```
import pandas as pd
```

```
df = pd.read_csv('/content/Unemployment in India.csv')
```

```
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

```
df.info()
df.isnull().sum()
```

```
df.columns
```

```
df['Region'].unique()
```

```
<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
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',
       'Uttar Pradesh', 'Uttarakhand', 'West Bengal', nan, 'Chandigarh'],
      dtype=object)
```

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

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

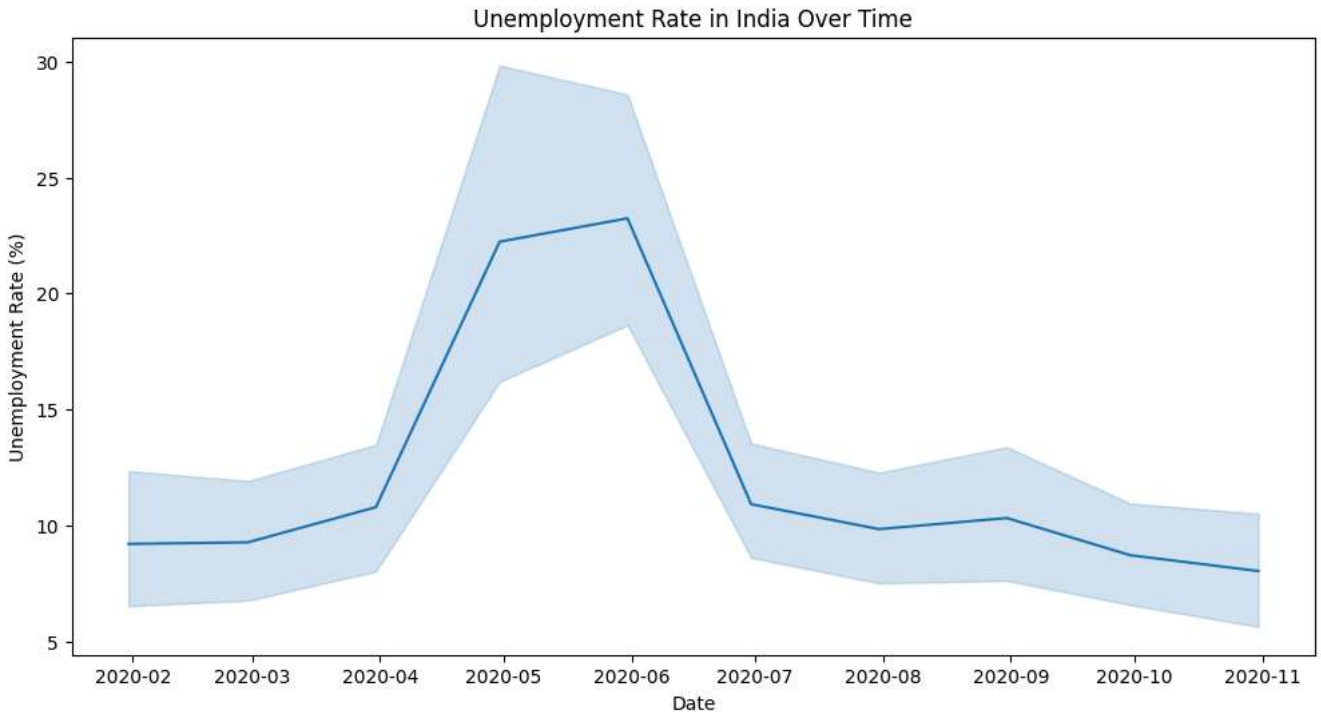
```
df.rename(columns=lambda x: x.strip(), inplace=True)
```

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

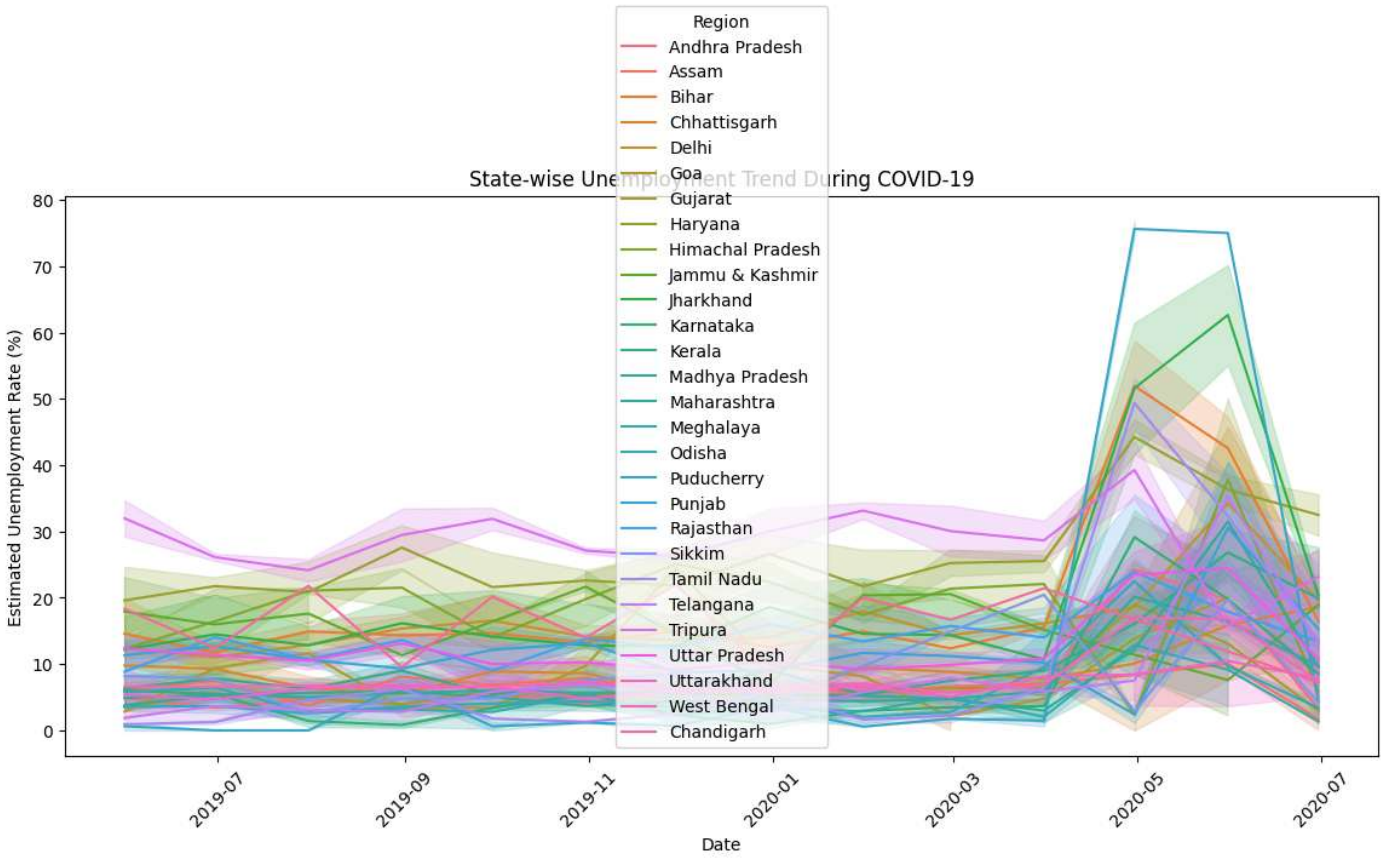
```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', data=df)
plt.title('Unemployment Rate in India Over Time')
```

```
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.show()
```



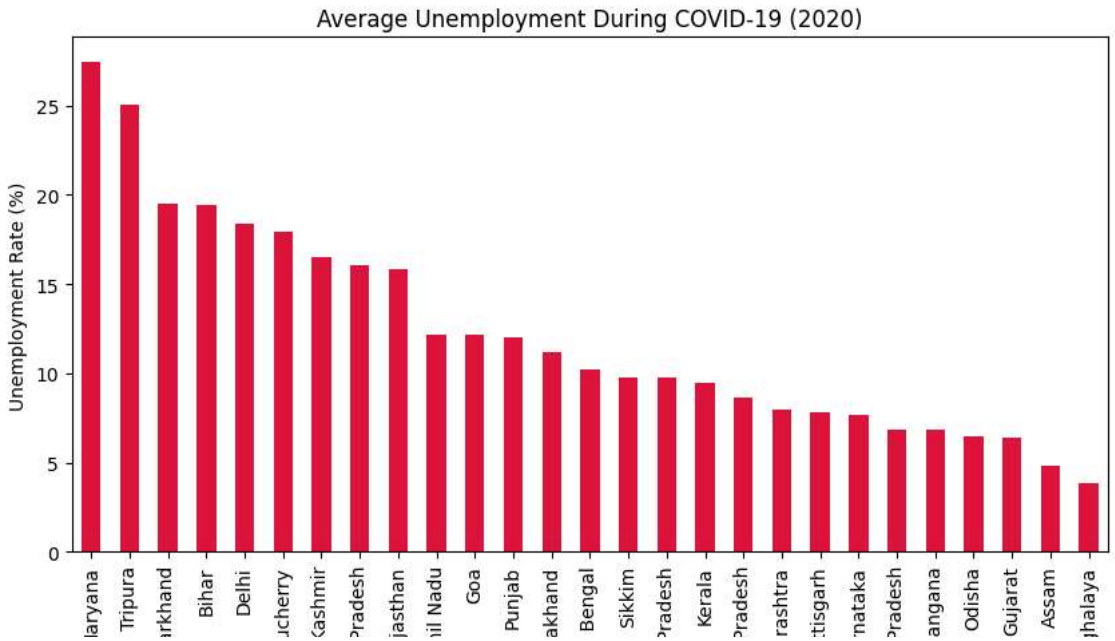

```
plt.figure(figsize=(14,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', hue='Region', data=df)
plt.title('State-wise Unemployment Trend During COVID-19')
plt.xticks(rotation=45)
plt.show()
```




```
covid_df = df[df['Date'].dt.year == 2020]


avg_2020 = covid_df.groupby('Region')['Estimated Unemployment Rate (%)'].mean().sort_values(ascending=False)

plt.figure(figsize=(10,5))
avg_2020.plot(kind='bar', color='crimson')
plt.title("Average Unemployment During COVID-19 (2020)")
plt.ylabel("Unemployment Rate (%)")
plt.xticks(rotation=90)
plt.show()
```




```
import pandas as pd

df = pd.read_csv('/content/Unemployment_Rate_upto_11_2020.csv')

print(df.columns)

df.head()
```

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

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.74
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.74
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.74

```
df.columns = df.columns.str.strip()

print(df.columns)
```

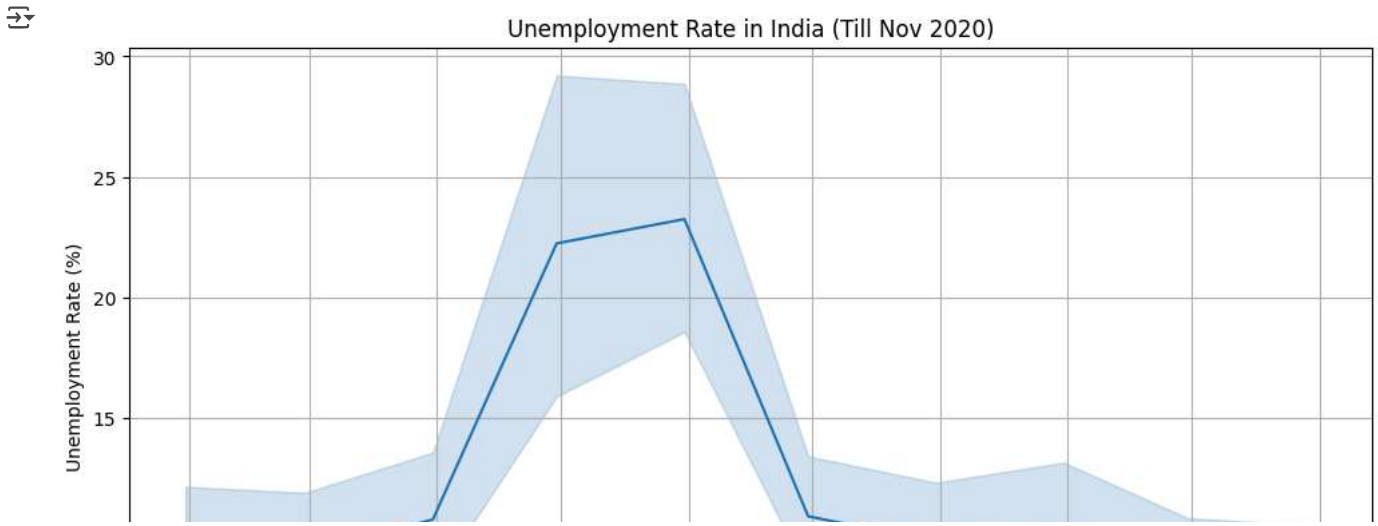
```
Index(['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)',
       'Estimated Employed', 'Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

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

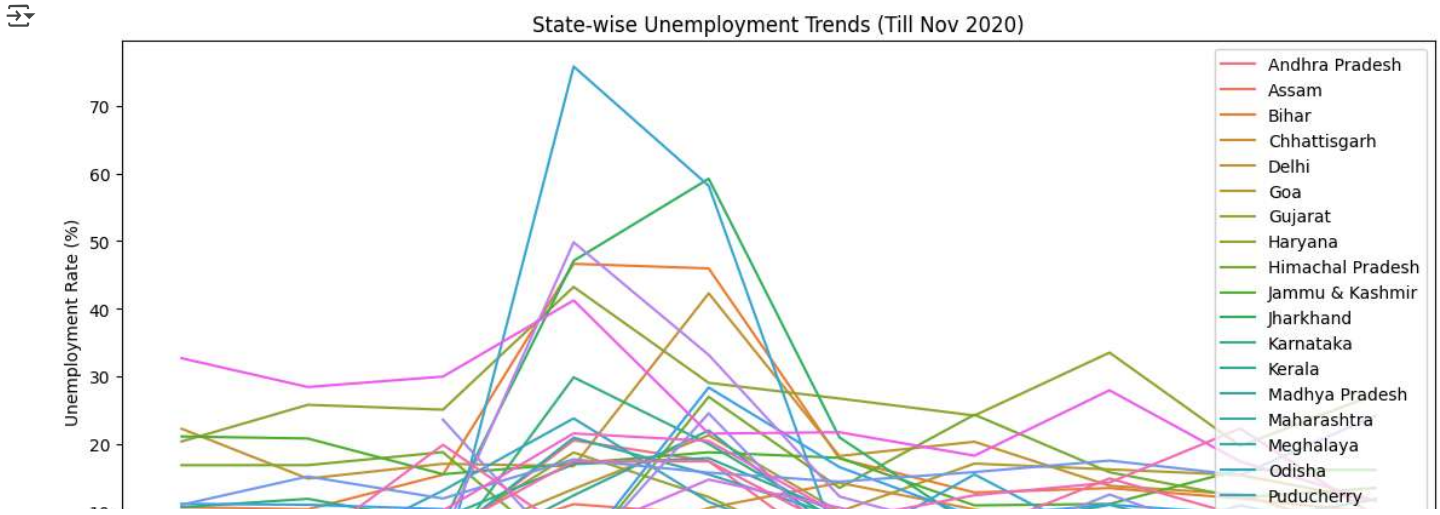
```
<ipython-input-23-b9c6864122be>:1: UserWarning: Parsing dates in  %d-%m-%Y format when dayfirst=False (the default) was specified. Pass
  df['Date'] = pd.to_datetime(df['Date'])
```

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', data=df)
plt.title('Unemployment Rate in India (Till Nov 2020)')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.grid(True)
plt.show()
```




```
plt.figure(figsize=(14,6))
sns.lineplot(x='Date', y='Estimated Unemployment Rate (%)', hue='Region', data=df)
plt.title('State-wise Unemployment Trends (Till Nov 2020)')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.xticks(rotation=45)
plt.legend(loc='upper right')
plt.show()
```




```
df_2020 = df[df['Date'].dt.year == 2020]
```