

Oasis Infobyte

Data Science Intern

Author : Gajjala Devendra Reddy

Project Name: 

Unemploment Analysis with python

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

In [3]:

```
data=pd.read_csv("C://Users//gajja//Downloads//Unemployment in India.csv")
data1=pd.read_csv("C://Users//gajja//Downloads//Unemployment_Rate_upto_11_2020.csv")
```

In [4]:

```
print(data.head())
```

	Region	Date	Frequency	Estimated Unemployment Rate
(%) \				
0	Andhra Pradesh	31-05-2019	Monthly	3.6
5				
1	Andhra Pradesh	30-06-2019	Monthly	3.0
5				
2	Andhra Pradesh	31-07-2019	Monthly	3.7
5				
3	Andhra Pradesh	31-08-2019	Monthly	3.3
2				
4	Andhra Pradesh	30-09-2019	Monthly	5.1
7				

	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	11999139.0	43.24	Rural
1	11755881.0	42.05	Rural
2	12086707.0	43.50	Rural
3	12285693.0	43.97	Rural
4	12256762.0	44.68	Rural

In [5]:

```
print(data.info())
```

```
<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
None
```

In [6]:

```
print(data1.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
None
```

In [7]:

```
print(data1.head())
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)
0	Andhra Pradesh	31-01-2020	M	5.4
1	Andhra Pradesh	29-02-2020	M	5.8
2	Andhra Pradesh	31-03-2020	M	5.7
3	Andhra Pradesh	30-04-2020	M	20.5
4	Andhra Pradesh	31-05-2020	M	17.4

	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1
0	16635535	41.02	South
1	16545652	40.90	South
2	15881197	39.18	South
3	11336911	33.10	South
4	12988845	36.46	South

	longitude	latitude
0	15.9129	79.74
1	15.9129	79.74
2	15.9129	79.74
3	15.9129	79.74
4	15.9129	79.74

In [8]:

```
print(data.describe())
```

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

	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

In [58]:

```
print(data1.describe())
```

	Estimated Unemployment Rate (%)	Estimated Employed \	
count	267.000000	2.670000e+02	
mean	12.236929	1.396211e+07	
std	10.803283	1.336632e+07	
min	0.500000	1.175420e+05	
25%	4.845000	2.838930e+06	
50%	9.650000	9.732417e+06	
75%	16.755000	2.187869e+07	
max	75.850000	5.943376e+07	

	Estimated Labour Participation Rate (%)	longitude	latitude
count	267.000000	267.000000	267.000000
mean	41.681573	22.826048	80.532425
std	7.845419	6.270731	5.831738
min	16.770000	10.850500	71.192400
25%	37.265000	18.112400	76.085600
50%	40.390000	23.610200	79.019300
75%	44.055000	27.278400	85.279900
max	69.690000	33.778200	92.937600

In [10]:

```
print(data.isnull().sum())
```

```
Region      28
Date        28
Frequency   28
Estimated Unemployment Rate (%)  28
Estimated Employed      28
Estimated Labour Participation Rate (%)  28
Area            28
dtype: int64
```

In [11]:

```
print(data1.isnull().sum())
```

```
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 [12]:

```
print(data.fillna(0))
```

	Region	Date	Frequency	Estimated Unemployment Rate			
(%) \							
0	Andhra Pradesh	31-05-2019	Monthly				
3.65							
1	Andhra Pradesh	30-06-2019	Monthly				
3.05							
2	Andhra Pradesh	31-07-2019	Monthly				
3.75							
3	Andhra Pradesh	31-08-2019	Monthly				
3.32							
4	Andhra Pradesh	30-09-2019	Monthly				
5.17							
..				
...							
763	0	0	0				
0.00							
764	0	0	0				
0.00							
765	0	0	0				
0.00							
766	0	0	0				
0.00							
767	0	0	0				
0.00							
	Estimated Employed	Estimated Labour Participation Rate (%)	Area				
0	11999139.0	43.24	Rural				
1	11755881.0	42.05	Rural				
2	12086707.0	43.50	Rural				
3	12285693.0	43.97	Rural				
4	12256762.0	44.68	Rural				
..				
763	0.0	0.00	0				
764	0.0	0.00	0				
765	0.0	0.00	0				
766	0.0	0.00	0				
767	0.0	0.00	0				

[768 rows x 7 columns]

In [12]:

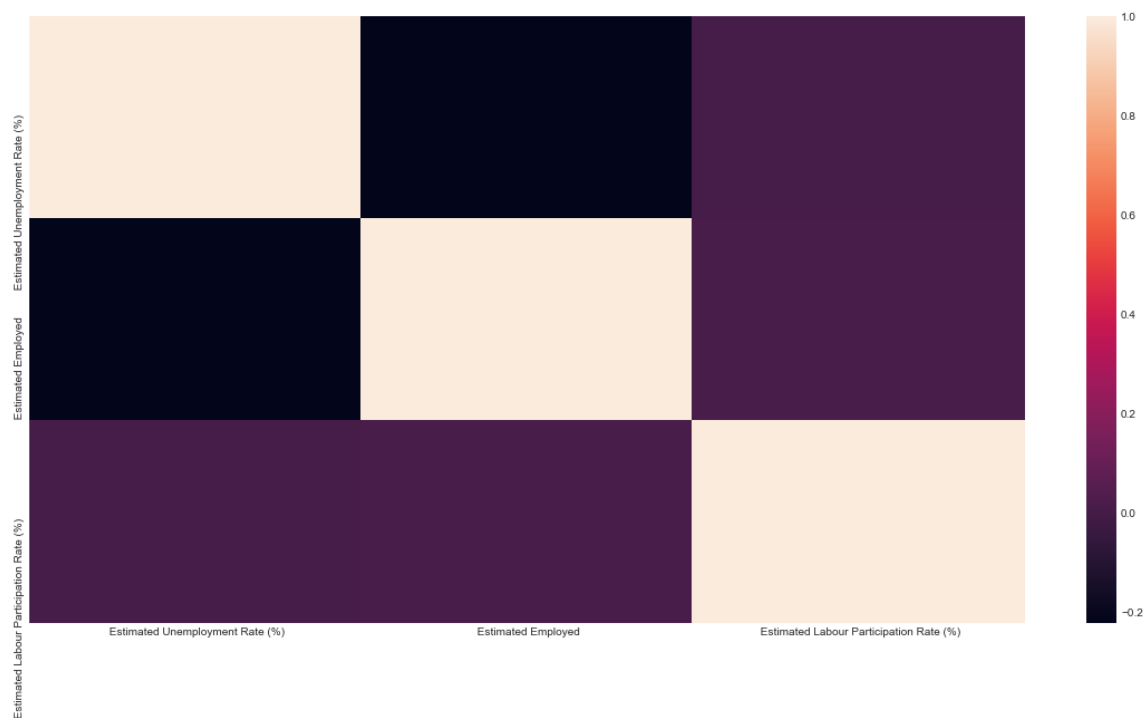
```
data.columns=["States","Date","Freq","Estimated Unemployment Rate","Estimated Employed",  
             "Estimated Labour Participation Rate","Region"]
```

In [13]:

```
data1.columns=["States","Date","Freq","Estimated Unemployment Rate","Estimated Employed",  
              "Estimated Labour Participation Rate","Region","Longi","Lati"]
```

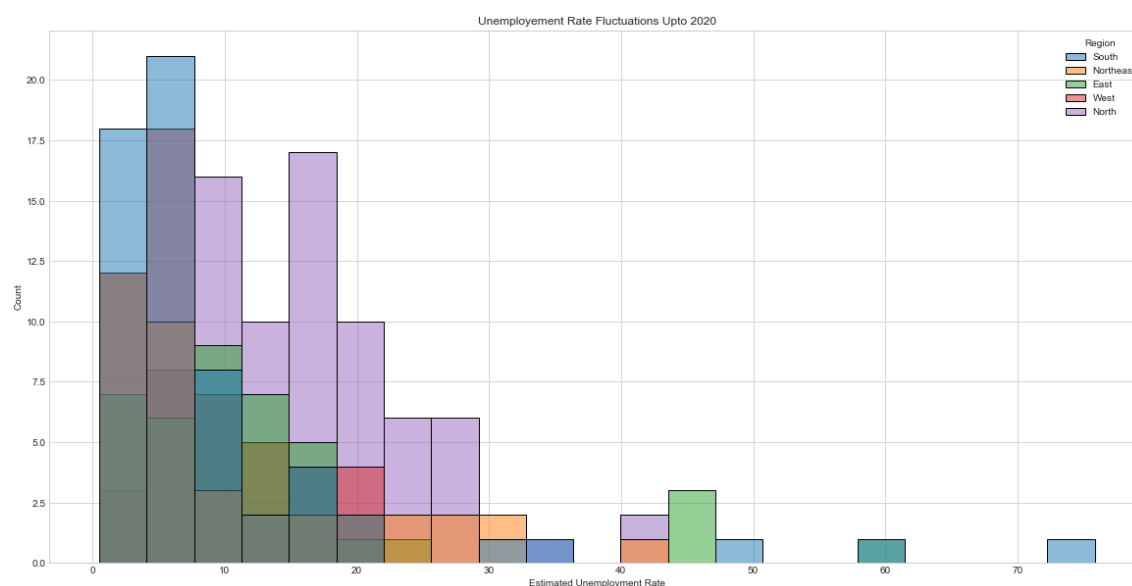
In [9]:

```
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(20, 10))
sns.heatmap(data.corr())
plt.show()
```



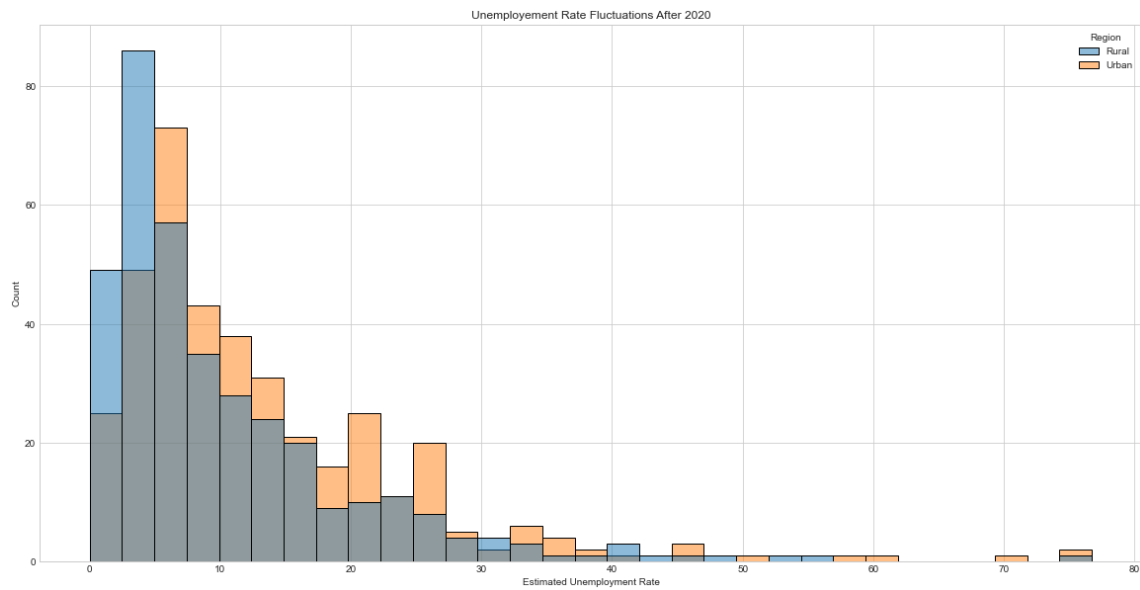
In [14]:

```
plt.figure(figsize=(20, 10))
plt.title("Unemployment Rate Fluctuations Upto 2020")
sns.histplot(data1,x="Estimated Unemployment Rate",hue="Region");
```



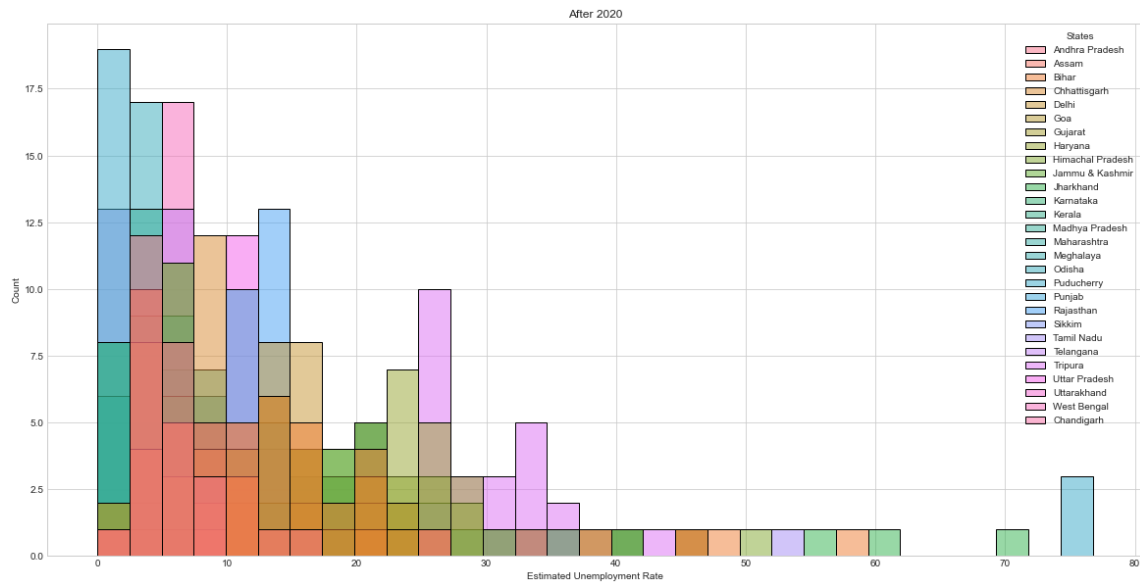
In [15]:

```
plt.figure(figsize=(20, 10))
plt.title("Unemployment Rate Fluctuations After 2020")
sns.histplot(data,x="Estimated Unemployment Rate",hue="Region");
```



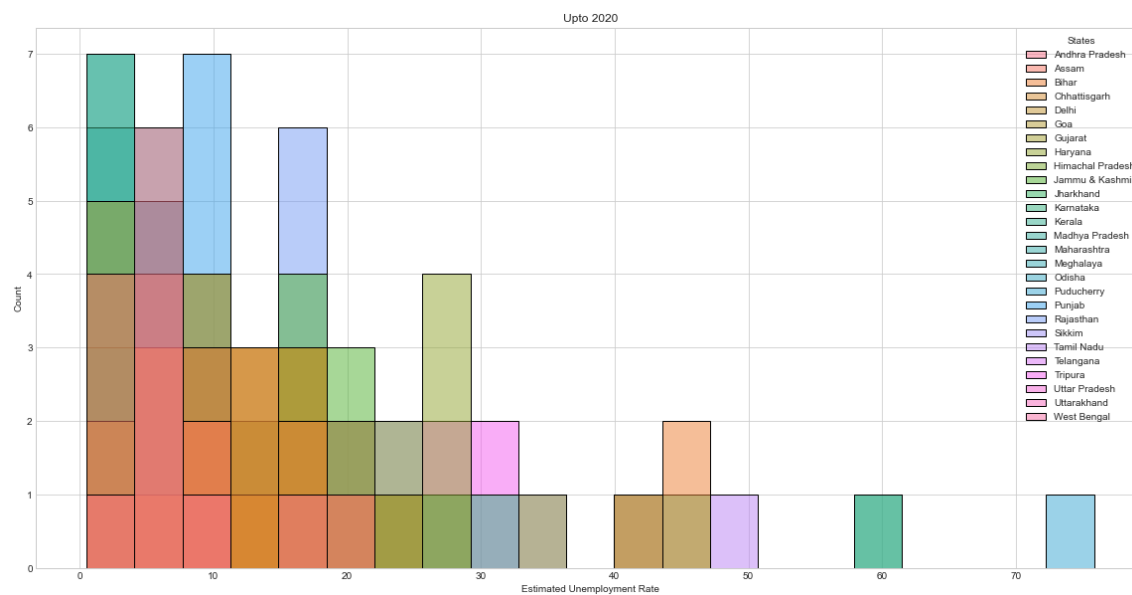
In [16]:

```
plt.figure(figsize=(20, 10))
plt.title("After 2020")
sns.histplot(data,x="Estimated Unemployment Rate",hue="States");
```



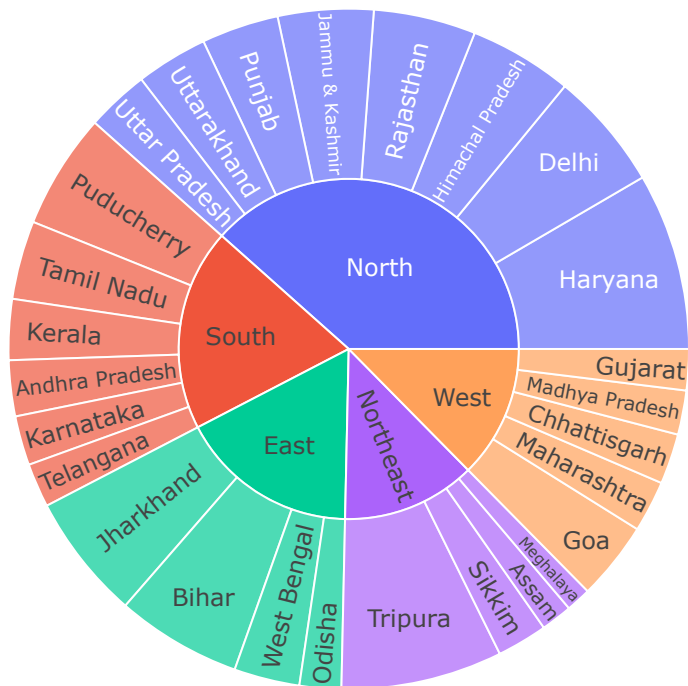
In [38]:

```
plt.figure(figsize=(20, 10))  
plt.title("Upto 2020")  
sns.histplot(data1,x="Estimated Unemployment Rate",hue="States");
```



In [17]:

```
emp = data1[["Estimated Unemployment Rate","States","Region"]]  
fig = px.sunburst(emp,path=["Region","States"],values="Estimated Unemployment Rate",width  
fig.show()
```



In []: