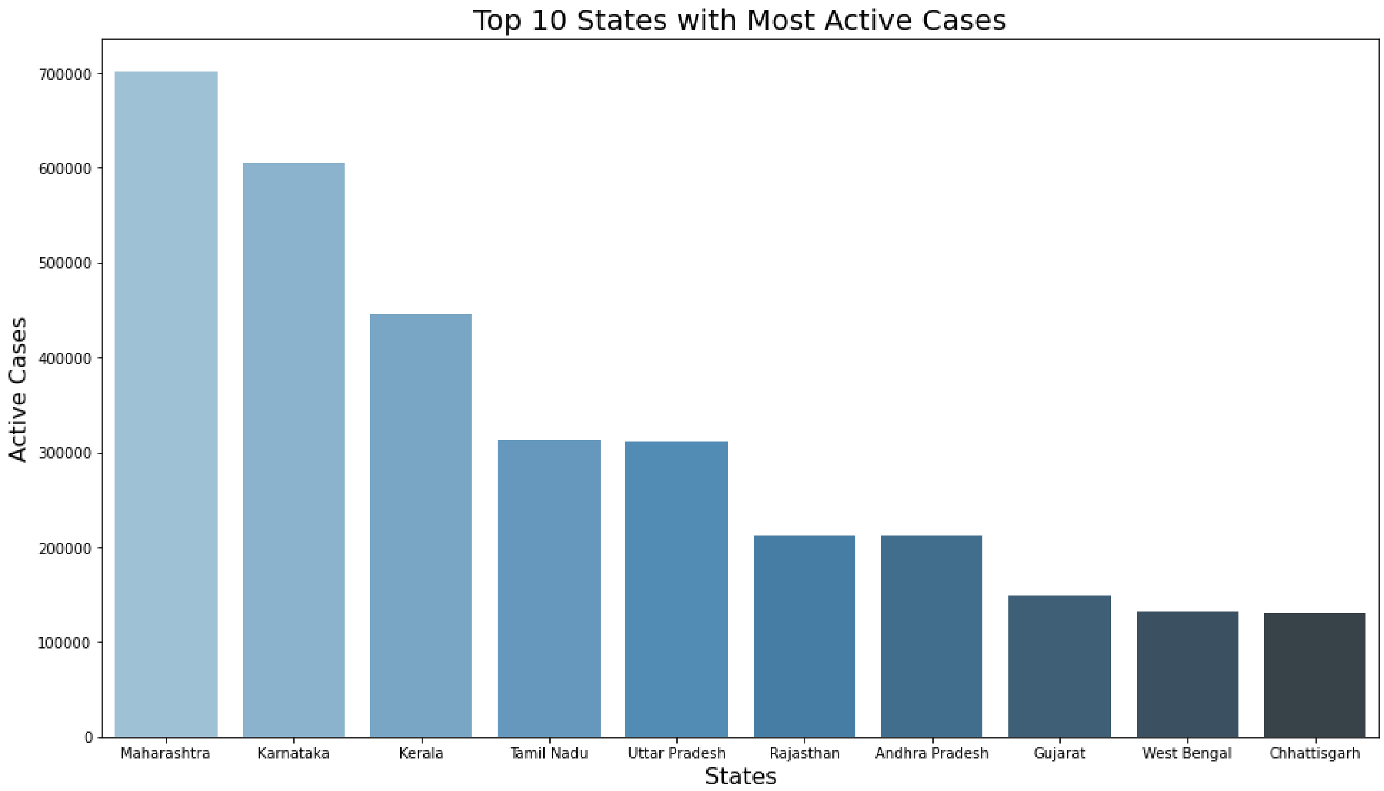
In [22]:

```
#Top 10 Active Cases by States(All code together).

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', size = 20)

ax = sns.barplot(x = 'State/UnionTerritory', y = 'Active_Cases', data = top_10_active_ca

plt.xlabel('States', fontsize = 16)
plt.ylabel('Active Cases', fontsize = 16)
plt.show()
```

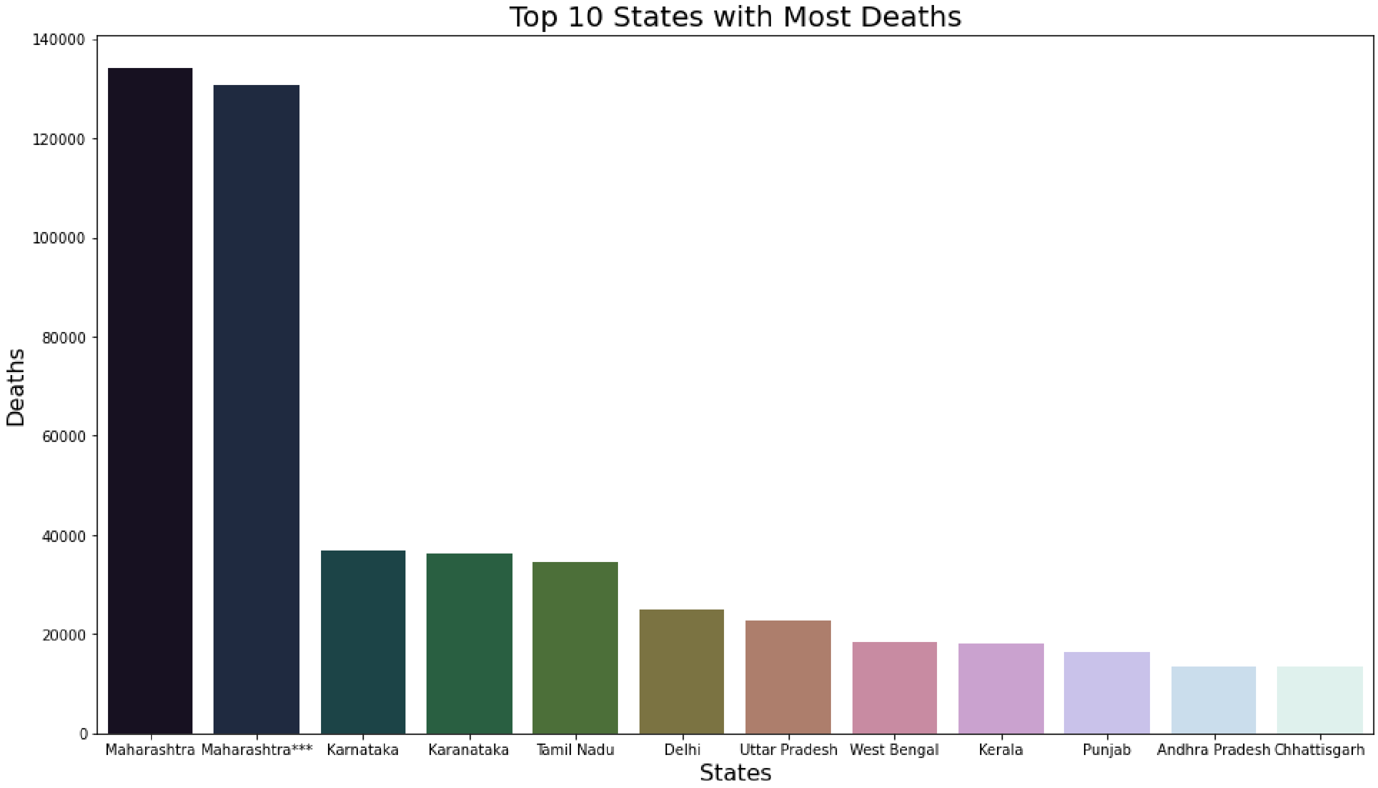


In [23]:

```
# Top 10 States with highest deaths

top_10_deaths = covid_df.groupby(by = 'State/UnionTerritory').max()[['Deaths','Date']].s
fig = plt.figure(figsize = (16,9))
plt.title('Top 10 States with Most Deaths', size = 20)

ax = sns.barplot(x = 'State/UnionTerritory', y = 'Deaths', data = top_10_deaths.iloc[:12
plt.xlabel('States', fontsize = 16)
plt.ylabel('Deaths', fontsize = 16)
plt.show()
```

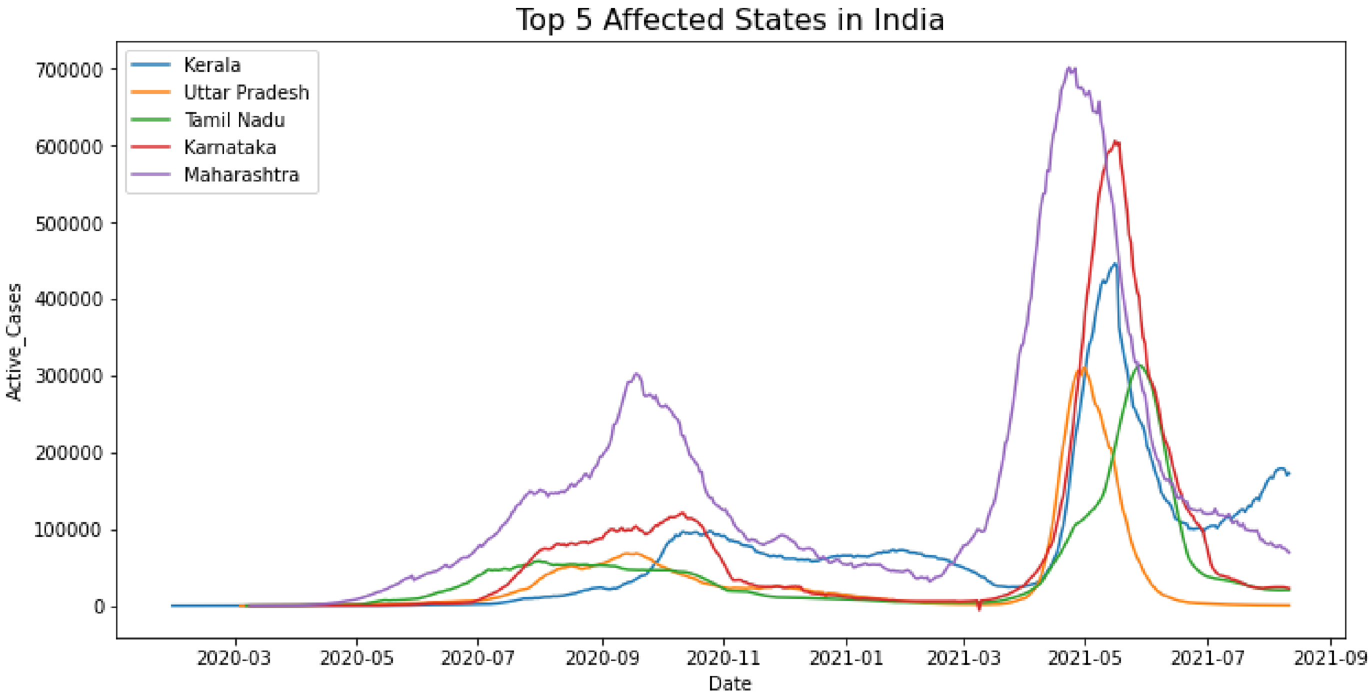


In [24]:
```
#Growth Trend

fig = plt.figure(figsize = (12,6))

ax = sns.lineplot(data = covid_df[covid_df['State/UnionTerritory'].isin(['Maharashtra','
ax.set_title("Top 5 Affected States in India", size = 16)
plt.legend(loc = 'upper left')
```

Out[24]:
```
<matplotlib.legend.Legend at 0x7fa2080c1b50>
```



In [25]:
```
#2nd Dataset

vaccine_df.head()
```

Out[25]:

Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	I

									Adminis
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 [26]:
```
vaccine_df.rename(columns = {'Updated On':'Vaccine_Date'}, inplace = True)
```

In [27]:
```
vaccine_df.head(10)
```

Out[27]:

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Adm
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	24/01/2021	India	835058.0	49851.0	18111.0	835058.0	0.0	NaN	
9	25/01/2021	India	1277104.0	55151.0	19682.0	1277104.0	0.0	NaN	

10 rows × 24 columns

In [28]:
```
vaccine_df.isnull().sum()
```

Out[28]:
```
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 [29]:

```
#dropping columns

vaccination = vaccine_df.drop(columns = ['Sputnik V (Doses Administered)','AEFI','18-44
vaccination.head()
```

Out[29]:

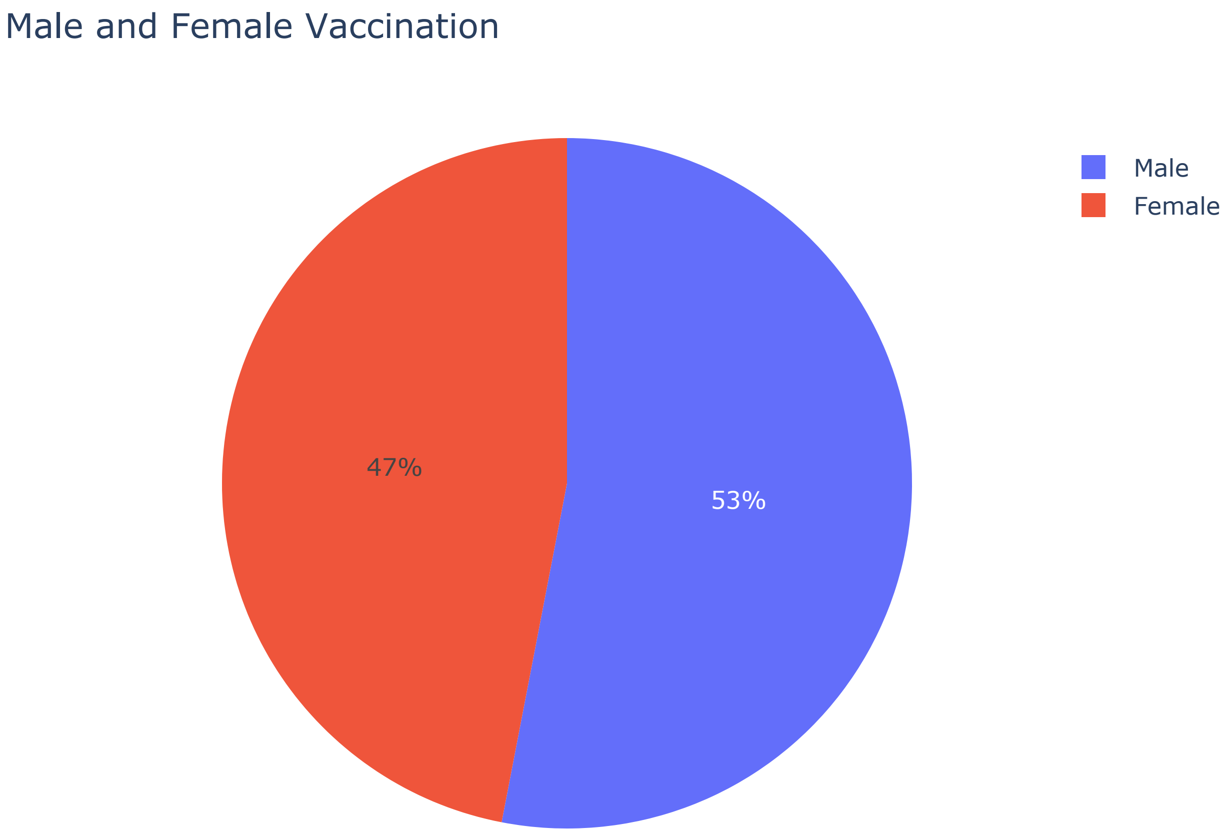
	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Adm
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	NaN	
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	NaN	
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	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	

In [30]:

```
#Male vs Female Vaccination

male = vaccination['Male(Individuals Vaccinated)'].sum()
female = vaccination['Female(Individuals Vaccinated)'].sum()

px.pie(names = ['Male','Female'], values = [male,female], title = "Male and Female Vacci
```



In [31]:

```
#Dropping rows where state = India

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

Out[31]:

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)
212	16/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0
213	17/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0
214	18/01/2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	29.0
215	19/01/2021	Andaman and Nicobar Islands	89.0	12.0	2.0	89.0	0.0	53.0
216	20/01/2021	Andaman and Nicobar Islands	124.0	16.0	3.0	124.0	0.0	67.0
...	...	...	...	...	...	...	...	...
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 [32]:

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

```
/var/folders/fz/ncjczqgd731g059fgz_1vlc80000gn/T/ipykernel_17001/2152126061.py:1: Settin
gWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
guide/indexing.html#returning-a-view-versus-a-copy
```

Out[32]:

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	A

212	16/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0
213	17/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0
214	18/01/2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	29.0
215	19/01/2021	Andaman and Nicobar Islands	89.0	12.0	2.0	89.0	0.0	53.0
216	20/01/2021	Andaman and Nicobar Islands	124.0	16.0	3.0	124.0	0.0	67.0

5 rows × 24 columns

In [33]:

```
#Most Vaccinated State

max_vac = vaccine.groupby(by = 'State').sum()[['Total']].sort_values(by = 'Total', ascen
max_vac
```

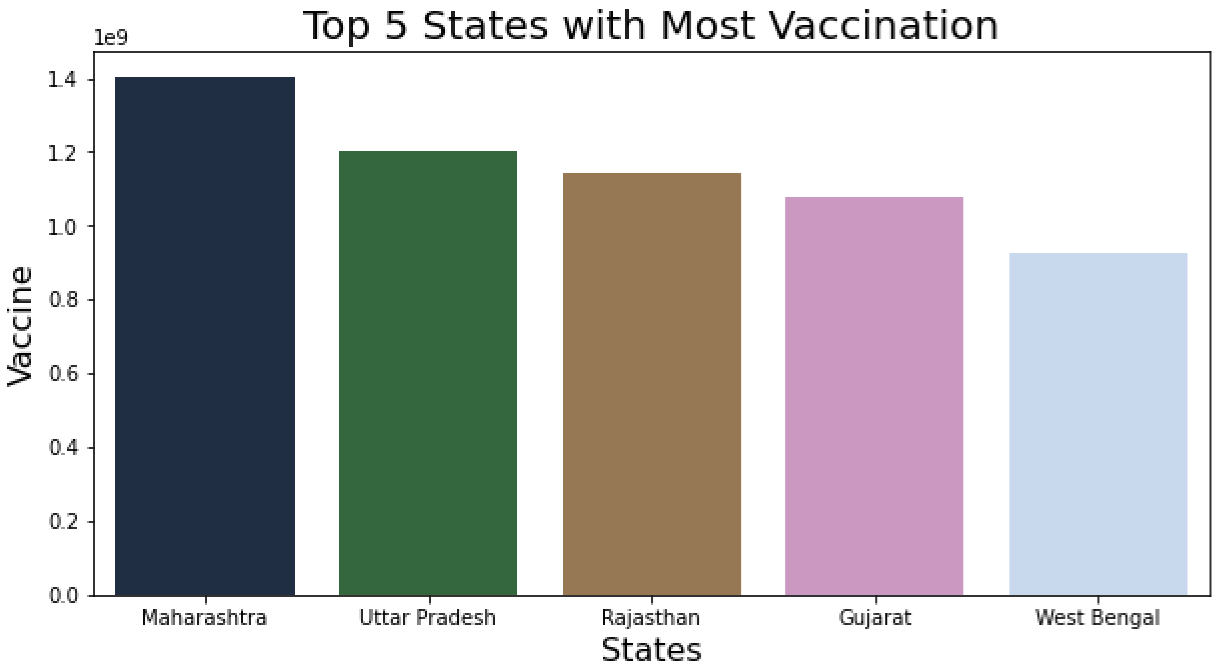
Out[33]:

	Total
State	
Maharashtra	1.403075e+09
Uttar Pradesh	1.200575e+09
Rajasthan	1.141163e+09
Gujarat	1.078261e+09
West Bengal	9.250227e+08

In [34]:

```
fig = plt.figure(figsize = (10,5))
plt.title('Top 5 States with Most Vaccination', size = 20)

ax = sns.barplot(x = max_vac.index, y = 'Total', data = max_vac, linewidth = 2, palette 
plt.xlabel('States', fontsize = 16)
plt.ylabel('Vaccine', fontsize = 16)
plt.show()
```

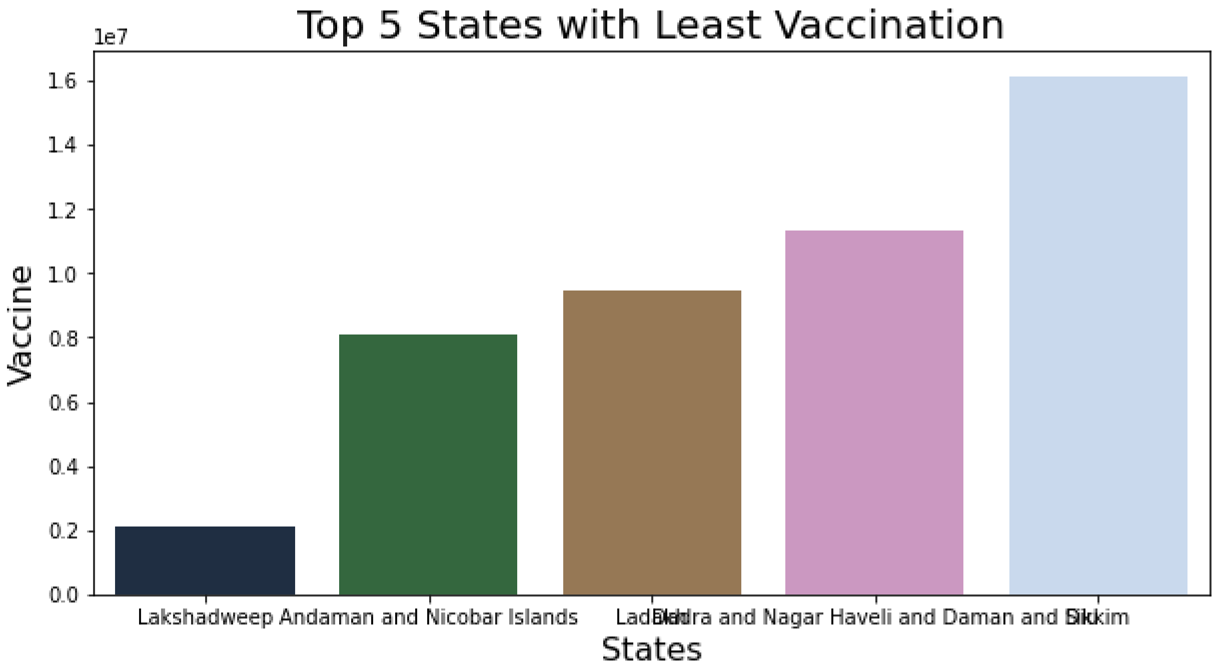


```
In [35]: #Least Vaccinated State

         least_vac = vaccine.groupby(by = 'State').sum()[['Total']].sort_values(by = 'Total')[:5]

         fig = plt.figure(figsize = (10,5))
         plt.title('Top 5 States with Least Vaccination', size = 20)

         ax = sns.barplot(x = least_vac.index, y = 'Total', data = least_vac, linewidth = 2, pale
         plt.xlabel('States', fontsize = 16)
         plt.ylabel('Vaccine', fontsize = 16)
         plt.show()
```


Top 5 States with Least Vaccination
1e7
Vaccine
Lakshadweep
Andaman and Nicobar Islands
Ladakh
Dadra and Nagar Haveli and Daman and Diu
Sikkim
States