codealpha-task-2

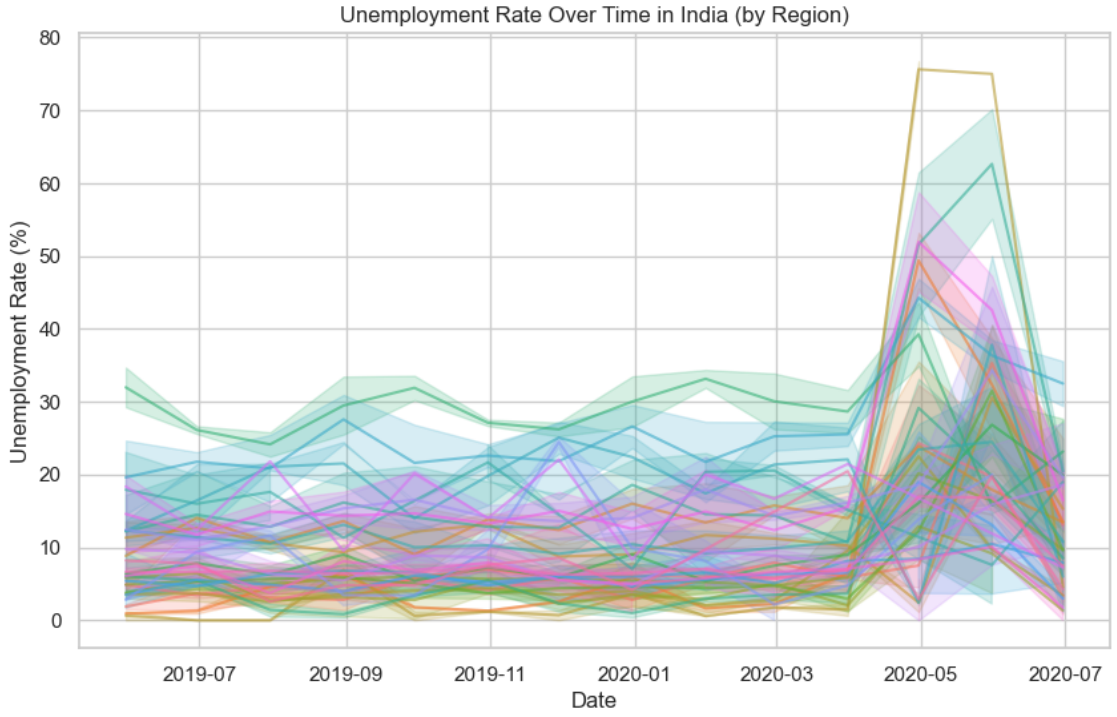
September 17, 2024

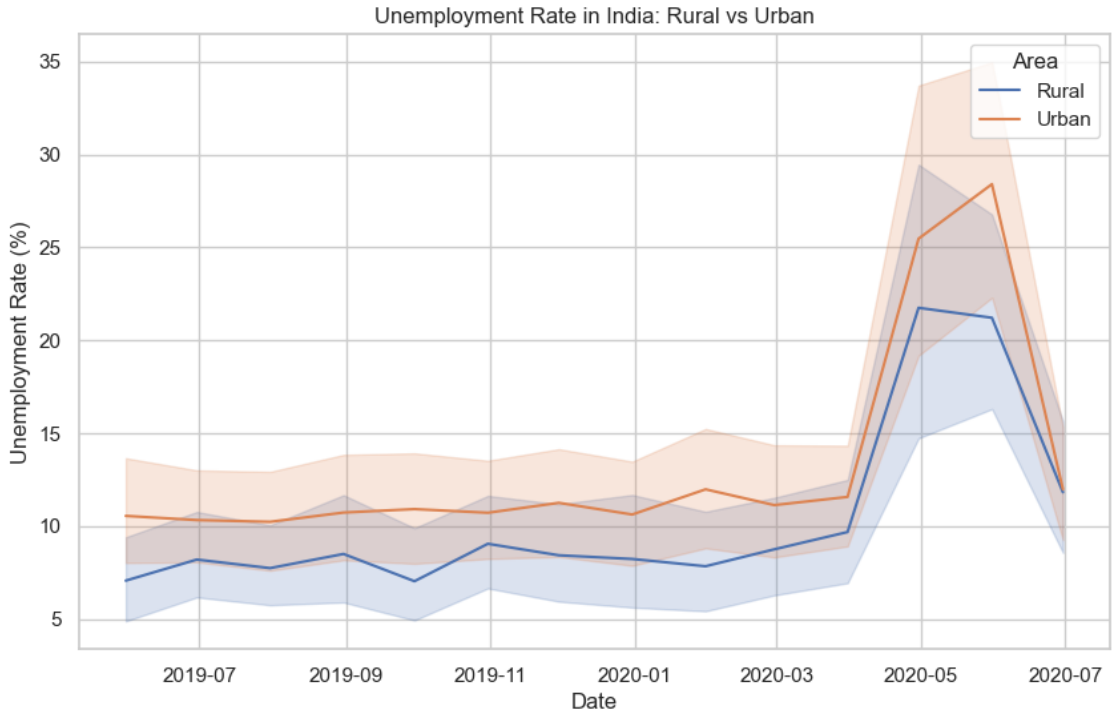
TASK -02

Unemployment is measured by the unemployment rate which is the number of people who are unemployed as a percentage of the total labour force. We have seen a sharp increase in the unemployment rate during Covid-19, so analyzing the unemployment rate can be a good data science project.

[6]:
```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = "C:\\Users\\Harshika k\\Downloads\\Unemployment in India.csv"
df = pd.read_csv(file_path)

# Clean the data
df.columns = df.columns.str.strip()

# Strip spaces from the Date column and then convert it to datetime format
df['Date'] = df['Date'].str.strip()
df['Date'] = pd.to_datetime(df['Date'], format='%d-%m-%Y')
df = df.sort_values(by='Date')

# Set Seaborn style for better visualizations
sns.set(style="whitegrid")

# Plot the unemployment rate over time across all regions
plt.figure(figsize=(10,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', data=df,␣
 ↪hue='Region', legend=False, alpha=0.6)
plt.title('Unemployment Rate Over Time in India (by Region)')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.grid(True)
plt.show()

# Plot the unemployment rate for rural vs urban areas
```

```
plt.figure(figsize=(10,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', data=df, hue='Area')
plt.title('Unemployment Rate in India: Rural vs Urban')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.grid(True)
plt.show()

# Basic statistics
print(df[['Estimated Unemployment Rate (%)', 'Estimated Labour Participation␣
 ↪Rate (%)']].describe())
```





	Estimated Unemployment Rate (%)	\
count	740.000000	
mean	11.787946	
std	10.721298	
min	0.000000	
25%	4.657500	
50%	8.350000	
75%	15.887500	
max	76.740000	

	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

[ ]: