Indian Vaccine Data Analysis

In [1]:
```
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]:
```
cov19_vacc_df=pd.read_csv("C:\\Users\\Ankit\\Desktop\\Data Science\\Covid-19 Data Analysis Using Python\\India Dataset\\Dataset\\covid_vaccine_statewise.csv")
```

In [3]:
```
cov19_vacc_df.head()
```

Out[3]:

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

5 rows × 24 columns

In [4]:
```
cov19_vacc_df.rename(columns={'Updated On' : 'Vaccine_Date'},inplace=True)
```

In [5]:
```
cov19_vacc_df.head(10)
```

Out[5]:

	Vaccine_Date	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	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)	M
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
5	21/01/2021	India	365965.0	32226.0	12600.0	365965.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
6	22/01/2021	India	549381.0	36988.0	14115.0	549381.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
7	23/01/2021	India	759008.0	43076.0	15605.0	759008.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
8	24/01/2021	India	835058.0	49851.0	18111.0	835058.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
9	25/01/2021	India	1277104.0	55151.0	19682.0	1277104.0	0.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	

10 rows × 24 columns

In [6]:
```
cov19_vacc_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
```

In [7]:
```
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
cov19_vacc_df=pd.read_csv("C:\\Users\\Ankit\\Desktop\\Data Science\\Covid-19 Data Analysis Using Python\\India Dataset\\Dataset\\covid_vaccine_statewise.csv")
cov19_vacc_df.isnull().sum()
```

Out[7]:
```
Updated On                            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 [8]:
```
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

cov19_vacc_df = pd.read_csv("C:\\Users\\Ankit\\Desktop\\Data Science\\Covid-19 Data Analysis Using Python\\India Dataset\\Dataset\\covid_vaccine_statewise.csv")

vaccination = cov19_vacc_df.drop(columns=['Sputnik V (Doses Administered)', 'AEFI', '18-44 Years (Doses Administered)', '45-60 Years (Doses Administered)', '60+ Years (Doses Administered)'], axis=1)
```

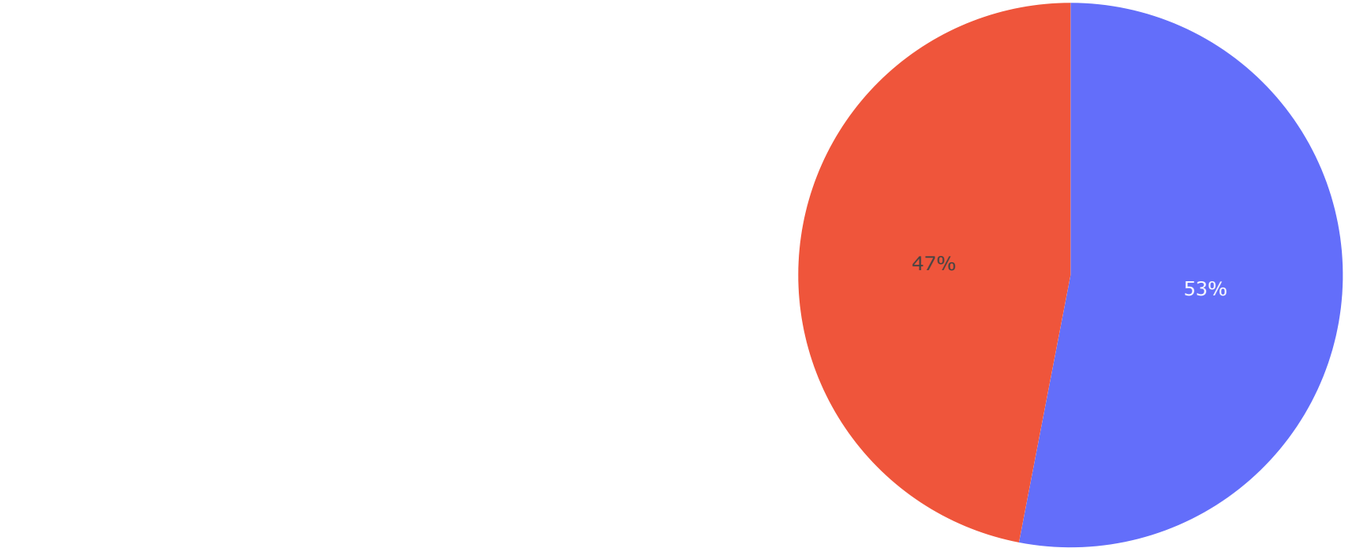
In [9]:
```
cov19_vacc_df.head()
```

Out[9]:

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

5 rows × 24 columns

In [10]:
```
## 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")
```



In [11]:
```
#remove rows where state=India
vaccine = cov19_vacc_df[cov19_vacc_df.State!='India']
vaccine
```

Out[11]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	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)
212	16/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
213	17/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
214	18/01/2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	29.0	13.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
215	19/01/2021	Andaman and Nicobar Islands	89.0	12.0	2.0	89.0	0.0	53.0	36.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
216	20/01/2021	Andaman and Nicobar Islands	124.0	16.0	3.0	124.0	0.0	67.0	57.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7840	11/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
7841	12/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
7842	13/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
7843	14/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
7844	15/08/2021	West Bengal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN

7633 rows × 24 columns

In [12]:
```
vaccine.rename(columns={ "Total Individuals Vaccinated":"Total"},inplace=True)
vaccine.head()
```

```
C:\Users\Ankit\AppData\Local\Temp\ipykernel_15392\1273045866.py:1: SettingWithCopyWarning:


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[12]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	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)
212	16/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
213	17/01/2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
214	18/01/2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	29.0	13.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
215	19/01/2021	Andaman and Nicobar Islands	89.0	12.0	2.0	89.0	0.0	53.0	36.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN
216	20/01/2021	Andaman and Nicobar Islands	124.0	16.0	3.0	124.0	0.0	67.0	57.0	0.0	...	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 24 columns

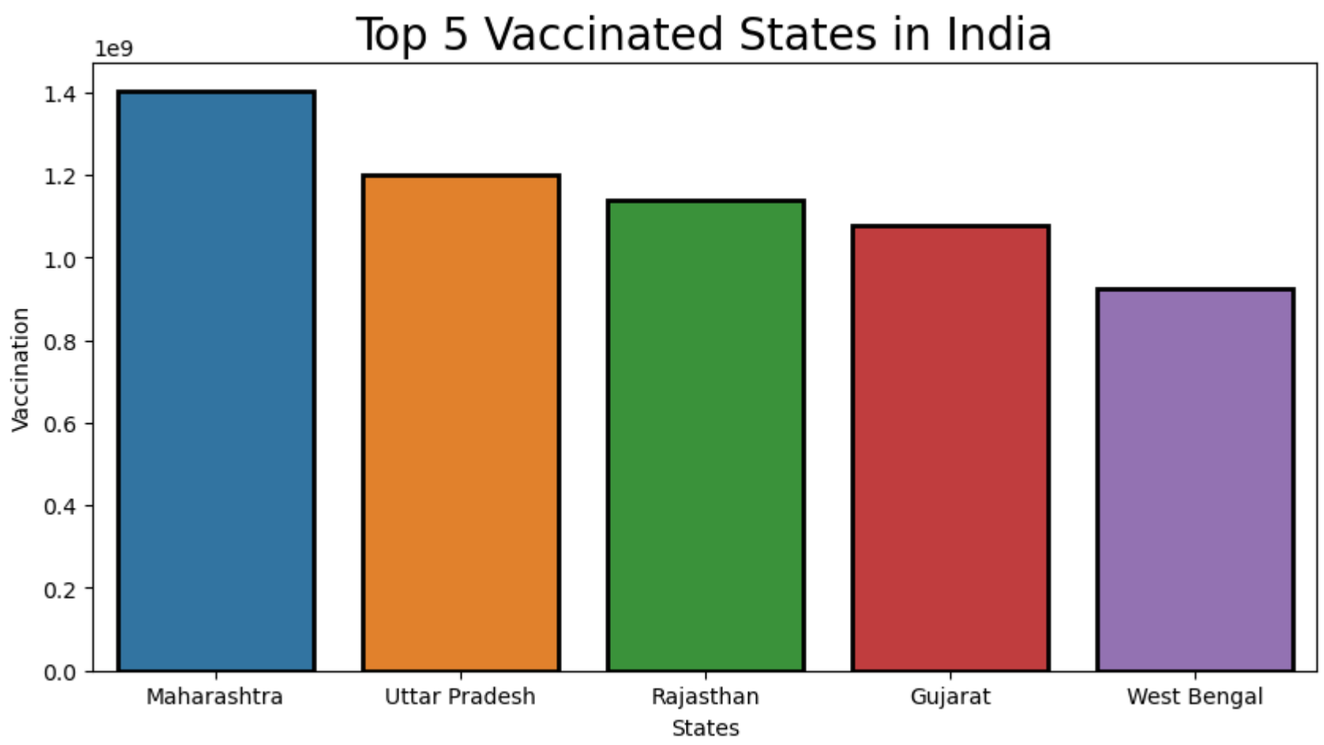
In [13]:
```
# Group by 'State' and calculate the total vaccinations, then sort in descending order and take the top 5.
max_vac = vaccine.groupby('State')['Total'].sum().to_frame('Total')
max_vac =max_vac.sort_values('Total' , ascending=False)[:5]
max_vac
```

Out[13]:

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

In [15]:
```
# Create a figure and axis with a specific size
fig, ax = plt.subplots(figsize=(10, 5))

# Plot the top 10 vaccinated states
sns.barplot(
    x=max_vac.head(5).index,
    y=max_vac.head(5)['Total'],
    ax=ax,
    linewidth=2,
    edgecolor='black')
```

```
# Set the title and axis labels
ax.set_title('Top 5 Vaccinated States in India', size=20)
ax.set_xlabel('States')
ax.set_ylabel('Vaccination')

# Show the plot
plt.show()
```

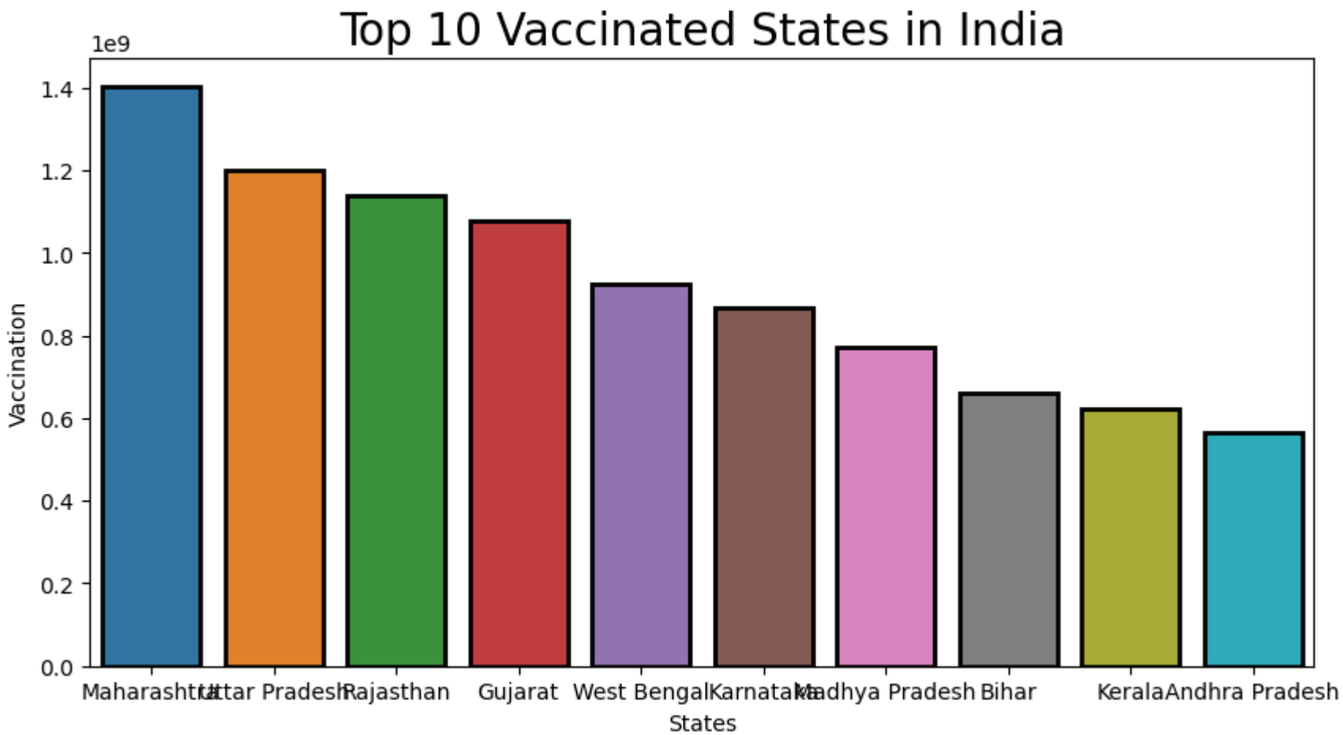


In [16]:
```
# Group by 'State' and calculate the total vaccinations, then sort in descending order and take the top 5.
max_vac = vaccine.groupby('State')['Total'].sum().to_frame('Total')
max_vac =max_vac.sort_values('Total' , ascending=False)[:10]
max_vac
```

Out[16]:

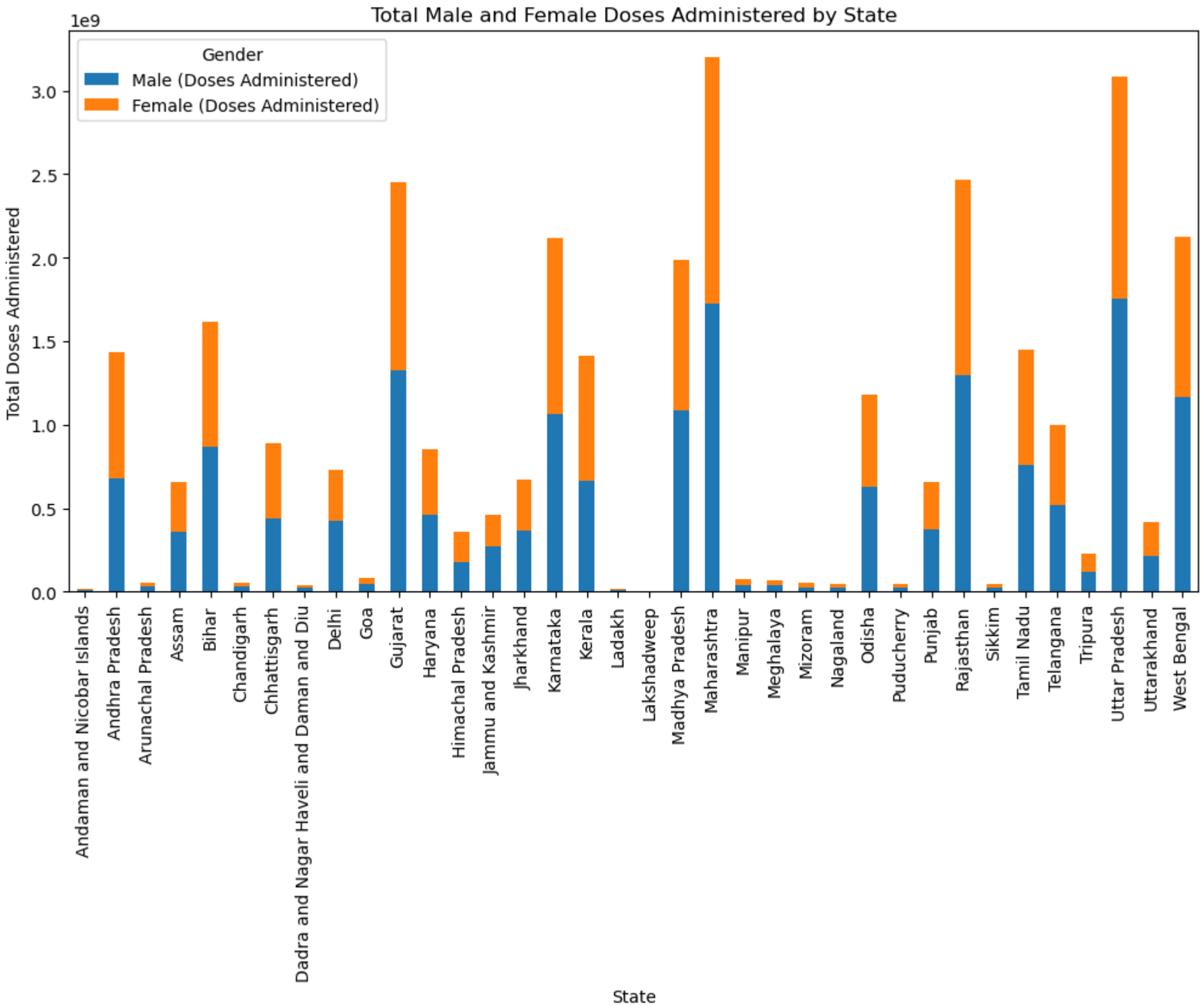
State	Total
Maharashtra	1.403075e+09
Uttar Pradesh	1.200575e+09
Rajasthan	1.141163e+09
Gujarat	1.078261e+09
West Bengal	9.250227e+08
Karnataka	8.685235e+08
Madhya Pradesh	7.718640e+08
Bihar	6.608479e+08
Kerala	6.208252e+08
Andhra Pradesh	5.645911e+08

In [17]:
```
# Create a figure and axis with a specific size
fig, ax = plt.subplots(figsize=(10, 5))

# Plot the top 10 vaccinated states
sns.barplot(
    x=max_vac.head(10).index,
    y=max_vac.head(10)['Total'],
    ax=ax,
    linewidth=2,
    edgecolor='black')

# Set the title and axis labels
ax.set_title('Top 10 Vaccinated States in India', size=20)
ax.set_xlabel('States')
ax.set_ylabel('Vaccination')

# Show the plot
plt.show()
```



In [18]:

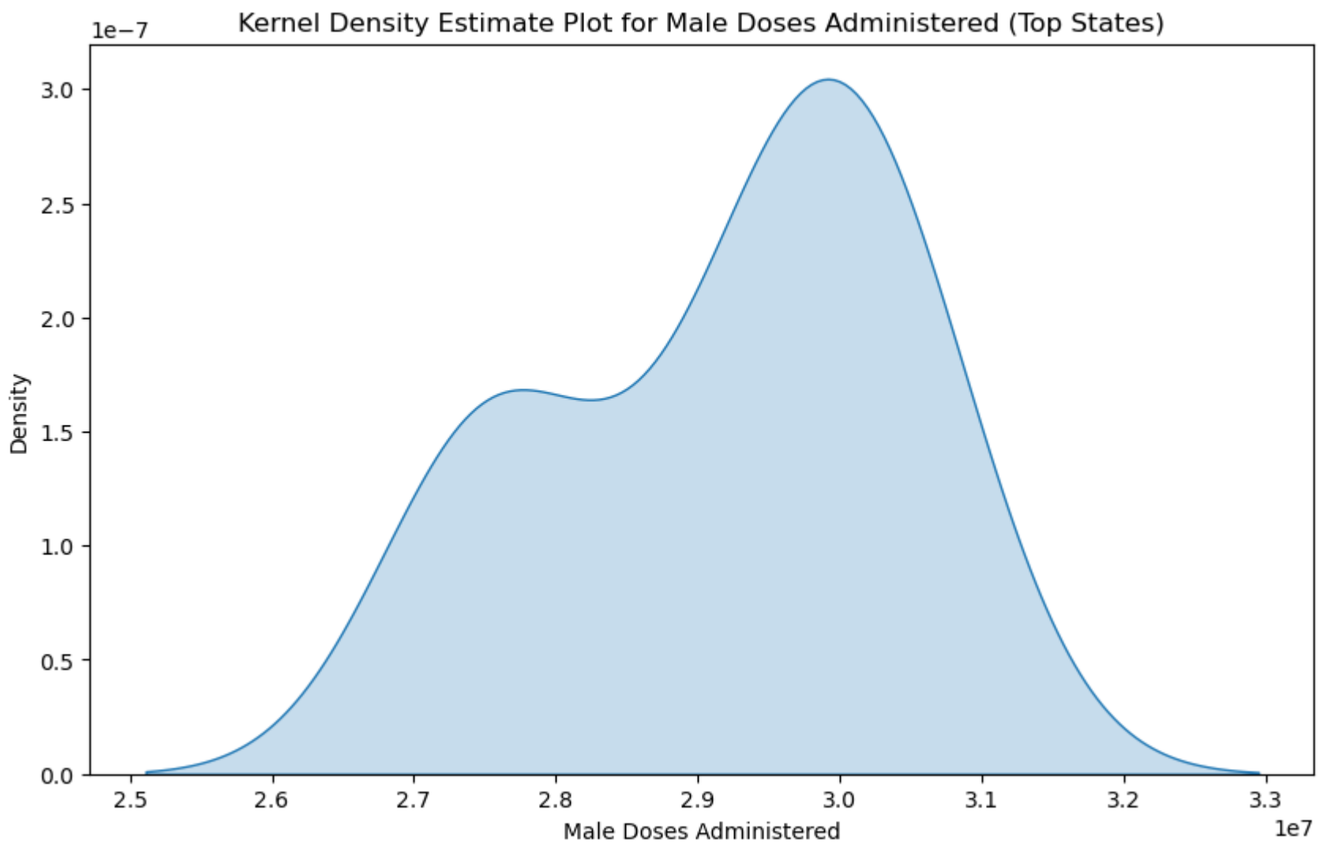
```
state_gender_totals = vaccine.groupby('State')[['Male (Doses Administered)', 'Female (Doses Administered)']].sum()

# Create a stacked bar chart
ax = state_gender_totals.plot(kind='bar', stacked=True, figsize=(12, 6))

# Set labels and title
plt.title('Total Male and Female Doses Administered by State')
plt.xlabel('State')
plt.ylabel('Total Doses Administered')

# Show the legend
plt.legend(title='Gender')

# Show the plot
plt.show()
```



In [19]:
```
# Sort the DataFrame by 'Male (Doses Administered)' in descending order to find the top states
top_male_states = vaccine.sort_values(by='Male (Doses Administered)', ascending=False).head(10)

# Create a KDE plot for Male (Doses Administered)
plt.figure(figsize=(10, 6))
sns.kdeplot(data=top_male_states, x='Male (Doses Administered)', shade=True)
```

```
# Set labels and title
plt.title('Kernel Density Estimate Plot for Male Doses Administered (Top States)')
plt.xlabel('Male Doses Administered')

# Show the plot
plt.show()
```

```
C:\Users\Ankit\AppData\Local\Temp\ipykernel_15392\3925145687.py:6: FutureWarning:


`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.
```



In [ ]: