Sakshi Kharat

Oasis Infobyte (Data Science) - Task2

June-P2 Batch Oasis Infobyte SIP

Unemployment Analysis With Python

Importing the Libraries

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import plotly.express as px
```

```
import seaborn as sns
```

Loading the Data

```
from google.colab import files


uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving Unemployment_Rate_upto_11_2020.csv to
Unemployment_Rate_upto_11_2020.csv
```

```
df = pd.read_csv("Unemployment_Rate_upto_11_2020.csv")
```

Viewing the Data

```
df
```

```
             States         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
..              ...          ...        ...
...
```

```
749              NaN          NaN        NaN
NaN
750              NaN          NaN        NaN
NaN
751              NaN          NaN        NaN
NaN
752              NaN          NaN        NaN
NaN
753              NaN          NaN        NaN
NaN

      Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
0             16635535.0                                    41.02
South
1             16545652.0                                    40.90
South
2             15881197.0                                    39.18
South
3             11336911.0                                    33.10
South
4             12988845.0                                    36.46
South
..                   ...                                      ...
...
749                  NaN                                      NaN
NaN
750                  NaN                                      NaN
NaN
751                  NaN                                      NaN
NaN
752                  NaN                                      NaN
NaN
753                  NaN                                      NaN
NaN

     longitude  latitude   Area
0      15.9129     79.74  Rural
1      15.9129     79.74  Rural
2      15.9129     79.74  Rural
3      15.9129     79.74  Rural
4      15.9129     79.74  Rural
..         ...       ...    ...
749        NaN       NaN  Urban
750        NaN       NaN  Urban
751        NaN       NaN  Urban
752        NaN       NaN  Urban
753        NaN       NaN  Urban

[754 rows x 10 columns]
```

```
df.head()
```

```
           States         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

    Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
0           16635535.0                                    41.02
South
1           16545652.0                                    40.90
South
2           15881197.0                                    39.18
South
3           11336911.0                                    33.10
South
4           12988845.0                                    36.46
South

   longitude  latitude   Area
0    15.9129     79.74  Rural
1    15.9129     79.74  Rural
2    15.9129     79.74  Rural
3    15.9129     79.74  Rural
4    15.9129     79.74  Rural
```

Returns First 3 Entries

```
df.head(3)
```

```
           States         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

    Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
0           16635535.0                                    41.02
```

```
South
1           16545652.0                                          40.90
South
2           15881197.0                                          39.18
South

   longitude  latitude   Area
0    15.9129     79.74  Rural
1    15.9129     79.74  Rural
2    15.9129     79.74  Rural
```

Display last 3 entries

```
df.tail(3)
```

```
    States  Date  Frequency   Estimated Unemployment Rate (%)  \
751    NaN   NaN        NaN                                NaN
752    NaN   NaN        NaN                                NaN
753    NaN   NaN        NaN                                NaN

      Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
751                  NaN                                       NaN
NaN
752                  NaN                                       NaN
NaN
753                  NaN                                       NaN
NaN

     longitude  latitude   Area
751        NaN       NaN  Urban
752        NaN       NaN  Urban
753        NaN       NaN  Urban
```

```
df.shape
```

```
(754, 10)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 754 entries, 0 to 753
Data columns (total 10 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   States                                    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    float64
 5    Estimated Labour Participation Rate (%)  267 non-null    float64
```

```
 6   Region                                    267 non-null    object
 7   longitude                                 267 non-null    float64
 8   latitude                                  267 non-null    float64
 9   Area                                      740 non-null    object
dtypes: float64(5), object(5)
memory usage: 59.0+ KB
```

Displaying the Numerical Description of the Dataset

```
df.describe()
```

```
        Estimated Unemployment Rate (%)   Estimated Employed  \
count                         267.000000         2.670000e+02
mean                           12.236929         1.396211e+07
std                            10.803283         1.336632e+07
min                             0.500000         1.175420e+05
25%                             4.845000         2.838930e+06
50%                             9.650000         9.732417e+06
75%                            16.755000         2.187869e+07
max                            75.850000         5.943376e+07

        Estimated Labour Participation Rate (%)   longitude
latitude
count                                267.000000  267.000000
267.000000
mean                                  41.681573   22.826048
80.532425
std                                    7.845419    6.270731
5.831738
min                                   16.770000   10.850500
71.192400
25%                                   37.265000   18.112400
76.085600
50%                                   40.390000   23.610200
79.019300
75%                                   44.055000   27.278400
85.279900
max                                   69.690000   33.778200
92.937600
```

Checking the Missing Values

```
df.isnull()
```

```
     States   Date   Frequency   Estimated Unemployment Rate (%)  \
0     False  False       False                              False
1     False  False       False                              False
2     False  False       False                              False
3     False  False       False                              False
4     False  False       False                              False
..      ...    ...         ...                                ...
```

```
749    True   True        True                                  True
750    True   True        True                                  True
751    True   True        True                                  True
752    True   True        True                                  True
753    True   True        True                                  True

      Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
0                   False                                   False
False
1                   False                                   False
False
2                   False                                   False
False
3                   False                                   False
False
4                   False                                   False
False
..                    ...                                     ...
...
749                  True                                    True
True
750                  True                                    True
True
751                  True                                    True
True
752                  True                                    True
True
753                  True                                    True
True

     longitude  latitude   Area
0        False     False  False
1        False     False  False
2        False     False  False
3        False     False  False
4        False     False  False
..         ...       ...    ...
749       True      True  False
750       True      True  False
751       True      True  False
752       True      True  False
753       True      True  False

[754 rows x 10 columns]
```

Removing the Rows and Columns contains Missing or null Value

```
df_cleaned = df.dropna()
```

```
print(df_cleaned)
```

```
             States         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  \
0             16635535.0                                   41.02
South
1             16545652.0                                   40.90
South
2             15881197.0                                   39.18
South
3             11336911.0                                   33.10
South
4             12988845.0                                   36.46
South
..                   ...                                     ...
...
262           30726310.0                                   40.39
East
263           35372506.0                                   46.17
East
264           33298644.0                                   47.48
East
265           35707239.0                                   47.73
East
266           33962549.0                                   45.63
```

```
East

     longitude  latitude   Area
0      15.9129    79.740  Rural
1      15.9129    79.740  Rural
2      15.9129    79.740  Rural
3      15.9129    79.740  Rural
4      15.9129    79.740  Rural
..         ...       ...    ...
262    22.9868    87.855  Rural
263    22.9868    87.855  Rural
264    22.9868    87.855  Rural
265    22.9868    87.855  Rural
266    22.9868    87.855  Rural

[267 rows x 10 columns]
```

```
df_cleaned.shape
```

```
(267, 10)
```

Checking the number of missing values of each column

```
df_cleaned.isnull().sum()
```

```
States                                    0
 Date                                     0
 Frequency                                0
 Estimated Unemployment Rate (%)          0
 Estimated Employed                       0
 Estimated Labour Participation Rate (%)  0
Region                                    0
longitude                                 0
latitude                                  0
Area                                      0
dtype: int64
```

Retrieving the Specific Row

```
df_cleaned.iloc[3]
```

```
States                                   Andhra Pradesh
 Date                                        30-04-2020
 Frequency                                            M
 Estimated Unemployment Rate (%)                  20.51
 Estimated Employed                          11336911.0
 Estimated Labour Participation Rate (%)           33.1
Region                                            South
longitude                                       15.9129
latitude                                          79.74
Area                                              Rural
Name: 3, dtype: object
```

```
df_cleaned["States"].value_counts()
```

```
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: States, dtype: int64
```

```
df_cleaned["Region"].value_counts()
```

```
North        79
South        60
West         50
East         40
Northeast    38
Name: Region, dtype: int64
```

```
sum = df_cleaned["latitude"].sum()
print(sum)
```

```
21502.157399999996
```

```
df_cleaned.isnull()
```

```
     States   Date   Frequency   Estimated Unemployment Rate (%)  \
0     False  False       False                              False
1     False  False       False                              False
2     False  False       False                              False
3     False  False       False                              False
```

```
4     False  False       False                                False
..      ...    ...         ...                                  ...
262   False  False       False                                False
263   False  False       False                                False
264   False  False       False                                False
265   False  False       False                                False
266   False  False       False                                False

      Estimated Employed   Estimated Labour Participation Rate (%)
Region  \
0                   False                                    False
False
1                   False                                    False
False
2                   False                                    False
False
3                   False                                    False
False
4                   False                                    False
False
..                    ...                                      ...
...
262                 False                                    False
False
263                 False                                    False
False
264                 False                                    False
False
265                 False                                    False
False
266                 False                                    False
False

     longitude  latitude   Area
0        False     False  False
1        False     False  False
2        False     False  False
3        False     False  False
4        False     False  False
..         ...       ...    ...
262      False     False  False
263      False     False  False
264      False     False  False
265      False     False  False
266      False     False  False

[267 rows x 10 columns]
```

Data Visualization

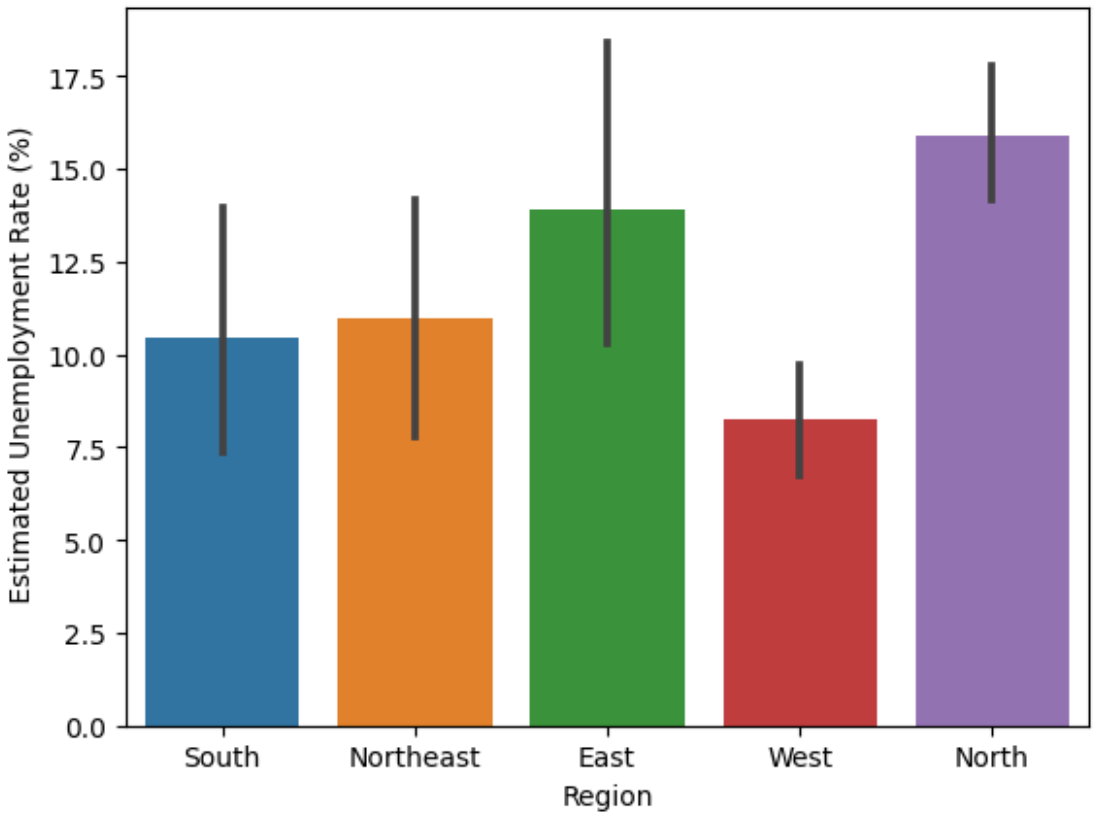
```
x = df_cleaned["States"]

x

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: States, Length: 267, dtype: object

y = df_cleaned[" Estimated Labour Participation Rate (%)"]

y

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

data = x = df_cleaned.iloc[:,3]

data

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
```

Data Visualization

Bar Graph

```
sns.barplot(data = df_cleaned, x='Region', y=' Estimated Unemployment
Rate (%)')
```

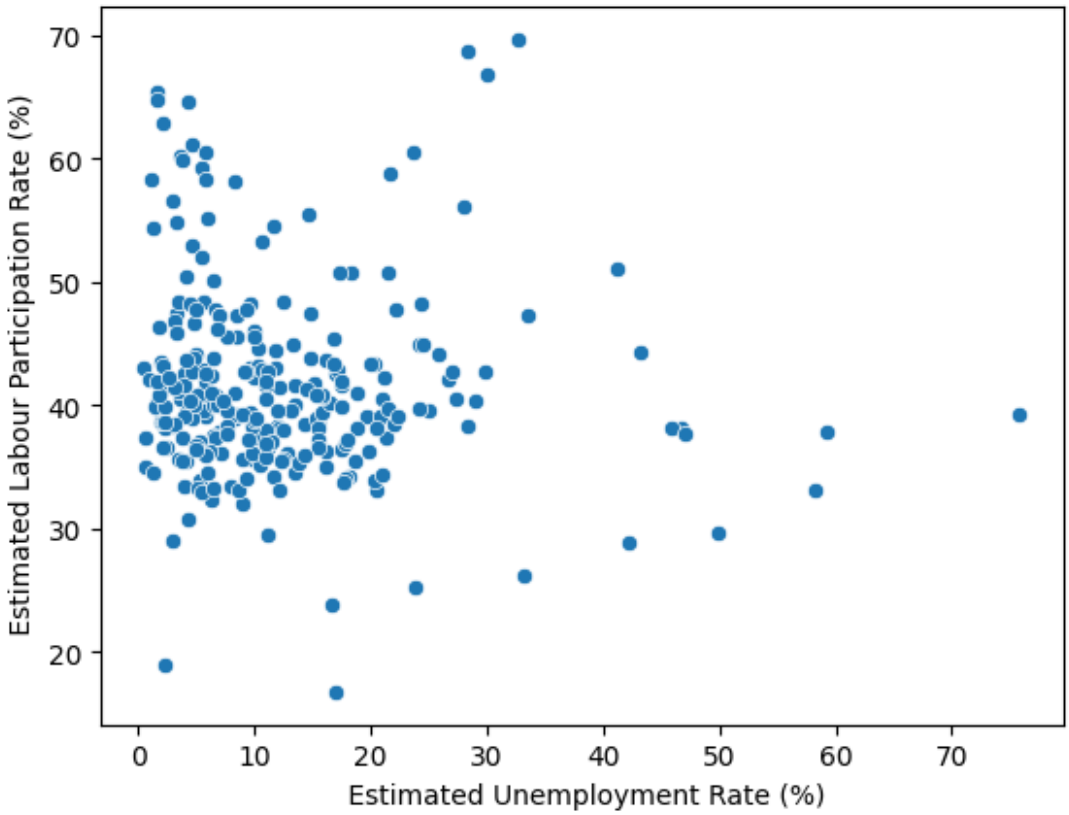
```
<Axes: xlabel='Region', ylabel=' Estimated Unemployment Rate (%)'>
```



ScatterPlot

```
sns.scatterplot(data=df_cleaned, x=' Estimated Unemployment Rate (%)',
y=' Estimated Labour Participation Rate (%)')
```

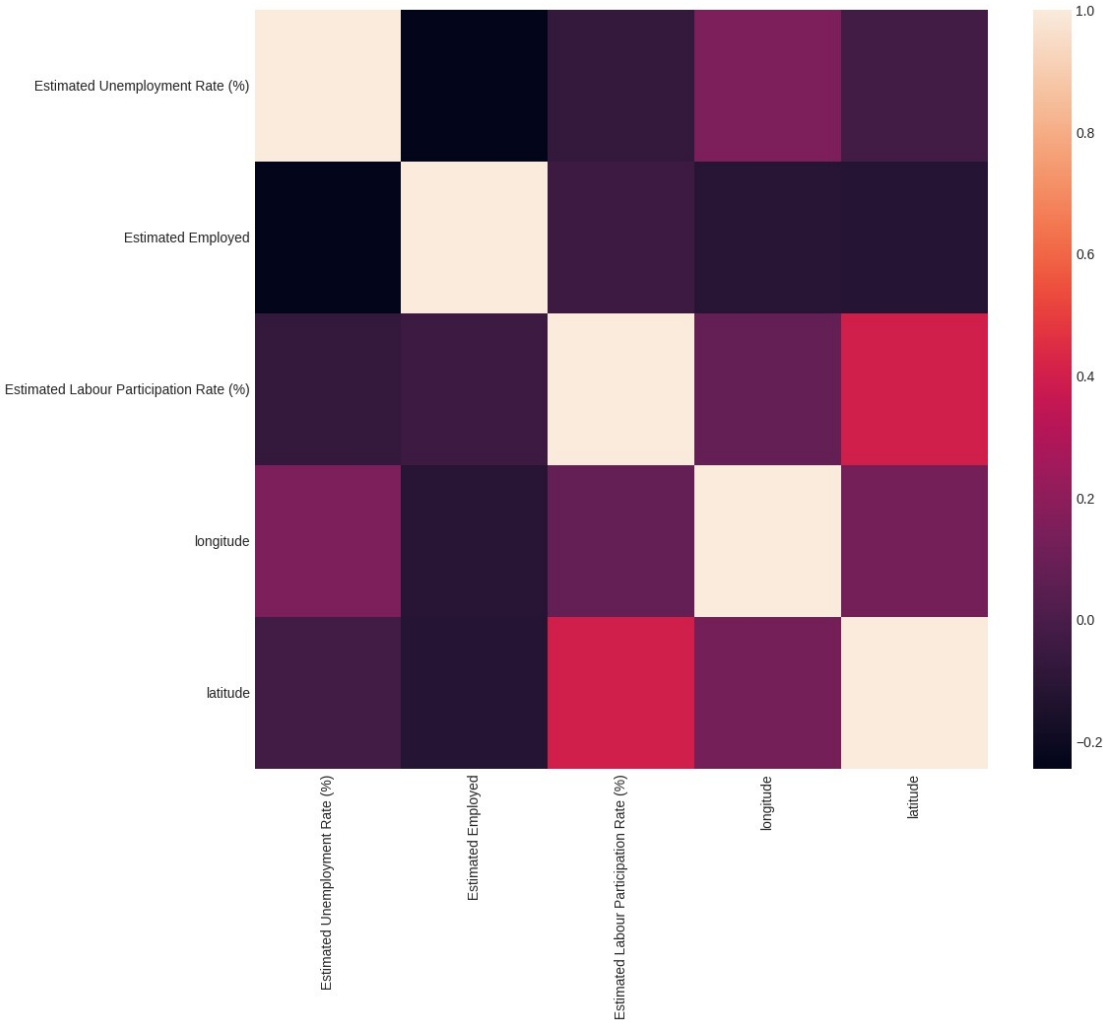
```
<Axes: xlabel=' Estimated Unemployment Rate (%)', ylabel=' Estimated
Labour Participation Rate (%)'>
```



Heat Map

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

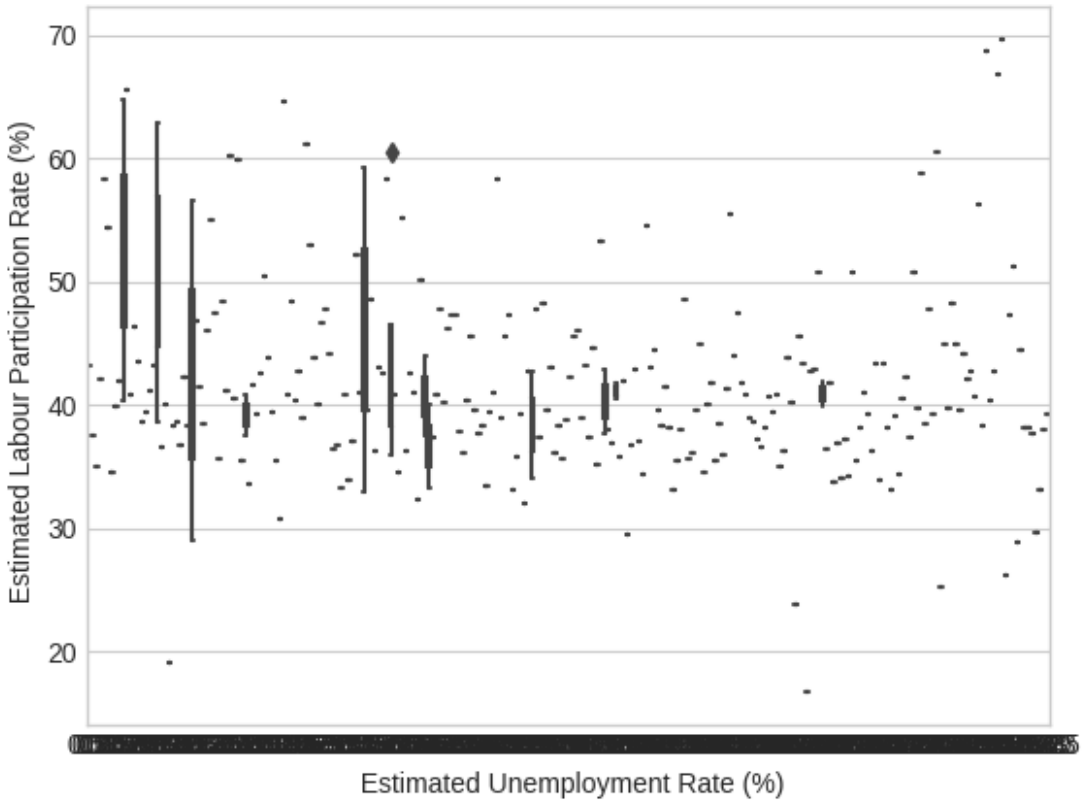
```
<ipython-input-94-9b7034756a82>:1: MatplotlibDeprecationWarning:

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 seaborn API instead.

<ipython-input-94-9b7034756a82>:3: FutureWarning:

The default value of numeric_only in DataFrame.corr is deprecated. In
a future version, it will default to False. Select only valid columns
or specify the value of numeric_only to silence this warning.
```



Box Plot

```
sns.boxplot(data=df_cleaned, x=' Estimated Unemployment Rate (%)', y=' Estimated Labour Participation Rate (%)')
```

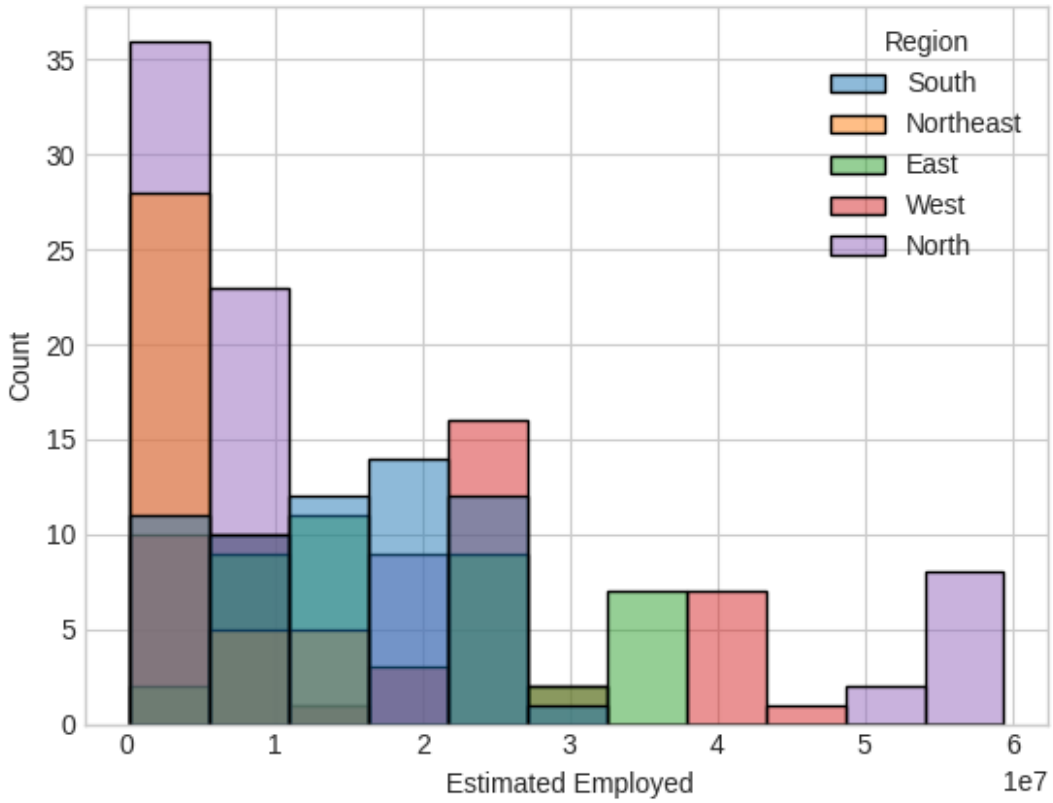
```
<Axes: xlabel=' Estimated Unemployment Rate (%)', ylabel=' Estimated Labour Participation Rate (%)'>
```



Histplot

```
sns.histplot(x=' Estimated Employed' ,hue='Region', data=df_cleaned)
```

```
<Axes: xlabel=' Estimated Employed', ylabel='Count'>
```




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

fig.show()
```

According to the Corresponding Graph North part of India has the Highest Unemployment Rate

```
fig = px.bar(df_cleaned, x='States', y=' Estimated Unemployment Rate
(%)' , color = "States",
title="Unemployment Rate By States in India")

fig.show()
```

According to the Corresponding Graph Haryana has the Highest Unemployment Rate

Box Plot

```
fig = px.box(df_cleaned, x='Region', y=' Estimated Unemployment Rate
(%)', color='Region')

fig.show()

fig = px.box(df_cleaned, x='States', y=' Estimated Unemployment Rate
(%)', color="States")
```

```
fig.show()
```

Histogram

```
fig = px.histogram(df_cleaned, x='Region', y=' Estimated Unemployment
Rate (%)', color='Region')

fig.show()

fig = px.histogram(df_cleaned, x='States', y=' Estimated Unemployment
Rate (%)', color='States')

fig.show()
```

End of the Code