UNEMPLOYMENT ANALYSIS USING PYTHON

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.

Import required modules

In [28]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
%matplotlib inline
```

Loading UNEMPLOYMENT Dataset

In [3]:
```
data = pd.read_csv('Unemployment_2020.csv')
data
```

Out[3]:

	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	79.740
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740
...	...	...	...	...	...	...	...	...	...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

267 rows × 9 columns

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

Out[4]:

	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
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.74
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.74

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

Out[5]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

In [7]:
```
data.describe()
```

Out[7]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	longitude	latitude
count	267.000000	2.670000e+02	267.000000	267.000000	267.000000
mean	12.236929	1.396211e+07	41.681573	22.826048	80.532425
std	10.803283	1.336632e+07	7.845419	6.270731	5.831738
min	0.500000	1.175420e+05	16.770000	10.850500	71.192400
25%	4.845000	2.839930e+06	37.265000	18.112400	76.085600
50%	9.650000	9.732417e+06	40.390000	23.610200	79.019300
75%	16.755000	2.187959e+07	44.055000	27.278400	85.279900
max	75.850000	5.943376e+07	69.690000	33.778200	92.937600

In [8]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 9 columns):
 #   Column                                   Non-Null Count  Dtype  
---  ------                                   --------------  -----  
 0   Region                                   267 non-null    object 
 1   Date                                     267 non-null    object 
 2   Frequency                                267 non-null    object 
 3   Estimated Unemployment Rate (%)          267 non-null    float64
 4   Estimated Employed                       267 non-null    int64  
 5   Estimated Labour Participation Rate (%)  267 non-null    float64
 6   Region.1                                 267 non-null    object 
 7   longitude                                267 non-null    float64
 8   latitude                                 267 non-null    float64
dtypes: float64(4), int64(1), object(4)
memory usage: 18.9+ KB
```

In [13]:
```
data.size
```

Out[13]:
```
2403
```

In [15]:
```
data.shape
```

Out[15]:
```
(267, 9)
```

In [9]:
```
data.columns
```

Out[9]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

In [10]:
```
data['Region'].value_counts()
```

Out[10]:
```
Jharkhand           10
Uttar Pradesh       10
Maharashtra         10
Goa                 10
Telangana           10
Madhya Pradesh      10
Bihar               10
Gujarat             10
Punjab              10
Tamil Nadu          10
Meghalaya           10
Puducherry          10
Assam               10
Himachal Pradesh    10
Tripura             10
Rajasthan           10
Odisha              10
Karnataka           10
Chhattisgarh        10
Delhi               10
Kerala              10
West Bengal         10
Andhra Pradesh      10
Uttarakhand         10
Haryana             10
Jammu & Kashmir      9
Sikkim               8
Name: Region, dtype: int64
```

Data Preprocessing

In [11]:
```
data.isnull().sum() #glad no null values
```

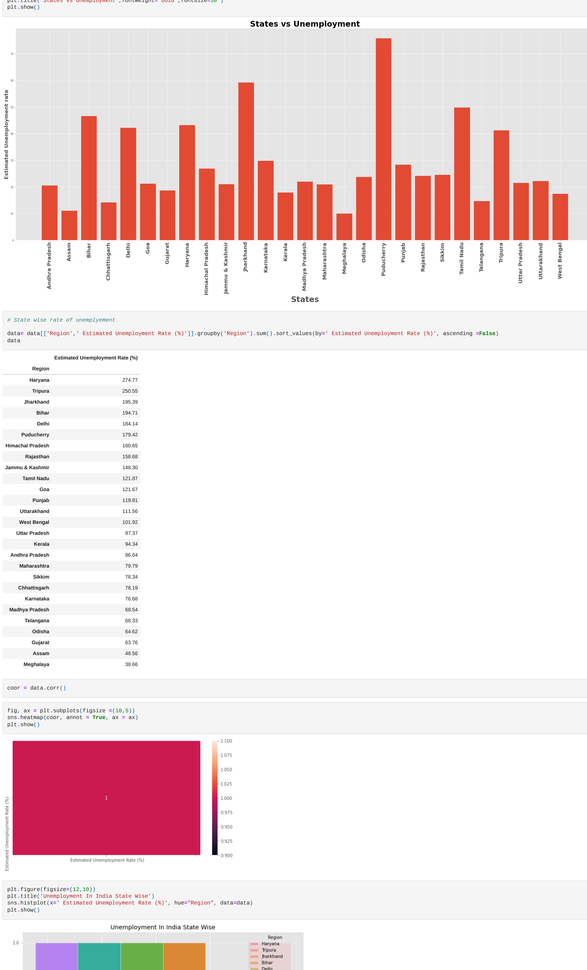
Out[11]:
```
Region                                     0
 Date                                      0
 Frequency                                 0
 Estimated Unemployment Rate (%)           0
 Estimated Employed                        0
 Estimated Labour Participation Rate (%)   0
Region.1                                   0
longitude                                  0
latitude                                   0
dtype: int64
```

In [17]:
```
#for checking duplicacy
data.duplicated().sum() #glad no duplicates in data
```

Out[17]:
```
0
```

Data Visualization

In [38]:
```
style.use('ggplot')
color = ['pink', 'yellow','orange', 'blue','royalblue','pink', 'yellow', 'violet', 'blue','red','blue','pink', 'yellow','orange', 'blue','violet', 'blue','pink', 'yellow','royalblu
data =pd.DataFrame(data)
y=data[' Estimated Unemployment Rate (%)']
x=data['Region']
pr= plt.figure(figsize=(40, 15))
plt.xlabel("States",fontweight='bold',fontsize=30)
plt.ylabel("Estimated Unemployment rate",fontweight='bold',fontsize=20)
plt.xticks(fontweight='bold',rotation='vertical',fontsize=20)
plt.bar(x,y,align='center')
plt.title('States vs Unemployment',fontweight='bold',fontsize=30 )
plt.show()
```



In [39]:
```
# State wise rate of unemployment

data= data[['Region',' Estimated Unemployment Rate (%)']].groupby('Region').sum().sort_values(by=' Estimated Unemployment Rate (%)', ascending =False)
data
```

Out[39]:

Region	Estimated Unemployment Rate (%)
Haryana	274.77
Tripura	250.55
Jharkhand	195.39
Bihar	194.71
Delhi	184.14
Puducherry	179.42
Himachal Pradesh	160.65
Rajasthan	158.68
Jammu & Kashmir	148.30
Tamil Nadu	121.87
Goa	121.67
Punjab	119.81
Uttarakhand	111.56
West Bengal	101.92
Uttar Pradesh	97.37
Kerala	94.34
Andhra Pradesh	86.64
Maharashtra	79.79
Sikkim	70.34
Chhattisgarh	78.19
Karnataka	76.68
Madhya Pradesh	68.54
Telangana	68.33
Odisha	64.62
Gujarat	63.76
Assam	48.56
Meghalaya	38.66

In [54]:
```
coor = data.corr()
```

In [55]:
```
fig, ax = plt.subplots(figsize =(10,5))
sns.heatmap(coor, annot = True, ax = ax)
plt.show()
```


In [59]:
```
plt.figure(figsize=(12,10))
plt.title('Unemployment In India State Wise')
sns.histplot(x=' Estimated Unemployment Rate (%)', hue="Region", data=data)
plt.show()
```