

```
In [28]: #importing required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline
```

```
In [34]: #uploading the data
data_set1=pd.read_csv(r"C:\Users\DB_Tech\Documents\Unemployment in India.csv")
```

```
In [35]: data_set1.head()
```

Out[35]:

	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

```
In [36]: data_set1.tail()
```

Out[36]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
In [37]: data_set1.shape
```

Out[37]: (768, 7)

```
In [38]: data_set1.describe()
```

Out[38]:

	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 [39]: data_set1.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
```

In []: #Data cleaning and Modifying

In [40]: data_set1.isnull().sum() #finding the missing value

```
Out[40]: Region          28
          Date           28
          Frequency       28
          Estimated Unemployment Rate (%) 28
          Estimated Employed      28
          Estimated Labour Participation Rate (%) 28
          Area            28
          dtype: int64
```

In [41]: data_set1.isnull().sum().sum() #Total missing values

Out[41]: 196

In [42]: data_set=data_set1.fillna(method='pad') #filling the null values with the previous value
data_set

Out[42]:

	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
...
763	West Bengal	30-06-2020	Monthly	9.86	9088931.0		37.57 Urban
764	West Bengal	30-06-2020	Monthly	9.86	9088931.0		37.57 Urban
765	West Bengal	30-06-2020	Monthly	9.86	9088931.0		37.57 Urban
766	West Bengal	30-06-2020	Monthly	9.86	9088931.0		37.57 Urban
767	West Bengal	30-06-2020	Monthly	9.86	9088931.0		37.57 Urban

768 rows × 7 columns

In [43]: data_set.isnull().sum()

```
Out[43]: Region          0
          Date           0
          Frequency       0
          Estimated Unemployment Rate (%) 0
          Estimated Employed      0
          Estimated Labour Participation Rate (%) 0
          Area            0
          dtype: int64
```

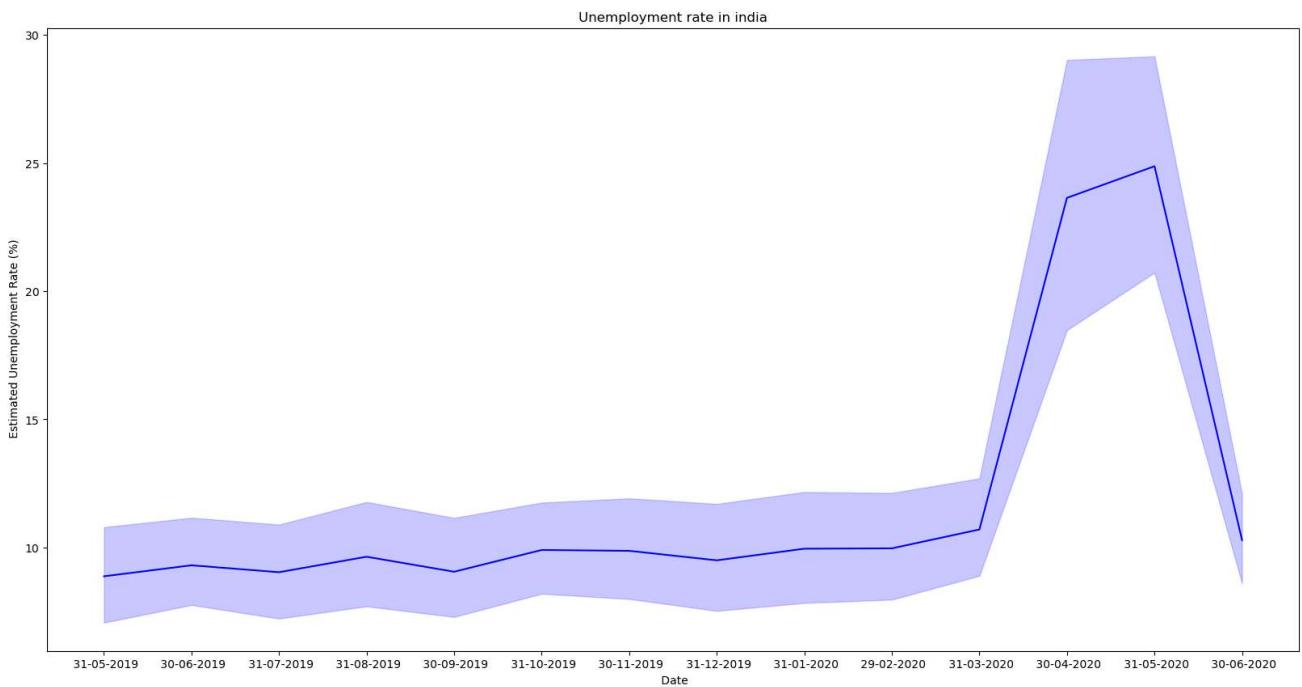
In [44]: print(data_set.columns.tolist())

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

In []: #Data Analysis Part

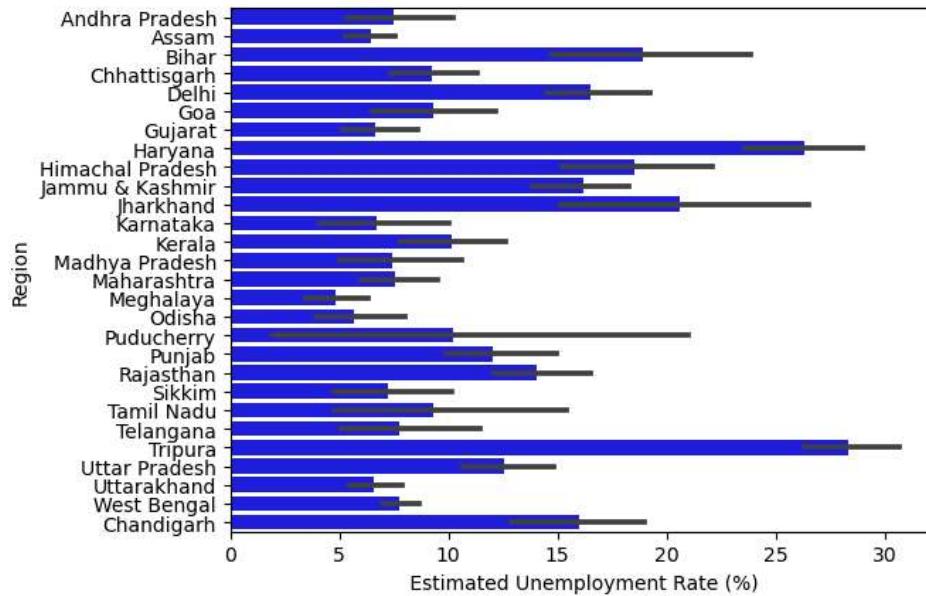
```
In [45]: plt.figure(figsize=(20,10)) #Unemployment rate in india
sns.lineplot(data=data_set,x=' Date',y=' Estimated Unemployment Rate (%)',color='blue')
plt.title("Unemployment rate in india")
```

Out[45]: Text(0.5, 1.0, 'Unemployment rate in india')



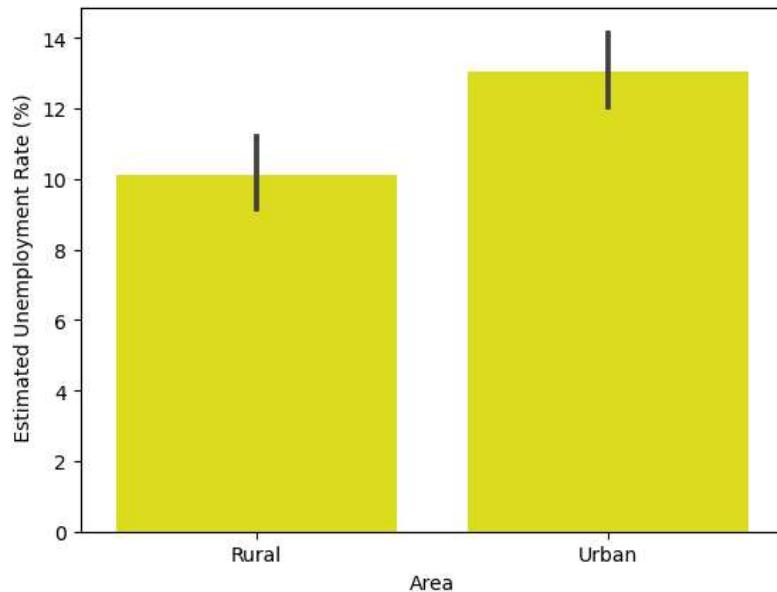
```
In [47]: sns.barplot(data=data_set,x=' Estimated Unemployment Rate (%)',y='Region',color='blue')
```

Out[47]: <Axes: xlabel=' Estimated Unemployment Rate (%)', ylabel='Region'>

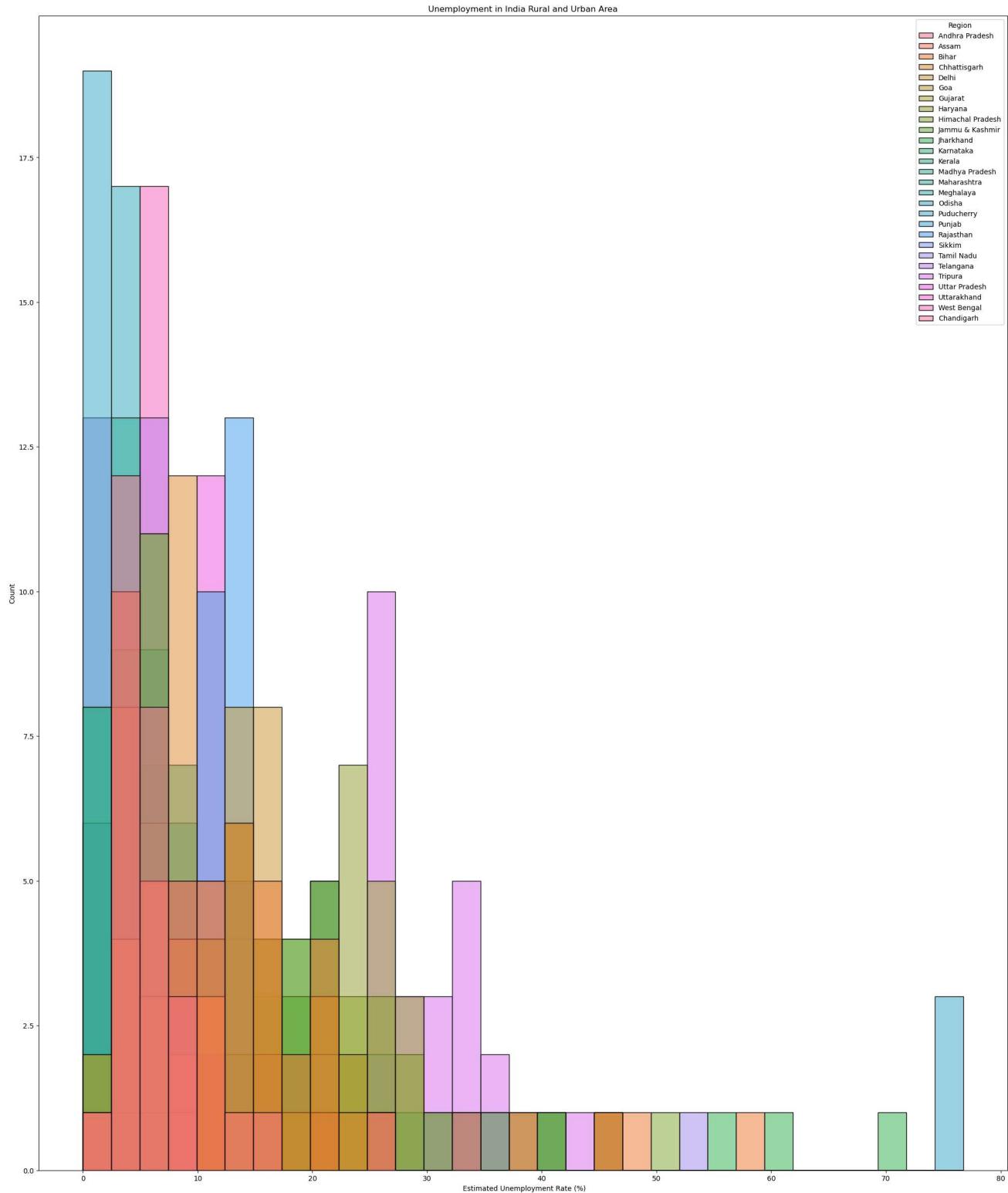


```
In [48]: sns.barplot(data=data_set,x='Area',y=' Estimated Unemployment Rate (%)',color='yellow')
```

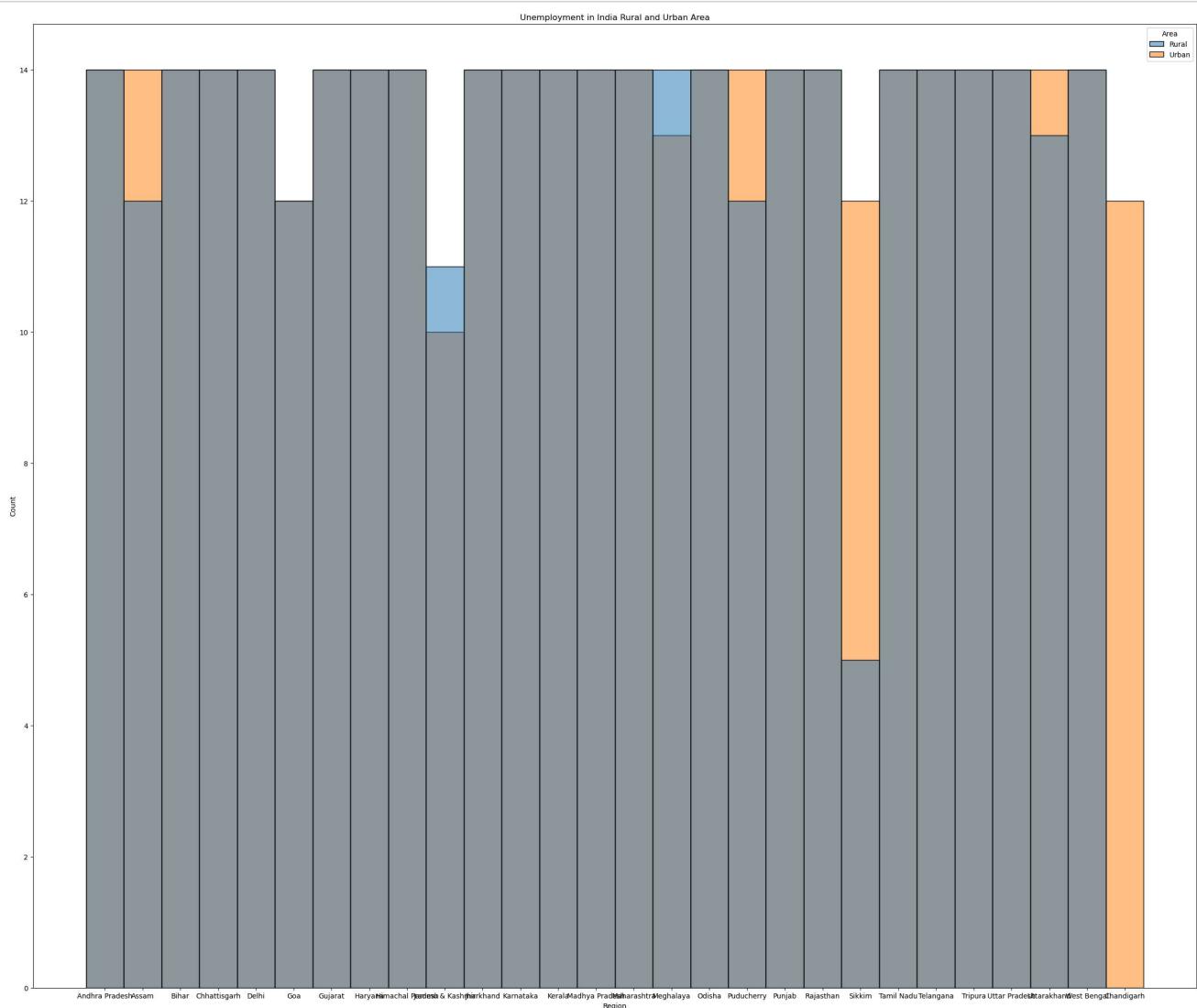
```
Out[48]: <Axes: xlabel='Area', ylabel=' Estimated Unemployment Rate (%)'>
```



```
In [49]: plt.figure(figsize=(25,30))
data_set.columns=['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed', 'Estimated Unemployed']
plt.title("Unemployment in India Rural and Urban Area")
sns.histplot(x='Estimated Unemployment Rate (%)',hue='Region',data=data_set)
plt.show()
```



```
In [50]: plt.figure(figsize=(30,25))
data_set.columns=['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed', 'Estimated Unemployed']
plt.title("Unemployment in India Rural and Urban Area")
sns.histplot(x='Region',hue='Area',data=data_set)
plt.show()
```



```
In [51]: sns.heatmap(data_set.corr(), annot=True)
```

C:\Users\DB Tech\AppData\Local\Temp\ipykernel_12012\3060311531.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
 sns.heatmap(data_set.corr(), annot=True)

Out[51]: <Axes: >



```
In [ ]: #Unemployment Rate upto 1/11/2020(November)
```

```
In [52]: data_set2=pd.read_csv(r"C:\Users\DB Tech\Documents\Unemployment_Rate_upto_11_2020.csv")
```

```
In [53]: data_set2.head()
```

Out[53]:

	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 [54]: `data_set2.tail()`

Out[54]:

	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 [55]: `print(data_set2.columns.tolist())`

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

In [56]: `data_set2.isnull().sum()`

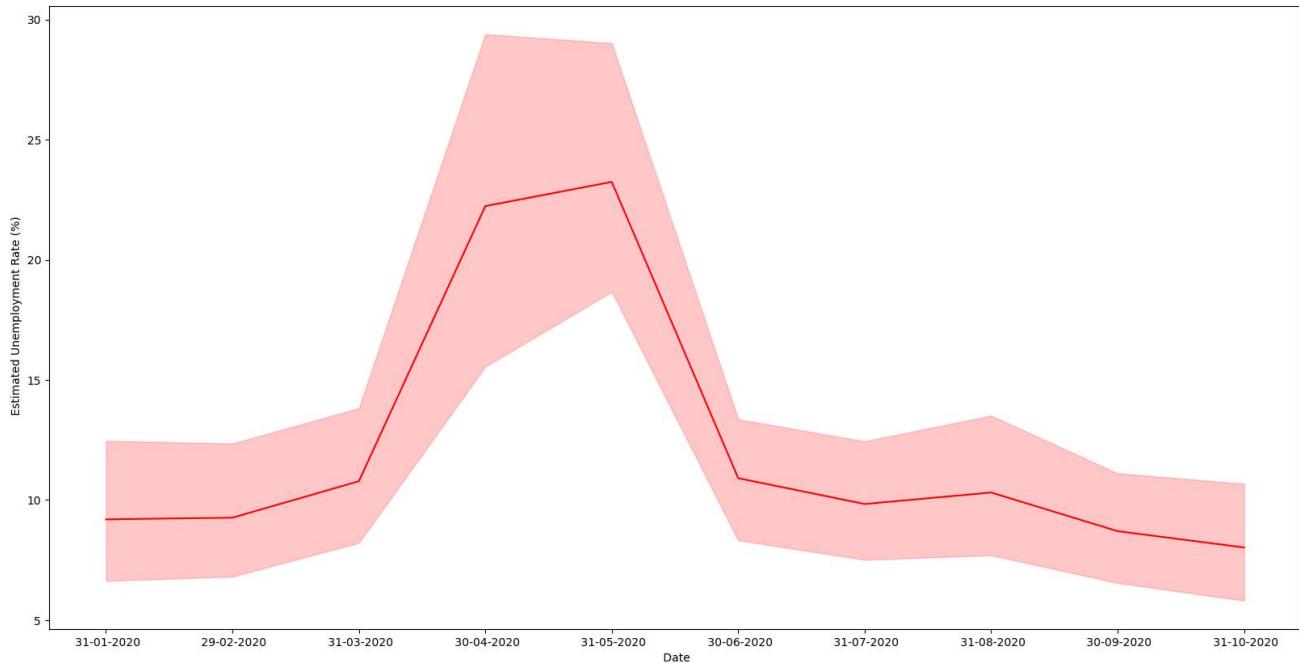
```
Out[56]: Region          0
         Date           0
         Frequency       0
         Estimated Unemployment Rate (%) 0
         Estimated Employed      0
         Estimated Labour Participation Rate (%) 0
         Region.1        0
         longitude       0
         latitude        0
         dtype: int64
```

In []: #Data Analysis Part

In [57]: #Unemployment rate in india

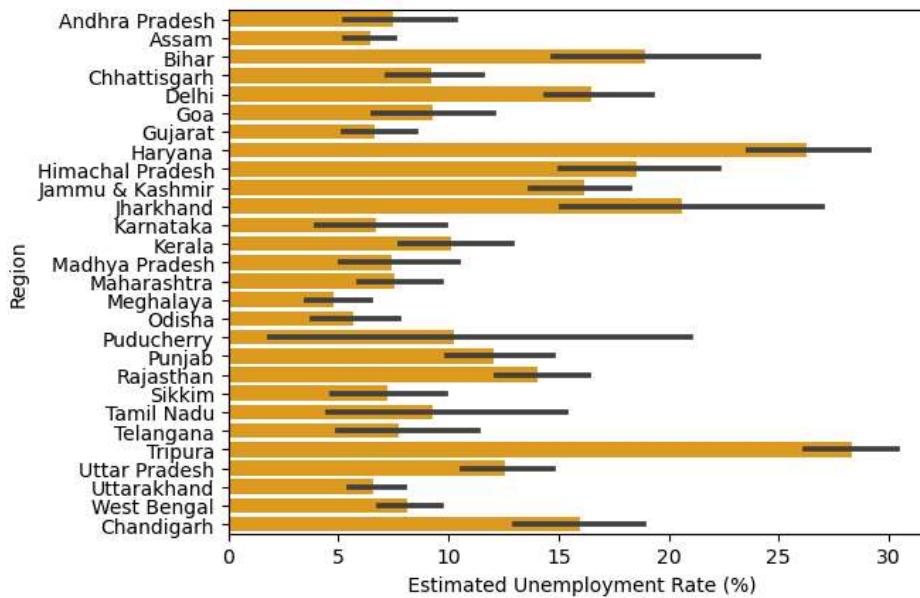
```
plt.figure(figsize=(20,10))
sns.lineplot(data=data_set2,x=' Date',y=' Estimated Unemployment Rate (%)',color='red')
```

Out[57]: <Axes: xlabel=' Date', ylabel=' Estimated Unemployment Rate (%)'>



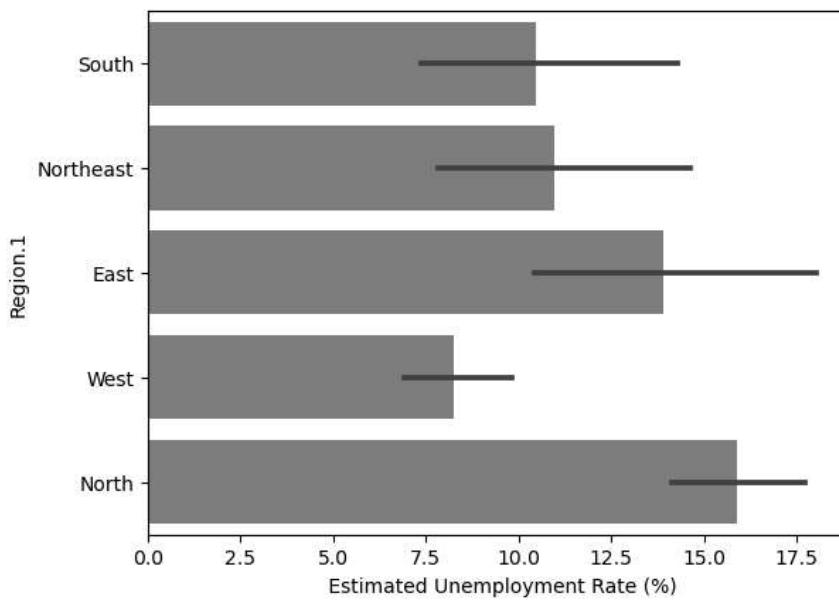
```
In [58]: #Unemployment rate in india
sns.barplot(data=data_set1,x=' Estimated Unemployment Rate (%)',y='Region',color='orange')
```

Out[58]: <Axes: xlabel=' Estimated Unemployment Rate (%)', ylabel='Region'>

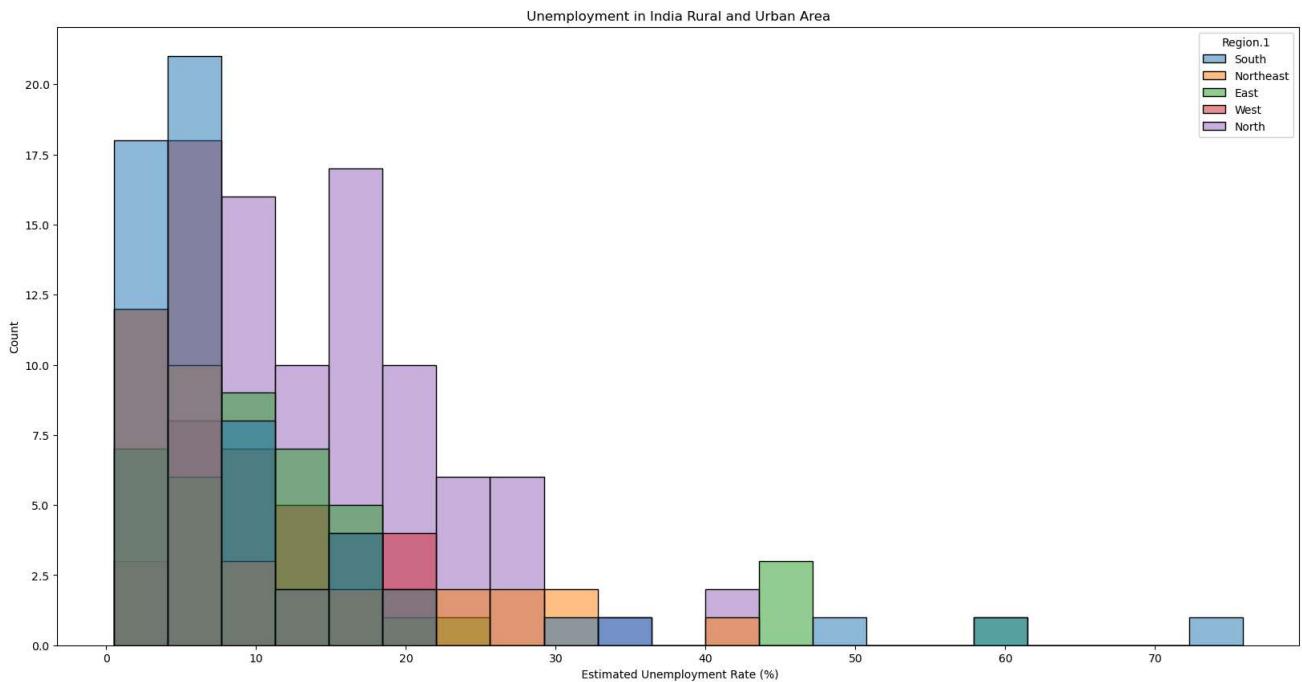


```
In [59]: sns.barplot(data=data_set2,x=' Estimated Unemployment Rate (%)',y='Region.1',color='gray')
```

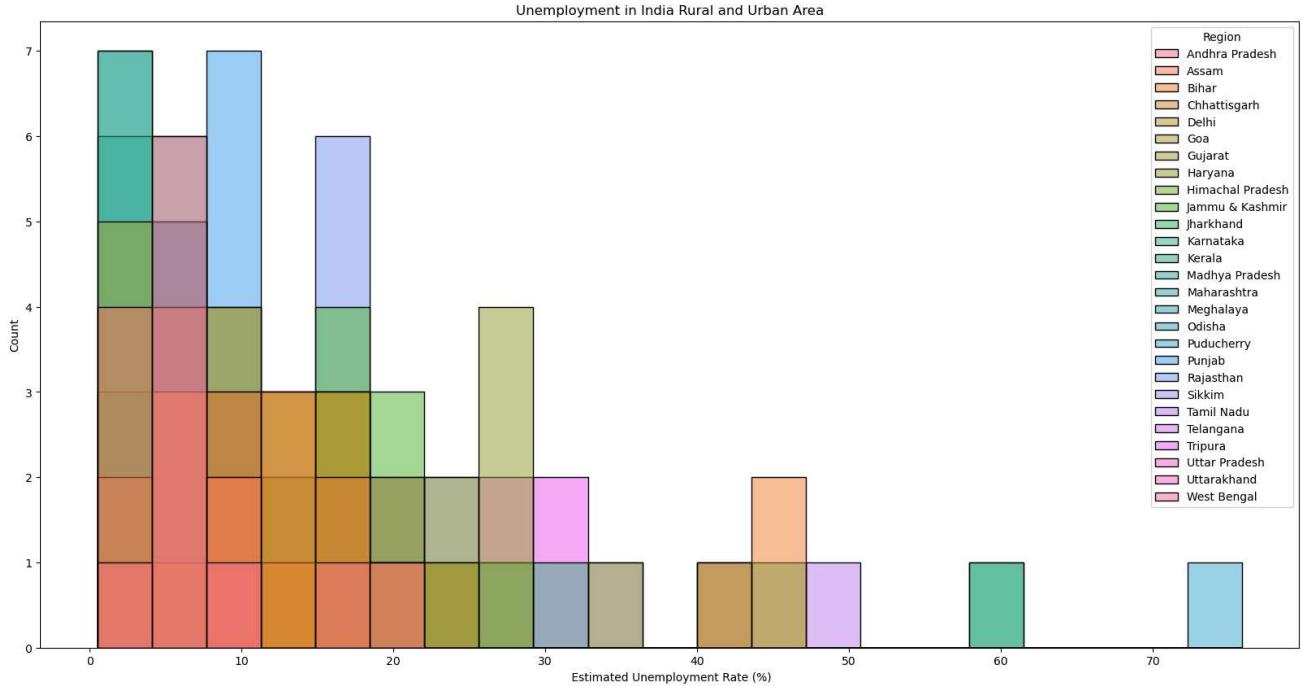
Out[59]: <Axes: xlabel=' Estimated Unemployment Rate (%)', ylabel='Region.1'>



```
In [60]: plt.figure(figsize=(20,10))
data_set2.columns=['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed', 'Estimated Unemployed']
plt.title("Unemployment in India Rural and Urban Area")
sns.histplot(x='Estimated Unemployment Rate (%)', hue='Region.1', data=data_set2)
plt.show()
```



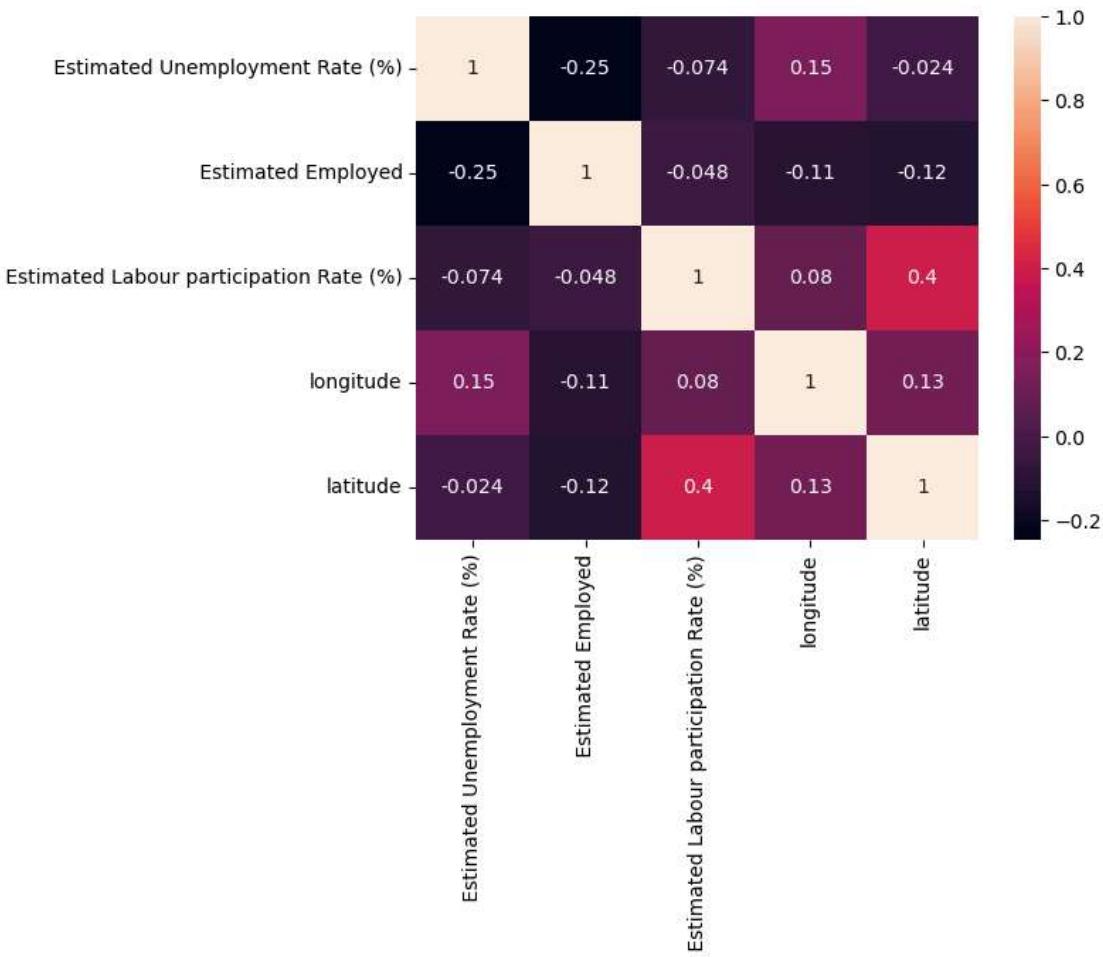
```
In [34]: plt.figure(figsize=(20,10))
data_set2.columns=['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed', 'Estimated Unemployed']
plt.title("Unemployment in India Rural and Urban Area")
sns.histplot(x='Estimated Unemployment Rate (%)', hue='Region', data=data_set2)
plt.show()
```



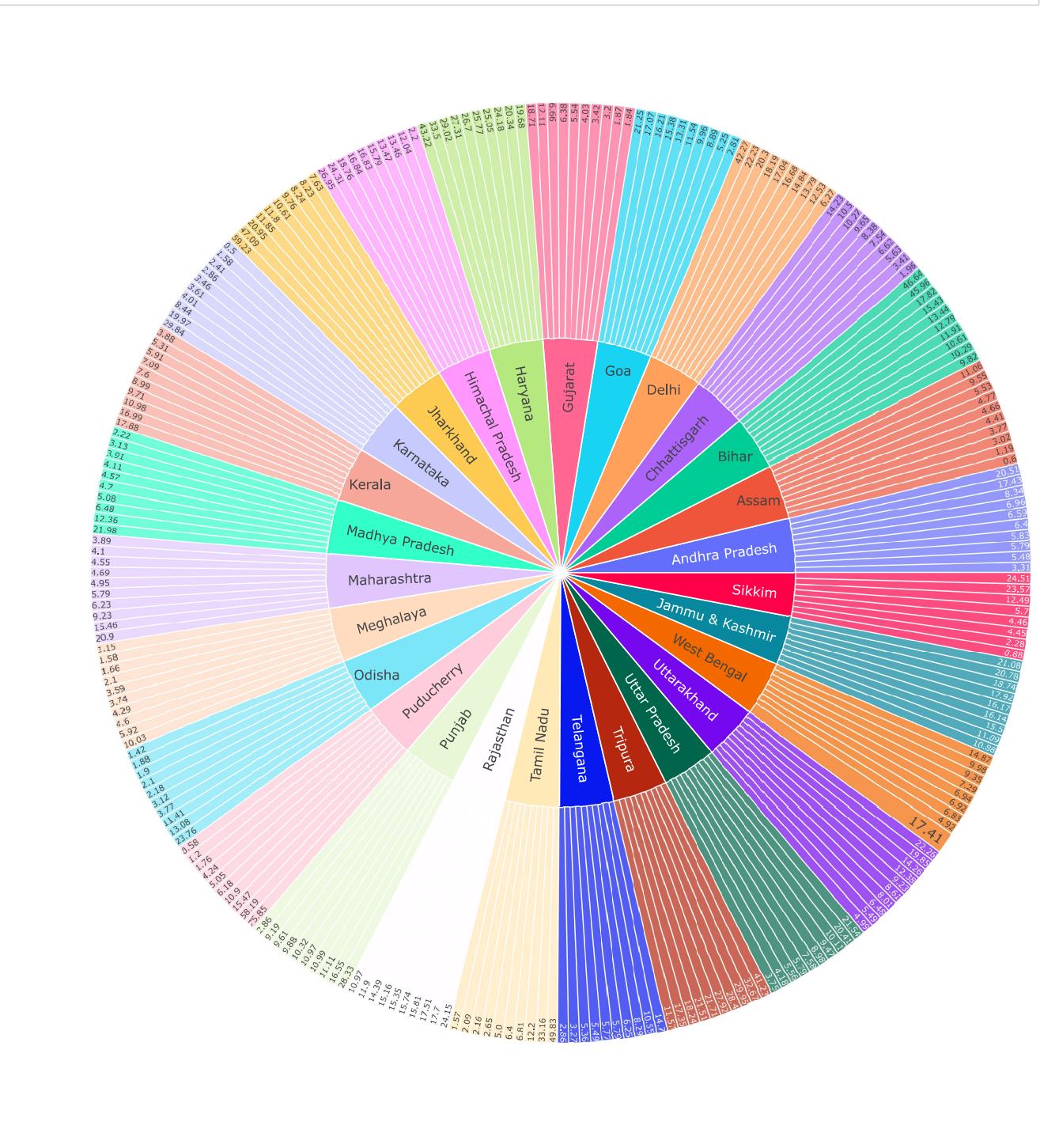
```
In [61]: sns.heatmap(data_set2.corr(), annot=True)
```

C:\Users\DB Tech\AppData\Local\Temp\ipykernel_12012\3135793750.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
sns.heatmap(data_set2.corr(), annot=True)

Out[61]: <Axes: >



```
In [62]: unemployment=data_set2[['Region', 'Date', 'Frequency', 'Estimated Unemployment Rate (%)', 'Estimated Employed']]
figure=px.sunburst(unemployment,path=['Region','Estimated Unemployment Rate (%)'],
width=1000,height=1000,color_continuous_scale="RdV1Gn")
)
figure.show()
```



In []: