UNEMPLOYMENT ANALYSIS WITH PYTHON

May 15, 2023

1 UNEMPLOYMENT ANALYSIS WITH PYTHON

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

[5]:
```
data=pd.read_csv("C:/Users/MyPc/Downloads/task 2/Unemployment in India.csv")
data
```

[5]:

	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]

[6]:
```
data.tail(10)
```

[6]:
```
     Region  Date  Frequency   Estimated Unemployment Rate (%)  \
758    NaN   NaN        NaN                                NaN
759    NaN   NaN        NaN                                NaN
760    NaN   NaN        NaN                                NaN
761    NaN   NaN        NaN                                NaN
762    NaN   NaN        NaN                                NaN
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
758                  NaN                                       NaN  NaN
759                  NaN                                       NaN  NaN
760                  NaN                                       NaN  NaN
761                  NaN                                       NaN  NaN
762                  NaN                                       NaN  NaN
763                  NaN                                       NaN  NaN
764                  NaN                                       NaN  NaN
765                  NaN                                       NaN  NaN
766                  NaN                                       NaN  NaN
767                  NaN                                       NaN  NaN
```

[7]:
```
new=pd.DataFrame({'null':(data.isnull().sum()),'no_unique':(data.
↪nunique()),'data_type':(data.dtypes)})
new
```

[7]:
```
                                          null  no_unique data_type
Region                                      28         28    object
 Date                                       28         14    object
 Frequency                                  28          2    object
 Estimated Unemployment Rate (%)            28        624   float64
 Estimated Employed                         28        740   float64
 Estimated Labour Participation Rate (%)    28        626   float64
Area                                        28          2    object
```

[8]:
```
data=data.dropna()
data.tail()
```

[8]:
```
          Region         Date  Frequency   Estimated Unemployment Rate (%)  \
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
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
```

[9]:
```
data.columns
```

[9]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

[10]:
```
type(data[' Date'])
```

[10]:
```
pandas.core.series.Series
```

[11]:
```
import datetime
data[' Date']= pd.to_datetime(data[' Date'])
```

```
C:\Users\MyPc\AppData\Local\Temp\ipykernel_8804\1075674774.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data[' Date']= pd.to_datetime(data[' Date'])
```

[12]:
```
data[' Date'].info()
```

```
<class 'pandas.core.series.Series'>
Int64Index: 740 entries, 0 to 753
Series name:  Date
Non-Null Count  Dtype
--------------  -----
740 non-null    datetime64[ns]
dtypes: datetime64[ns](1)
memory usage: 11.6 KB
```

[13]:
```
data=data.set_index(' Date')
data
```

[13]:
```
                    Region  Frequency   Estimated Unemployment Rate (%)  \
 Date
2019-05-31  Andhra Pradesh    Monthly                              3.65
```

```
2019-06-30  Andhra Pradesh    Monthly                                3.05
2019-07-31  Andhra Pradesh    Monthly                                3.75
2019-08-31  Andhra Pradesh    Monthly                                3.32
2019-09-30  Andhra Pradesh    Monthly                                5.17
…                      …        …                                    …
2020-02-29     West Bengal    Monthly                                7.55
2020-03-31     West Bengal    Monthly                                6.67
2020-04-30     West Bengal    Monthly                               15.63
2020-05-31     West Bengal    Monthly                               15.22
2020-06-30     West Bengal    Monthly                                9.86

             Estimated Employed   Estimated Labour Participation Rate (%)  \
 Date
2019-05-31           11999139.0                                      43.24
2019-06-30           11755881.0                                      42.05
2019-07-31           12086707.0                                      43.50
2019-08-31           12285693.0                                      43.97
2019-09-30           12256762.0                                      44.68
…                           …                                       …
2020-02-29           10871168.0                                      44.09
2020-03-31           10806105.0                                      43.34
2020-04-30            9299466.0                                      41.20
2020-05-31            9240903.0                                      40.67
2020-06-30            9088931.0                                      37.57

             Area
 Date
2019-05-31  Rural
2019-06-30  Rural
2019-07-31  Rural
2019-08-31  Rural
2019-09-30  Rural
…           …
2020-02-29  Urban
2020-03-31  Urban
2020-04-30  Urban
2020-05-31  Urban
2020-06-30  Urban

[740 rows x 6 columns]
```

[14]:
```python
from sklearn.preprocessing import LabelEncoder

# create a LabelEncoder object
le = LabelEncoder()

# fit and transform the 'category' column using LabelEncoder
```

```
data['Region'] = le.fit_transform(data['Region'])
data['Area'] = le.fit_transform(data['Area'])
data[' Frequency'] = le.fit_transform(data[' Frequency'])
# print the result
print(data)
```

```
            Region   Frequency   Estimated Unemployment Rate (%)  \
 Date
2019-05-31       0           0                              3.65
2019-06-30       0           0                              3.05
2019-07-31       0           0                              3.75
2019-08-31       0           0                              3.32
2019-09-30       0           0                              5.17
…              …           …                              …
2020-02-29      27           1                              7.55
2020-03-31      27           1                              6.67
2020-04-30      27           1                             15.63
2020-05-31      27           1                             15.22
2020-06-30      27           1                              9.86

             Estimated Employed   Estimated Labour Participation Rate (%)  \
 Date
2019-05-31           11999139.0                                     43.24
2019-06-30           11755881.0                                     42.05
2019-07-31           12086707.0                                     43.50
2019-08-31           12285693.0                                     43.97
2019-09-30           12256762.0                                     44.68
…                           …                                     …
2020-02-29           10871168.0                                     44.09
2020-03-31           10806105.0                                     43.34
2020-04-30            9299466.0                                     41.20
2020-05-31            9240903.0                                     40.67
2020-06-30            9088931.0                                     37.57

            Area
 Date
2019-05-31     0
2019-06-30     0
2019-07-31     0
2019-08-31     0
2019-09-30     0
…            …
2020-02-29     1
2020-03-31     1
2020-04-30     1
2020-05-31     1
2020-06-30     1
```

```
[740 rows x 6 columns]
```

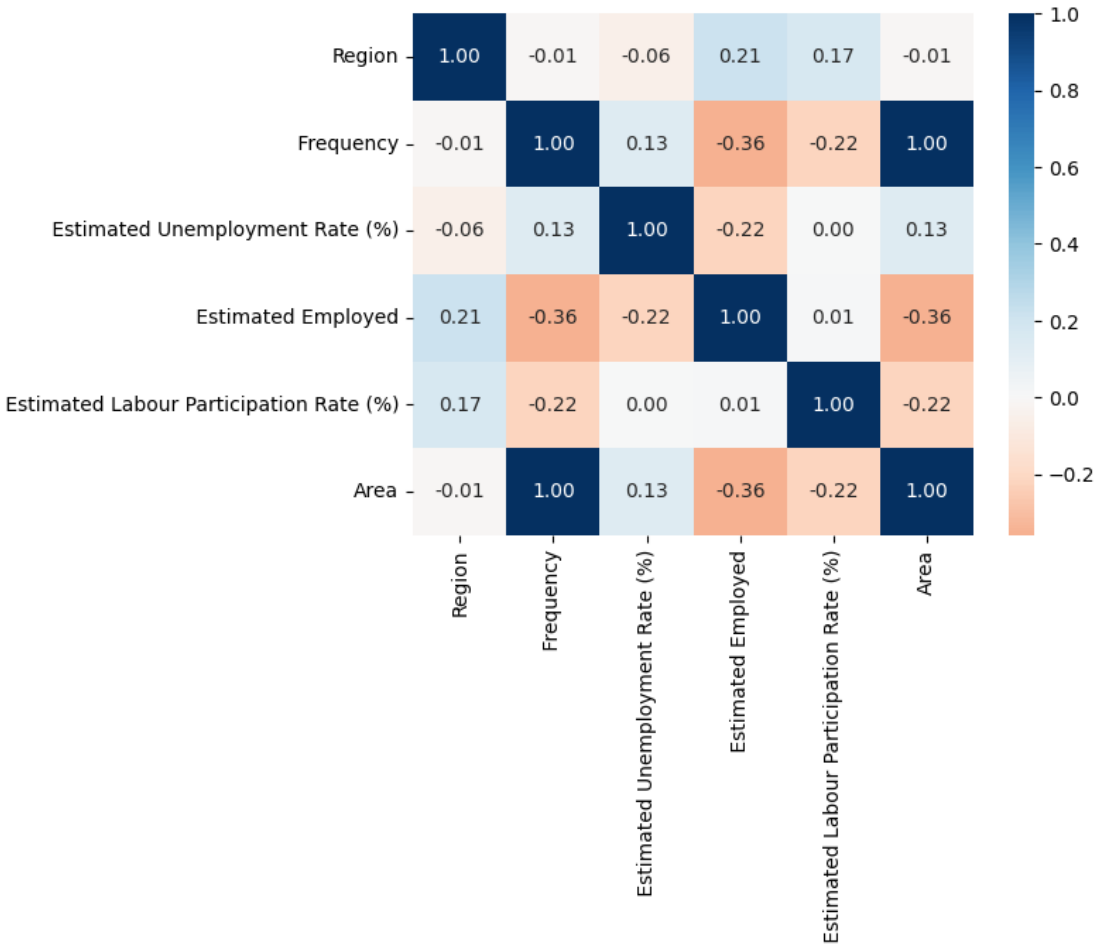
[15]:
```
data[' Frequency']
```

[15]:
```
 Date
2019-05-31    0
2019-06-30    0
2019-07-31    0
2019-08-31    0
2019-09-30    0
              ..
2020-02-29    1
2020-03-31    1
2020-04-30    1
2020-05-31    1
2020-06-30    1
Name:  Frequency, Length: 740, dtype: int32
```

[19]:
```
import seaborn as sns
# Calculate the correlation matrix
corr_matrix = data.corr()

# Create a heatmap using Seaborn
sns.heatmap(corr_matrix, cmap='RdBu', center=0, annot=True, fmt='.2f')
```

[19]:
```
<AxesSubplot:>
```



[22]:
```
data2=pd.read_csv("C:/Users/MyPc/Downloads/task 2/
↪Unemployment_Rate_upto_11_2020.csv")
data2.head()
```

[22]:

	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

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

[23]:
```
new2=pd.DataFrame({'null':(data2.isnull().sum()),'no_unique':(data2.
 ↪nunique()),'data_type':(data2.dtypes)})
new2
```

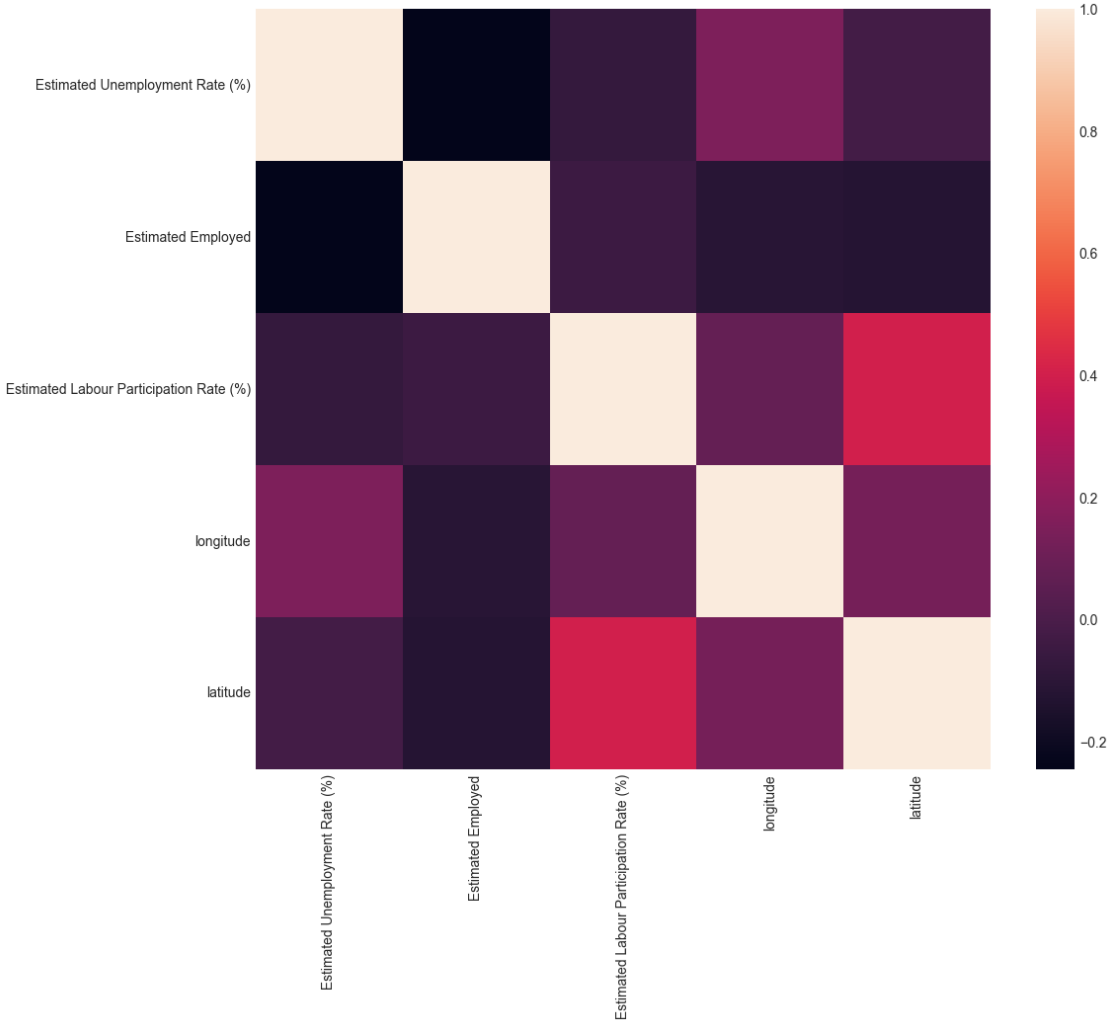
[23]:
```
                                        null  no_unique data_type
Region                                     0         27    object
 Date                                      0         10    object
 Frequency                                 0          1    object
 Estimated Unemployment Rate (%)           0        252   float64
 Estimated Employed                        0        267     int64
 Estimated Labour Participation Rate (%)   0        248   float64
Region.1                                   0          5    object
longitude                                  0         27   float64
latitude                                   0         24   float64
```

[24]:
```
data2.columns
```

[24]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

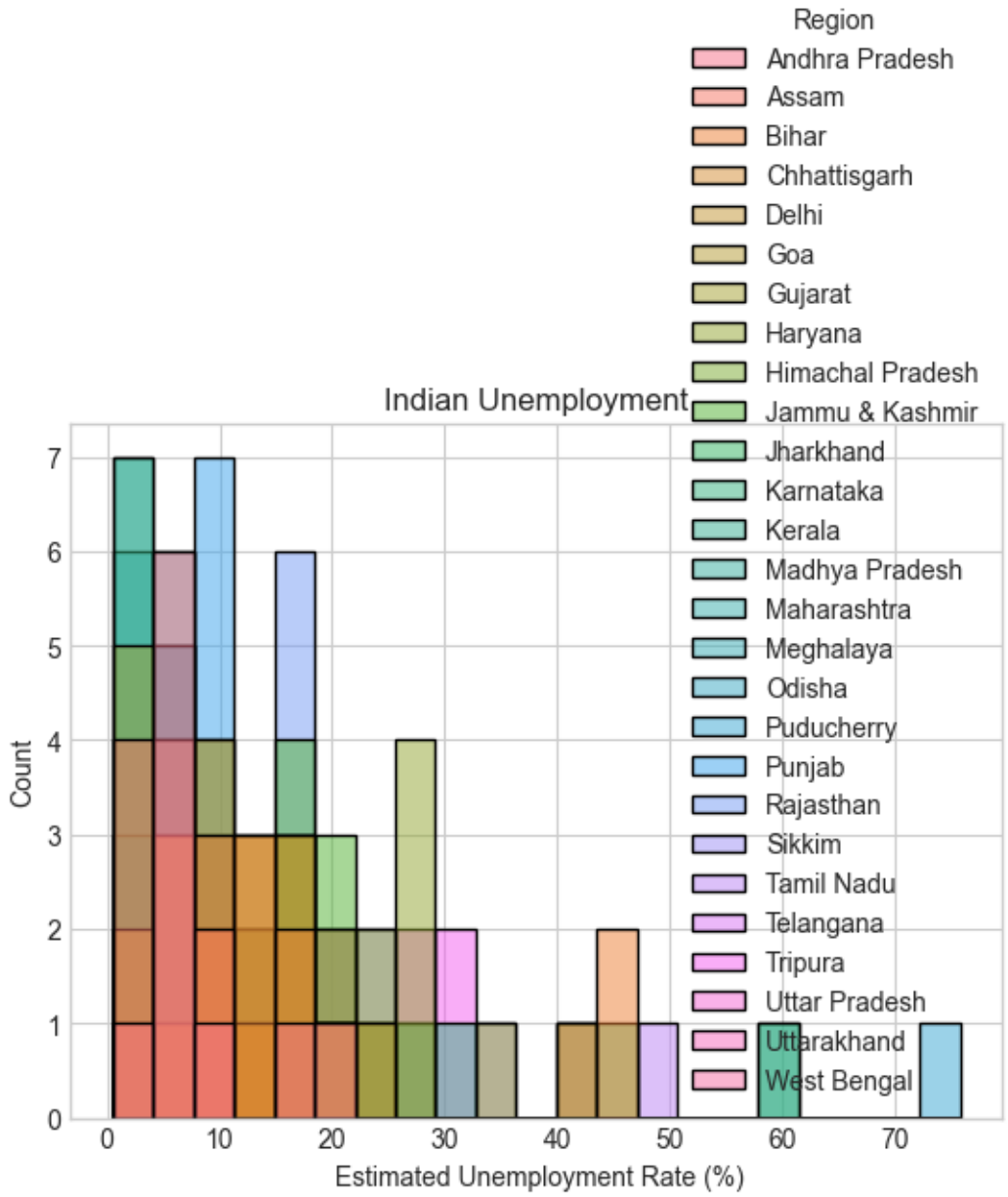
[25]:
```
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(12, 10))
sns.heatmap(data2.corr())
plt.show()
```

```
C:\python37\lib\site-packages\matplotlib\animation.py:887: UserWarning:
Animation was deleted without rendering anything. This is most likely not
intended. To prevent deletion, assign the Animation to a variable, e.g. `anim`,
that exists until you have outputted the Animation using `plt.show()` or
`anim.save()`.
  warnings.warn(
```



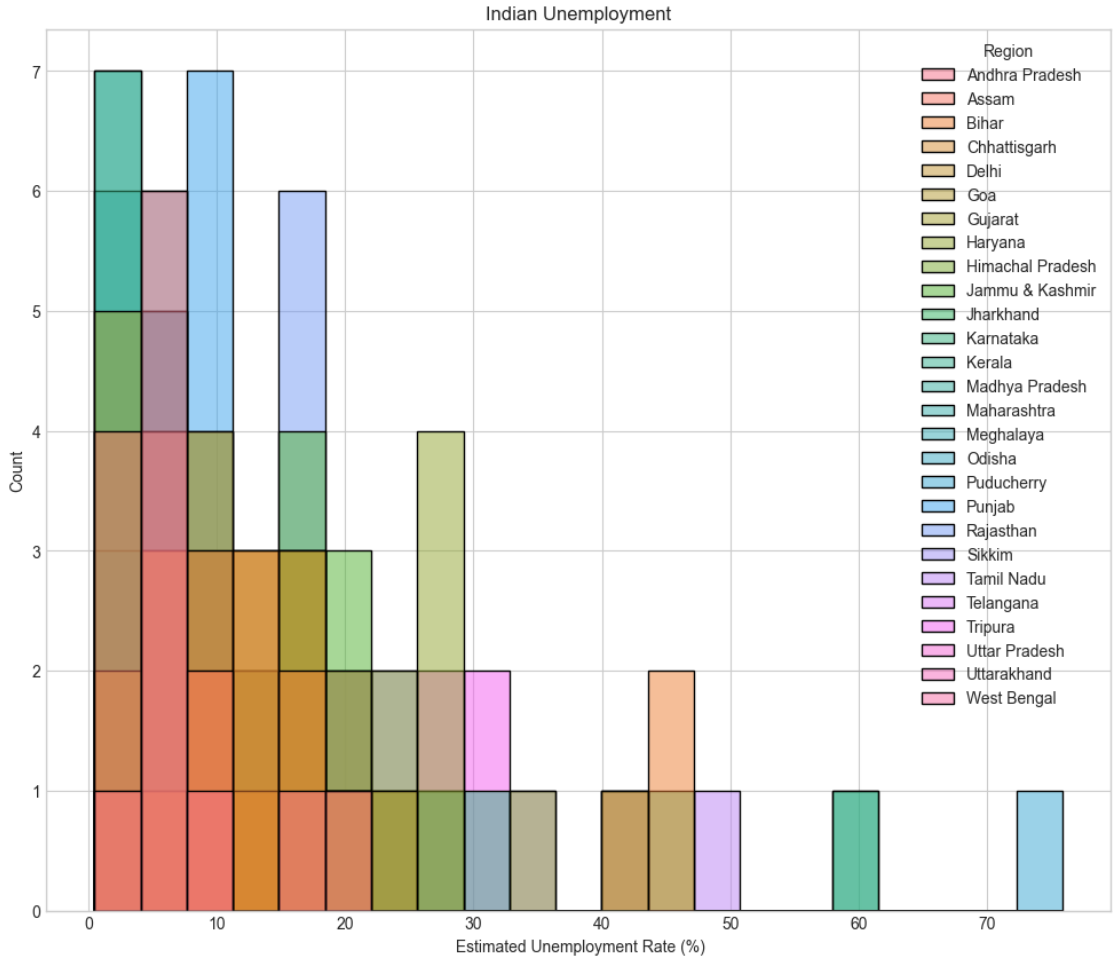
2 Unemployment Rate Analysis: Data Visualization

[26]:
```
plt.title("Indian Unemployment")
sns.histplot(x=' Estimated Unemployment Rate (%)', hue="Region", data=data2)
plt.show()
```



[27]:
```
print(data2.columns)
plt.figure(figsize=(12, 10))
plt.title("Indian Unemployment")
sns.histplot(x=' Estimated Unemployment Rate (%)', hue="Region", data=data2)
plt.show()
```

```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```



[28]:
```
print(data2.columns)
unemploment = data2[['Region.1', "Region", ' Estimated Unemployment Rate (%)']]
figure = px.sunburst(unemploment, path=['Region.1',"Region"],
                     values=' Estimated Unemployment Rate (%)',
                     width=700, height=700, color_continuous_scale="RdY1Gn",
                     title="Unemployment Rate in India")
figure.show()
```

```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

[29]:
```
# Calculate the correlation matrix
corr_matrix = data2.corr()

# Create a heatmap using Seaborn
```

```
sns.heatmap(corr_matrix, cmap='RdBu', center=0, annot=True, fmt='.2f')
```

[29]: <AxesSubplot:>

3 Summary

So this is how you can analyze the unemployment rate by using the Python programming language. Unemployment is measured by the unemployment rate which is the number of people who are unemployed as a percentage of the total labour force. I hope you liked this article on unemployment rate analysis with Python. Feel free to ask your valuable questions in the comments section below.