OASIS_INFOBYTE_TASK - 2

July 6, 2023

1 OASIS INFOBYTE INTERNSHIP

2 TASK 2 - UNIMPLOYMENT ANALYSIS WITH PYTHON

3 PROBLEM STATEMENT :-

Analyzing Unemployment Rate During the COVID-19 Pandemic

The objective of this data science project is to analyze the unemployment rate during the COVID-19 pandemic and gain insights into its trends, patterns, and potential factors influencing its fluctuations. By examining the unemployment rate data over a specific period, we aim to understand the impact of the pandemic on employment and identify any correlations or drivers related to changes in the unemployment rate.

4 Importing Libraries

[1]:
```
#importing required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

5 Load the Dataset

[2]:
```
data = pd.read_csv('Unemployment in India.csv')
```

[3]:
```
# pd.set_option('display.max_rows',None)
```

[4]:
```
data
```

[4]:

	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]
```

[5]:
```
#top 5 rows
data.head()
```

[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

    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
```

[6]:
```
#last 5 rows
data.tail()
```

[6]:
```
    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
```

6 Data Preprocessing:

[7]:
```
#all data information
data.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
```

[8]:
```
#describe the data
data.describe()
```

[8]:
```
        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
```

[9]:
```
#cheking null values in dataset
data.isna().sum()
```

[9]:
```
Region                                   28
 Date                                    28
 Frequency                               28
 Estimated Unemployment Rate (%)         28
 Estimated Employed                      28
 Estimated Labour Participation Rate (%) 28
Area                                     28
dtype: int64
```

[10]:
```
#cheking duplicates in data
data.duplicated().sum()
```

[10]:
```
27
```

[11]:
```
data.dtypes
```

[11]:
```
Region                                    object
 Date                                     object
 Frequency                                object
 Estimated Unemployment Rate (%)         float64
 Estimated Employed                      float64
 Estimated Labour Participation Rate (%) float64
Area                                      object
dtype: object
```

[12]:
```
data.dropna(axis=0,inplace=True)
```

[13]:
```
data.isna().sum()
```

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

[14]:
```
data.duplicated().sum()
```

[14]:
```
0
```

[15]:
```
data.head(8)
```

[15]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural
5	Andhra Pradesh	31-10-2019	Monthly	3.52	12017412.0	43.01	Rural
6	Andhra Pradesh	30-11-2019	Monthly	4.12	11397681.0	41.00	Rural
7	Andhra Pradesh	31-12-2019	Monthly	4.38	12528395.0	45.14	Rural

[16]:
```
data.dtypes
```

[16]:
```
Region                                    object
 Date                                     object
 Frequency                                object
 Estimated Unemployment Rate (%)         float64
 Estimated Employed                      float64
 Estimated Labour Participation Rate (%)    float64
Area                                      object
dtype: object
```

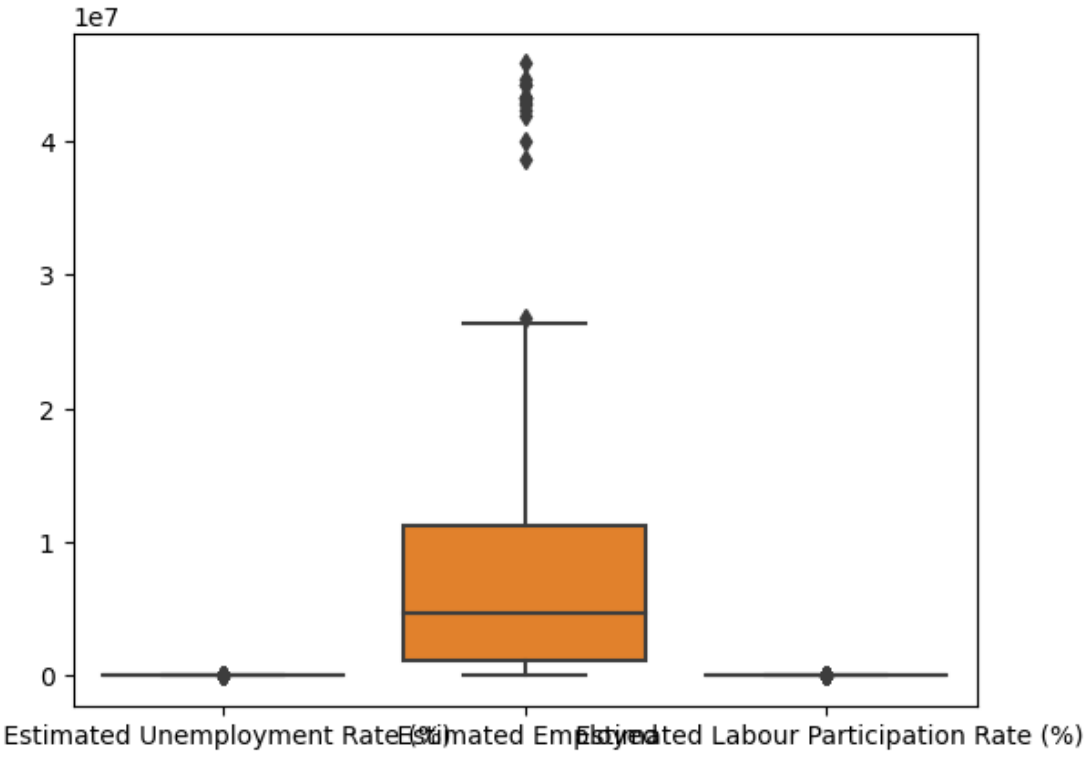
[17]:
```
data.columns
```

[17]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

[18]:
```
# import pandas as pd
data[' Date'] = pd.to_datetime(data[' Date'])
```

[19]:
```
sns.boxplot(data)
```

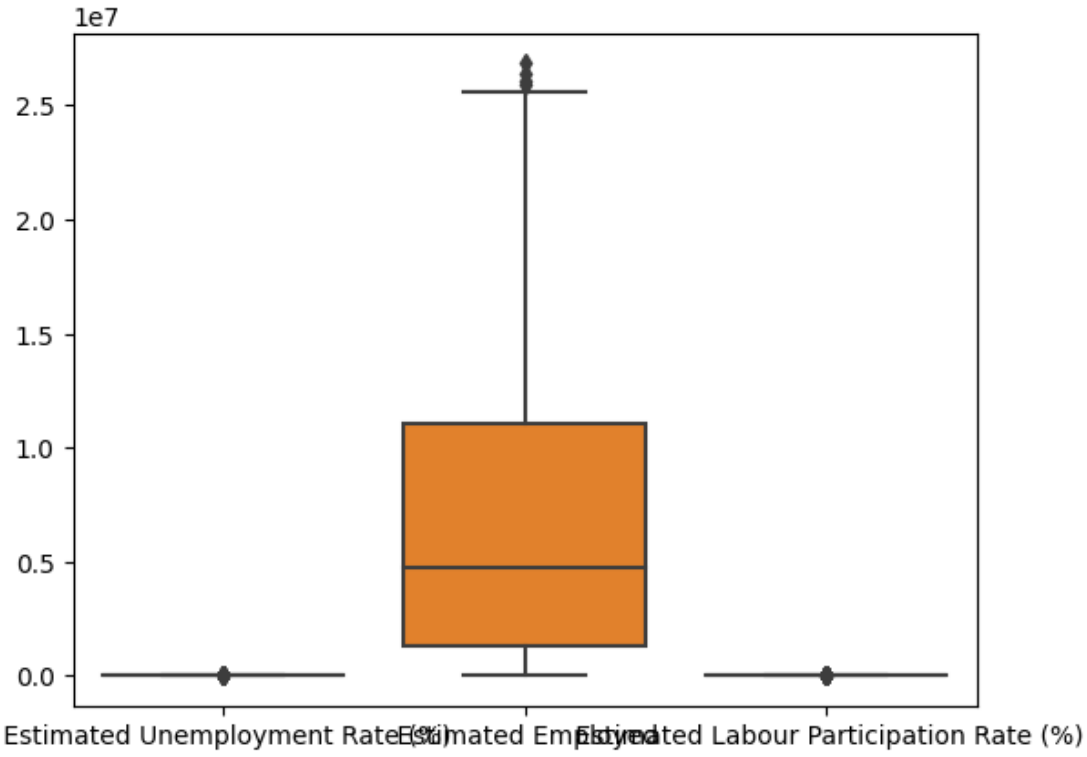
[19]:
```
<Axes: >
```



[20]:
```
from scipy import stats
numeric_columns = data.select_dtypes(include=np.number).columns
z_scores = stats.zscore(data[numeric_columns])
data = data[(np.abs(z_scores) < 3).all(axis=1)]
```

[21]:
```
sns.boxplot(data)
```

[21]: <Axes: >



[22]:
```
data.head()
```

[22]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	2019-05-31	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	2019-06-30	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	2019-07-31	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	2019-08-31	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	2019-09-30	Monthly	5.17	12256762.0	44.68	Rural

[23]:
```
data['Region'].unique()
```

[23]:
```
array(['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', 'Goa',
       'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',
       'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh',
       'Maharashtra', 'Meghalaya', 'Odisha', 'Puducherry', 'Punjab',
```

```
       'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Tripura',
       'Uttarakhand', 'West Bengal', 'Chandigarh', 'Uttar Pradesh'],
      dtype=object)
```

[24]:
```
data['Area'].unique()
```

[24]:
```
array(['Rural', 'Urban'], dtype=object)
```

[25]:
```
data.groupby('Region').size()
```

[25]:
```
Region
Andhra Pradesh      28
Assam               26
Bihar               25
Chandigarh          12
Chhattisgarh        28
Delhi               27
Goa                 24
Gujarat             28
Haryana             27
Himachal Pradesh    27
Jammu & Kashmir     21
Jharkhand           25
Karnataka           28
Kerala              27
Madhya Pradesh      28
Maharashtra         28
Meghalaya           24
Odisha              28
Puducherry          23
Punjab              28
Rajasthan           28
Sikkim              17
Tamil Nadu          26
Telangana           25
Tripura             21
Uttar Pradesh       14
Uttarakhand         27
West Bengal         28
dtype: int64
```

[26]:
```
region_stats = data.groupby(['Region'])[[' Estimated Unemployment Rate (%)','␣
↪Estimated Employed', ' Estimated Labour Participation Rate (%)']].mean().
↪reset_index()
```

[27]:
```
region_stats =round(region_stats,2)
```

[28]:
```
region_stats
```

[28]:

	Region	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
0	Andhra Pradesh	7.48	8154093.18	39.38
1	Assam	6.43	5354772.15	44.87
2	Bihar	15.14	12646269.60	38.25
3	Chandigarh	15.99	316831.25	39.34
4	Chhattisgarh	9.24	4303498.57	42.81
5	Delhi	15.41	2638021.37	39.32
6	Goa	9.27	226308.33	39.25
7	Gujarat	6.66	11402012.79	46.10
8	Haryana	25.52	3629312.93	43.01
9	Himachal Pradesh	17.38	1094081.33	44.25
10	Jammu & Kashmir	16.19	1799931.67	41.03
11	Jharkhand	15.59	4797540.72	41.97
12	Karnataka	6.68	10667119.29	41.35
13	Kerala	10.10	4524852.44	35.67
14	Madhya Pradesh	7.41	11115484.32	
15	Maharashtra	7.56	19990195.86	
16	Meghalaya	5.24	624798.71	
17	Odisha	5.66	6545746.96	
18	Puducherry	1.71	232050.00	
19	Punjab	12.03	4539362.00	
20	Rajasthan	14.06	10041064.75	
21	Sikkim	7.25	106880.71	
22	Tamil Nadu	6.20	12839543.92	
23	Telangana	8.02	7423194.00	
24	Tripura	27.66	652083.19	
25	Uttar Pradesh	14.89	13322807.07	
26	Uttarakhand	6.58	1390228.11	
27	West Bengal	8.12	17198538.00	

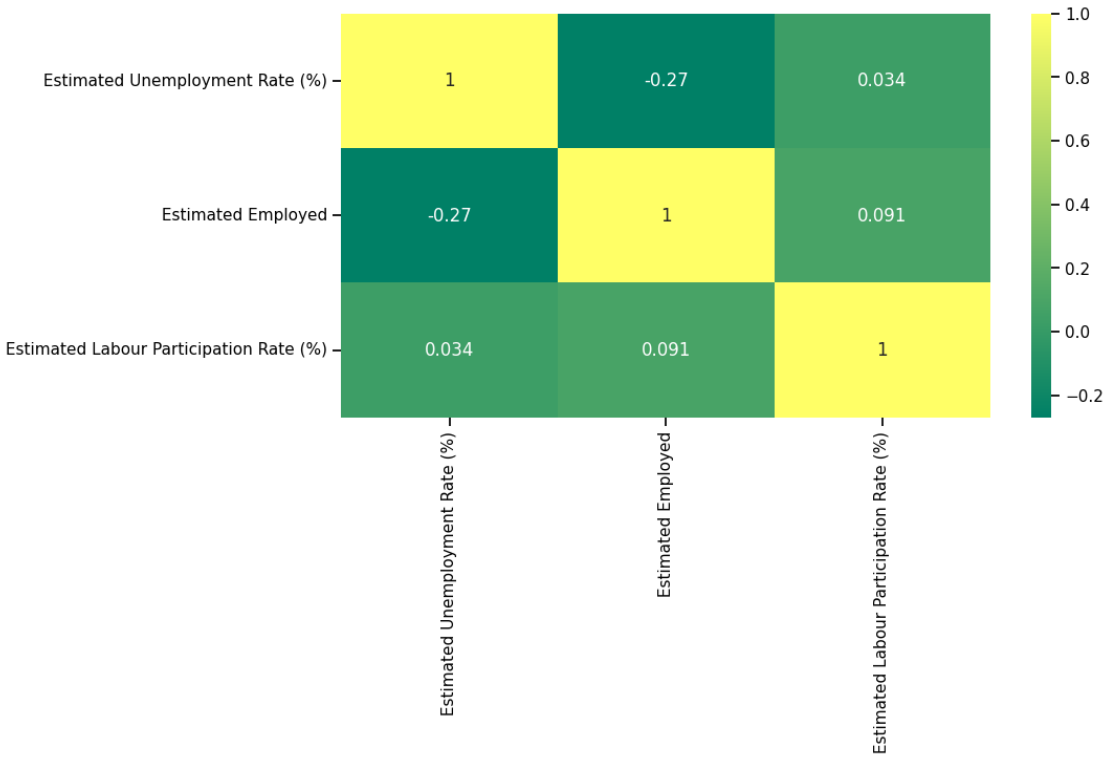
```
14                                      38.82
15                                      42.30
16                                      55.63
17                                      38.93
18                                      39.14
19                                      41.14
20                                      39.97
21                                      46.07
22                                      41.74
23                                      50.98
24                                      59.38
25                                      39.86
26                                      33.78
27                                      45.42
```

[29]:
```
heat_map = data[[' Estimated Unemployment Rate (%)',' Estimated Employed', '␣
↪Estimated Labour Participation Rate (%)']]
heat_map = heat_map.corr()
```

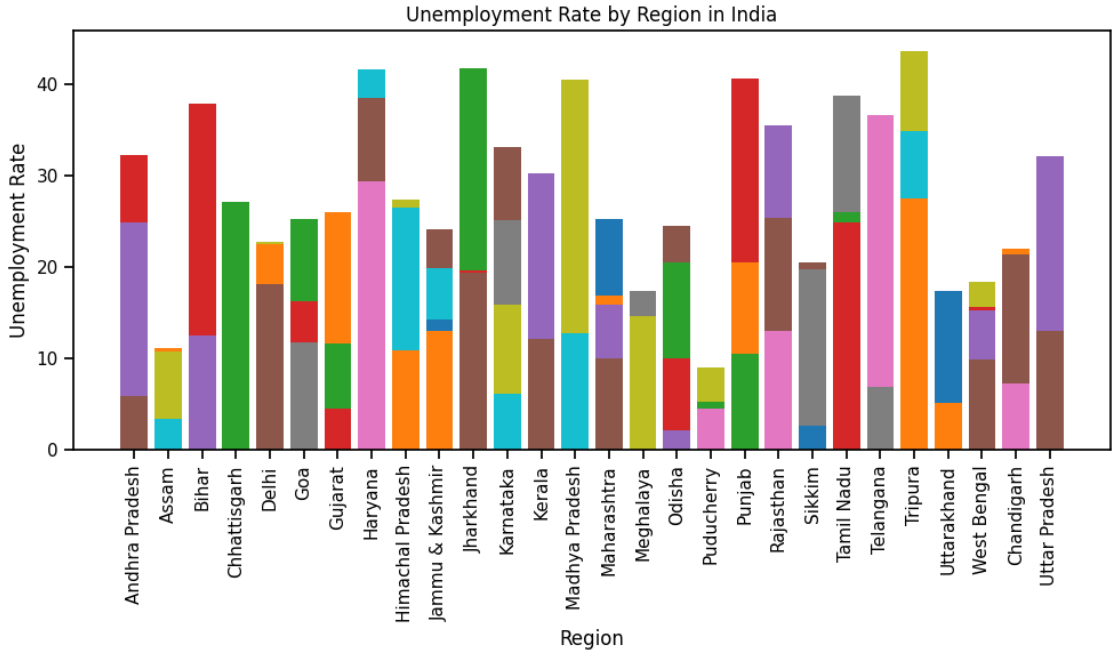
[30]:
```
heat_map
```

[30]:
```
                                        Estimated Unemployment Rate (%)  \
 Estimated Unemployment Rate (%)                                1.000000
 Estimated Employed                                            -0.269090
 Estimated Labour Participation Rate (%)                        0.033643

                                        Estimated Employed  \
 Estimated Unemployment Rate (%)                -0.269090
 Estimated Employed                              1.000000
 Estimated Labour Participation Rate (%)         0.091031

                                        Estimated Labour Participation Rate
(%)
 Estimated Unemployment Rate (%)
0.033643
 Estimated Employed
0.091031
 Estimated Labour Participation Rate (%)
1.000000
```

[31]:
```
plt.figure(figsize=(10,5))
sns.set_context('notebook',font_scale=1)
sns.heatmap(heat_map, annot=True,cmap='summer');
```



[32]:
```
import matplotlib.pyplot as plt
# Define a color palette with 28 distinct colors for each region
color_palette = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd',␣
 ↪'#8c564b', '#e377c2', '#7f7f7f',
                  '#bcbd22', '#17becf', '#1f77b4', '#ff7f0e', '#2ca02c',␣
 ↪'#d62728', '#9467bd', '#8c564b',
                  '#e377c2', '#7f7f7f', '#bcbd22', '#17becf', '#1f77b4',␣
 ↪'#ff7f0e', '#2ca02c', '#d62728',
                  '#9467bd', '#8c564b', '#e377c2', '#7f7f7f']
# Create a bar plot of unemployment rate by region with different colors
plt.figure(figsize=(10, 6))  # Adjust the figure size as needed
plt.bar(data['Region'], data[' Estimated Unemployment Rate (%)'],color =␣
 ↪color_palette)
plt.title('Unemployment Rate by Region in India')
plt.xlabel('Region')
plt.ylabel('Unemployment Rate')
plt.xticks(rotation=90)  # Rotate x-axis labels if needed
plt.tight_layout()  # Adjust spacing between plot elements
plt.show()
```



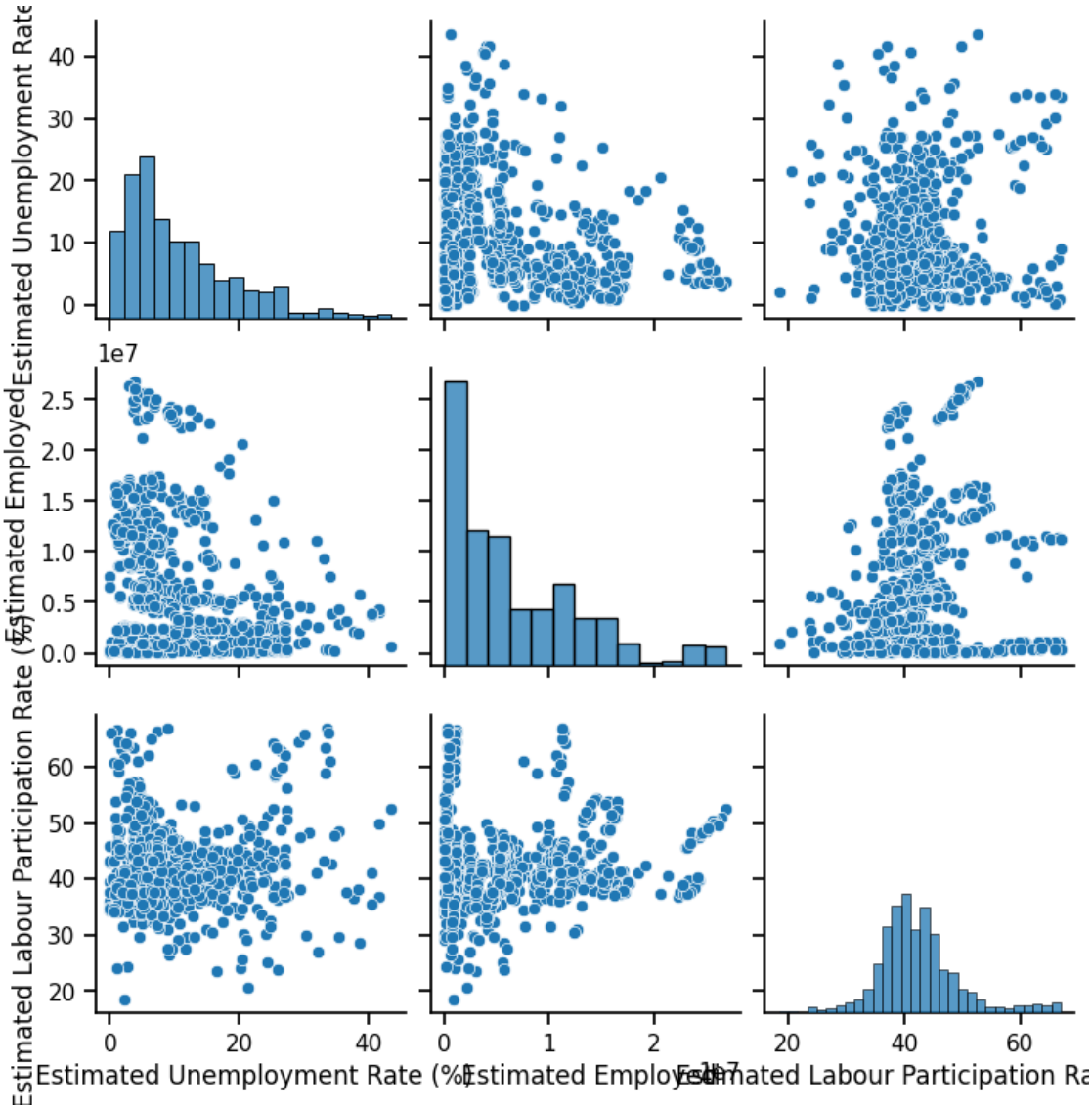
[33]:
```
data.columns
```

[33]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

[34]:
```
import plotly.express as px

region = data.groupby(["Region"])[[' Estimated Unemployment Rate (%)','␣
 ↪Estimated Employed', ' Estimated Labour Participation Rate (%)']].mean()
region = pd.DataFrame(region).reset_index()
fig = px.bar(region, x="Region", y=' Estimated Unemployment Rate (%)',␣
 ↪color="Region", title="Average Unemployment Rate by Region")
fig.update_layout(xaxis={'categoryorder':'total descending'})
fig.show()
```

[35]:
```
unemployment = data[["Region", ' Estimated Unemployment Rate (%)']]
fig = px.sunburst(unemployment, path=['Region'], values=' Estimated␣
 ↪Unemployment Rate (%)',
                  title= 'Unemployment rate in every State and Region',␣
 ↪height=650)
fig.show()
```

[36]:
```
sns.pairplot(data)
plt.show()
```



7 Dataset 2: Unemployment_Rate_upto_11_2020

[37]:
```
#importing required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[38]:
```
data2 = pd.read_csv('Unemployment_Rate_upto_11_2020.csv')
```

[39]:
```
data2
```

[39]:
```
            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]
```

[40]:
```
#top 5 rows
data2.head()
```

[40]:
```
            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
```

[41]:
```
#last 5 rows
data2.tail()
```

[41]:
```
           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
```

[42]:
```
#all data information
data2.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
```

[43]:
```
#describe the data
data2.describe()
```

[43]:

	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

[44]:
```
#cheking null values in dataset
data2.isna().sum()
```

[44]:
```
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
```

[45]:
```
#cheking duplicates in data
data2.duplicated().sum()
```

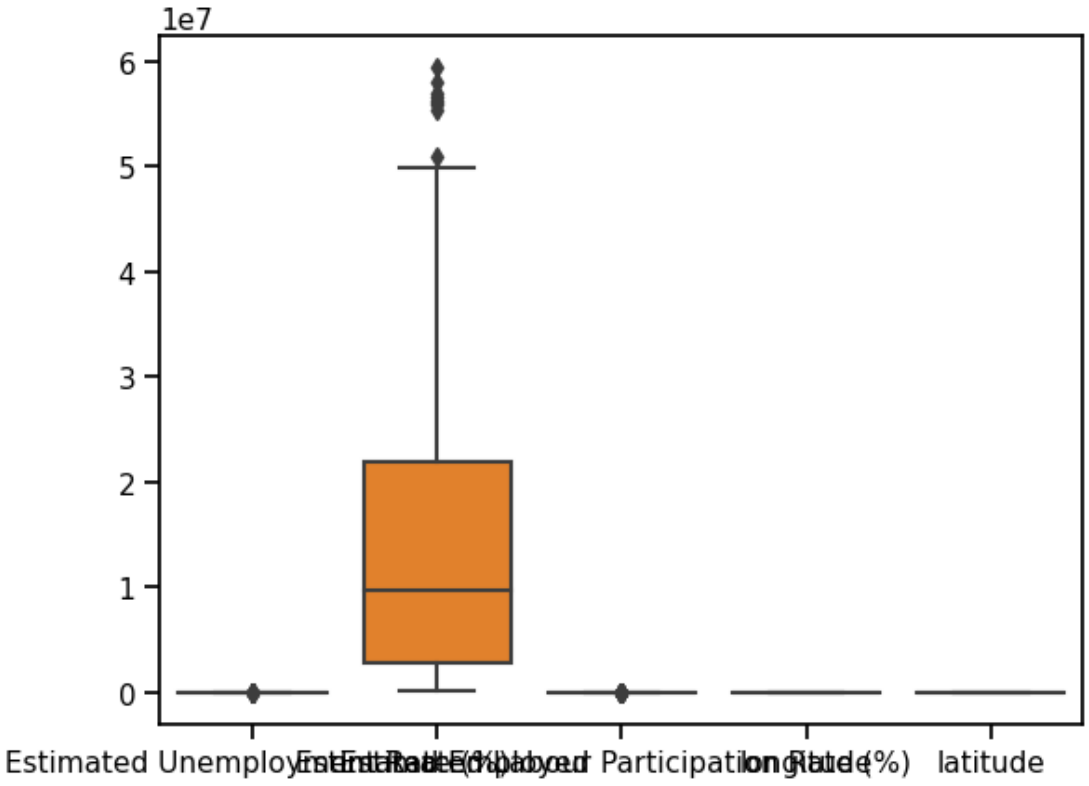
[45]:
```
0
```

[46]:
```
data2.dropna(axis=0,inplace=True)
```

[47]:
```
import pandas as pd
data2[' Date'] = pd.to_datetime(data2[' Date'])
```

[48]:
```
sns.boxplot(data2)
```

[48]:
```
<Axes: >
```

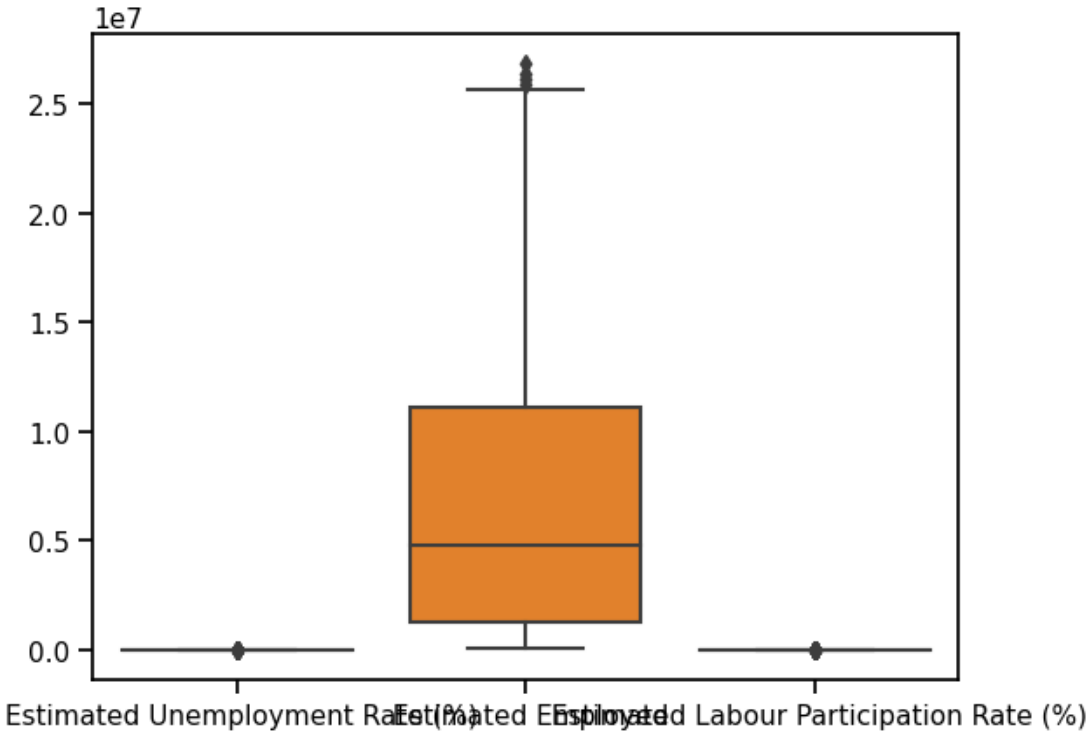


[49]:
```
from scipy import stats
numeric_columns = data2.select_dtypes(include=np.number).columns
```

```
z_scores = stats.zscore(data2[numeric_columns])
data2= data2[(np.abs(z_scores) < 3).all(axis=1)]
```

[50]:
```
sns.boxplot(data)
```

[50]: <Axes: >

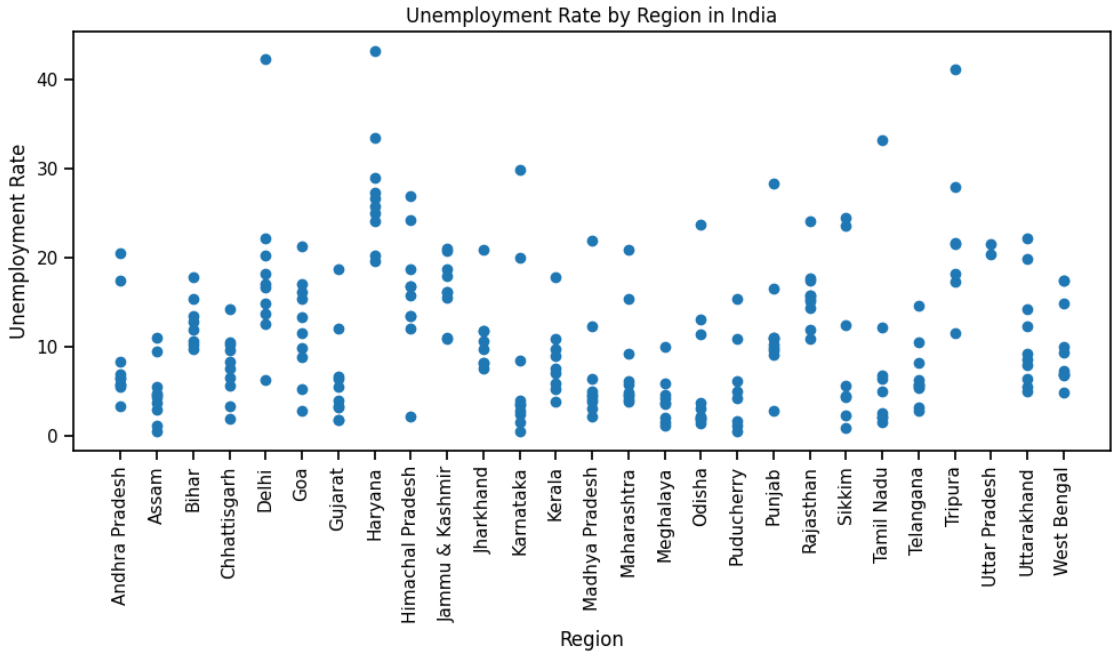


[51]:
```
data['Region'].unique()
```

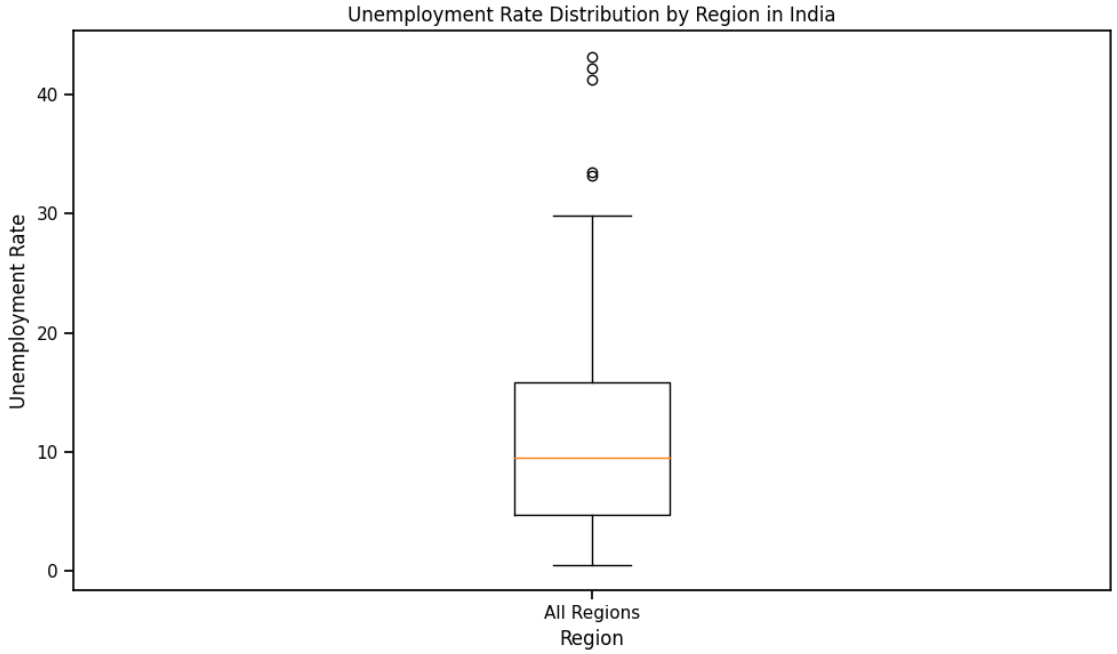
[51]:
```
array(['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', 'Goa',
       'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',
       'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh',
       'Maharashtra', 'Meghalaya', 'Odisha', 'Puducherry', 'Punjab',
       'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Tripura',
       'Uttarakhand', 'West Bengal', 'Chandigarh', 'Uttar Pradesh'],
      dtype=object)
```

[52]:
```
plt.figure(figsize=(10, 6))
plt.scatter(data2['Region'], data2[' Estimated Unemployment Rate (%)'])
plt.title('Unemployment Rate by Region in India')
plt.xlabel('Region')
plt.ylabel('Unemployment Rate')
plt.xticks(rotation=90)
```

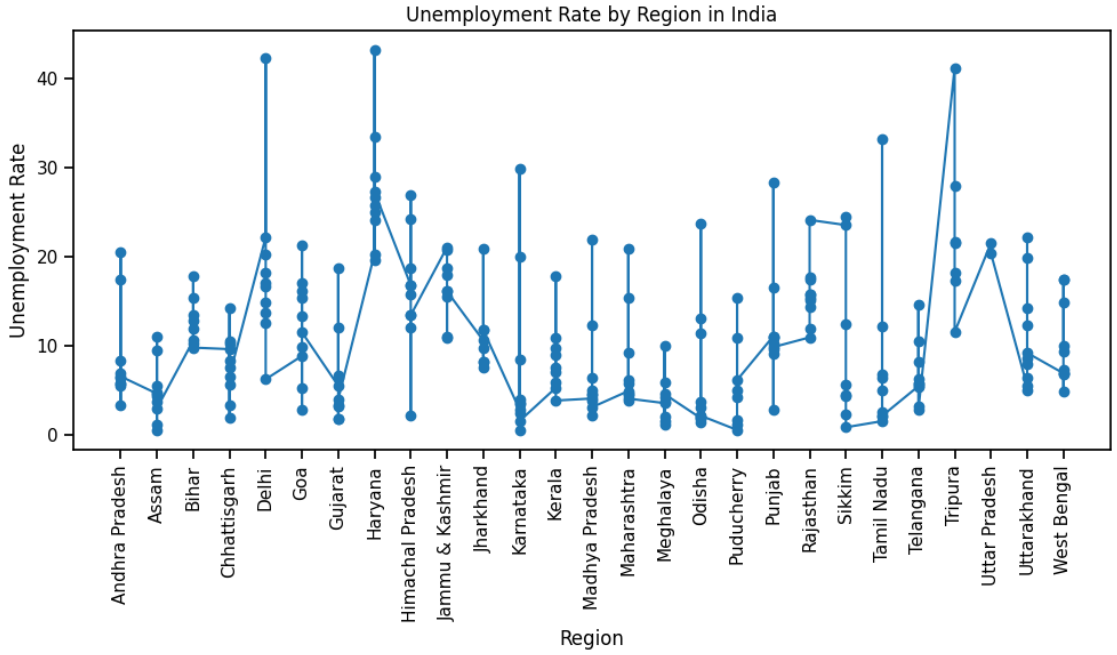
```
plt.tight_layout()
plt.show()
```



[53]:
```
plt.figure(figsize=(10, 6))
plt.boxplot(data2[' Estimated Unemployment Rate (%)'])
plt.title('Unemployment Rate Distribution by Region in India')
plt.xlabel('Region')
plt.ylabel('Unemployment Rate')
plt.xticks(ticks=[1], labels=['All Regions'])
plt.tight_layout()
plt.show()
```

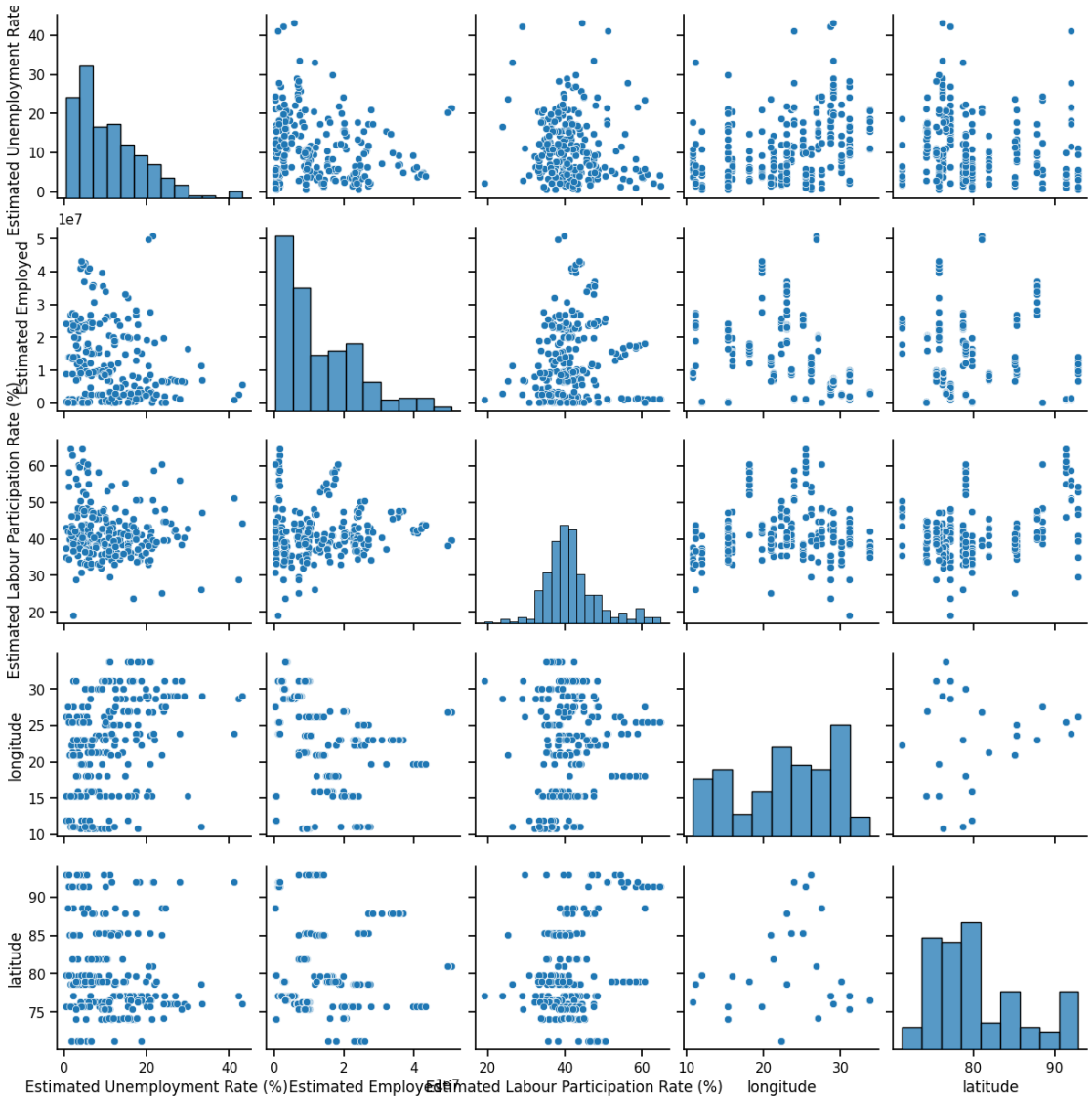


[54]:
```
plt.figure(figsize=(10, 6))
plt.plot(data2['Region'], data2[' Estimated Unemployment Rate (%)'], marker='o')
plt.title('Unemployment Rate by Region in India')
plt.xlabel('Region')
plt.ylabel('Unemployment Rate')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



[55]:
```
sns.pairplot(data2)
```

[55]: <seaborn.axisgrid.PairGrid at 0x1a2316cbe20>


Estimated Unemployment Rate (%)
Estimated Employed
Estimated Labour Participation Rate (%)
longitude
latitude
1e7
1e7

8 Thank You!