OASIS INFOBYTE DATA SCIENCE INTERN

Ranjeeta Kumari

TASK - 2 UNEMPLOYMENT ANALYSIS WITH PYTHON

In [18]:

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

```python
df = pd.read_csv(R'C:\Users\ajeet singh\OneDrive\Desktop\Unemployment.csv')
print(df)
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)
0	Andhra Pradesh	31-05-2019	Monthly	3.65
1	Andhra Pradesh	30-06-2019	Monthly	3.05
2	Andhra Pradesh	31-07-2019	Monthly	3.75
3	Andhra Pradesh	31-08-2019	Monthly	3.32
4	Andhra Pradesh	30-09-2019	Monthly	5.17
..	...	...	...	...
763	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN

	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	11999139.0	43.24	Rural
1	11755881.0	42.05	Rural
2	12086707.0	43.50	Rural
3	12285693.0	43.97	Rural
4	12256762.0	44.68	Rural
..	...	...	...
763	NaN	NaN	NaN
764	NaN	NaN	NaN
765	NaN	NaN	NaN
766	NaN	NaN	NaN
767	NaN	NaN	NaN

[768 rows x 7 columns]

In [20]:

```
df.head()
```

Out[20]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural

In [21]:

```
df.tail()
```

Out[21]:

|	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [22]:

```
df.shape
```

Out[22]:

```
(768, 7)
```

```
#now, we are checking start with a pairplot, and check for missing values
sns.heatmap(df.isnull(),cbar=False)
```

```
<AxesSubplot:>
```
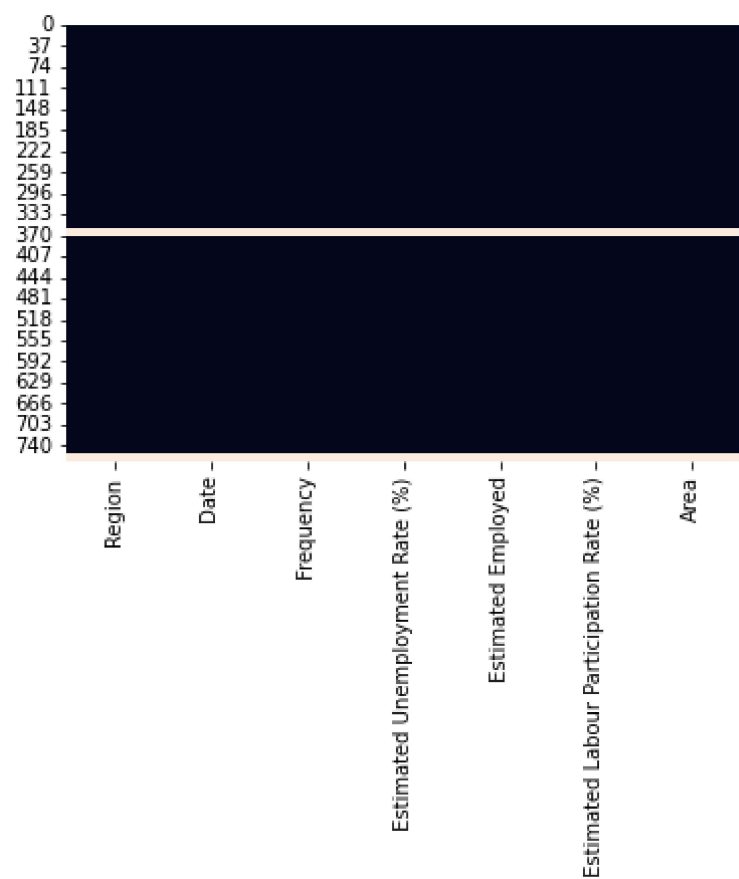
In [26]:

```python
#Which state has the most data
color = sns.color_palette()
cnt_srs = df.Region.value_counts()

plt.figure(figsize=(12,8))
sns.barplot(cnt_srs.index, cnt_srs.values, alpha=0.8, color=color[1])
plt.ylabel('Number of Occurrences', fontsize=12)
plt.xlabel('States', fontsize=12)
plt.title('Count the states', fontsize=15)
plt.xticks(rotation='vertical')
plt.show()
```

D:\anacondaa1\lib\site-packages\seaborn_decorators.py:36: FutureWarning:
Pass the following variables as keyword args: x, y. From version 0.12, the
only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
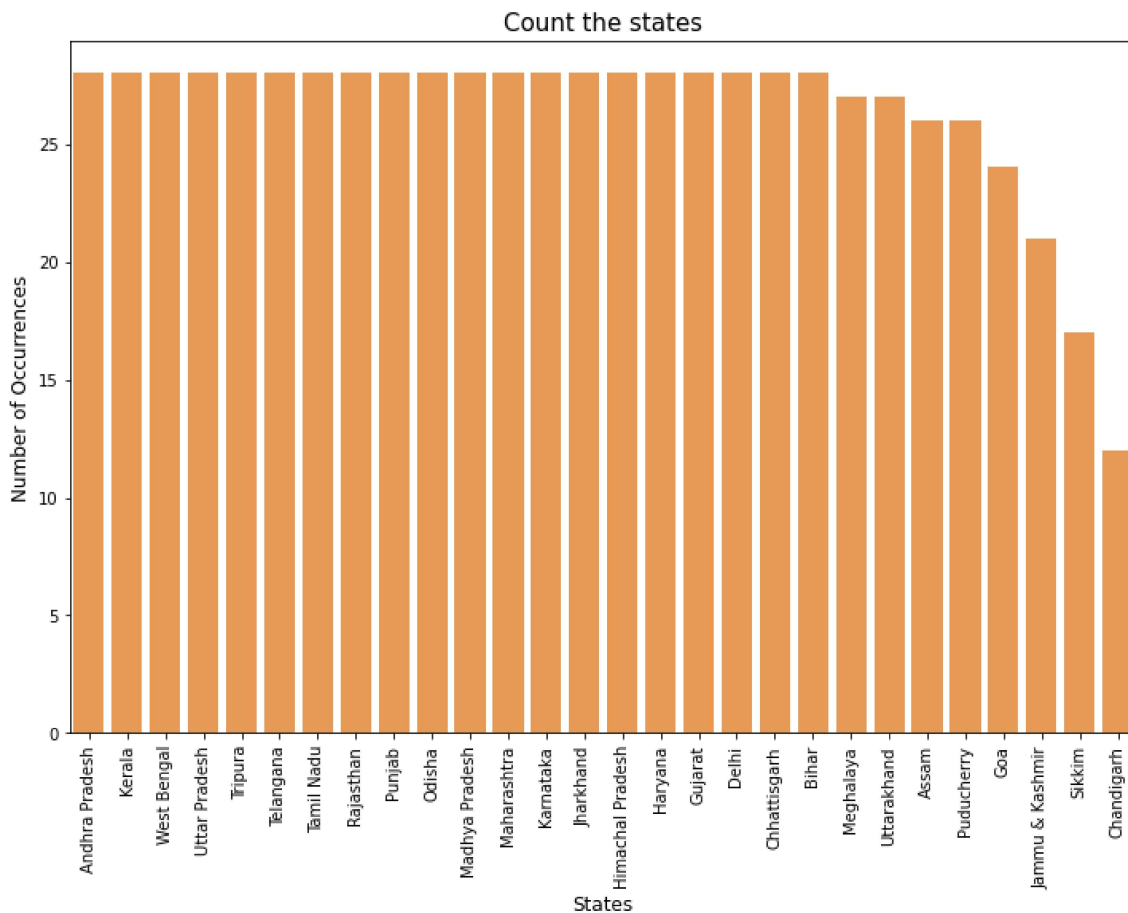
In [37]:

```python
#see the number of unique states
df.Region.nunique()
```

Out[37]:

28

In [ ]: