```
# ------------------------------
# Task 2: Unemployment Analysis
# ------------------------------

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load Dataset
df = pd.read_csv("Unemployment in India.csv")

# Display first few rows
df.head()
```

	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

Next steps: Generate code with df | New interactive sheet

```
# Checking dataset info
df.info()

# Checking null values
df.isnull().sum()

# Summary statistics
df.describe()
```

```
<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
```

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
count	740.000000	7.400000e+02	740.000000
mean	11.787946	7.204460e+06	42.630122
std	10.721298	8.087988e+06	8.111094
min	0.000000	4.942000e+04	13.330000
25%	4.657500	1.190404e+06	38.062500
50%	8.350000	4.744178e+06	41.160000
75%	15.887500	1.127549e+07	45.505000
max	76.740000	4.577751e+07	72.570000

```
print(df.columns)
```

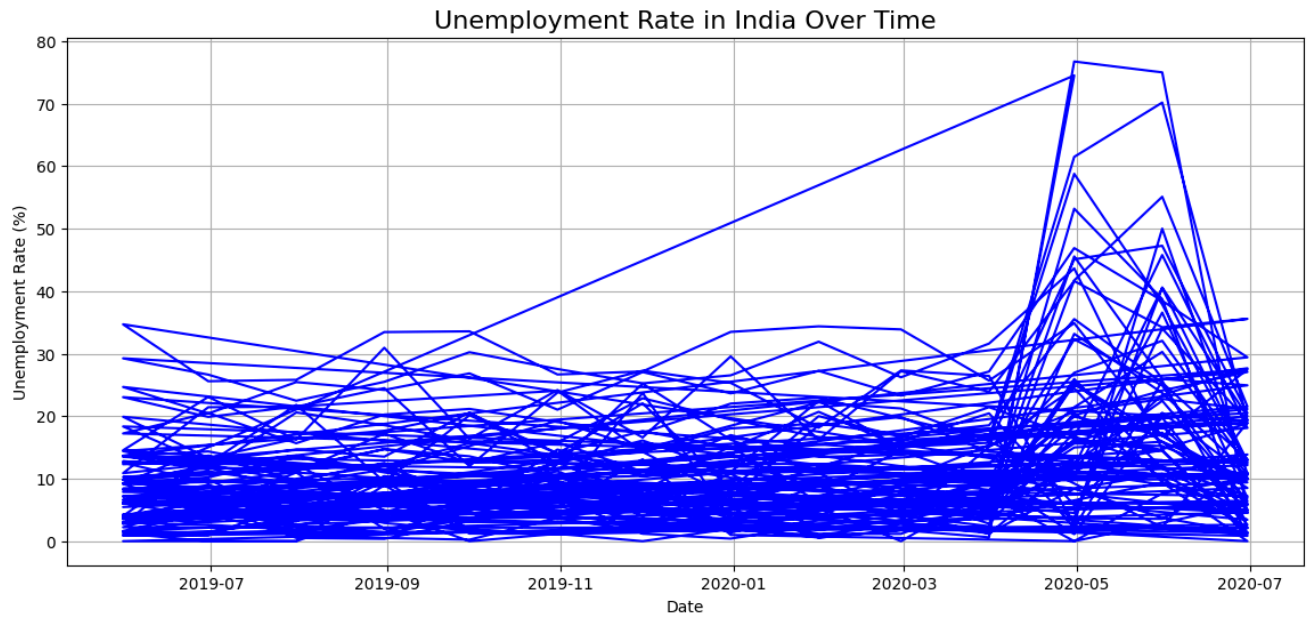
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

```
df.rename(columns={'Date ': 'Date'}, inplace=True)
```

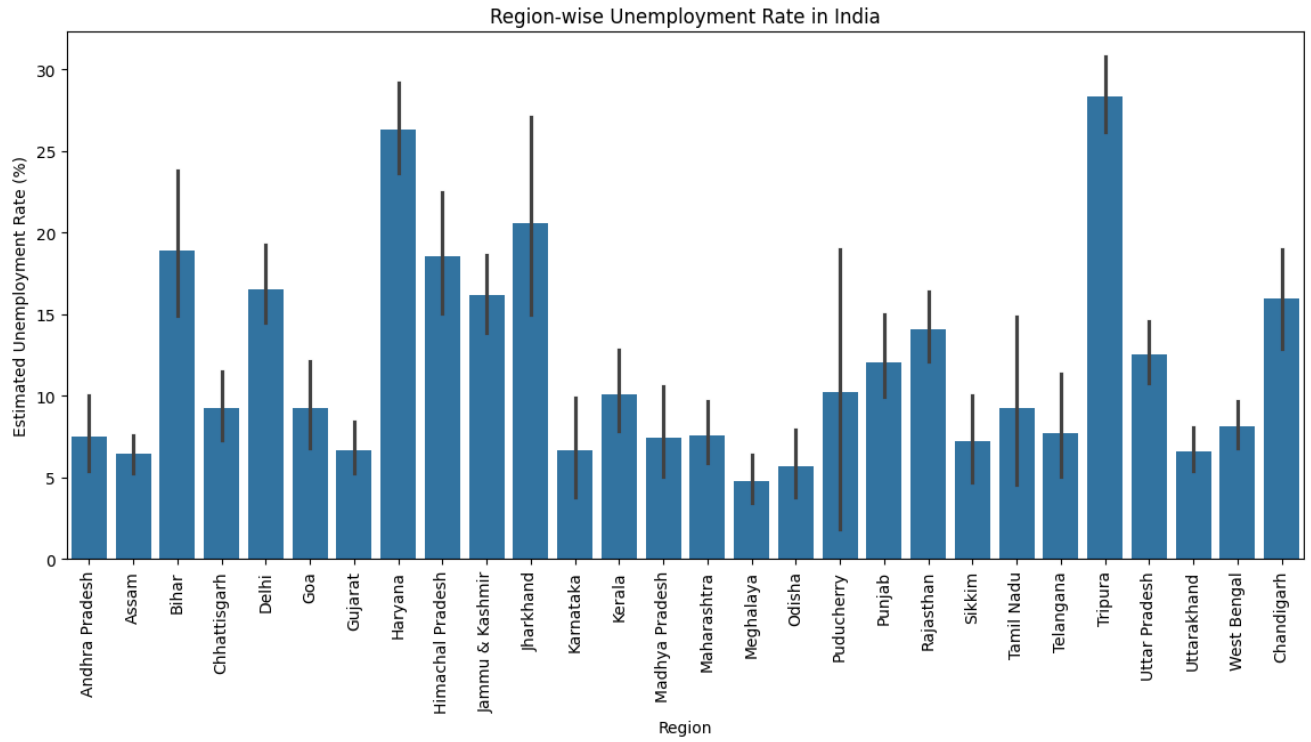
```
df.columns = df.columns.str.strip()
df['Date'] = pd.to_datetime(df['Date'])
```

```
/tmp/ipython-input-536003683.py:2: UserWarning: Parsing dates in  %d-%m-%Y format when dayfirst=False (the default) was specifie
  df['Date'] = pd.to_datetime(df['Date'])
```

```
plt.figure(figsize=(14,6))
plt.plot(df['Date'], df['Estimated Unemployment Rate (%)'], color='blue')
plt.title("Unemployment Rate in India Over Time", fontsize=16)
plt.xlabel("Date")
plt.ylabel("Unemployment Rate (%)")
plt.grid(True)
plt.show()
```

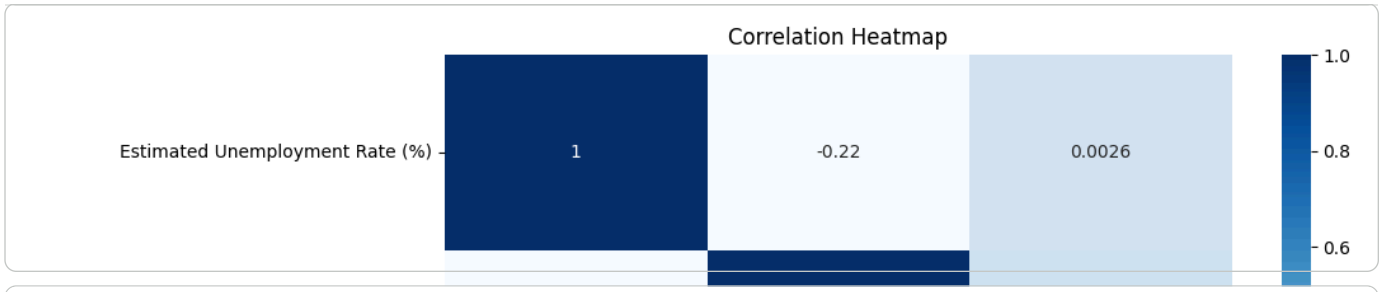



```
plt.figure(figsize=(14,6))
sns.barplot(data=df, x='Region', y='Estimated Unemployment Rate (%)')
plt.xticks(rotation=90)
plt.title("Region-wise Unemployment Rate in India")
plt.show()
```




```
# Select only numeric columns
numeric_df = df.select_dtypes(include=['int64', 'float64'])

plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='Blues')
plt.title("Correlation Heatmap")
plt.show()
```




```
plt.figure(figsize=(14,6))
sns.boxplot(data=df, x='Region', y='Estimated Unemployment Rate (%)')
plt.xticks(rotation=90)
plt.title("State-wise Variation in Unemployment Rate")
plt.show()
```