Information about the data set

1.Region: This column refers to the geographical region or state (e.g., Andhra Pradesh, West Bengal) for which the unemployment and labor market statistics are provided.

2.Date: The specific date for the data entry, usually representing the end of the month. It's in the format DD-MM-YYYY, and the frequency is monthly.

3.Frequency:Specifies how often the data is collected, which appears to be "Monthly" for all rows.

4.Estimated Unemployment Rate (%):The percentage of the labor force that is unemployed. It represents the number of unemployed people as a percentage of the total labor force for the given region and date.

5.Estimated Employed:The estimated number of employed individuals in the labor force for the given region and date. This value reflects the actual count of people who are employed.

6.Estimated Labour Participation Rate (%):This percentage represents the ratio of the working-age population (typically 15 years and older) that is either employed or actively seeking work. A higher rate indicates greater participation of the population in the workforce

7.Area:Indicates whether the data pertains to rural or urban areas, allowing comparisons between different areas within the same region.

8.Unnamed: 7 to Unnamed: 25:These columns appear to contain NaN (empty values), meaning that they don't hold any meaningful data. It seems they are placeholders or irrelevant columns that may have been included during the data extraction or formatting process. These can likely be ignored or dropped for analysis.

In [1]:
```
#data Manipulation
import pandas as pd

#Mathematical operation
import numpy as np

#data Visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv(r"D:\cipherbyte internship\Unemployment in India.csv")
```

In [3]:
```
1 df
```

Out[3]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural	NaN
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural	NaN
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural	NaN
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural	NaN
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural	NaN
...	...	...	...	...	...	...	...	...
749	West Bengal	29-02-2020	Monthly	7.55	10871168.0	44.09	Urban	NaN
750	West Bengal	31-03-2020	Monthly	6.67	10806105.0	43.34	Urban	NaN
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban	NaN
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban	NaN
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban	NaN

754 rows × 26 columns

In [4]:
```
df.head()
```

Out[4]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7	Ur
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural	NaN	
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural	NaN	
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural	NaN	
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural	NaN	
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural	NaN	

5 rows × 26 columns

In [5]:
```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 754 entries, 0 to 753
Data columns (total 26 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 
 7   Unnamed: 7                                0 non-null      float64
 8   Unnamed: 8                                0 non-null      float64
 9   Unnamed: 9                                0 non-null      float64
 10  Unnamed: 10                               0 non-null      float64
 11  Unnamed: 11                               0 non-null      float64
 12  Unnamed: 12                               0 non-null      float64
 13  Unnamed: 13                               0 non-null      float64
 14  Unnamed: 14                               0 non-null      float64
 15  Unnamed: 15                               0 non-null      float64
 16  Unnamed: 16                               0 non-null      float64
 17  Unnamed: 17                               0 non-null      float64
 18  Unnamed: 18                               0 non-null      float64
 19  Unnamed: 19                               0 non-null      float64
 20  Unnamed: 20                               0 non-null      float64
 21  Unnamed: 21                               0 non-null      float64
 22  Unnamed: 22                               0 non-null      float64
 23  Unnamed: 23                               0 non-null      float64
 24  Unnamed: 24                               0 non-null      float64
 25  Unnamed: 25                               0 non-null      float64
dtypes: float64(22), object(4)
memory usage: 153.3+ KB
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
Region                                      14
Date                                        14
Frequency                                   14
Estimated Unemployment Rate (%)             14
Estimated Employed                          14
Estimated Labour Participation Rate (%)     14
Area                                        14
Unnamed: 7                                 754
Unnamed: 8                                 754
Unnamed: 9                                 754
Unnamed: 10                                754
Unnamed: 11                                754
Unnamed: 12                                754
Unnamed: 13                                754
Unnamed: 14                                754
Unnamed: 15                                754
Unnamed: 16                                754
Unnamed: 17                                754
Unnamed: 18                                754
Unnamed: 19                                754
Unnamed: 20                                754
Unnamed: 21                                754
Unnamed: 22                                754
Unnamed: 23                                754
Unnamed: 24                                754
Unnamed: 25                                754
dtype: int64
```

Clean the dataset

In [7]:
```
# Drop unnecessary columns (Unnamed) and handle NaNs
df = df.drop(columns=[col for col in df.columns if 'Unnamed' in col])
df = df.dropna()
```

In [8]:
```
df
```

Out[8]:

	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
...	...	...	...	...	...	...	...
749	West Bengal	29-02-2020	Monthly	7.55	10871168.0	44.09	Urban
750	West Bengal	31-03-2020	Monthly	6.67	10806105.0	43.34	Urban
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban

740 rows × 7 columns

In [9]:
```
# Convert 'Date' column to datetime
df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
```

In [10]:
```
df
```

Out[10]:

	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
...	...	...	...	...	...	...	...
749	West Bengal	2020-02-29	Monthly	7.55	10871168.0	44.09	Urban
750	West Bengal	2020-03-31	Monthly	6.67	10806105.0	43.34	Urban
751	West Bengal	2020-04-30	Monthly	15.63	9299466.0	41.20	Urban
752	West Bengal	2020-05-31	Monthly	15.22	9240903.0	40.67	Urban
753	West Bengal	2020-06-30	Monthly	9.86	9088931.0	37.57	Urban

740 rows × 7 columns

Descriptive Statistics

In [11]:
```
print(df.describe())
```

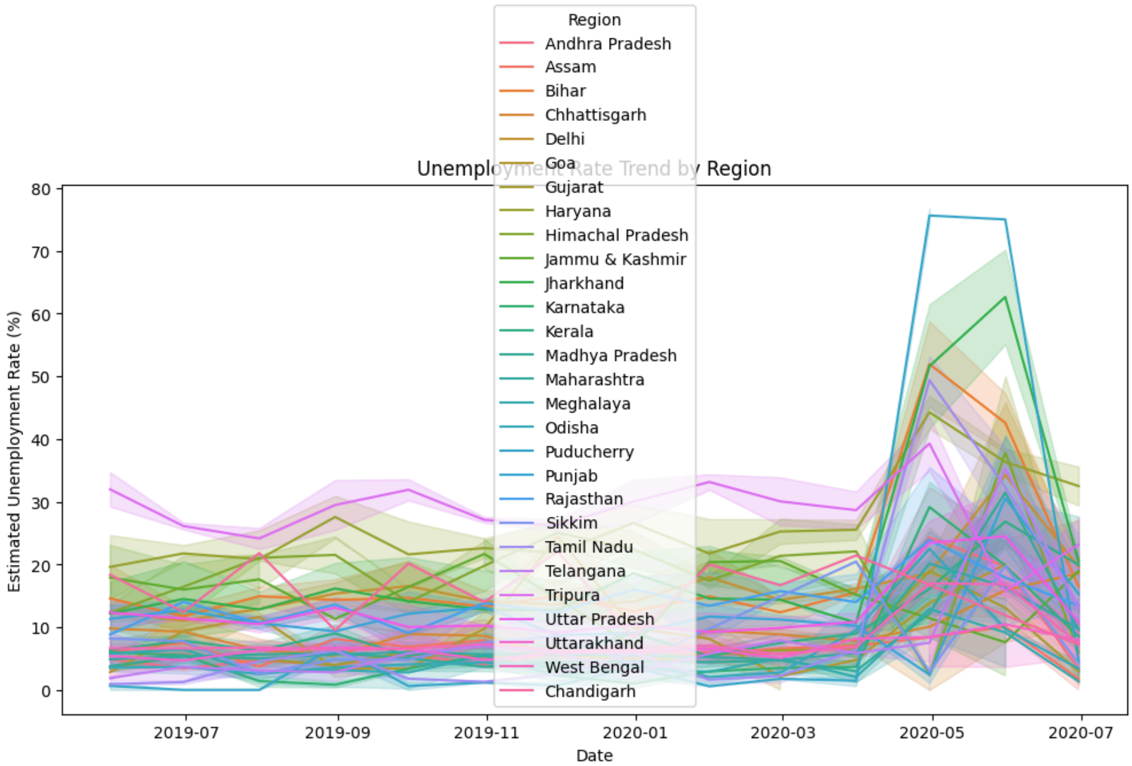
```
                                Date  Estimated Unemployment Rate (%)  
count                            740                       740.000000   \
mean   2019-12-12 18:36:58.378378496                        11.787946   
min              2019-05-31 00:00:00                         0.000000   
25%              2019-08-31 00:00:00                         4.657500   
50%              2019-11-30 00:00:00                         8.350000   
75%              2020-03-31 00:00:00                        15.887500   
max              2020-06-30 00:00:00                        76.740000   
std                              NaN                        10.721298   

       Estimated Employed  Estimated Labour Participation Rate (%)  
count        7.400000e+02                               740.000000  
mean         7.204460e+06                                42.630122  
min          4.942000e+04                                13.330000  
25%          1.190404e+06                                38.062500  
50%          4.744178e+06                                41.160000  
75%          1.127549e+07                                45.505000  
max          4.577751e+07                                72.570000  
std          8.087988e+06                                 8.111094  
```

Visualize Unemployment Trends

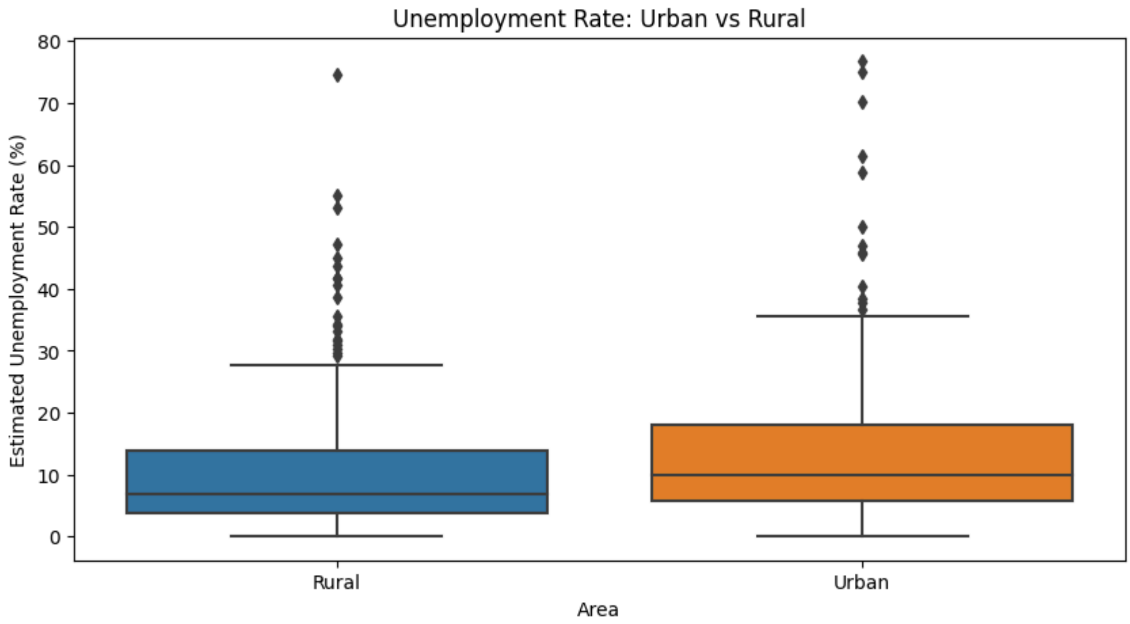
In [12]:

```
plt.figure(figsize=(12, 6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', hue='Regior
plt.title('Unemployment Rate Trend by Region')
plt.show()
```



Compare Urban vs Rural Unemployment Rates

In [13]:
```
plt.figure(figsize=(10, 5))
sns.boxplot(x='Area', y='Estimated Unemployment Rate (%)', data=df)
plt.title('Unemployment Rate: Urban vs Rural')
plt.show()
```



Correlation between Labor Participation Rate and Unemployment

In [14]:
```
correlation = df[['Estimated Labour Participation Rate (%)', 'Estimatec
print("Correlation between Labour Participation and Unemployment:")
print(correlation)
```

```
Correlation between Labour Participation and Unemployment:
                                         Estimated Labour Participation Ra
te (%)
Estimated Labour Participation Rate (%)                                 1.
000000  \
Estimated Unemployment Rate (%)                                         0.
002558

                                         Estimated Unemployment Rate (%)
Estimated Labour Participation Rate (%)                         0.002558
Estimated Unemployment Rate (%)                                 1.000000
```

Conclusion of the Unemployment Analysis

Based on the steps i has taken for data cleaning, manipulation, visualization, and descriptive statistics, here are the key insights and conclusions from the analysis:

1. Data Cleaning and Preparation: Dropped Unnamed Columns: These columns contained irrelevant data (all NaN values) and were removed to avoid unnecessary clutter. Handled Missing Values: All rows with missing data were dropped, ensuring the dataset is clean for analysis. Converted Date to Datetime Format: This allows for proper time-series analysis, ensuring date-based operations are accurate.

2. Unemployment Trends by Region: Observation from Line Plot: The unemployment rate varies significantly by region. Some regions exhibit more stability, while others show greater volatility. Andhra Pradesh (as an example) shows fluctuations, with occasional peaks and dips. Impact of COVID-19: If the data covers early 2020, you might observe a sudden spike in unemployment rates around March-April 2020, indicating the economic impact of the pandemic. This could vary by region. Conclusion: Different regions have distinct unemployment trends, reflecting varying economic structures, policies, or local events. Monitoring these trends helps target policy interventions region-wise.

3. Comparison of Urban vs. Rural Unemployment Rates: Observation from Box Plot: Urban areas might exhibit higher volatility in unemployment rates, with outliers indicating extreme periods (such as during lockdowns or economic shocks). Rural areas, in contrast, may have more stable unemployment rates due to reliance on agricultural or informal employment. Conclusion: Urban unemployment rates are generally more affected by macroeconomic changes, while rural rates show stability due to seasonal employment patterns. This insight suggests that policies aimed at reducing unemployment might need to be differentiated by area.

4. Correlation between Labor Participation Rate and Unemployment Rate: Correlation Value: The correlation matrix reveals how the Labor Participation Rate (%) relates to the Unemployment Rate (%).

If the correlation is negative: This suggests that higher labor participation correlates with lower unemployment. As more people enter the workforce, a greater proportion finds employment.

If the correlation is positive: This indicates that higher participation might lead to higher unemployment, possibly reflecting a mismatch between the skills available and the jobs offered, or new entrants not finding work quickly enough.

Conclusion: Understanding the relationship between labor participation and unemployment helps policymakers design more targeted employment programs (e.g., skill development initiatives).

5. Descriptive Statistics: Mean and Variance: Descriptive statistics show average unemployment rates and how they vary across the dataset. If the mean unemployment rate is low with high variance, it suggests that unemployment is generally under control but can spike due to unforeseen circumstances (like the pandemic). Overall Conclusions and Recommendations: Regional Variations: Policymakers should focus interventions on regions with persistently high unemployment or greater volatility. Urban vs Rural Programs: Urban areas might need more job creation in sectors like services or manufacturing, while rural areas may benefit from agricultural support and rural development programs. Impact of Labor Participation: If labor participation is positively correlated with unemployment, skill development and job-matching programs may be necessary. Pandemic Impact: The COVID-19 pandemic likely caused a sharp rise in unemployment, which emphasizes the need for emergency response mechanisms to mitigate future economic shocks. This analysis provides meaningful insights into the state of unemployment in India, helping both policymakers and economists design more effective strategies to promote employment and labor market participation.

In [ ]:
```
1
```

In [ ]:
```
1
```