In [1]:
```python
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]:
```python
df = pd.read_csv('C:\\Users\\saswa\\OneDrive\\Desktop\\Pinaki-unemployment-analy
```

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

Out[3]:

| | Region | Date | Frequency | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Region.1 |
|---|---|---|---|---|---|---|---|
| 0 | Andhra Pradesh | 31-01-2020 | M | 5.48 | 16635535 | 41.02 | South |
| 1 | Andhra Pradesh | 29-02-2020 | M | 5.83 | 16545652 | 40.90 | South |
| 2 | Andhra Pradesh | 31-03-2020 | M | 5.79 | 15881197 | 39.18 | South |
| 3 | Andhra Pradesh | 30-04-2020 | M | 20.51 | 11336911 | 33.10 | South |
| 4 | Andhra Pradesh | 31-05-2020 | M | 17.43 | 12988845 | 36.46 | South |

In [4]:
```python
df
```

Out[4]:

| | Region | Date | Frequency | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Region.1 |
|---|---|---|---|---|---|---|---|
| 0 | Andhra Pradesh | 31-01-2020 | M | 5.48 | 16635535 | 41.02 | South |
| 1 | Andhra Pradesh | 29-02-2020 | M | 5.83 | 16545652 | 40.90 | South |
| 2 | Andhra Pradesh | 31-03-2020 | M | 5.79 | 15881197 | 39.18 | South |
| 3 | Andhra Pradesh | 30-04-2020 | M | 20.51 | 11336911 | 33.10 | South |
| 4 | Andhra Pradesh | 31-05-2020 | M | 17.43 | 12988845 | 36.46 | South |
| ... | ... | ... | ... | ... | ... | ... | . |
| 262 | West Bengal | 30-06-2020 | M | 7.29 | 30726310 | 40.39 | Eas |
| 263 | West Bengal | 31-07-2020 | M | 6.83 | 35372506 | 46.17 | Eas |
| 264 | West Bengal | 31-08-2020 | M | 14.87 | 33298644 | 47.48 | Eas |
| 265 | West Bengal | 30-09-2020 | M | 9.35 | 35707239 | 47.73 | Eas |
| 266 | West Bengal | 31-10-2020 | M | 9.98 | 33962549 | 45.63 | Eas |

267 rows × 9 columns

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

Out[5]:

| | Region | Date | Frequency | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Region.1 |
|---|---|---|---|---|---|---|---|
| 264 | West Bengal | 31-08-2020 | M | 14.87 | 33298644 | 47.48 | East |
| 265 | West Bengal | 30-09-2020 | M | 9.35 | 35707239 | 47.73 | East |
| 266 | West Bengal | 31-10-2020 | M | 9.98 | 33962549 | 45.63 | East |

In [6]: `df.shape`

Out[6]: (267, 9)

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 9 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   Region                                  267 non-null    object
 1   Date                                    267 non-null    object
 2   Frequency                               267 non-null    object
 3   Estimated Unemployment Rate (%)         267 non-null    float64
 4   Estimated Employed                      267 non-null    int64
 5   Estimated Labour Participation Rate (%) 267 non-null    float64
 6   Region.1                                267 non-null    object
 7   Longitude                               267 non-null    float64
 8   Latitude                                267 non-null    float64
dtypes: float64(4), int64(1), object(4)
memory usage: 18.9+ KB
```

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

Out[8]:

|  | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Longitude | Latitude |
|---|---|---|---|---|---|
| count | 267.000000 | 2.670000e+02 | 267.000000 | 267.000000 | 267.000000 |
| mean | 12.236929 | 1.396211e+07 | 41.681573 | 22.826048 | 80.532425 |
| std | 10.803283 | 1.336632e+07 | 7.845419 | 6.270731 | 5.831738 |
| min | 0.500000 | 1.175420e+05 | 16.770000 | 10.850500 | 71.192400 |
| 25% | 4.845000 | 2.838930e+06 | 37.265000 | 18.112400 | 76.085600 |
| 50% | 9.650000 | 9.732417e+06 | 40.390000 | 23.610200 | 79.019300 |
| 75% | 16.755000 | 2.187869e+07 | 44.055000 | 27.278400 | 85.279900 |
| max | 75.850000 | 5.943376e+07 | 69.690000 | 33.778200 | 92.937600 |

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

Out[9]:

| | Region | Date | Frequency | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Region.1 | Lon |
|---|---|---|---|---|---|---|---|---|
| 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 | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| 262 | False | False | False | False | False | False | False | |
| 263 | False | False | False | False | False | False | False | |
| 264 | False | False | False | False | False | False | False | |
| 265 | False | False | False | False | False | False | False | |
| 266 | False | False | False | False | False | False | False | |

267 rows × 9 columns

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

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

```
               Region        Date Frequency  Estimated Unemployment Rate (%)  \
0      Andhra Pradesh  31-01-2020         M                             5.48
1      Andhra Pradesh  29-02-2020         M                             5.83
2      Andhra Pradesh  31-03-2020         M                             5.79
3      Andhra Pradesh  30-04-2020         M                            20.51
4      Andhra Pradesh  31-05-2020         M                            17.43
..                ...         ...       ...                              ...
262       West Bengal  30-06-2020         M                             7.29
263       West Bengal  31-07-2020         M                             6.83
264       West Bengal  31-08-2020         M                            14.87
265       West Bengal  30-09-2020         M                             9.35
266       West Bengal  31-10-2020         M                             9.98

     Estimated Employed  Estimated Labour Participation Rate (%) Region.1  \
0              16635535                                    41.02    South
1              16545652                                    40.90    South
2              15881197                                    39.18    South
3              11336911                                    33.10    South
4              12988845                                    36.46    South
..                  ...                                      ...      ...
262            30726310                                    40.39     East
263            35372506                                    46.17     East
264            33298644                                    47.48     East
265            35707239                                    47.73     East
266            33962549                                    45.63     East

     Longitude  Latitude
0      15.9129    79.740
1      15.9129    79.740
2      15.9129    79.740
3      15.9129    79.740
4      15.9129    79.740
..         ...       ...
262    22.9868    87.855
263    22.9868    87.855
264    22.9868    87.855
265    22.9868    87.855
266    22.9868    87.855

[267 rows x 9 columns]
```

In [12]: `df_cleaned.shape`

Out[12]: `(267, 9)`

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
Region.1                                   0
Longitude                                  0
Latitude                                   0
dtype: int64
```

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

```
Out[14]:  Region                                    Andhra Pradesh
          Date                                          30-04-2020
          Frequency                                              M
          Estimated Unemployment Rate (%)                    20.51
          Estimated Employed                              11336911
          Estimated Labour Participation Rate (%)             33.1
          Region.1                                           South
          Longitude                                        15.9129
          Latitude                                           79.74
          Name: 3, dtype: object
```

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

```
Out[15]:  Andhra Pradesh       10
          Assam                10
          Uttarakhand          10
          Uttar Pradesh        10
          Tripura              10
          Telangana            10
          Tamil Nadu           10
          Rajasthan            10
          Punjab               10
          Puducherry           10
          Odisha               10
          Meghalaya            10
          Maharashtra          10
          Madhya Pradesh       10
          Kerala               10
          Karnataka            10
          Jharkhand            10
          Himachal Pradesh     10
          Haryana              10
          Gujarat              10
          Goa                  10
          Delhi                10
          Chhattisgarh         10
          Bihar                10
          West Bengal          10
          Jammu & Kashmir       9
          Sikkim                8
          Name: Region, dtype: int64
```

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

```
Out[16]: Andhra Pradesh       10
         Assam                10
         Uttarakhand          10
         Uttar Pradesh        10
         Tripura              10
         Telangana            10
         Tamil Nadu           10
         Rajasthan            10
         Punjab               10
         Puducherry           10
         Odisha               10
         Meghalaya            10
         Maharashtra          10
         Madhya Pradesh       10
         Kerala               10
         Karnataka            10
         Jharkhand            10
         Himachal Pradesh     10
         Haryana              10
         Gujarat              10
         Goa                  10
         Delhi                10
         Chhattisgarh         10
         Bihar                10
         West Bengal          10
         Jammu & Kashmir       9
         Sikkim                8
         Name: Region, dtype: int64
```

In [17]:
```python
sum = df_cleaned["Latitude"].sum()
print(sum)
```

```
21502.157399999996
```

In [18]:
```python
df_cleaned.isnull()
```

Out[18]:

| | Region | Date | Frequency | Estimated Unemployment Rate (%) | Estimated Employed | Estimated Labour Participation Rate (%) | Region.1 | Lon |
|---|---|---|---|---|---|---|---|---|
| **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 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | |
| **262** | False | False | False | False | False | False | False | |
| **263** | False | False | False | False | False | False | False | |
| **264** | False | False | False | False | False | False | False | |
| **265** | False | False | False | False | False | False | False | |
| **266** | False | False | False | False | False | False | False | |

267 rows × 9 columns

```python
In [19]: x = df_cleaned["Region"]
```

```python
In [20]: x
```

```
Out[20]: 0      Andhra Pradesh
         1      Andhra Pradesh
         2      Andhra Pradesh
         3      Andhra Pradesh
         4      Andhra Pradesh
                   ...
         262       West Bengal
         263       West Bengal
         264       West Bengal
         265       West Bengal
         266       West Bengal
         Name: Region, Length: 267, dtype: object
```

```python
In [21]: y = df_cleaned["Estimated Labour Participation Rate (%)"]
```

```python
In [22]: y
```

```
Out[22]:  0       41.02
          1       40.90
          2       39.18
          3       33.10
          4       36.46
                   ...
          262     40.39
          263     46.17
          264     47.48
          265     47.73
          266     45.63
          Name: Estimated Labour Participation Rate (%), Length: 267, dtype: float64
```

```
In [23]:  data = x = df_cleaned.iloc[:,3]
```
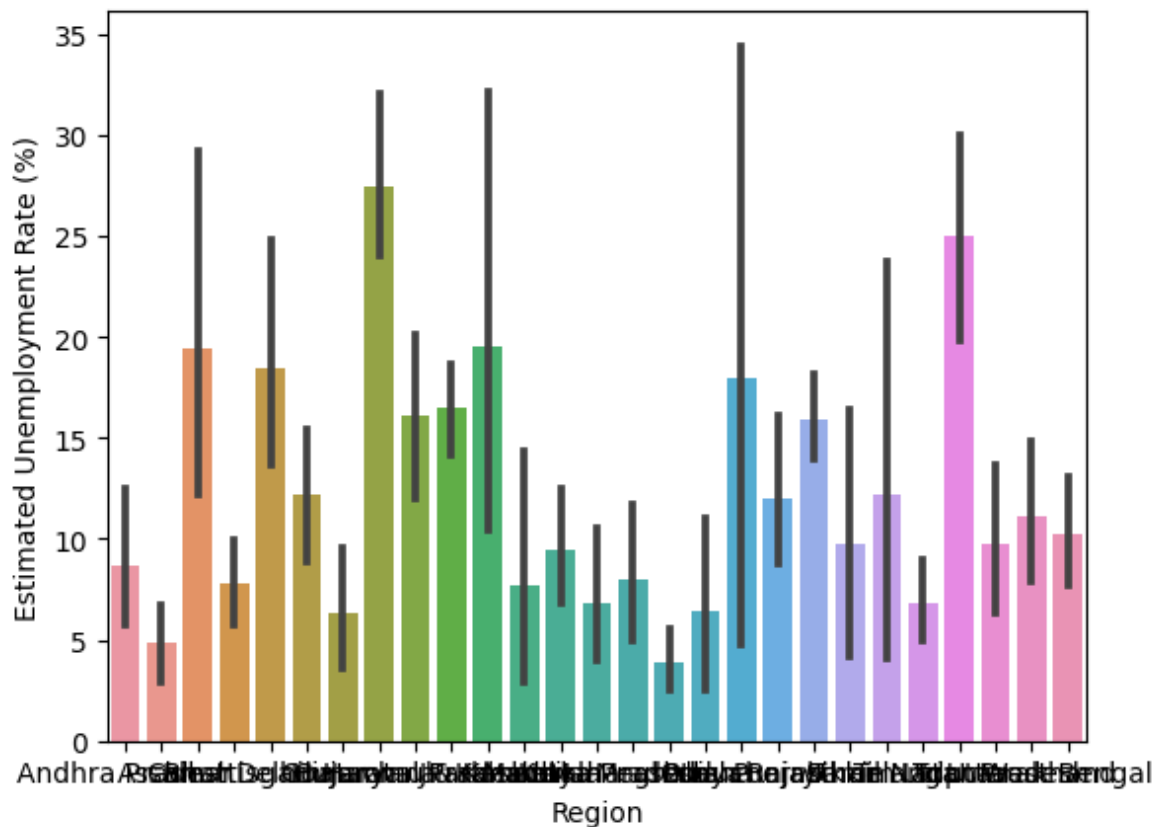
```
In [24]:  data
```

```
Out[24]:  0        5.48
          1        5.83
          2        5.79
          3       20.51
          4       17.43
                   ...
          262      7.29
          263      6.83
          264     14.87
          265      9.35
          266      9.98
          Name: Estimated Unemployment Rate (%), Length: 267, dtype: float64
```

```
In [25]:  sns.barplot(data = df_cleaned, x = 'Region', y = 'Estimated Unemployment Rate (%
```
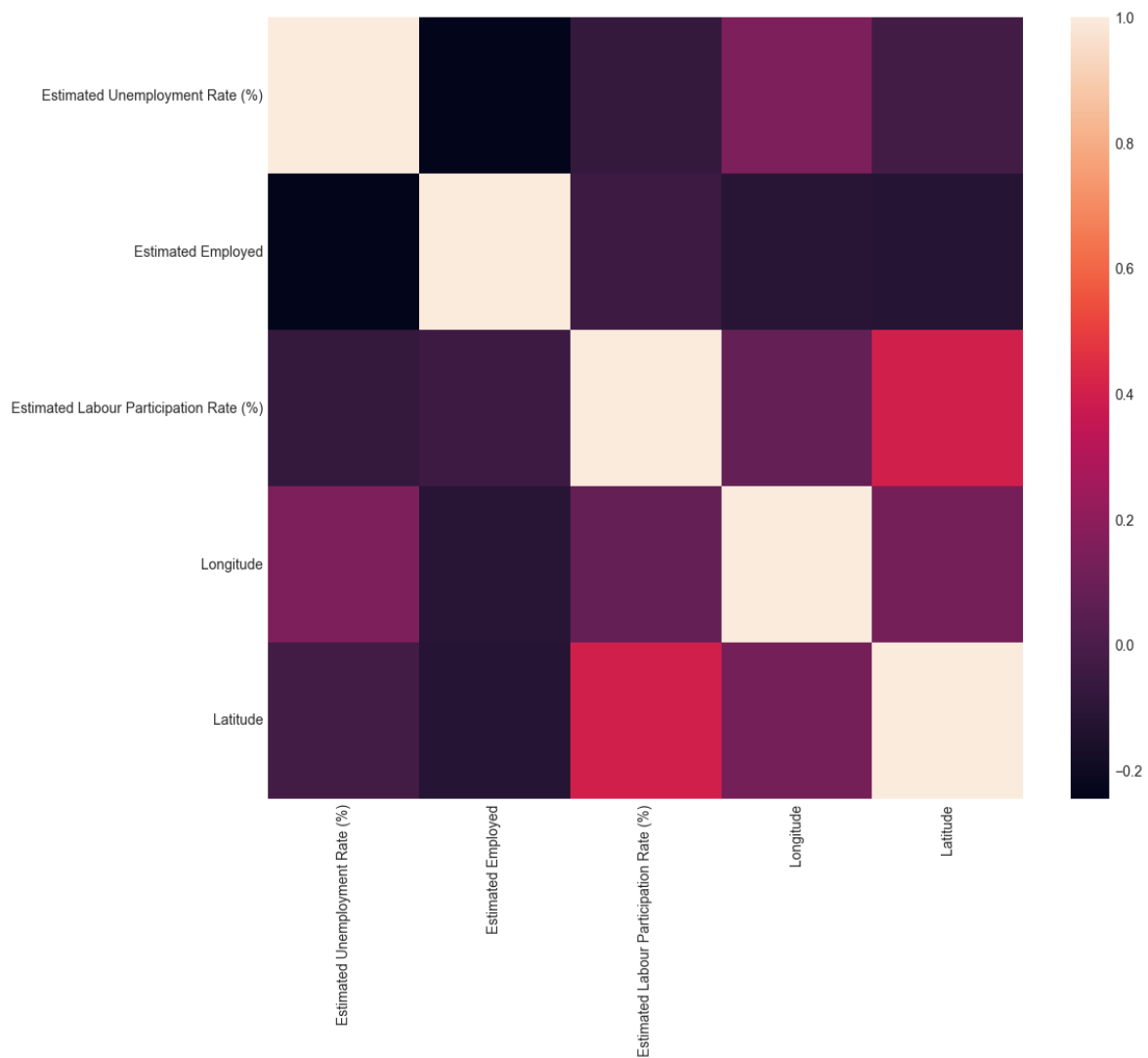
```
Out[25]:  <Axes: xlabel='Region', ylabel='Estimated Unemployment Rate (%)'>
```

In [26]: `sns.scatterplot(data=df_cleaned, x='Estimated Unemployment Rate (%)', y='Estimat`

Out[26]: `<Axes: xlabel='Estimated Unemployment Rate (%)', ylabel='Estimated Labour Participation Rate (%)'>`
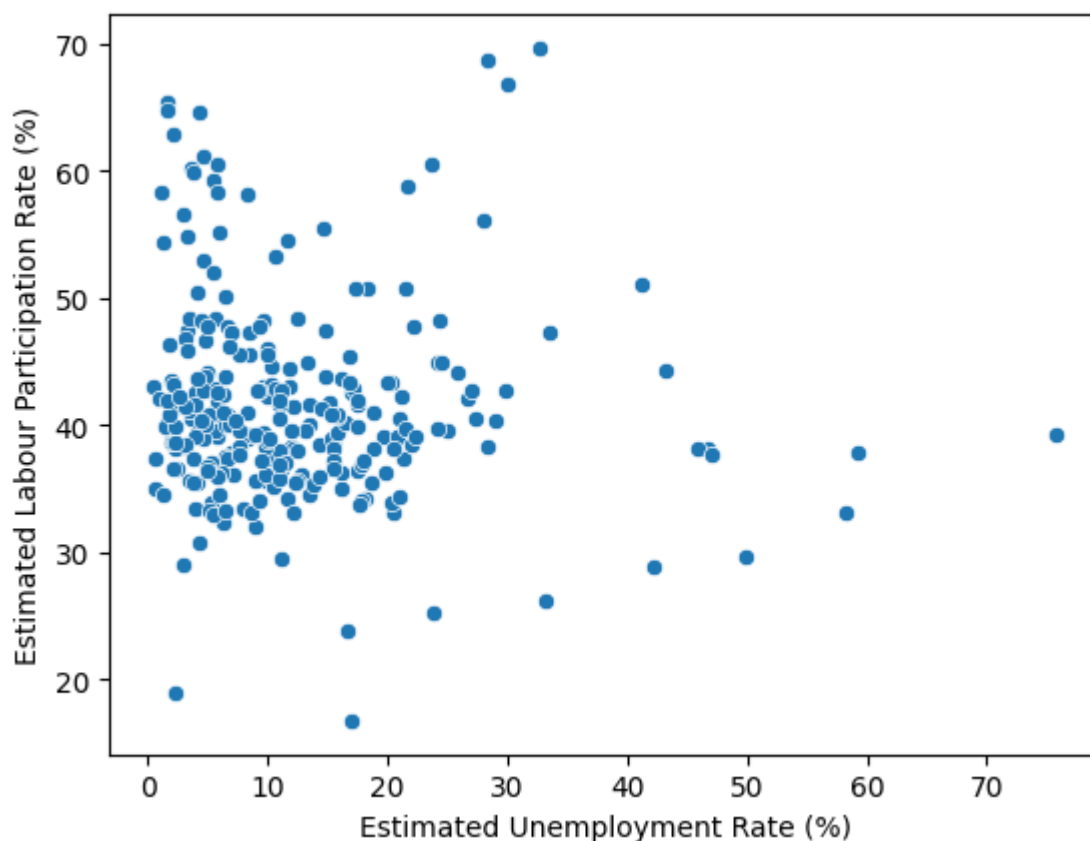


In [27]:
```python
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(12, 10))
sns.heatmap(df_cleaned.corr())
plt.show()
```

```
C:\Users\saswa\AppData\Local\Temp\ipykernel_3240\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_3240\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())
```
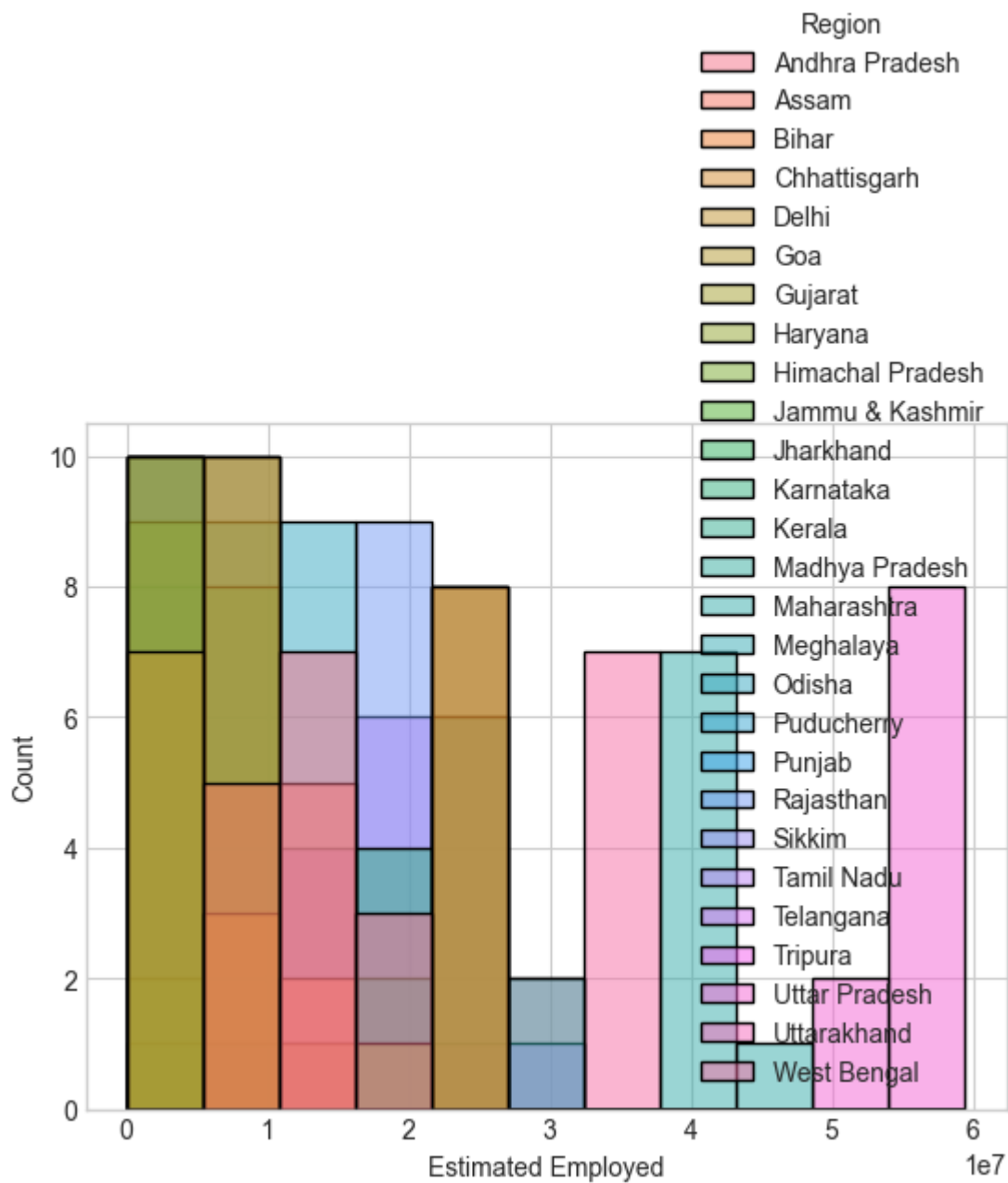
```
In [28]: sns.boxplot(data=df_cleaned, x='Estimated Unemployment Rate (%)', y='Estimated L
```

Out[28]: `<Axes: xlabel='Estimated Unemployment Rate (%)', ylabel='Estimated Labour Participation Rate (%)'>`
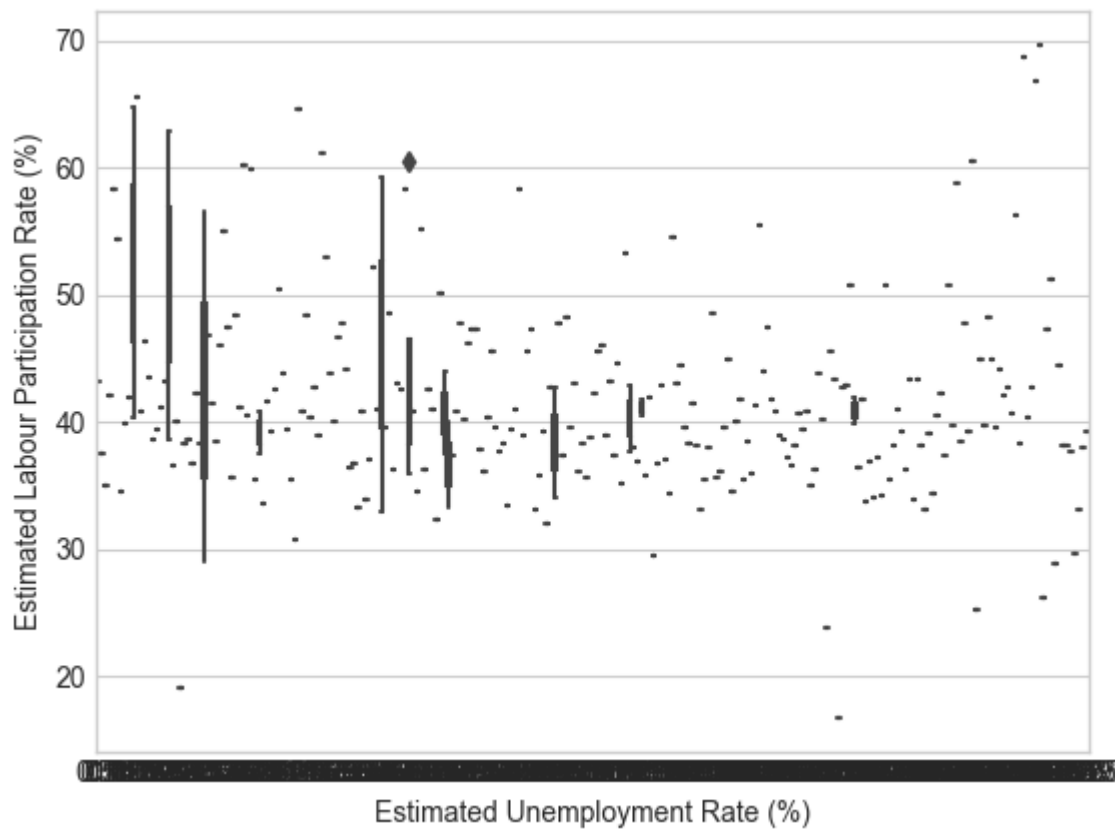
```
In [29]:  sns.histplot(x='Estimated Employed' ,hue='Region', data=df_cleaned)
```

Out[29]:  <Axes: xlabel='Estimated Employed', ylabel='Count'>

```
In [30]: fig = px.bar(df_cleaned, x="Region", y='Estimated Unemployment Rate (%)', color
                 title="Unemployment Rates Region wise")
```

```
In [31]: fig.show()
```

In [32]:
```python
fig = px.bar(df_cleaned, x='Region', y='Estimated Unemployment Rate (%)' , color
title="Unemployment Rate By States in India")
```

In [33]:
```python
fig.show()
```

In [34]:
```python
fig = px.box(df_cleaned, x='Region', y='Estimated Unemployment Rate (%)', color=
```

In [35]:
```python
fig.show()
```

In [36]: 
```python
fig = px.histogram(df_cleaned, x='Region', y='Estimated Unemployment Rate (%)',
```

In [37]: 
```python
fig.show()
```

In [104…

In [ ]:

In [ ]: