UNEMPLOYMENT ANALYSIS IN INDIA DURING COVID PANDAMIC

About the Dataset

This datasets give informations about how lock-down affects employment opportunities and how the unemployment rate increases during the Covid-19.

The unemployment rate experienced a significant rise during the period of the COVID-19 pandemic.

The aim is to analyze the unemployment rate using Machine Learning.

Features

Region: Various states constituting the Indian subcontinent(States)

Date: The specific dates of unemployment rate recordings

Frequency: The regularity of measurement collection (Monthly)

Estimated Unemployment Rate (%): The proportion of unemployed individuals in each Indian state

Estimated Employed: The tally of presently engaged individuals

Estimated Labour Participation Rate (%): The percentage of the working-age population (16-64 years) actively involved in the job market.

Region.1: Regions like North,South,East,West

Longitude:Longitude of each region

Latitude:Latitude of each region

Area: areas like rural and urban

Importing Python Library

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
import calendar
```

_

```
#Loading Unemployment_Rate_upto_11_2020.csv dataset
df=pd.read_csv('/home/anusha/Downloads/Unemployment_Rate_upto_11_2020.
csv')
df
```

```
              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]
```

Data Preprocessing

```
df.head()
```

```
            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

    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

   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
```

_

```
df.tail()
```

```
           Region         Date  Frequency   Estimated Unemployment Rate
(%)  \
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  \
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
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
```

_

```
df.shape
```

```
(267, 9)
```

_

```
df.columns
```

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

_

```
#Printing datatypes
df.dtypes
```

```
Region                                      object
 Date                                       object
 Frequency                                  object
 Estimated Unemployment Rate (%)           float64
 Estimated Employed                          int64
 Estimated Labour Participation Rate (%)   float64
Region.1                                    object
longitude                                  float64
latitude                                   float64
dtype: object
```

_

```
#Checking for missing values
df.isna().sum()
```

```
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
```

_

```
#Checking for null values
df.isin(['null']).sum()
```

```
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
```

_

```
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
```

_

```
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
```

_

```
df.describe(include= 'object')
```

```
                Region         Date  Frequency Region.1
count              267          267        267      267
unique              27           10          1        5
top     Andhra Pradesh   31-03-2020          M    North
freq                10           27        267       79
```

_

```
#Checking for duplicates
df.duplicated().any()
```

```
False
```

_

```
df.nunique()
```

```
Region                                      27
 Date                                       10
 Frequency                                   1
 Estimated Unemployment Rate (%)           252
 Estimated Employed                        267
 Estimated Labour Participation Rate (%)   248
Region.1                                     5
longitude                                   27
latitude                                    24
dtype: int64
```

_

```
# Dropping unwanted columns
df.drop(' Frequency',axis=1,inplace=True)
```

_

```
#Renaming some features
df.columns = ['state','date','estimated unemployment rate','estimated
employed','estimated labour participation
rate','region','longitude','latitude']
df.head()
```

```
            state         date  estimated unemployment rate  \
0  Andhra Pradesh   31-01-2020                          5.48
1  Andhra Pradesh   29-02-2020                          5.83
2  Andhra Pradesh   31-03-2020                          5.79
3  Andhra Pradesh   30-04-2020                         20.51
4  Andhra Pradesh   31-05-2020                         17.43

   estimated employed  estimated labour participation rate region
longitude  \
0            16635535                                41.02  South
15.9129
1            16545652                                40.90  South
15.9129
2            15881197                                39.18  South
15.9129
3            11336911                                33.10  South
15.9129
4            12988845                                36.46  South
15.9129

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

_

```
df['month_int'] = pd.to_datetime(df['date']).dt.month
```

```
/tmp/ipykernel_6006/1883822796.py:1: UserWarning: Parsing dates in
%d-%m-%Y format when dayfirst=False (the default) was specified. Pass
`dayfirst=True` or specify a format to silence this warning.
  df['month_int'] = pd.to_datetime(df['date']).dt.month
```

_

```
df['month'] = df['month_int'].apply(lambda x: calendar.month_name[x])
```

_

```
df.head()
```

```
            state         date  estimated unemployment rate  \
0  Andhra Pradesh   31-01-2020                          5.48
1  Andhra Pradesh   29-02-2020                          5.83
2  Andhra Pradesh   31-03-2020                          5.79
3  Andhra Pradesh   30-04-2020                         20.51
4  Andhra Pradesh   31-05-2020                         17.43

   estimated employed  estimated labour participation rate region
longitude  \
0            16635535                                   41.02  South
15.9129
1            16545652                                   40.90  South
15.9129
2            15881197                                   39.18  South
15.9129
3            11336911                                   33.10  South
15.9129
4            12988845                                   36.46  South
15.9129

   latitude  month_int     month
0     79.74          1   January
1     79.74          2  February
2     79.74          3     March
3     79.74          4     April
4     79.74          5       May
```

DATA VISUALIZATION

```
df['estimated unemployment rate'].value_counts()
```

```
estimated unemployment rate
5.79     4
2.86     3
1.58     2
10.97    2
10.61    2
        ..
21.08    1
20.78    1
15.50    1
18.74    1
9.98     1
Name: count, Length: 252, dtype: int64
```

_

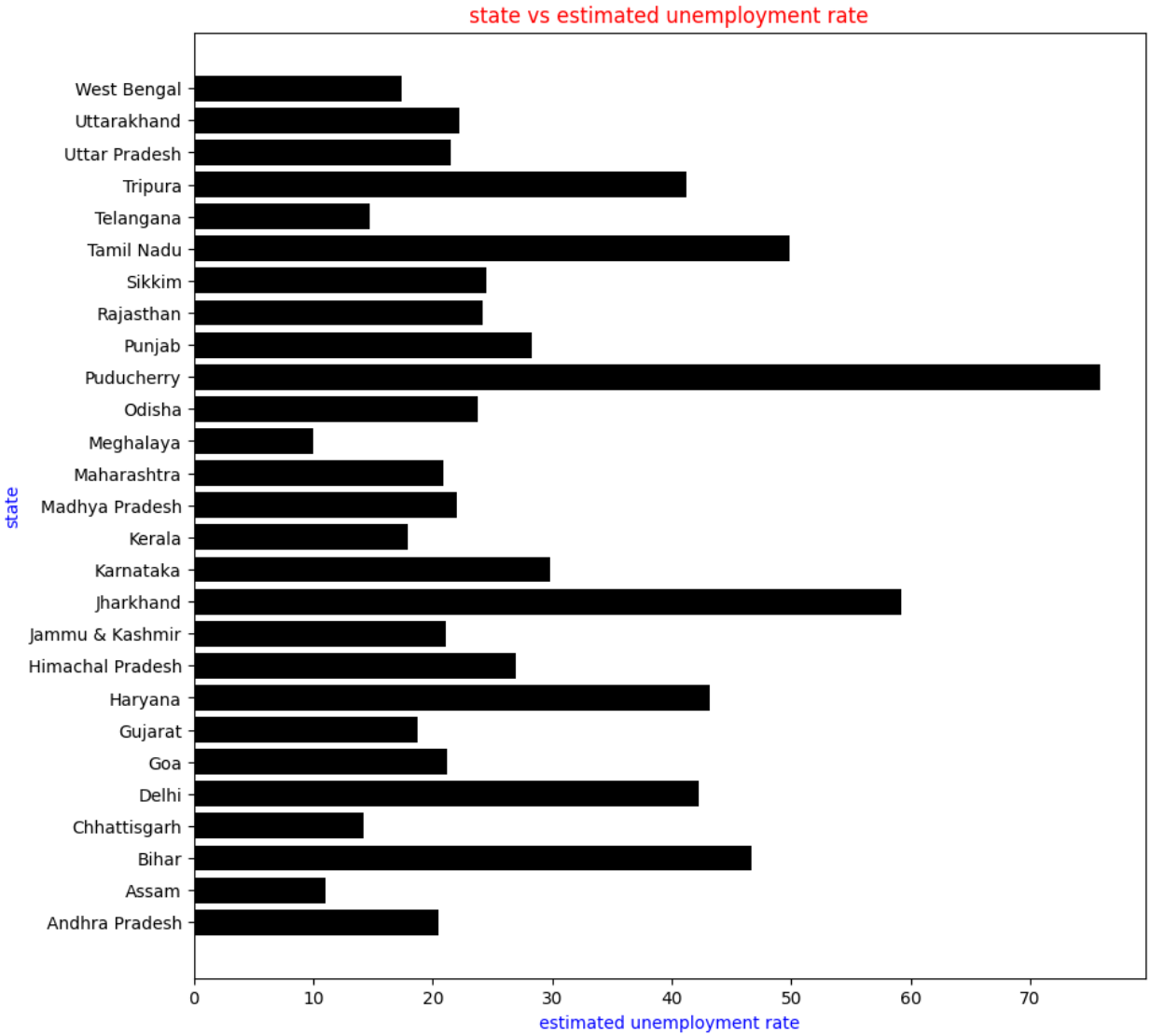
```
df['state'].value_counts()
```

```
state
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: count, dtype: int64
```

_

```
plt.figure(figsize=(10,10))
plt.barh(df['state'],df['estimated unemployment rate'],color='black')
plt.xlabel('estimated unemployment rate',color='blue')
plt.ylabel('state',color='blue')
plt.title('state vs estimated unemployment rate',color='red')
```

```
Text(0.5, 1.0, 'state vs estimated unemployment rate')
```



Estimated unemployment rate is very high in Puducherry and very low in Meghalaya, compared to other states

```
df.head()
```

```
            state         date  estimated unemployment rate  \
0  Andhra Pradesh   31-01-2020                          5.48
1  Andhra Pradesh   29-02-2020                          5.83
2  Andhra Pradesh   31-03-2020                          5.79
3  Andhra Pradesh   30-04-2020                         20.51
4  Andhra Pradesh   31-05-2020                         17.43

   estimated employed  estimated labour participation rate region
longitude  \
0            16635535                                41.02  South
15.9129
```

```
1            16545652                                  40.90  South
15.9129
2            15881197                                  39.18  South
15.9129
3            11336911                                  33.10  South
15.9129
4            12988845                                  36.46  South
15.9129

   latitude  month_int     month
0     79.74          1   January
1     79.74          2  February
2     79.74          3     March
3     79.74          4     April
4     79.74          5       May
```

_

```
df['estimated employed'].value_counts()
```

```
estimated employed
16635535    1
6872938     1
457950      1
493961      1
421028      1
           ..
3202336     1
3558889     1
3429950     1
3210281     1
33962549    1
Name: count, Length: 267, dtype: int64
```

_

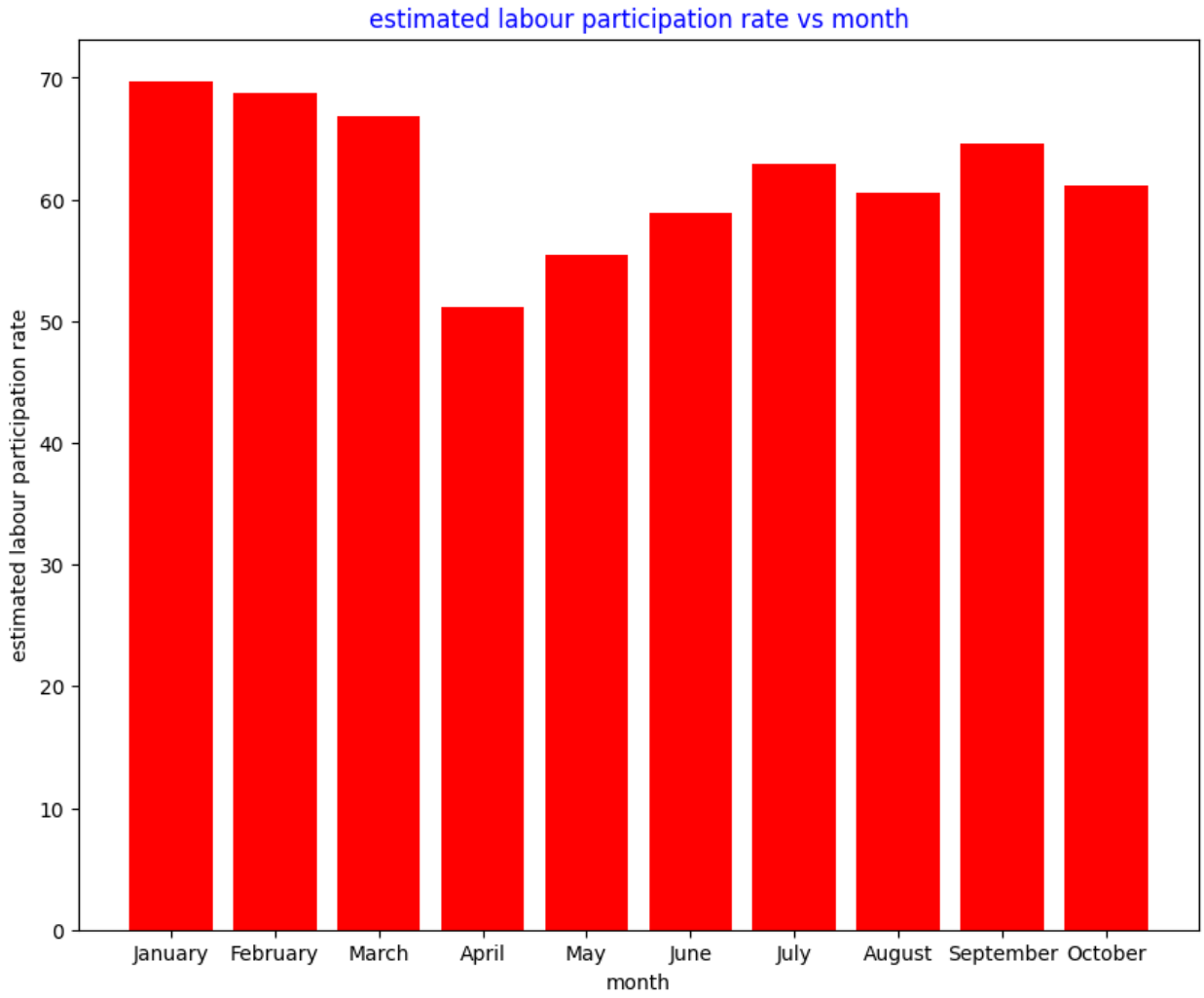
```
df['estimated labour participation rate'].value_counts()
```

```
estimated labour participation rate
38.48    2
40.88    2
60.59    2
40.39    2
39.20    2
        ..
38.03    1
36.31    1
35.05    1
42.92    1
```

```
45.63    1
Name: count, Length: 248, dtype: int64
```

_

```
plt.figure(figsize=(10,8))
plt.bar(df['month'],df['estimated labour participation
rate'],color='red')
plt.xlabel('month',color='black')
plt.ylabel('estimated labour participation rate',color='black')
plt.title('estimated labour participation rate vs month',color='blue')

Text(0.5, 1.0, 'estimated labour participation rate vs month')
```

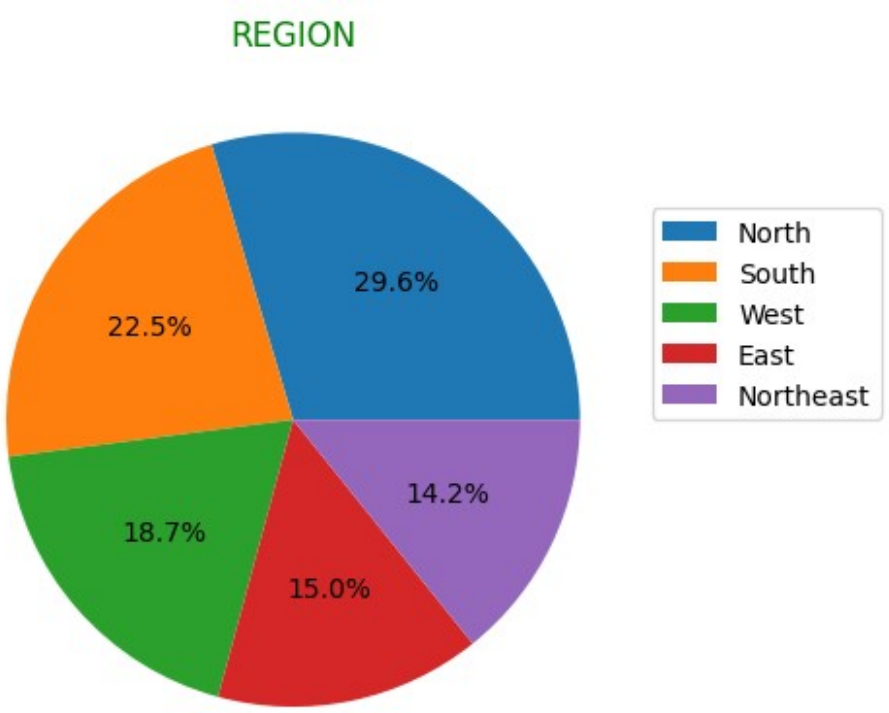



```
df['region'].value_counts()

region
North        79
```

```
South         60
West          50
East          40
Northeast     38
Name: count, dtype: int64
```

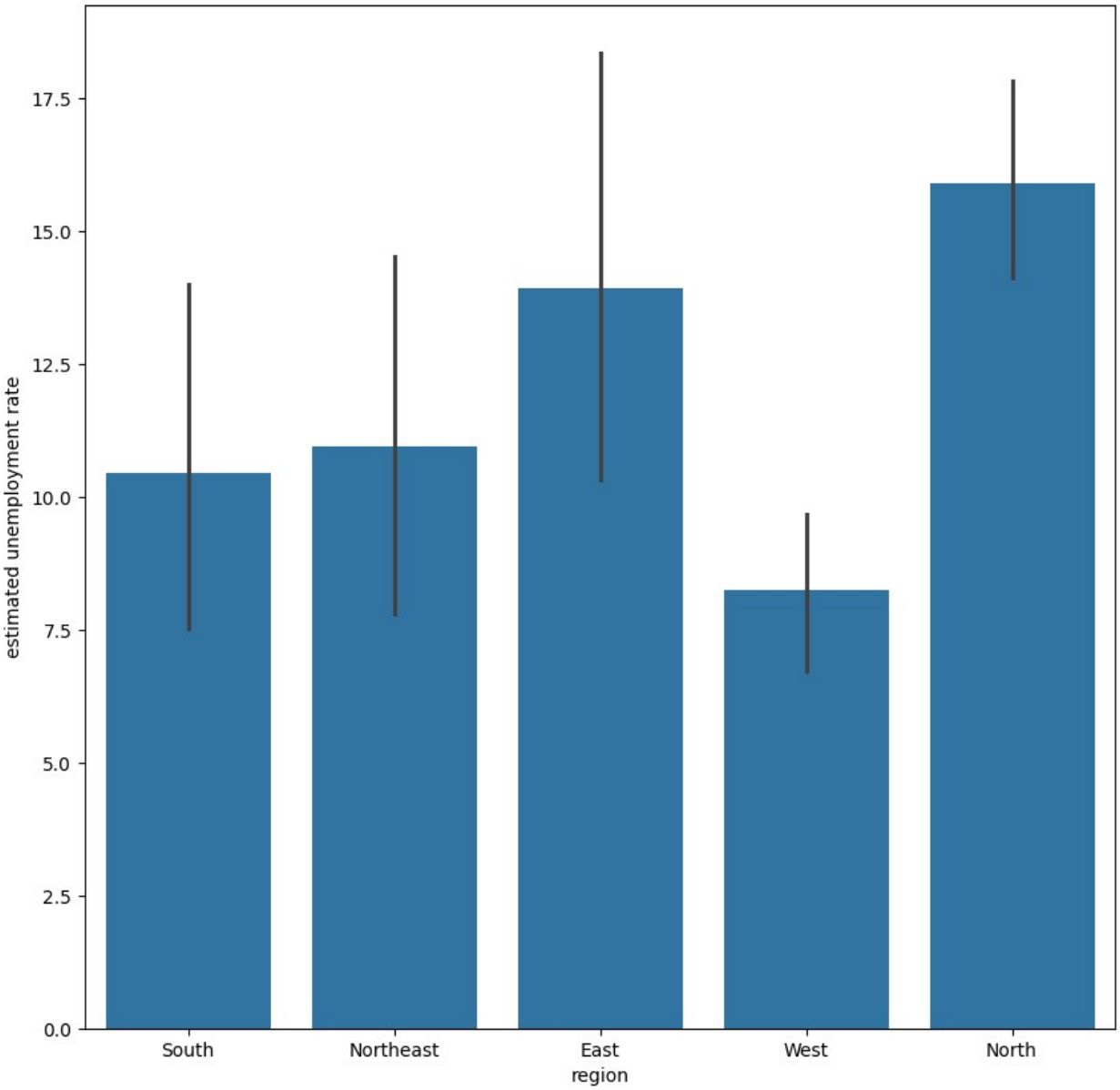
_

```
plt.pie(df['region'].value_counts(),autopct='%1.1f%%')
plt.legend(df['region'].value_counts().index,loc=(1,0.5))
plt.title('REGION',color='green')

Text(0.5, 1.0, 'REGION')
```



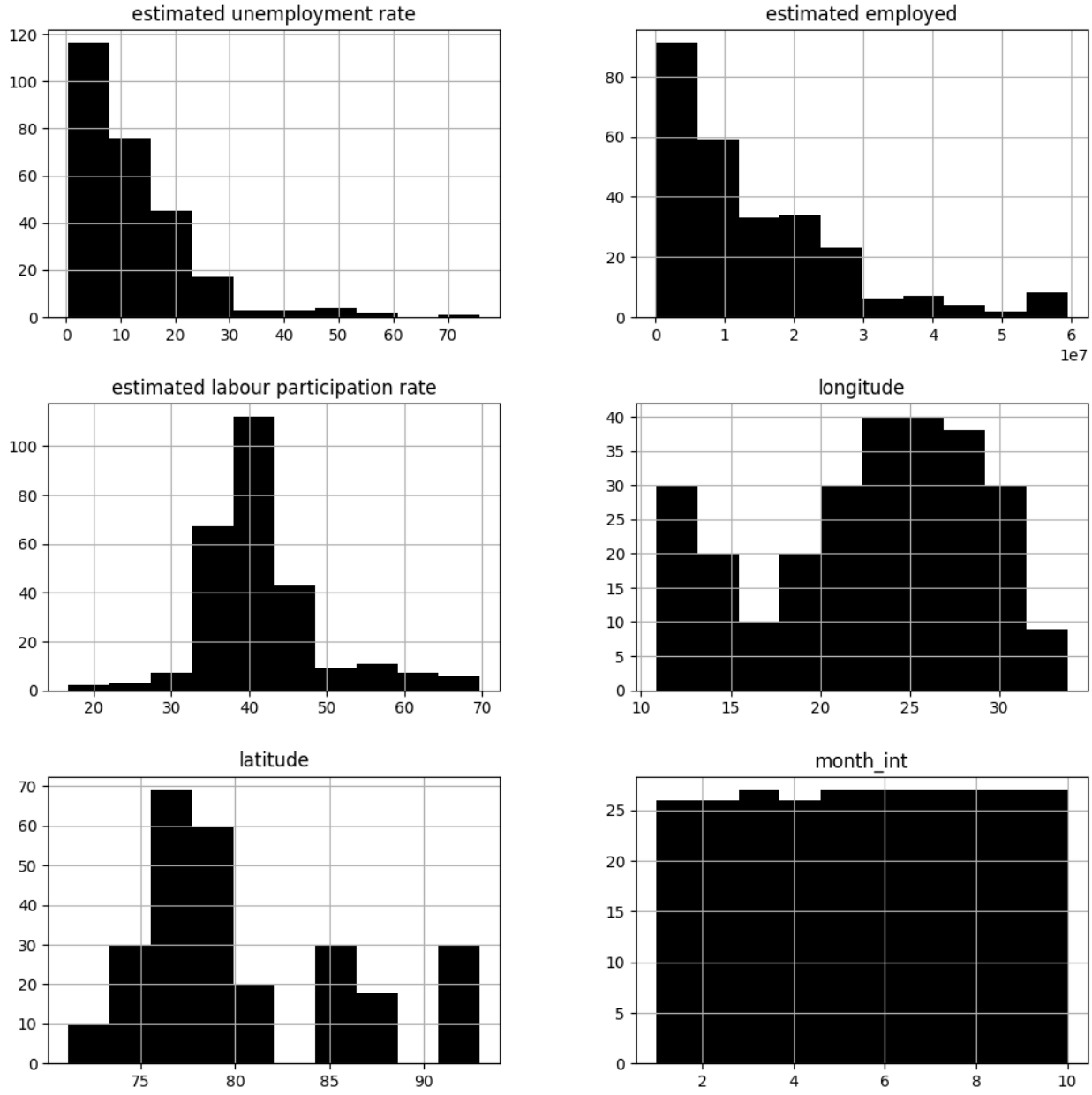
_

```
#Graph of estimated unemployment rate vs region
plt.figure(figsize=(10,10))
sns.barplot(x='region',y='estimated unemployment rate',data=df)

<Axes: xlabel='region', ylabel='estimated unemployment rate'>
```



—

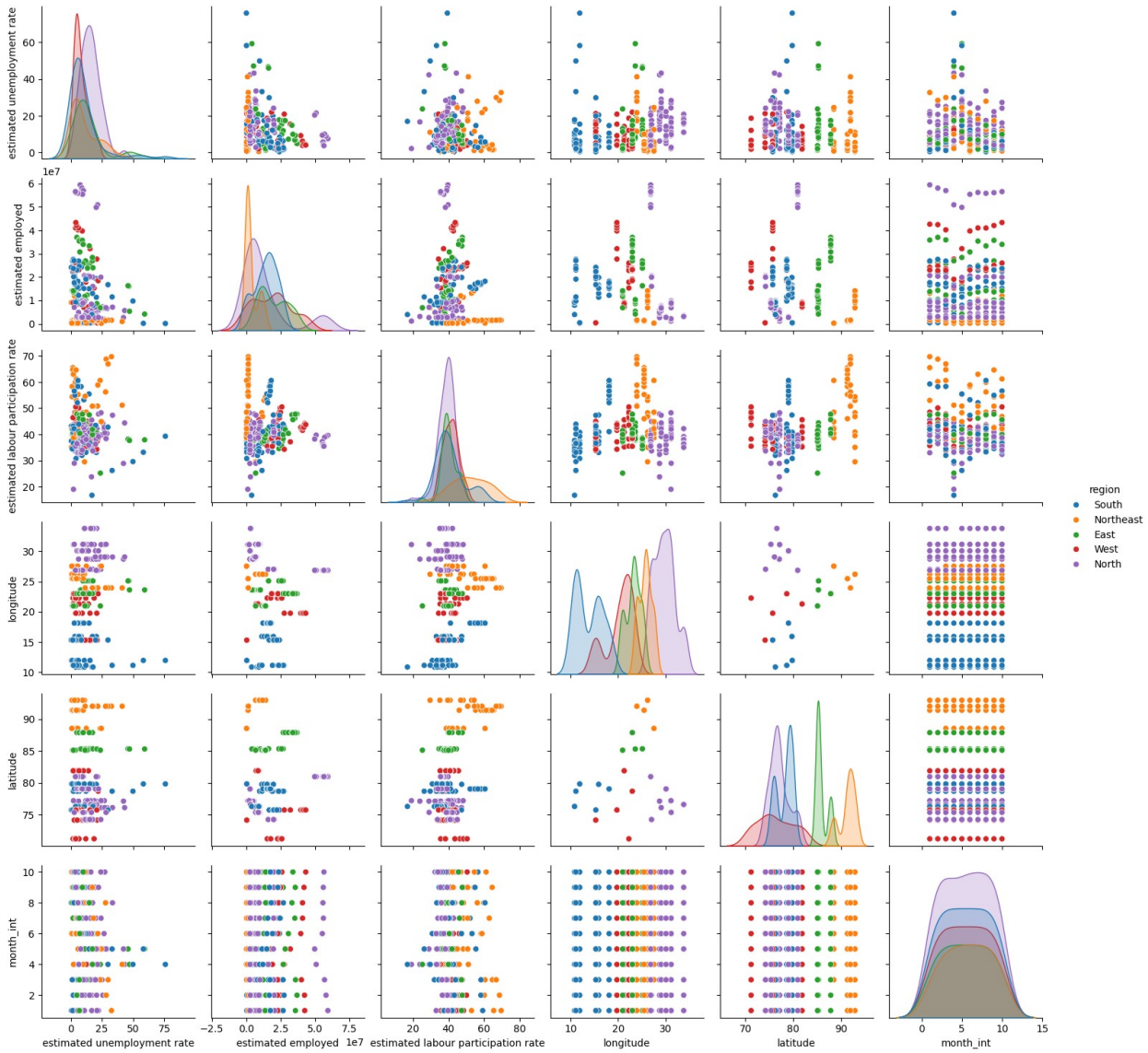
```
df.hist(figsize=(12,12),color='black')
plt.show()
```



—

```
sns.pairplot(data=df,hue='region')
```

```
<seaborn.axisgrid.PairGrid at 0x7f1d305fc280>
```



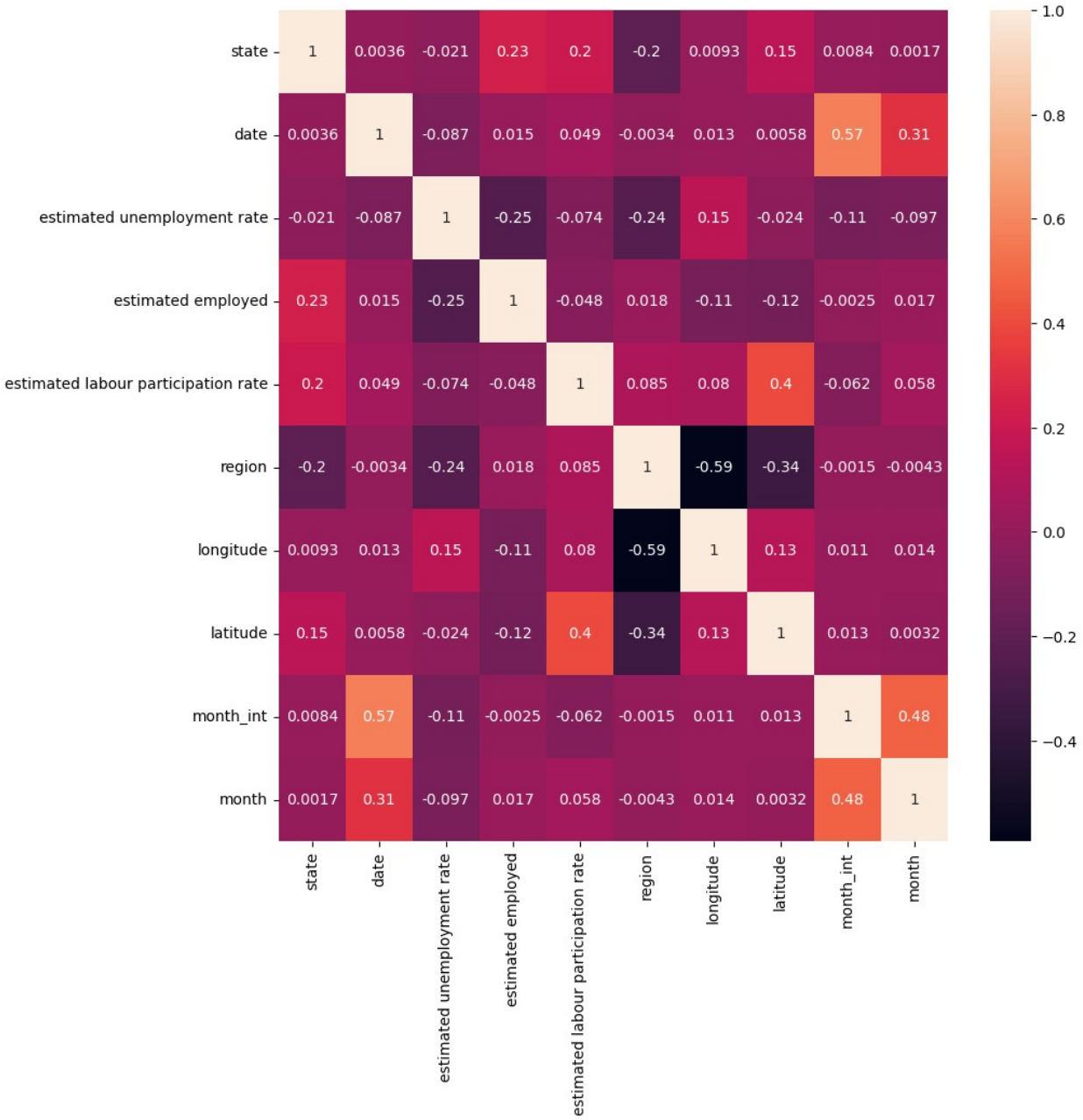
—

```
#Label encoding
from sklearn.preprocessing import LabelEncoder
lab=LabelEncoder()
for i in df.columns:
  if df[i].dtype=='object':
    df[i]=lab.fit_transform(df[i])

#Correlation of data
df.corr()
```

```
                                      state      date  \
state                              1.000000  0.003568
date                               0.003568  1.000000
estimated unemployment rate       -0.020823 -0.087194
```

```
estimated employed                    0.234531  0.014531
estimated labour participation rate  0.197804  0.049004
region                               -0.201465 -0.003375
longitude                             0.009293  0.012913
latitude                              0.145440  0.005847
month_int                             0.008432  0.573959
month                                 0.001725  0.312992

                                      estimated unemployment rate  \
state                                                    -0.020823
date                                                     -0.087194
estimated unemployment rate                               1.000000
estimated employed                                       -0.245176
estimated labour participation rate                      -0.073540
region                                                   -0.236189
longitude                                                 0.149976
latitude                                                 -0.023976
month_int                                                -0.109654
month                                                    -0.097408

                                      estimated employed  \
state                                           0.234531
date                                            0.014531
estimated unemployment rate                    -0.245176
estimated employed                              1.000000
estimated labour participation rate            -0.047948
region                                          0.018463
longitude                                      -0.113664
latitude                                       -0.119321
month_int                                      -0.002507
month                                           0.016896

                                      estimated labour participation
rate  \
state
0.197804
date
0.049004
estimated unemployment rate                                        -
0.073540
estimated employed                                                 -
0.047948
estimated labour participation rate
1.000000
region
0.085358
longitude
0.080372
latitude
0.397836
```

```
month_int                                                                -
0.061983
month
0.057862

                                        region  longitude  latitude
month_int  \
state                                -0.201465   0.009293  0.145440
0.008432
date                                 -0.003375   0.012913  0.005847
0.573959
estimated unemployment rate          -0.236189   0.149976 -0.023976  -
0.109654
estimated employed                    0.018463  -0.113664 -0.119321  -
0.002507
estimated labour participation rate   0.085358   0.080372  0.397836  -
0.061983
region                                1.000000  -0.593279 -0.337161  -
0.001468
longitude                            -0.593279   1.000000  0.125895
0.011294
latitude                             -0.337161   0.125895  1.000000
0.013002
month_int                            -0.001468   0.011294  0.013002
1.000000
month                                -0.004327   0.014204  0.003165
0.475506

                                         month
state                                 0.001725
date                                  0.312992
estimated unemployment rate          -0.097408
estimated employed                    0.016896
estimated labour participation rate   0.057862
region                               -0.004327
longitude                             0.014204
latitude                              0.003165
month_int                             0.475506
month                                 1.000000

plt.figure(figsize=(10,10))
sns.heatmap(df.corr(),annot=True)

<Axes: >
```



```
#Importing Unemployment in India.csv Dataset
df1 = pd.read_csv('/home/anusha/Desktop/Unemployment in India.csv')
df1
```

```
               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]
```

_

```
df1.head()
```

```
            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

    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
```

_

```
df1.tail()
```

```
    Region  Date  Frequency   Estimated Unemployment Rate (%)  \
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
763                  NaN                                       NaN
NaN
764                  NaN                                       NaN
NaN
765                  NaN                                       NaN
NaN
766                  NaN                                       NaN
NaN
```

```
767                    NaN                                                  NaN
NaN
```

_

```
df1.shape
```

```
(768, 7)
```

_

```
df1.columns
```

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

_

```
#Printing datatypes
df1.dtypes
```

```
Region                                    object
 Date                                     object
 Frequency                                object
 Estimated Unemployment Rate (%)         float64
 Estimated Employed                      float64
 Estimated Labour Participation Rate (%)    float64
Area                                      object
dtype: object
```

_

```
#checking for missing value
df1.isna().sum()
```

```
Region                                    28
 Date                                     28
 Frequency                                28
 Estimated Unemployment Rate (%)          28
 Estimated Employed                       28
 Estimated Labour Participation Rate (%)    28
Area                                      28
dtype: int64
```

_

```
#Checking for null values
df1.isin(['null']).sum()
```

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

_

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
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: 42.1+ KB
```

_

```
df1.nunique()
```

```
Region                                    28
 Date                                     14
 Frequency                                 2
 Estimated Unemployment Rate (%)         624
 Estimated Employed                      740
 Estimated Labour Participation Rate (%)  626
Area                                       2
dtype: int64
```

_

```
df1[' Frequency'].unique()
```

```
array([' Monthly', nan, 'Monthly'], dtype=object)
```

_

```
df1['Area'].unique()
```

```
array(['Rural', nan, 'Urban'], dtype=object)
```

_

```
df1.describe()
```

```
        Estimated Unemployment Rate (%)   Estimated Employed  \
count                        740.000000         7.400000e+02
mean                          11.787946         7.204460e+06
std                           10.721298         8.087988e+06
min                            0.000000         4.942000e+04
25%                            4.657500         1.190404e+06
50%                            8.350000         4.744178e+06
75%                           15.887500         1.127549e+07
max                           76.740000         4.577751e+07

        Estimated Labour Participation Rate (%)
count                                740.000000
mean                                  42.630122
std                                    8.111094
min                                   13.330000
25%                                   38.062500
50%                                   41.160000
75%                                   45.505000
max                                   72.570000
```

_

```
df1.describe(include='object')
```

```
                Region         Date  Frequency   Area
count              740          740        740    740
unique              28           14          2      2
top     Andhra Pradesh   31-10-2019    Monthly  Urban
freq                28           55        381    381
```

_

```
df1.tail(15)
```

```
          Region         Date  Frequency   Estimated Unemployment Rate
(%)  \
753  West Bengal   30-06-2020    Monthly
9.86
754          NaN          NaN        NaN
NaN
```

```
755           NaN           NaN         NaN
NaN
756           NaN           NaN         NaN
NaN
757           NaN           NaN         NaN
NaN
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
753            9088931.0                                    37.57
Urban
754                  NaN                                      NaN
NaN
755                  NaN                                      NaN
NaN
756                  NaN                                      NaN
NaN
757                  NaN                                      NaN
NaN
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
```

_

```
#Removing last 14 columns which containg missing values

df1=df1.iloc[:753]
df1
```

```
              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
..              ...          ...        ...
...
748     West Bengal   31-01-2020    Monthly
7.27
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

      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
..                    ...                                      ...
...
748            11208617.0                                    45.39
Urban
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 rows x 7 columns]
```

_

```
# Dropping unwanted columns
df1.drop(' Frequency',axis=1,inplace=True)
```

```
/tmp/ipykernel_6006/1482455147.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df1.drop(' Frequency',axis=1,inplace=True)
```

_

```
#Renaming some columns

df1.columns = ['state','date','estimated unemployment rate','estimated
employed','estimated labour participation rate','area']
```

_

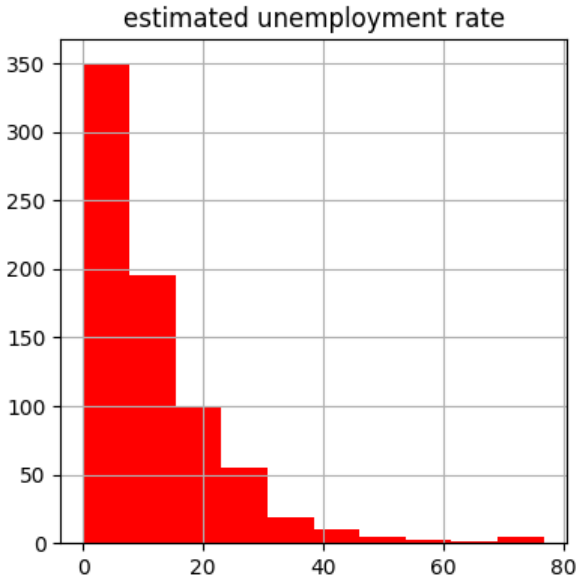
```
df1.head()
```

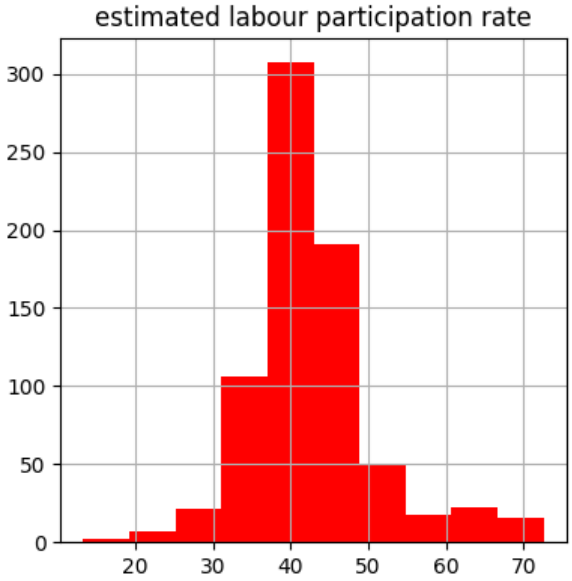
```
            state         date  estimated unemployment rate  \
0  Andhra Pradesh   31-05-2019                          3.65
1  Andhra Pradesh   30-06-2019                          3.05
2  Andhra Pradesh   31-07-2019                          3.75
```

```
3  Andhra Pradesh   31-08-2019                            3.32
4  Andhra Pradesh   30-09-2019                            5.17

   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
```

DATA VISUALIZATION

```
df1.hist(figsize=(10,10),color='red')

array([[<Axes: title={'center': 'estimated unemployment rate'}>,
        <Axes: title={'center': 'estimated employed'}>],
       [<Axes: title={'center': 'estimated labour participation
rate'}>,
        <Axes: >]], dtype=object)
```

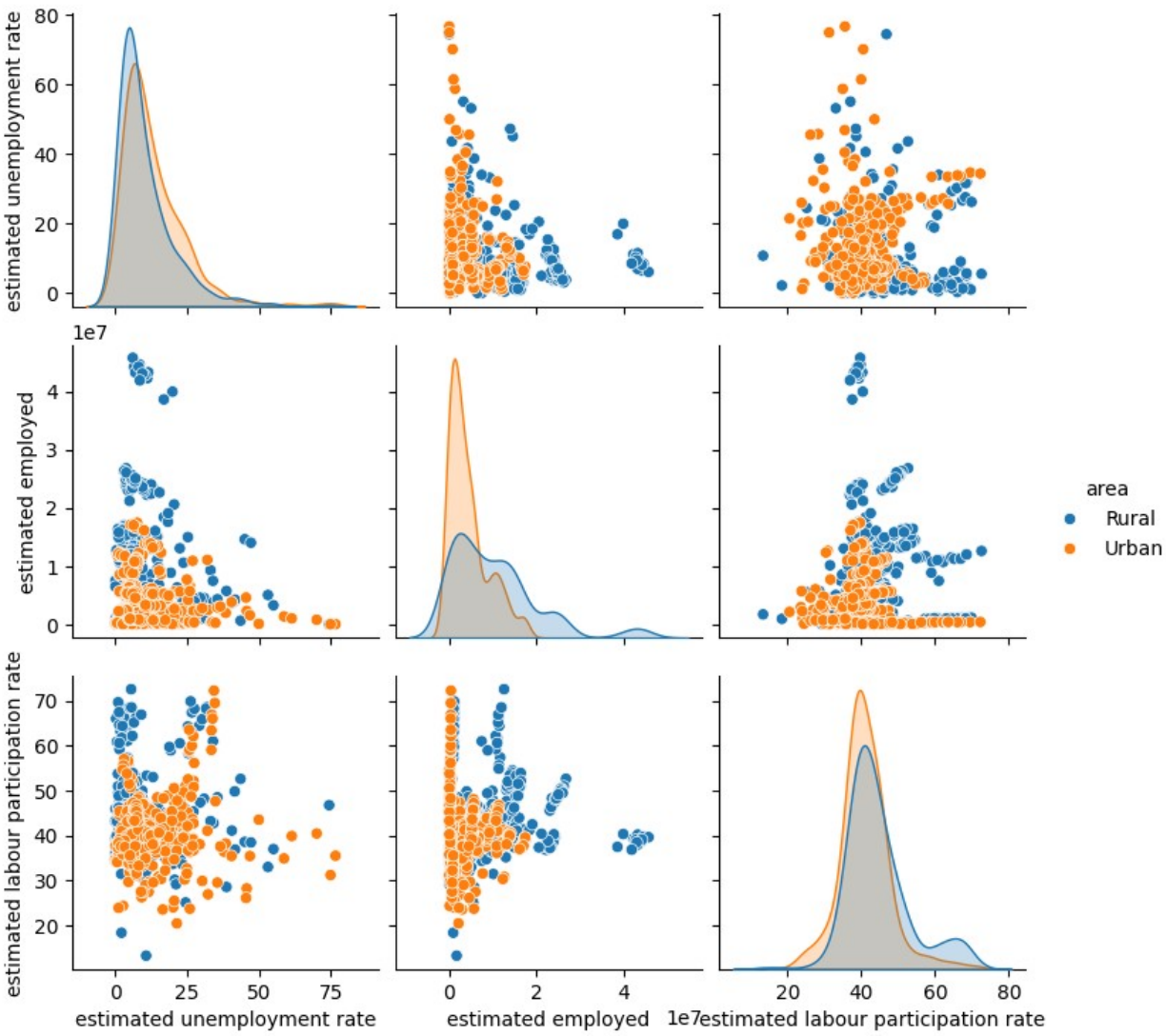





```
plt.figure(figsize=(20,20))
sns.pairplot(df1,hue='area')
```

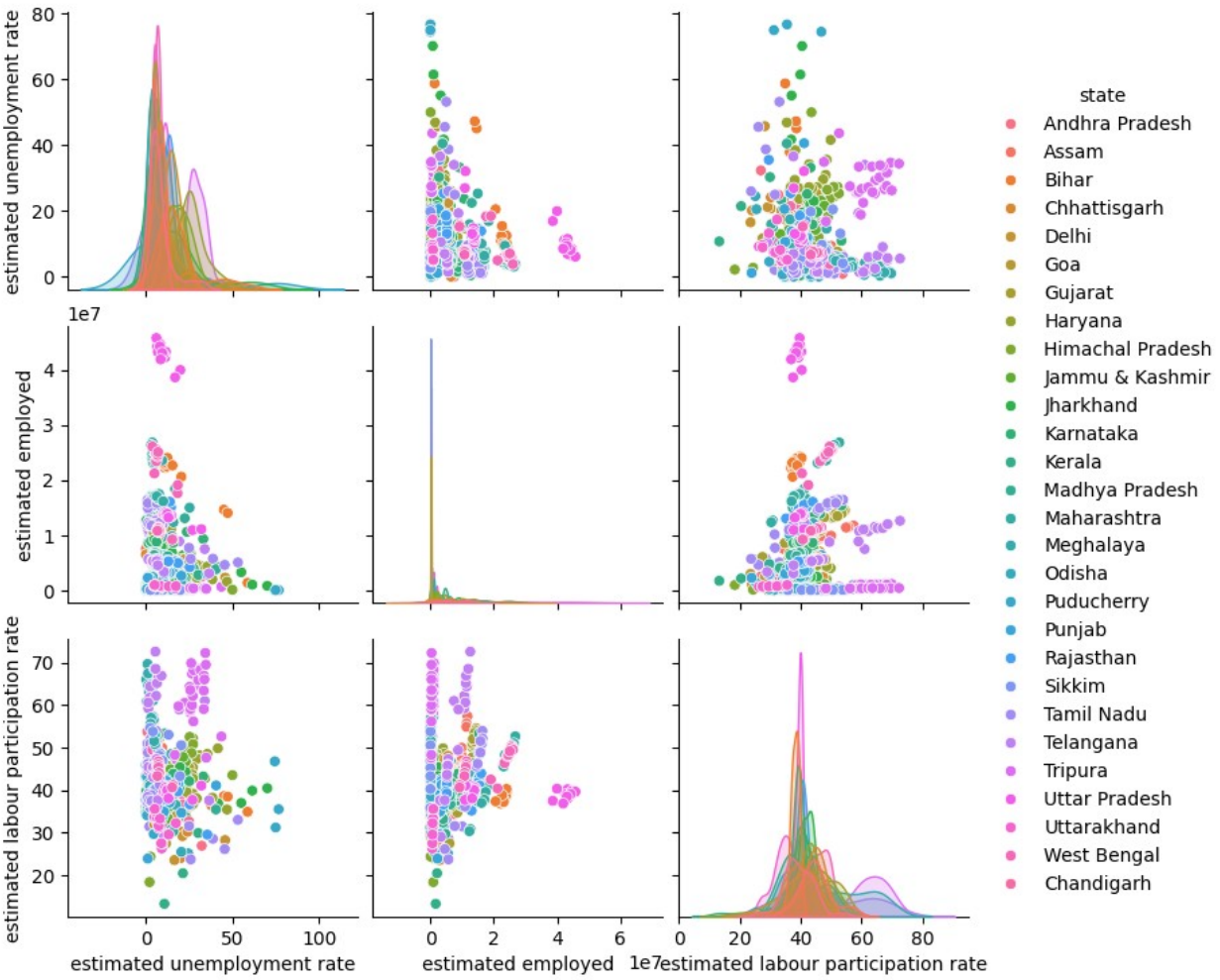
```
<seaborn.axisgrid.PairGrid at 0x7f1d2801ee80>
```

```
<Figure size 2000x2000 with 0 Axes>
```




```
sns.pairplot(df1, hue='state')
```

```
<seaborn.axisgrid.PairGrid at 0x7f1d27d78d30>
```




```
df1['estimated unemployment rate'].value_counts()
```

```
estimated unemployment rate
0.00     11
3.31      4
5.45      3
6.46      3
5.35      3
         ..
4.03      1
40.59     1
3.69      1
10.51     1
15.22     1
Name: count, Length: 623, dtype: int64
```

_

```
df1['state'].value_counts()
```

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

_

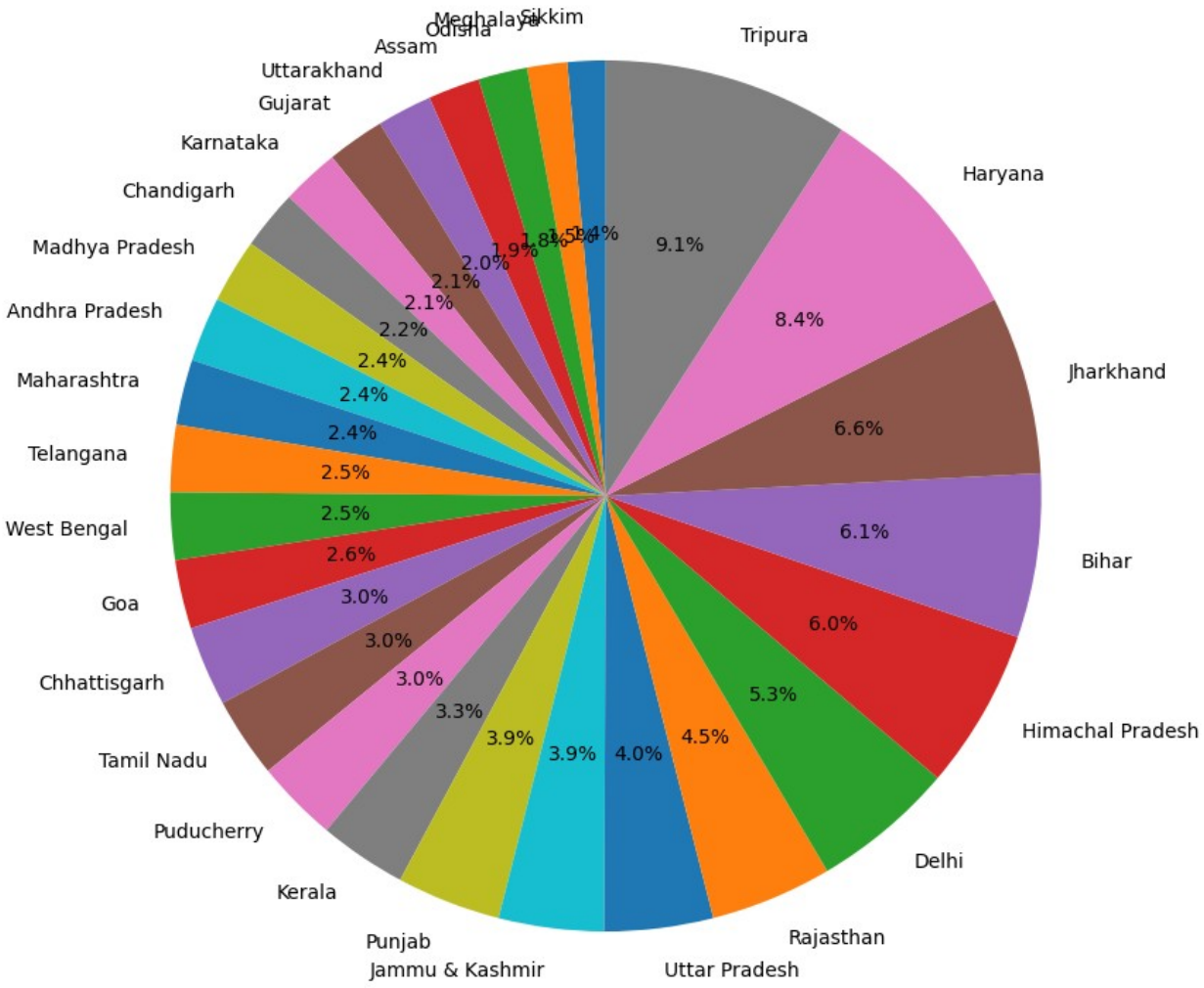
```
unemployment_rate_sta = df1.groupby('state')['estimated unemployment
rate'].sum()
unemployment_rate_sta =
unemployment_rate_sta.sort_values(ascending=True)
unemployment_rate_sta
```

```
state
Sikkim              123.24
Meghalaya           129.57
Odisha              158.42
Assam               167.13
Uttarakhand         177.74
Gujarat             186.59
Karnataka           186.93
Chandigarh          191.90
Madhya Pradesh      207.38
```

```
Andhra Pradesh      209.36
Maharashtra         211.61
Telangana           216.66
West Bengal         217.63
Goa                 222.58
Chhattisgarh        258.73
Tamil Nadu          259.96
Puducherry          265.59
Kerala              283.47
Punjab              336.87
Jammu & Kashmir     339.96
Uttar Pradesh       351.44
Rajasthan           393.63
Delhi               461.87
Himachal Pradesh    519.13
Bihar               529.71
Jharkhand           576.38
Haryana             735.93
Tripura             793.81
Name: estimated unemployment rate, dtype: float64
```

_

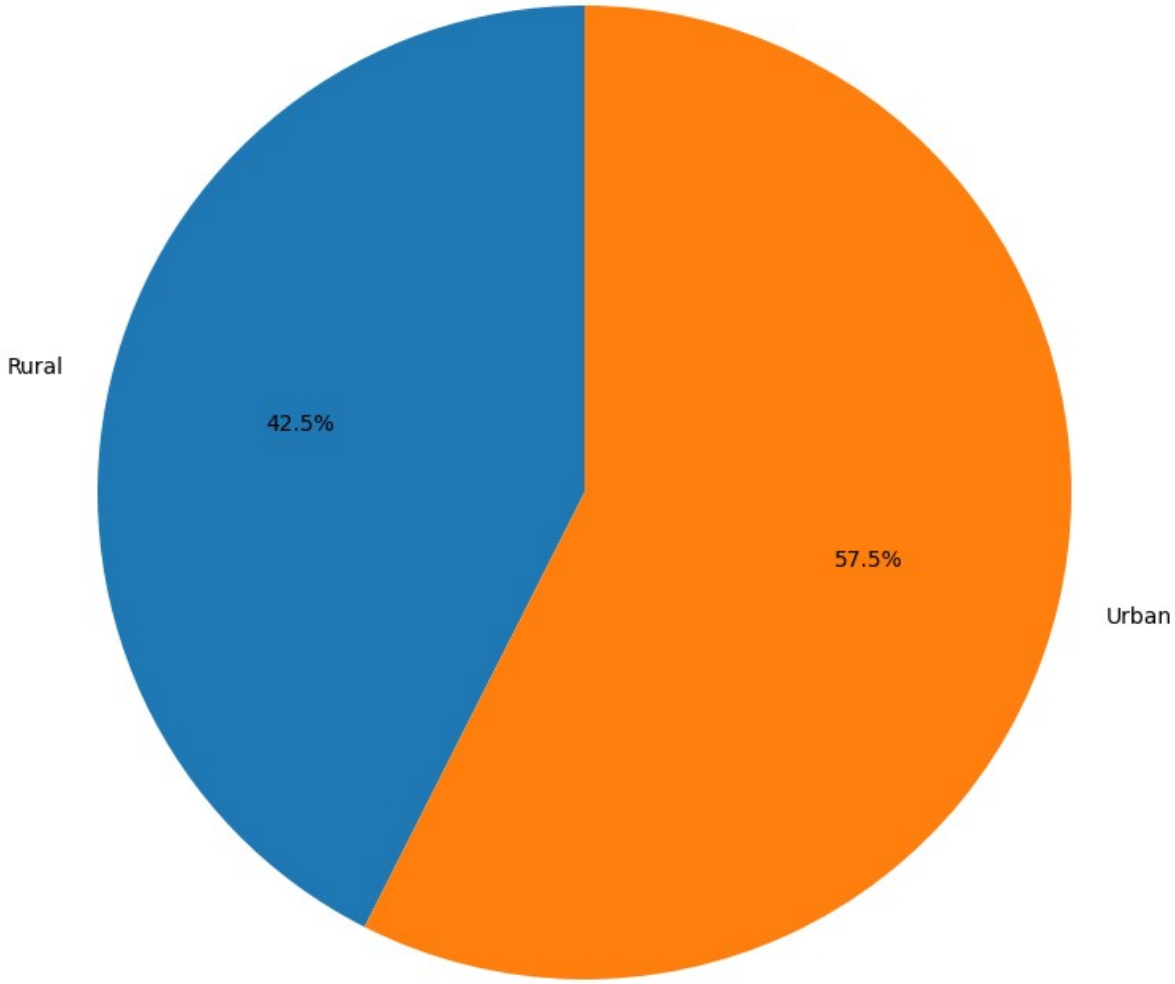
```
plt.figure(figsize=(10,10))
plt.pie(unemployment_rate_sta,labels=unemployment_rate_sta.index,autop
ct='%1.1f%%',startangle=90)
plt.show()
```




```
unemployment_rate_reg = df1.groupby('area')['estimated unemployment
rate'].sum()
unemployment_rate_reg =
unemployment_rate_reg.sort_values(ascending=True)
unemployment_rate_reg
```

```
area
Rural    3706.60
Urban    5006.62
Name: estimated unemployment rate, dtype: float64
```

```
plt.figure(figsize=(10,10))
plt.pie(unemployment_rate_reg,labels=unemployment_rate_reg.index,autop
ct='%1.1f%%',startangle=90)
plt.show()
```




```
#Encoding dataset

from sklearn.preprocessing import LabelEncoder
lab=LabelEncoder()

for i in df1.columns:
   if df1[i].dtype=='object':
      df1[i]=lab.fit_transform(df1[i])
```

```
/tmp/ipykernel_6006/3312773600.py:8: 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
  df1[i]=lab.fit_transform(df1[i])
/tmp/ipykernel_6006/3312773600.py:8: 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
  df1[i]=lab.fit_transform(df1[i])
/tmp/ipykernel_6006/3312773600.py:8: 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
  df1[i]=lab.fit_transform(df1[i])
```

```
df1.corr()
```

```
                                     state      date  \
state                                1.000000  0.059070
date                                 0.059070  1.000000
estimated unemployment rate         -0.056429 -0.079663
estimated employed                   0.210246  0.012952
estimated labour participation rate  0.170877  0.092125
area                                 0.074599  0.097448

                                     estimated unemployment rate  \
state                                                  -0.056429
date                                                   -0.079663
estimated unemployment rate                             1.000000
estimated employed                                     -0.222833
estimated labour participation rate                     0.002407
area                                                    0.132887

                                     estimated employed  \
state                                          0.210246
date                                           0.012952
estimated unemployment rate                   -0.222833
estimated employed                             1.000000
estimated labour participation rate            0.011501
area                                          -0.359450

                                     estimated labour participation
rate  \
state
```

```
0.170877
date
0.092125
estimated unemployment rate
0.002407
estimated employed
0.011501
estimated labour participation rate
1.000000
area                                                                -
0.219094
```

```
                                        area
state                                0.074599
date                                 0.097448
estimated unemployment rate          0.132887
estimated employed                  -0.359450
estimated labour participation rate -0.219094
area                                 1.000000
```

```
plt.figure(figsize=(10,10))
sns.heatmap(df1.corr(),annot=True)
```

```
<Axes: >
```


state
date
estimated unemployment rate
estimated employed
estimated labour participation rate
area
1
0.059
-0.056
0.21
0.17
0.075
0.059
1
-0.08
0.013
0.092
0.097
-0.056
-0.08
1
-0.22
0.0024
0.13
0.21
0.013
-0.22
1
0.012
-0.36
0.17
0.092
0.0024
0.012
1
-0.22
0.075
0.097
0.13
-0.36
-0.22
1
state
date
estimated unemployment rate
estimated employed
estimated labour participation rate
area
1.0
0.8
0.6
0.4
0.2
0.0
−0.2