

Covid Data Analysis

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

# Load Covid-19 India data
covid_df = pd.read_csv("/Users/anjalikumari/Downloads/covid
dataset/covid_19_india.csv")

# Load Covid Vaccine Statewise data
vaccine_df = pd.read_csv("/Users/anjalikumari/Downloads/covid
dataset/covid_vaccine_statewise.csv")

covid_df.head(10)
```

	Sno	Date	Time	State/UnionTerritory
ConfirmedIndianNational \				
0	1.0	2020-01-30	6:00 PM	Kerala
1				
1	2.0	2020-01-31	6:00 PM	Kerala
1				
2	3.0	2020-02-01	6:00 PM	Kerala
2				
3	4.0	2020-02-02	6:00 PM	Kerala
3				
4	5.0	2020-02-03	6:00 PM	Kerala
3				
5	6.0	2020-02-04	6:00 PM	Kerala
3				
6	7.0	2020-02-05	6:00 PM	Kerala
3				
7	8.0	2020-02-06	6:00 PM	Kerala
3				
8	9.0	2020-02-07	6:00 PM	Kerala
3				
9	10.0	2020-02-08	6:00 PM	Kerala
3				

	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	0	0.0	0.0	1.0
1	0	0.0	0.0	1.0
2	0	0.0	0.0	2.0
3	0	0.0	0.0	3.0
4	0	0.0	0.0	3.0

5	0	0.0	0.0	3.0
6	0	0.0	0.0	3.0
7	0	0.0	0.0	3.0
8	0	0.0	0.0	3.0
9	0	0.0	0.0	3.0

```
# Display info of covid_df
covid_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15114 entries, 0 to 15113
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                  15086 non-null  object
1   State/UnionTerritory  15086 non-null  object
2   Cured                 15086 non-null  float64
3   Deaths               15086 non-null  float64
4   Confirmed             15086 non-null  float64
dtypes: float64(3), object(2)
memory usage: 590.5+ KB
```

```
# Display descriptive statistics of covid_df
covid_df.describe()
```

	Cured	Deaths	Confirmed
count	1.508600e+04	15086.000000	1.508600e+04
mean	1.747937e+05	2721.084449	1.942820e+05
std	3.648330e+05	7182.672358	4.095184e+05
min	0.000000e+00	0.000000	0.000000e+00
25%	1.685000e+03	12.000000	2.935500e+03
50%	1.964700e+04	364.000000	2.608150e+04
75%	2.087552e+05	2170.000000	2.216012e+05
max	4.927480e+06	83777.000000	5.433506e+06

```
# Display first 7 rows of vaccine_df
vaccine_df.head(7)
```

	Updated On	State	Total Doses Administered	Sessions	Sites	\
0	16/01/2021	India	48276.0	3455.0	2957.0	
1	17/01/2021	India	58604.0	8532.0	4954.0	
2	18/01/2021	India	99449.0	13611.0	6583.0	
3	19/01/2021	India	195525.0	17855.0	7951.0	
4	20/01/2021	India	251280.0	25472.0	10504.0	
5	21/01/2021	India	365965.0	32226.0	12600.0	
6	22/01/2021	India	549381.0	36988.0	14115.0	

	First Dose Administered	Second Dose Administered	\
0	48276.0	0.0	
1	58604.0	0.0	

2	99449.0	0.0
3	195525.0	0.0
4	251280.0	0.0
5	365965.0	0.0
6	549381.0	0.0

	Male (Doses Administered)	Female (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	

	Transgender (Doses Administered)	... 18-44 Years (Doses Administered)	\
0	NaN	...	
NaN			
1	NaN	...	
NaN			
2	NaN	...	
NaN			
3	NaN	...	
NaN			
4	NaN	...	
NaN			
5	NaN	...	
NaN			
6	NaN	...	
NaN			

	45-60 Years (Doses Administered)	60+ Years (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	

	18-44 Years(Individuals Vaccinated)	45-60 Years(Individuals Vaccinated)	\
0	NaN		
NaN			
1	NaN		
NaN			
2	NaN		
NaN			
3	NaN		

NaN	
4	NaN
NaN	
5	NaN
NaN	
6	NaN
NaN	

	60+ Years(Individuals Vaccinated)	Male(Individuals Vaccinated)	\
0	NaN	23757.0	
1	NaN	27348.0	
2	NaN	41361.0	
3	NaN	81901.0	
4	NaN	98111.0	
5	NaN	132784.0	
6	NaN	193899.0	

	Female(Individuals Vaccinated)	Transgender(Individuals Vaccinated)	\
0	24517.0	2.0	
1	31252.0	4.0	
2	58083.0	5.0	
3	113613.0	11.0	
4	153145.0	24.0	
5	233143.0	38.0	
6	355402.0	80.0	

	Total Individuals Vaccinated
0	48276.0
1	58604.0
2	99449.0
3	195525.0
4	251280.0
5	365965.0
6	549381.0

[7 rows x 24 columns]

Display first 7 rows of vaccine_df
vaccine_df.head(7)

	Updated On	State	Total Doses Administered	Sessions	Sites	\
0	16/01/2021	India	48276.0	3455.0	2957.0	

1	17/01/2021	India	58604.0	8532.0	4954.0
2	18/01/2021	India	99449.0	13611.0	6583.0
3	19/01/2021	India	195525.0	17855.0	7951.0
4	20/01/2021	India	251280.0	25472.0	10504.0
5	21/01/2021	India	365965.0	32226.0	12600.0
6	22/01/2021	India	549381.0	36988.0	14115.0

	First Dose Administered	Second Dose Administered	\
0	48276.0	0.0	
1	58604.0	0.0	
2	99449.0	0.0	
3	195525.0	0.0	
4	251280.0	0.0	
5	365965.0	0.0	
6	549381.0	0.0	

	Male (Doses Administered)	Female (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	

	Transgender (Doses Administered)	... 18-44 Years (Doses Administered)	\
0	NaN	NaN	...
1	NaN	NaN	...
2	NaN	NaN	...
3	NaN	NaN	...
4	NaN	NaN	...
5	NaN	NaN	...
6	NaN	NaN	...

	45-60 Years (Doses Administered)	60+ Years (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	

	18-44 Years(Individuals Vaccinated)	45-60 Years(Individuals Vaccinated)
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
5	NaN	NaN
6	NaN	NaN

0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
5	NaN
6	NaN

	60+ Years(Individuals Vaccinated)	Male(Individuals Vaccinated)
0	NaN	23757.0
1	NaN	27348.0
2	NaN	41361.0
3	NaN	81901.0
4	NaN	98111.0
5	NaN	132784.0
6	NaN	193899.0

	Female(Individuals Vaccinated)	Transgender(Individuals Vaccinated)
0	24517.0	2.0
1	31252.0	4.0
2	58083.0	5.0
3	113613.0	11.0
4	153145.0	24.0
5	233143.0	38.0
6	355402.0	80.0

	Total Individuals Vaccinated
0	48276.0
1	58604.0
2	99449.0
3	195525.0
4	251280.0

```
5          365965.0
6          549381.0
```

```
[7 rows x 24 columns]
```

```
# Drop unnecessary columns from covid_df
```

```
covid_df.drop(['Sno', 'Time', 'ConfirmedIndianNational',  
'ConfirmedForeignNational'], inplace=True, axis=1)
```

```
covid_df.head()
```

	Date	State/UnionTerritory	Cured	Deaths	Confirmed
0	2020-01-30	Kerala	0.0	0.0	1.0
1	2020-01-31	Kerala	0.0	0.0	1.0
2	2020-02-01	Kerala	0.0	0.0	2.0
3	2020-02-02	Kerala	0.0	0.0	3.0
4	2020-02-03	Kerala	0.0	0.0	3.0

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

```
covid_df.head()
```

	Date	State/UnionTerritory	Cured	Deaths	Confirmed
0	2020-01-30	Kerala	0.0	0.0	1.0
1	2020-01-31	Kerala	0.0	0.0	1.0
2	2020-02-01	Kerala	0.0	0.0	2.0
3	2020-02-02	Kerala	0.0	0.0	3.0
4	2020-02-03	Kerala	0.0	0.0	3.0

```
covid_df['Active_Cases'] = covid_df['Confirmed'] - (covid_df['Cured']  
+ covid_df['Deaths'])
```

```
covid_df.tail()
```

	Date	State/UnionTerritory	Cured	Deaths	Confirmed
Active_Cases					
15109	NaT	NaN	NaN	NaN	NaN
NaN					
15110	NaT	NaN	NaN	NaN	NaN
NaN					
15111	NaT	NaN	NaN	NaN	NaN
NaN					
15112	NaT	NaN	NaN	NaN	NaN
NaN					
15113	NaT	NaN	NaN	NaN	NaN
NaN					

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

```

                                index='State/UnionTerritory',
aggfunc='max')

statewise['Recovery Rate'] = statewise['Cured'] * 100 /
statewise['Confirmed']

statewise['Mortality Rate'] = statewise['Deaths'] * 100 /
statewise['Confirmed']

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

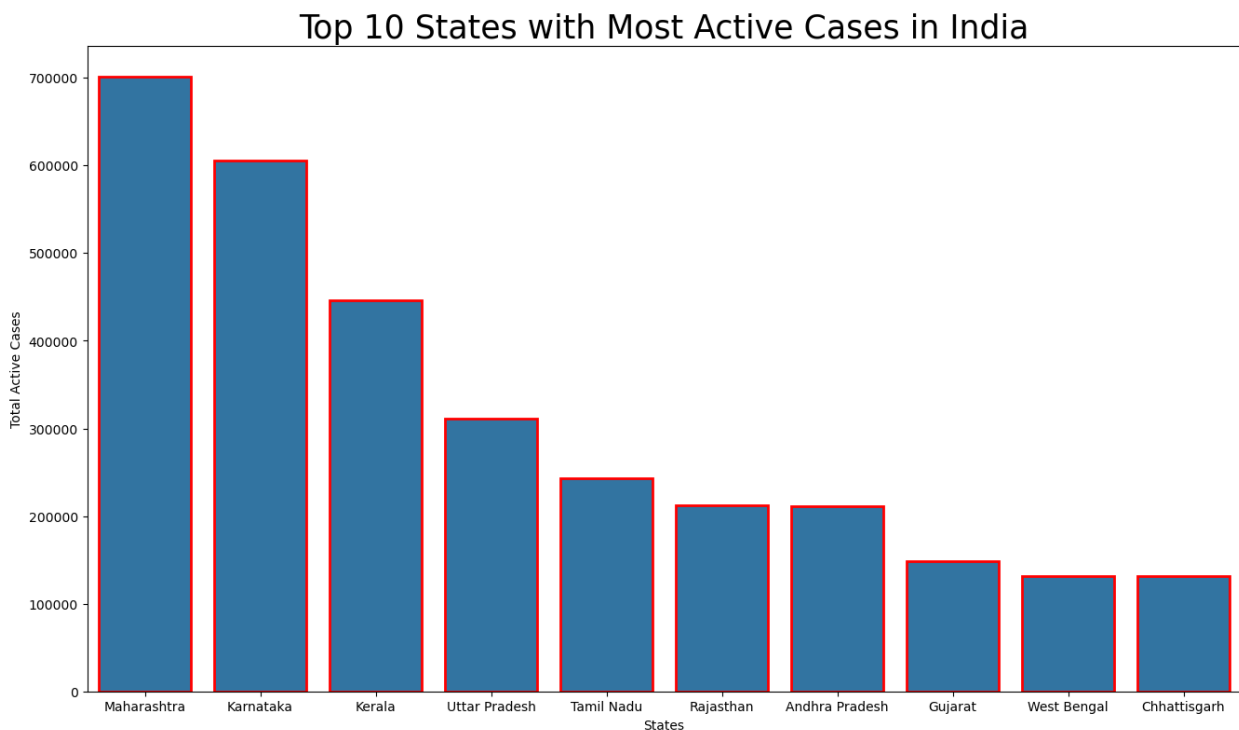
statewise.style.background_gradient(cmap='cubehelix')

<pandas.io.formats.style.Styler at 0x14a581160>

# Top 10 active cases states
top_10_active_cases = covid_df.groupby('State/UnionTerritory').max(
).sort_values(by=['Active_Cases'], ascending=False).reset_index()

fig = plt.figure(figsize=(16, 9))
plt.title("Top 10 States with Most Active Cases in India", size=25)
ax = sns.barplot(data=top_10_active_cases.iloc[:10], y='Active_Cases',
x='State/UnionTerritory',
                    linewidth=2, edgecolor='red')
plt.xlabel("States")
plt.ylabel("Total Active Cases")
plt.show()

```

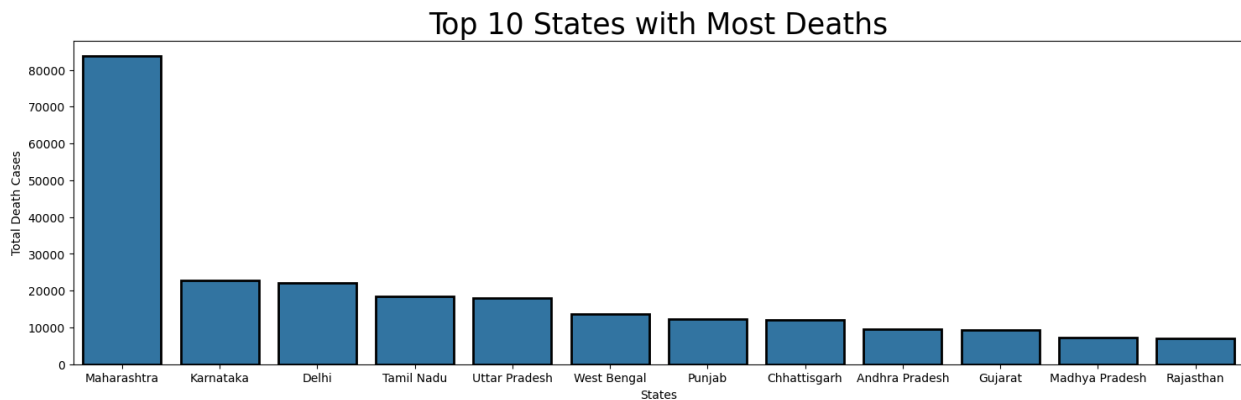



```

# Top states with highest deaths
top_10_deaths = covid_df.groupby('State/UnionTerritory').max(
).sort_values(by=['Deaths'], ascending=False).reset_index()

# Plotting the bar plot
plt.figure(figsize=(18, 5))
plt.title("Top 10 States with Most Deaths", size=25)
ax = sns.barplot(data=top_10_deaths.iloc[:12], y='Deaths',
x='State/UnionTerritory',
linewidth=2, edgecolor='black')
plt.xlabel("States")
plt.ylabel("Total Death Cases")
plt.show()

```

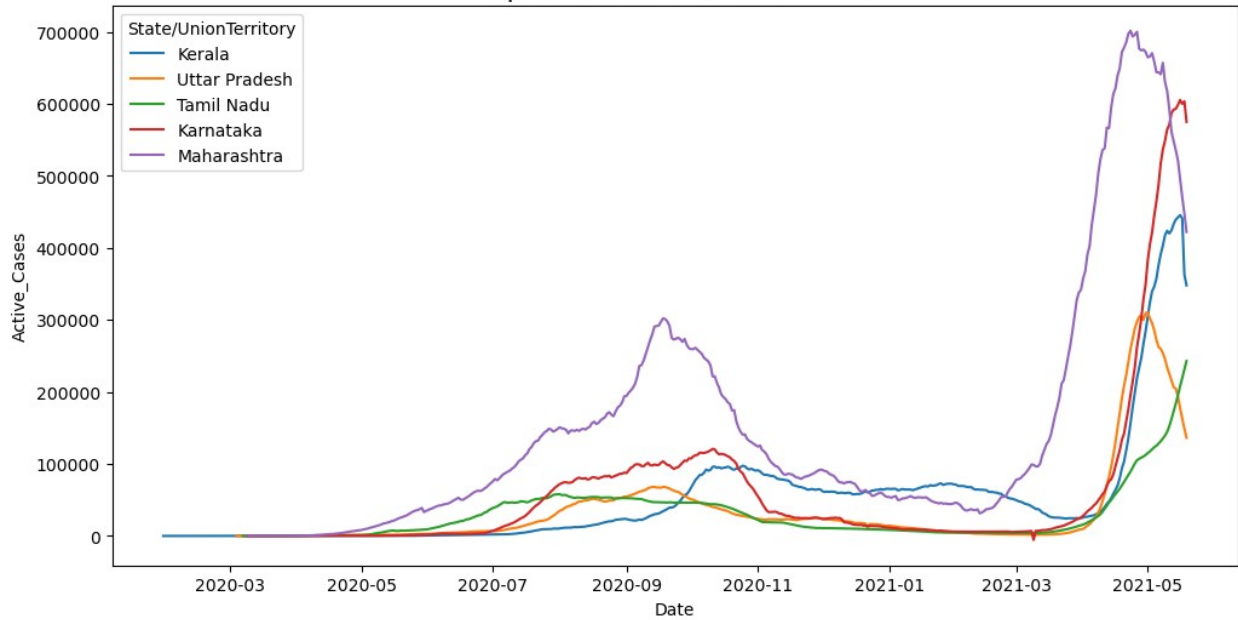


```

# Growth trend
plt.figure(figsize=(12, 6))
ax = sns.lineplot(data=covid_df[covid_df['State/UnionTerritory'].isin(
['Maharashtra', 'Karnataka', 'Kerala', 'Tamil Nadu', 'Uttar
Pradesh'])],
x='Date', y='Active_Cases',
hue='State/UnionTerritory')
ax.set_title("Top 5 Affected States in India", size=16)
plt.show()

```

Top 5 Affected States in India



```
vaccine_df.head()
```

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

	First Dose Administered	Second Dose Administered	\
0	48276.0	0.0	
1	58604.0	0.0	
2	99449.0	0.0	
3	195525.0	0.0	
4	251280.0	0.0	

	Male (Doses Administered)	Female (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	

	Transgender (Doses Administered)	...	18-44 Years (Doses Administered)	\
0	NaN	...	NaN	
1	NaN	...	NaN	
2	NaN	...	NaN	

NaN	
3	NaN ...
NaN	
4	NaN ...
NaN	

	45-60 Years (Doses Administered)	60+ Years (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	

	18-44 Years(Individuals Vaccinated)	45-60 Years(Individuals Vaccinated)	\
0	NaN		
NaN			
1	NaN		
NaN			
2	NaN		
NaN			
3	NaN		
NaN			
4	NaN		
NaN			

	60+ Years(Individuals Vaccinated)	Male(Individuals Vaccinated)	\
0	NaN	23757.0	
1	NaN	27348.0	
2	NaN	41361.0	
3	NaN	81901.0	
4	NaN	98111.0	

	Female(Individuals Vaccinated)	Transgender(Individuals Vaccinated)	\
0	24517.0	2.0	
1	31252.0	4.0	
2	58083.0	5.0	
3	113613.0	11.0	
4	153145.0	24.0	

	Total Individuals Vaccinated
0	48276.0
1	58604.0
2	99449.0

```
3          195525.0
4          251280.0
```

```
[5 rows x 24 columns]
```

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

```
vaccine_df.head(10)
```

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	\
0	16/01/2021	India	48276.0	3455.0	2957.0	
1	17/01/2021	India	58604.0	8532.0	4954.0	
2	18/01/2021	India	99449.0	13611.0	6583.0	
3	19/01/2021	India	195525.0	17855.0	7951.0	
4	20/01/2021	India	251280.0	25472.0	10504.0	
5	21/01/2021	India	365965.0	32226.0	12600.0	
6	22/01/2021	India	549381.0	36988.0	14115.0	
7	23/01/2021	India	759008.0	43076.0	15605.0	
8	24/01/2021	India	835058.0	49851.0	18111.0	
9	25/01/2021	India	1277104.0	55151.0	19682.0	

	First Dose Administered	Second Dose Administered	\
0	48276.0	0.0	
1	58604.0	0.0	
2	99449.0	0.0	
3	195525.0	0.0	
4	251280.0	0.0	
5	365965.0	0.0	
6	549381.0	0.0	
7	759008.0	0.0	
8	835058.0	0.0	
9	1277104.0	0.0	

	Male (Doses Administered)	Female (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	
7	NaN	NaN	
8	NaN	NaN	
9	NaN	NaN	

	Transgender (Doses Administered)	...	18-44 Years (Doses Administered)	\
0	NaN	...	NaN	

1	NaN ...
NaN	
2	NaN ...
NaN	
3	NaN ...
NaN	
4	NaN ...
NaN	
5	NaN ...
NaN	
6	NaN ...
NaN	
7	NaN ...
NaN	
8	NaN ...
NaN	
9	NaN ...
NaN	

	45-60 Years (Doses Administered)	60+ Years (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	
7	NaN	NaN	
8	NaN	NaN	
9	NaN	NaN	

	18-44 Years(Individuals Vaccinated)	45-60 Years(Individuals Vaccinated)	\
0	NaN		
NaN			
1	NaN		
NaN			
2	NaN		
NaN			
3	NaN		
NaN			
4	NaN		
NaN			
5	NaN		
NaN			
6	NaN		
NaN			
7	NaN		
NaN			

8		NaN
NaN		
9		NaN
NaN		
	60+ Years(Individuals Vaccinated)	Male(Individuals Vaccinated) \
0	NaN	23757.0
1	NaN	27348.0
2	NaN	41361.0
3	NaN	81901.0
4	NaN	98111.0
5	NaN	132784.0
6	NaN	193899.0
7	NaN	267856.0
8	NaN	296283.0
9	NaN	444137.0
	Female(Individuals Vaccinated)	Transgender(Individuals Vaccinated)
\		
0	24517.0	2.0
1	31252.0	4.0
2	58083.0	5.0
3	113613.0	11.0
4	153145.0	24.0
5	233143.0	38.0
6	355402.0	80.0
7	491049.0	103.0
8	538647.0	128.0
9	832766.0	201.0
	Total Individuals Vaccinated	
0	48276.0	
1	58604.0	
2	99449.0	
3	195525.0	
4	251280.0	
5	365965.0	
6	549381.0	
7	759008.0	
8	835058.0	
9	1277104.0	

```
[10 rows x 24 columns]
```

```
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	Vaccine_Date	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
```

```
vaccine_df.isnull().sum()
```

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

```

```

vaccination = vaccine_df.drop(columns=[
    'Sputnik V (Doses Administered)', 'AEFI', '18-44 Years (Doses
Administered)',
    '45-60 Years (Doses Administered)', '60+ Years (Doses
Administered)'
], axis=1)

```

```
vaccination.head()
```

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	\
0	16/01/2021	India	48276.0	3455.0	2957.0	
1	17/01/2021	India	58604.0	8532.0	4954.0	
2	18/01/2021	India	99449.0	13611.0	6583.0	
3	19/01/2021	India	195525.0	17855.0	7951.0	
4	20/01/2021	India	251280.0	25472.0	10504.0	

	First Dose Administered	Second Dose Administered	\
0	48276.0	0.0	
1	58604.0	0.0	
2	99449.0	0.0	
3	195525.0	0.0	
4	251280.0	0.0	

	Male (Doses Administered)	Female (Doses Administered)	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	

	Transgender (Doses Administered)	Covaxin (Doses Administered)	\
0	NaN	579.0	
1	NaN	635.0	
2	NaN	1299.0	
3	NaN	3017.0	

4	NaN	3946.0
---	-----	--------

CoviShield (Doses Administered)	18-44 Years(Individuals Vaccinated) \
---------------------------------	---------------------------------------

0	47697.0
---	---------

NaN	
-----	--

1	57969.0
---	---------

NaN	
-----	--

2	98150.0
---	---------

NaN	
-----	--

3	192508.0
---	----------

NaN	
-----	--

4	247334.0
---	----------

NaN	
-----	--

45-60 Years(Individuals Vaccinated)	60+ Years(Individuals Vaccinated) \
-------------------------------------	-------------------------------------

0	NaN
---	-----

NaN	
-----	--

1	NaN
---	-----

NaN	
-----	--

2	NaN
---	-----

NaN	
-----	--

3	NaN
---	-----

NaN	
-----	--

4	NaN
---	-----

NaN	
-----	--

Male(Individuals Vaccinated)	Female(Individuals Vaccinated) \
------------------------------	----------------------------------

0	23757.0	24517.0
---	---------	---------

1	27348.0	31252.0
---	---------	---------

2	41361.0	58083.0
---	---------	---------

3	81901.0	113613.0
---	---------	----------

4	98111.0	153145.0
---	---------	----------

Transgender(Individuals Vaccinated)	Total Individuals Vaccinated
-------------------------------------	------------------------------

0	2.0	48276.0
---	-----	---------

1	4.0	58604.0
---	-----	---------

2	5.0	99449.0
---	-----	---------

3	11.0	195525.0
---	------	----------

4	24.0	251280.0
---	------	----------

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 Vaccination")
```

Male and Female Vaccination



```
max_vac = vaccine.groupby('State')
['Total'].sum().to_frame('Total').sort_values('Total',
ascending=False)
max_vac = max_vac.iloc[:5]
max_vac
```

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

```
# Plotting the bar plot
plt.figure(figsize=(10, 5))
plt.title("Top 5 Vaccinated States in India", size=20)
x = sns.barplot(data=max_vac.iloc[:10], y=max_vac.Total,
x=max_vac.index,
                 linewidth=2, edgecolor='blue')
plt.xlabel("States")
plt.ylabel("Vaccination")
plt.show()
```

