

Importing Libraries

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Data Collection

```
In [2]: df=pd.read_csv(r"D:\UNEMPLOYMENT_IN_INDIA_dataset.csv")
df
```

	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

754 rows × 7 columns

```
In [3]: df.shape
```

Out[3]: (754, 7)

```
In [4]: df.head(10)
```

	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
8	Andhra Pradesh	31-01-2020	Monthly	4.84	12016676.0	43.46	Rural
9	Andhra Pradesh	29-02-2020	Monthly	5.91	11723617.0	42.83	Rural

```
In [5]: df.tail()
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
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

EDA Before Preprocessing

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

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...
749	False	False	False	False	False	False	False
750	False	False	False	False	False	False	False
751	False	False	False	False	False	False	False
752	False	False	False	False	False	False	False
753	False	False	False	False	False	False	False

754 rows × 7 columns

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 754 entries, 0 to 753
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: 41.4+ KB
```

```
In [8]: df.isnull().sum()
```

Region	14
Date	14
Frequency	14
Estimated Unemployment Rate (%)	14
Estimated Employed	14
Estimated Labour Participation Rate (%)	14
Area	14
dtype: int64	

```
In [9]: df.describe()
```

	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

```
In [10]: x = df.iloc[10:20]
x
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
10	Andhra Pradesh	31-03-2020	Monthly	4.06	11359660.0	40.66	Rural
11	Andhra Pradesh	30-04-2020	Monthly	16.29	8792827.0	36.03	Rural
12	Andhra Pradesh	31-05-2020	Monthly	14.46	9526902.0	38.16	Rural
13	Andhra Pradesh	30-06-2020	Monthly	0.85	15572975.0	53.76	Rural
14	Assam	31-05-2019	Monthly	4.29	11749334.0	57.39	Rural
15	Assam	30-06-2019	Monthly	5.08	8923222.0	43.87	Rural
16	Assam	31-07-2019	Monthly	4.26	9911534.0	48.21	Rural
17	Assam	31-08-2019	Monthly	5.79	9292039.0	45.83	Rural
18	Assam	30-09-2019	Monthly	4.46	11468349.0	55.67	Rural
19	Assam	31-10-2019	Monthly	4.65	8395906.0	40.76	Rural

```
In [11]: df.dropna(inplace=True)
```

```
In [12]: df.isnull().sum()
```

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

```
In [13]: df.shape
```

Out[13]: (740, 7)

```
In [14]: df.columns
```

Out[14]: Index(['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed', 'Estimated Labour Participation Rate (%)', 'Area'], dtype='object')

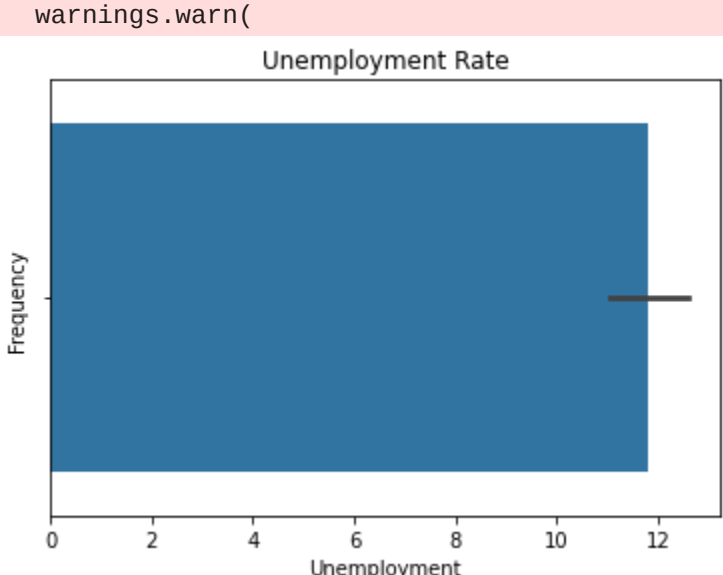
```
In [15]: pd.DataFrame(df.iloc[1:,3])
```

	Estimated Unemployment Rate (%)
1	3.05
2	3.75
3	3.32
4	5.17
5	3.52
...	...
749	7.55
750	6.67
751	15.63
752	15.22
753	9.86

739 rows × 1 columns

```
In [16]: sns.barplot(df.iloc[1:,3])
plt.title("Unemployment Rate")
plt.xlabel("Unemployment")
plt.ylabel("Frequency")
plt.show()
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From versi on 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or m isinterpretation.
warnings.warn(



```
In [18]: sns.histplot(df.iloc[1:,3])
plt.title("Unemployment Rate")
plt.xlabel("Unemployment Rate %")
plt.ylabel("Frequency")
plt.show()
```



Preprocessing

```
In [28]: df.drop(['Region', 'Frequency', 'Area'], axis=1, inplace=True)
```

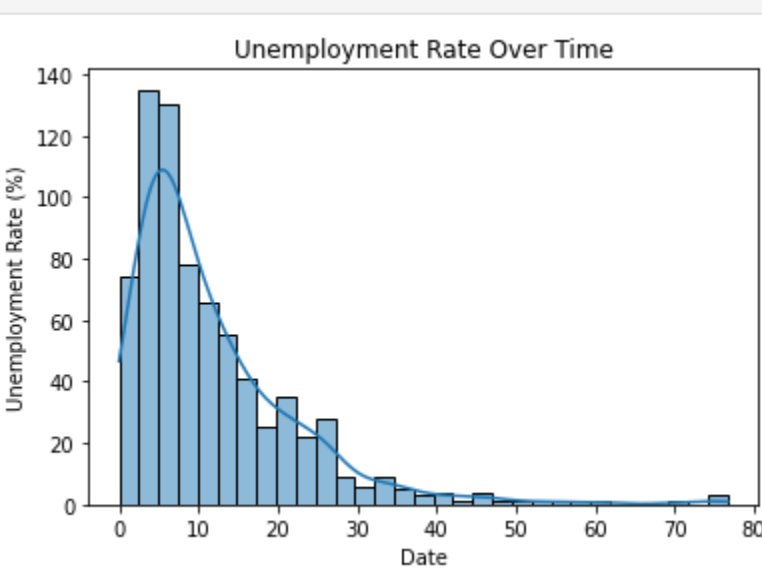
```
In [30]: df['Date'] = pd.to_datetime(df['Date'])
```

```
In [33]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 740 entries, 0 to 753
Data columns (total 4 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Date                                  740 non-null    datetime64[ns]
1   Estimated Unemployment Rate (%)       740 non-null    float64
2   Estimated Employed                    740 non-null    float64
3   Estimated Labour Participation Rate (%) 740 non-null    float64
dtypes: datetime64[ns](1), float64(3)
memory usage: 45.1 KB
```

```
In [34]: df.set_index('Date', inplace=True)
```

```
In [38]: sns.histplot(df['Estimated Unemployment Rate (%)'], kde=True)
plt.title('Unemployment Rate Over Time')
plt.xlabel('Date')
plt.ylabel('Unemployment Rate (%)')
plt.show()
```



```
In [ ]:
```

```
In [ ]:
```