In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import calendar
import datetime as dt
import plotly.io as pio
import plotly.express as px
import plotly.graph_objects as go
import plotly.figure_factory as ff
from IPython.display import HTML
```

In [2]:
```
df = pd.read_csv('C:\\Users\\saswa\\OneDrive\\Desktop\\Pinaki-unemployment-analy
```

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

Out[3]:

	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 [4]:
```
df
```

Out[4]:

	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
...	...	...	...	...	...	...	...
749	West Bengal	29-02-2020	Monthly	7.55	10871168.0	44.09	Urban
750	West Bengal	31-03-2020	Monthly	6.67	10806105.0	43.34	Urban
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban

754 rows × 7 columns

In [5]:
```
df.tail(3)
```

Out[5]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban

In [6]:
```
df.shape
```

Out[6]:
```
(754, 7)
```

In [7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 754 entries, 0 to 753
Data columns (total 7 columns):
 #   Column                                   Non-Null Count  Dtype  
---  ------                                   --------------  -----  
 0   Region                                   740 non-null    object 
 1   Date                                     740 non-null    object 
 2   Frequency                                740 non-null    object 
 3   Estimated Unemployment Rate (%)          740 non-null    float64
 4   Estimated Employed                       740 non-null    float64
 5   Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                     740 non-null    object 
dtypes: float64(3), object(4)
memory usage: 41.4+ KB
```

In [8]:
```
df.describe()
```

Out[8]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
count	740.000000	7.400000e+02	740.000000
mean	11.787946	7.204460e+06	42.630122
std	10.721298	8.087988e+06	8.111094
min	0.000000	4.942000e+04	13.330000
25%	4.657500	1.190404e+06	38.062500
50%	8.350000	4.744178e+06	41.160000
75%	15.887500	1.127549e+07	45.505000
max	76.740000	4.577751e+07	72.570000

In [9]:
```
df.isnull()
```

Out[9]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
749	False	False	False	False	False	False	False
750	False	False	False	False	False	False	False
751	False	False	False	False	False	False	False
752	False	False	False	False	False	False	False
753	False	False	False	False	False	False	False

754 rows × 7 columns

In [10]:
```
df_cleaned = df.dropna()
```

In [11]:
```
print(df_cleaned)
```

```
            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
..              ...         ...       ...                              ...
749     West Bengal  29-02-2020   Monthly                             7.55
750     West Bengal  31-03-2020   Monthly                             6.67
751     West Bengal  30-04-2020   Monthly                            15.63
752     West Bengal  31-05-2020   Monthly                            15.22
753     West Bengal  30-06-2020   Monthly                             9.86

     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
..                  ...                                   ...    ...
749          10871168.0                                 44.09  Urban
750          10806105.0                                 43.34  Urban
751           9299466.0                                 41.20  Urban
752           9240903.0                                 40.67  Urban
753           9088931.0                                 37.57  Urban

[740 rows x 7 columns]
```

In [12]:
```
df_cleaned.shape
```

Out[12]:
```
(740, 7)
```

In [13]:
```
df_cleaned.isnull().sum()
```

Out[13]:
```
Region                                     0
Date                                       0
Frequency                                  0
Estimated Unemployment Rate (%)            0
Estimated Employed                         0
Estimated Labour Participation Rate (%)    0
Area                                       0
dtype: int64
```

In [14]:
```
df_cleaned.iloc[3]
```

Out[14]:
```
Region                                     Andhra Pradesh
Date                                           31-08-2019
Frequency                                         Monthly
Estimated Unemployment Rate (%)                      3.32
Estimated Employed                             12285693.0
Estimated Labour Participation Rate (%)             43.97
Area                                                Rural
Name: 3, dtype: object
```

In [15]:
```
df_cleaned["Region"].value_counts()
```

Out[15]:
```
Andhra Pradesh      28
Kerala              28
West Bengal         28
Uttar Pradesh       28
Tripura             28
Telangana           28
Tamil Nadu          28
Rajasthan           28
Punjab              28
Odisha              28
Madhya Pradesh      28
Maharashtra         28
Karnataka           28
Jharkhand           28
Himachal Pradesh    28
Haryana             28
Gujarat             28
Delhi               28
Chhattisgarh        28
Bihar               28
Meghalaya           27
Uttarakhand         27
Assam               26
Puducherry          26
Goa                 24
Jammu & Kashmir     21
Sikkim              17
Chandigarh          12
Name: Region, dtype: int64
```

In [16]:
```
df_cleaned["Area"].value_counts()
```

Out[16]:
```
Urban    381
Rural    359
Name: Area, dtype: int64
```

In [17]:
```
df_cleaned.isnull()
```

Out[17]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
749	False	False	False	False	False	False	False
750	False	False	False	False	False	False	False
751	False	False	False	False	False	False	False
752	False	False	False	False	False	False	False
753	False	False	False	False	False	False	False

740 rows × 7 columns

In [18]:
```
x = df_cleaned["Region"]
```

In [19]:
```
x
```

Out[19]:
```
0      Andhra Pradesh
1      Andhra Pradesh
2      Andhra Pradesh
3      Andhra Pradesh
4      Andhra Pradesh
           ...
749       West Bengal
750       West Bengal
751       West Bengal
752       West Bengal
753       West Bengal
Name: Region, Length: 740, dtype: object
```

In [20]:
```
y = df_cleaned["Estimated Labour Participation Rate (%)"]
```

In [21]:
```
y
```

Out[21]:
```
0      43.24
1      42.05
2      43.50
3      43.97
4      44.68
       ...
749    44.09
750    43.34
751    41.20
752    40.67
753    37.57
Name: Estimated Labour Participation Rate (%), Length: 740, dtype: float64
```

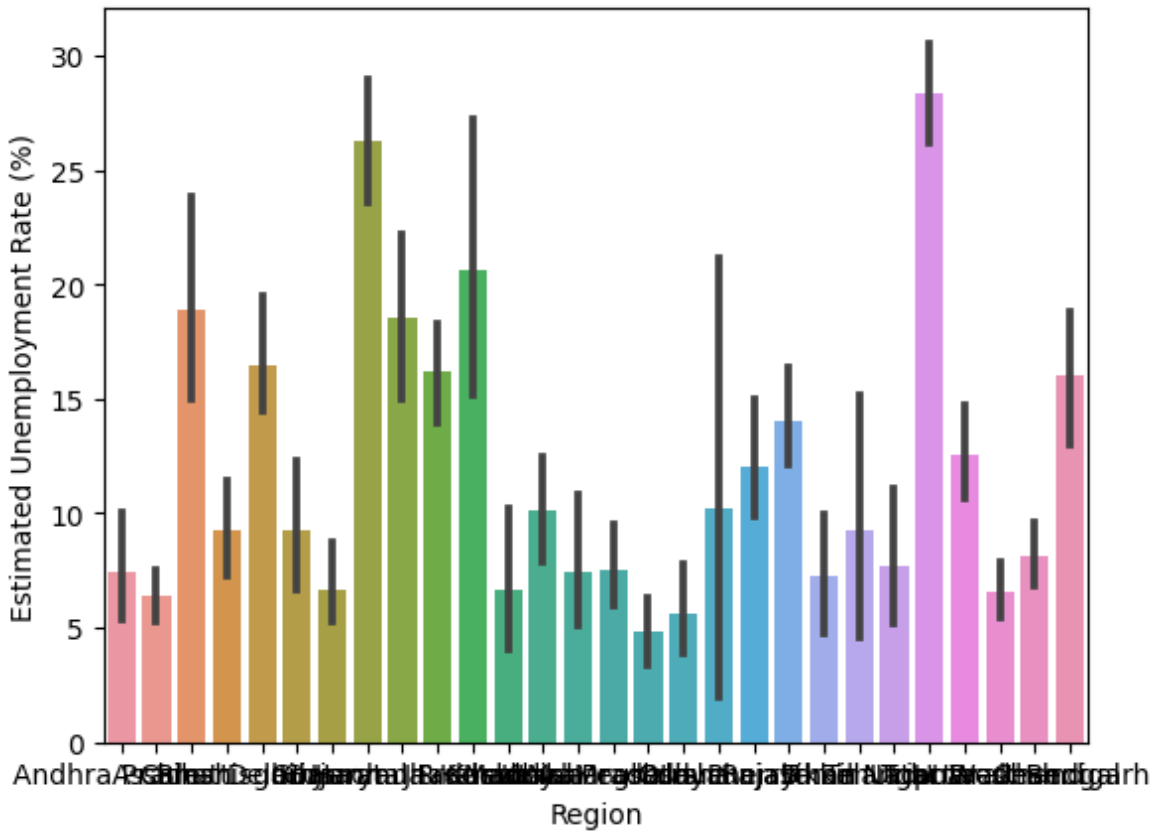
In [22]:
```
data = x = df_cleaned.iloc[:,3]
```

In [23]:
```
data
```

Out[23]:
```
0       3.65
1       3.05
2       3.75
3       3.32
4       5.17
       ...
749     7.55
750     6.67
751    15.63
752    15.22
753     9.86
Name: Estimated Unemployment Rate (%), Length: 740, dtype: float64
```

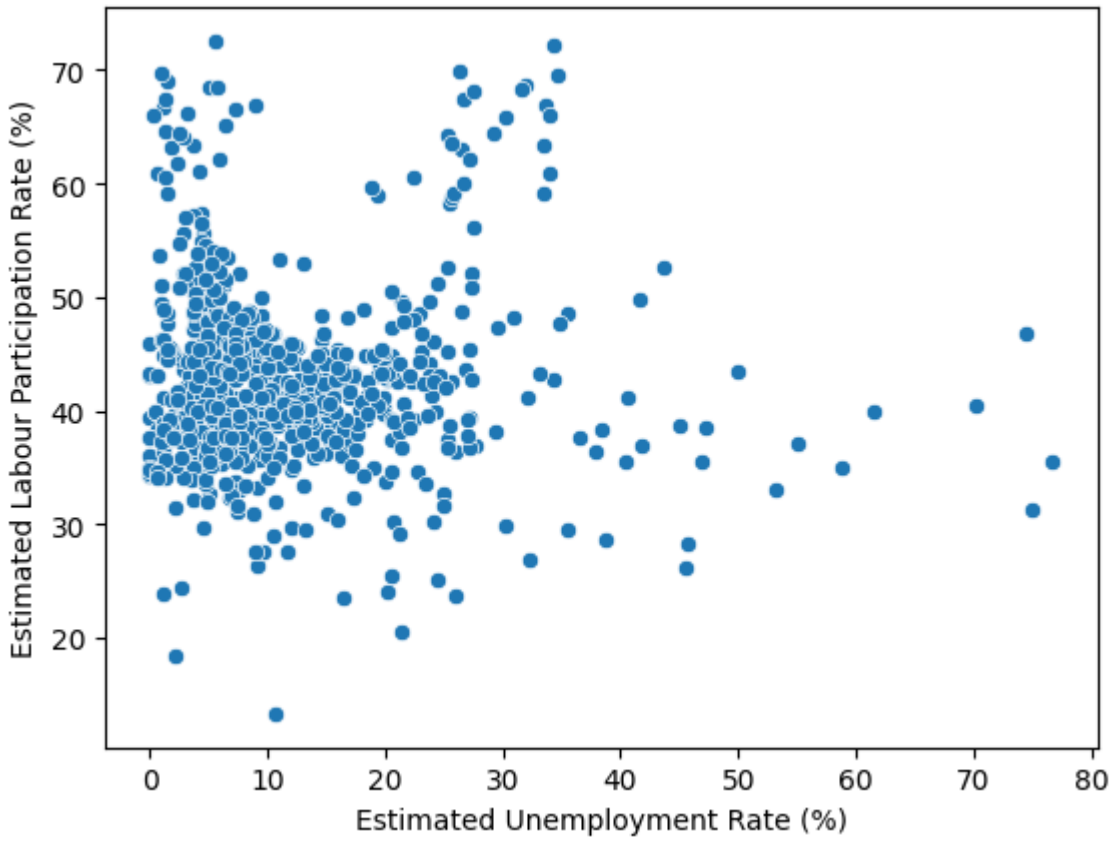
In [24]:
```
sns.barplot(data = df_cleaned, x = 'Region', y = 'Estimated Unemployment Rate (%
```

Out[24]:
```
<Axes: xlabel='Region', ylabel='Estimated Unemployment Rate (%)'>
```



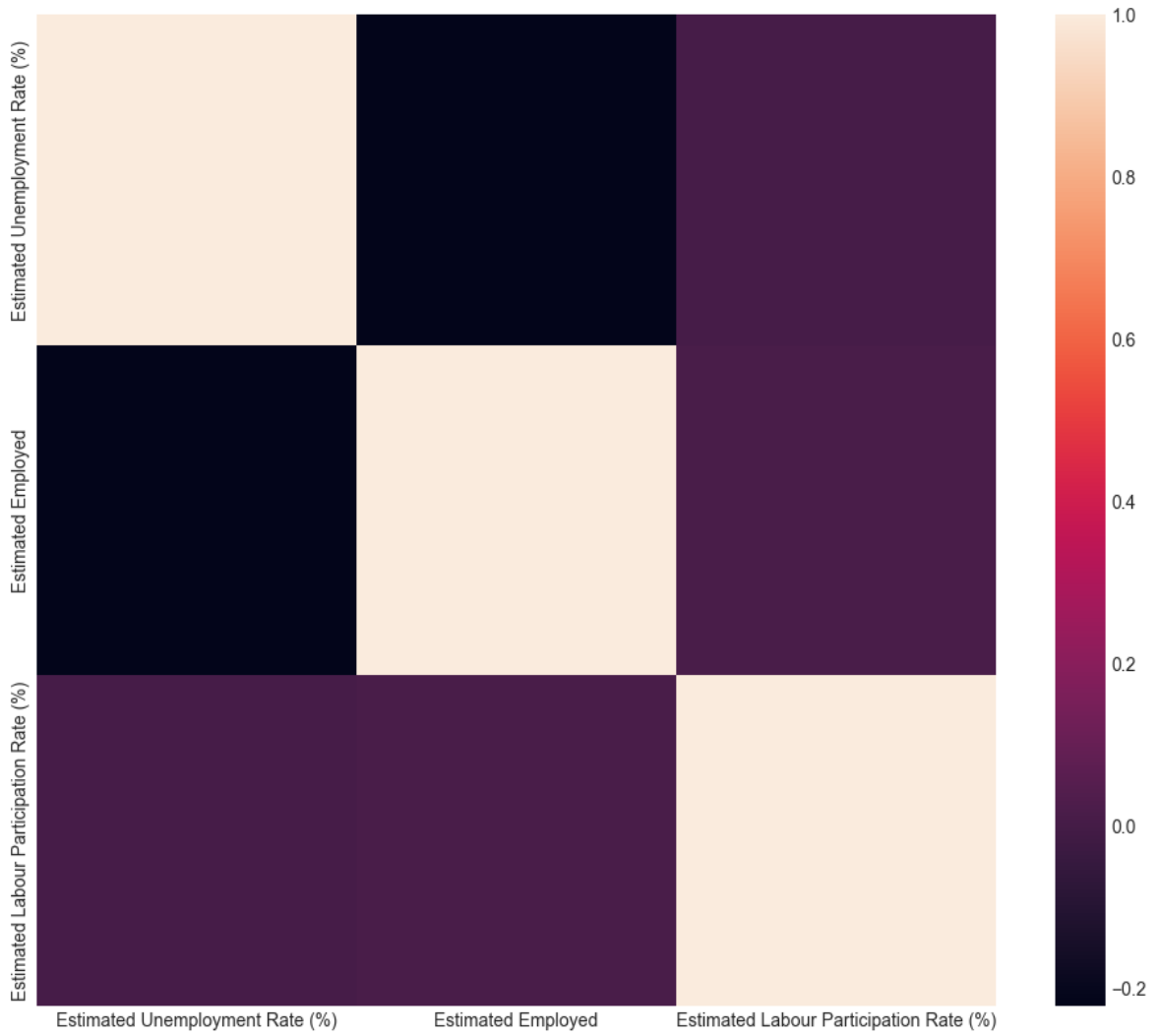
In [25]:
```
sns.scatterplot(data=df_cleaned, x='Estimated Unemployment Rate (%)', y='Estimat
```

Out[25]:
```
<Axes: xlabel='Estimated Unemployment Rate (%)', ylabel='Estimated Labour Parti
cipation Rate (%)'>
```



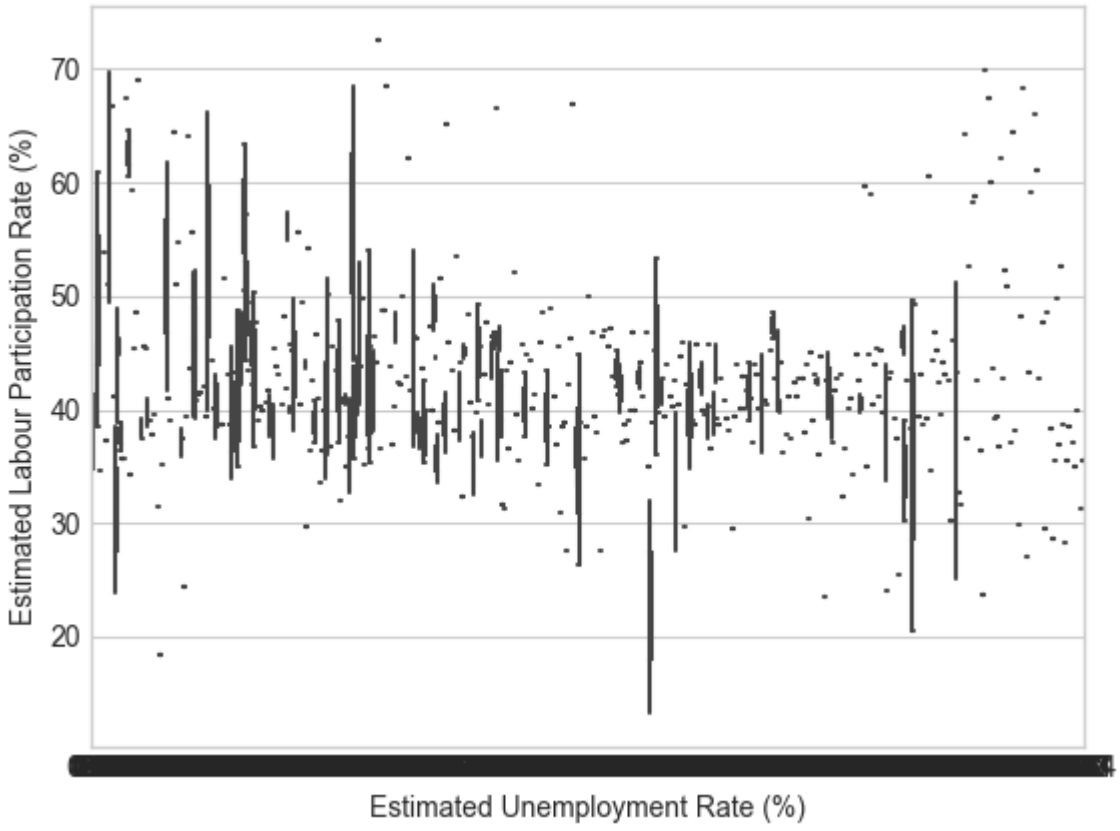
In [26]:
```
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(12, 10))
sns.heatmap(df_cleaned.corr())
plt.show()
```

```
C:\Users\saswa\AppData\Local\Temp\ipykernel_1516\3528222674.py:1: MatplotlibDepre
cationWarning: The seaborn styles shipped by Matplotlib are deprecated since 3.6,
as they no longer correspond to the styles shipped by seaborn. However, they will
remain available as 'seaborn-v0_8-<style>'. Alternatively, directly use the seabo
rn API instead.
  plt.style.use('seaborn-whitegrid')
C:\Users\saswa\AppData\Local\Temp\ipykernel_1516\3528222674.py:3: FutureWarning:
The default value of numeric_only in DataFrame.corr is deprecated. In a future ve
rsion, it will default to False. Select only valid columns or specify the value o
f numeric_only to silence this warning.
  sns.heatmap(df_cleaned.corr())
```



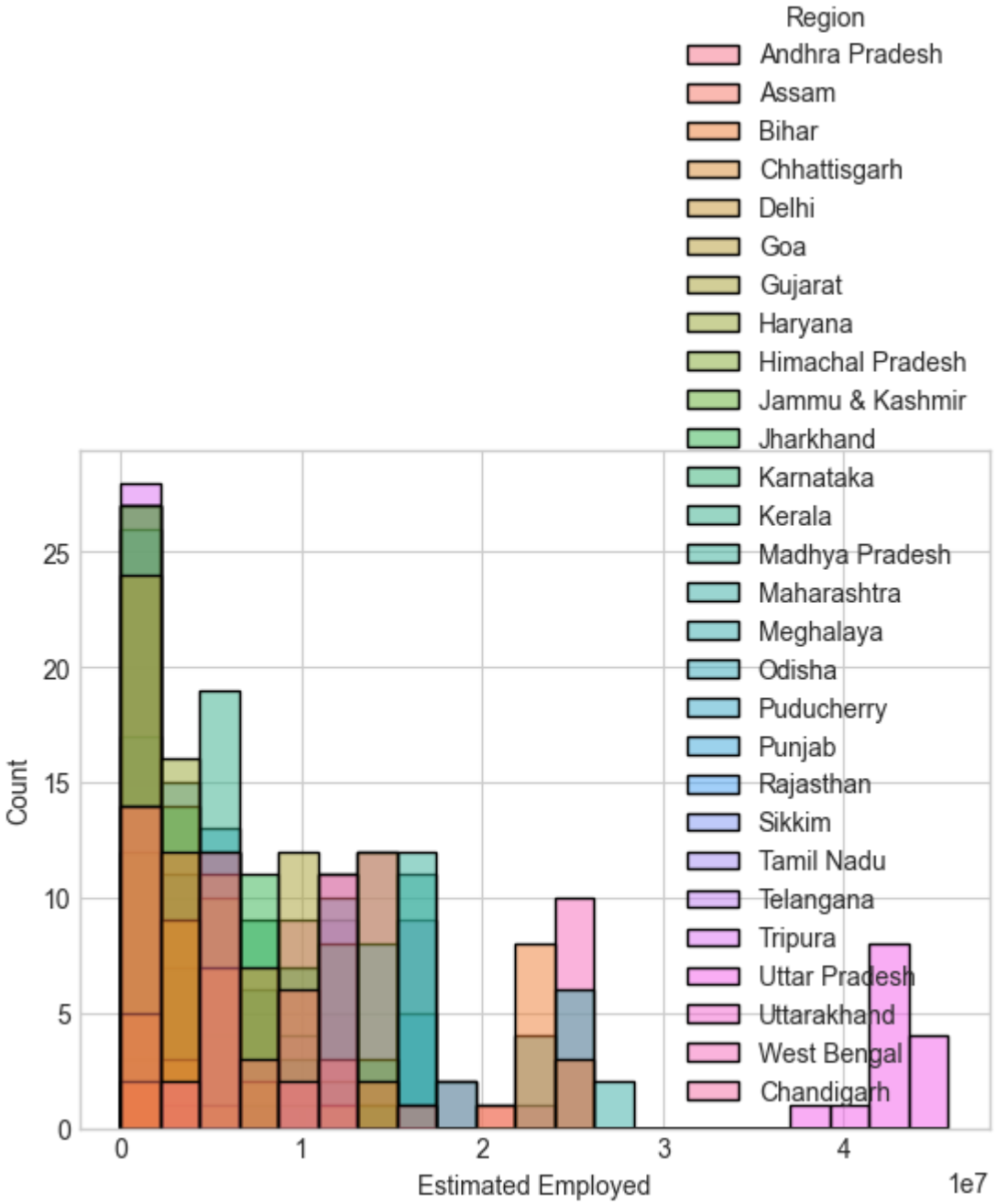
In [27]:
```
sns.boxplot(data=df_cleaned, x='Estimated Unemployment Rate (%)', y='Estimated L
```

Out[27]:
```
<Axes: xlabel='Estimated Unemployment Rate (%)', ylabel='Estimated Labour Parti
cipation Rate (%)'>
```



In [28]:
```
sns.histplot(x='Estimated Employed' ,hue='Region', data=df_cleaned)
```

Out[28]:
```
<Axes: xlabel='Estimated Employed', ylabel='Count'>
```



In [29]:
```
fig = px.bar(df_cleaned, x="Region", y='Estimated Unemployment Rate (%)', color
             title="Unemployment Rates Region wise")
```

In [30]:
```
fig.show()
```

In [31]:
```
fig = px.bar(df_cleaned, x='Region', y='Estimated Labour Participation Rate (%)'
title="Unemployment Rate By States in India")
```

In [32]:
```
fig.show()
```

In [33]:
```
fig = px.box(df_cleaned, x='Region', y='Estimated Unemployment Rate (%)', color=
```

In [34]:
```
fig.show()
```

In [35]:
```
fig = px.histogram(df_cleaned, x='Region', y='Estimated Unemployment Rate (%)',
```

In [36]:
```
fig.show()
```

In [ ]: