Code Written By: Ankit Kotnala

Import the important Libraries

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

Loading the dataset

In [2]:
```
data = pd.read_csv('Unemployment.csv')
data
```

Out[2]:

	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 [3]:
```
# checking dataset information
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 [4]:
```
# describing the dataset
data.describe()
```

Out[4]:

	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	2.838930e+06	37.265000	18.112400	76.085600
50%	9.650000	9.732417e+06	40.390000	23.610200	79.019300
75%	16.755000	2.187869e+07	44.055000	27.278400	85.279900
max	75.850000	5.943376e+07	69.690000	33.778200	92.937600

In [5]:
```
# check null/missing values
data.isnull().sum()
```

Out[5]:
```
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 [6]:
```
# rename columns
data.columns = ['States','Date','Frequency','Estimated Unemployment Rate',
                'Estimated Employed','Estimated Labour Participation Rate',
                'Region','Longitude','Latitude']
```

In [7]:
```
# analysing top rows of dataset
data.head()
```

Out[7]:

	States	Date	Frequency	Estimated Unemployment Rate	Estimated Employed	Estimated Labour Participation Rate	Region	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

CHECKING THE CORRELATION BETWEEN THE FEATURE OF DATASET

In [8]:
```
# Set the style
plt.style.use('seaborn-whitegrid')

# Set the figure size
plt.figure(figsize=(8, 6))

# Compute the correlation matrix
corr_matrix = data.corr()

# Set tick parameters
plt.tick_params(size=10, color='w', labelsize=10, labelcolor='w')

# Create the correlation heatmap
sns.heatmap(corr_matrix, annot=True, linewidth=3, cmap='coolwarm')

# Show the plot
plt.show()
```

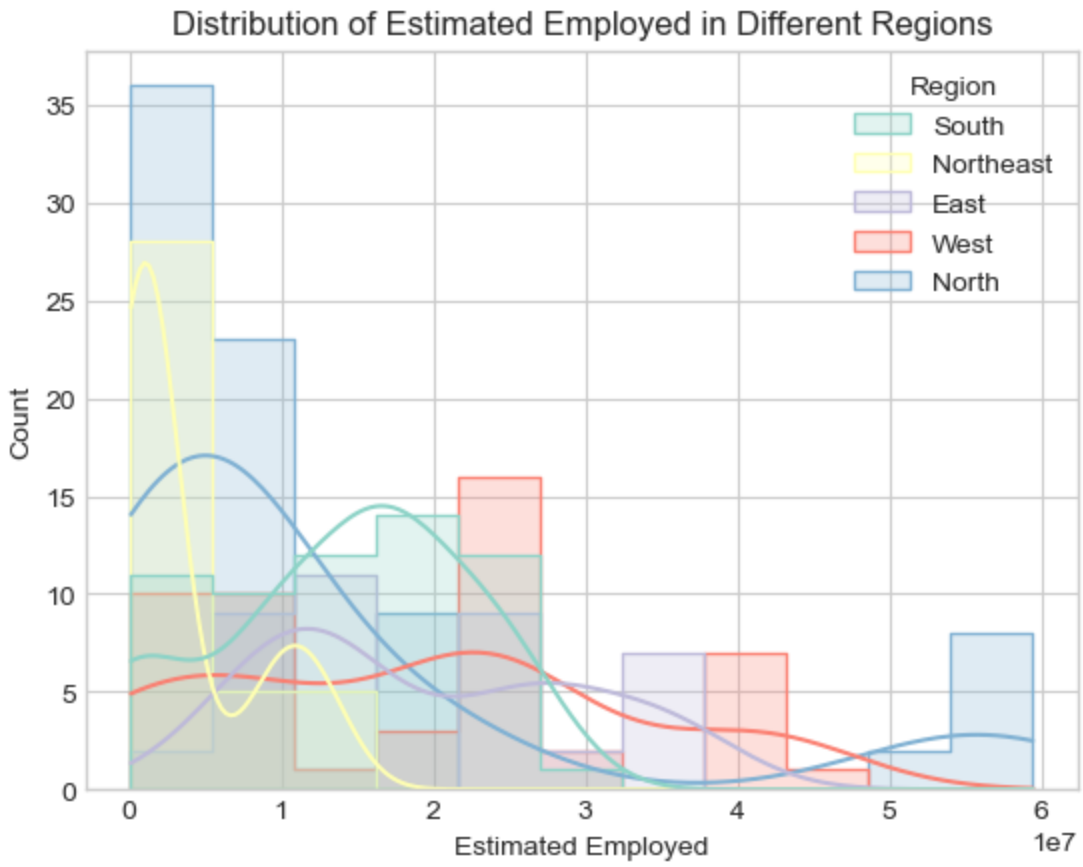
```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_16948\729498097.py:2: MatplotlibDeprecationWarning: The seaborn styles shipped by Matplotlib are deprecated since 3.6, as they no longer correspond to the styles shipped by seaborn. However, they will remain available as 'seaborn-v0_8-<style>'. Alternatively, directly use the seaborn API instead.
  plt.style.use('seaborn-whitegrid')
C:\Users\Administrator\AppData\Local\Temp\ipykernel_16948\729498097.py:8: 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.
  corr_matrix = data.corr()
```

Estimated no of employee according to different region of india

In [9]:
```
# Set the style
sns.set_style("whitegrid")

# Rename the columns
data.columns = ['States', 'Date', 'Frequency', 'Estimated Unemployment Rate',
                'Estimated Employed', 'Estimated Labour Participation Rate',
                'Region', 'Longitude', 'Latitude']

# Specify a custom color palette (e.g., 'Set3')
custom_palette = sns.color_palette("Set3")

# Create a histogram with separate bars for each region and set the color palette
sns.histplot(data=data, x='Estimated Employed', hue='Region', element="step", common_nor

# Set plot title
plt.title('Distribution of Estimated Employed in Different Regions')

# Show the plot
plt.show()
```

```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_16948\3026068327.py:13: UserWarning:
The palette list has more values (12) than needed (5), which may not be intended.
```

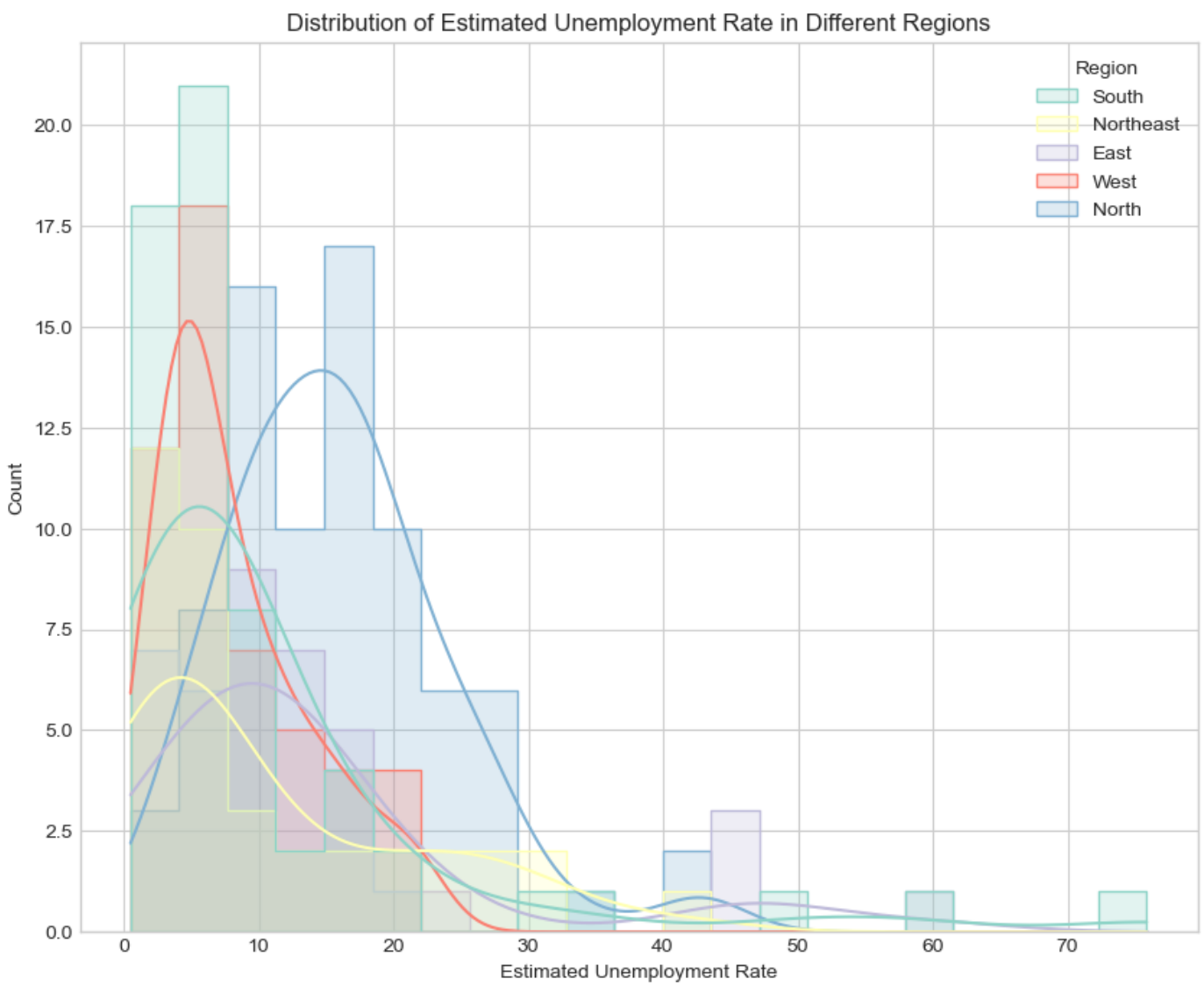
```
  sns.histplot(data=data, x='Estimated Employed', hue='Region', element="step", common_n
orm=False, kde=True, palette=custom_palette)
```



Unemployment rate according to different regions of india

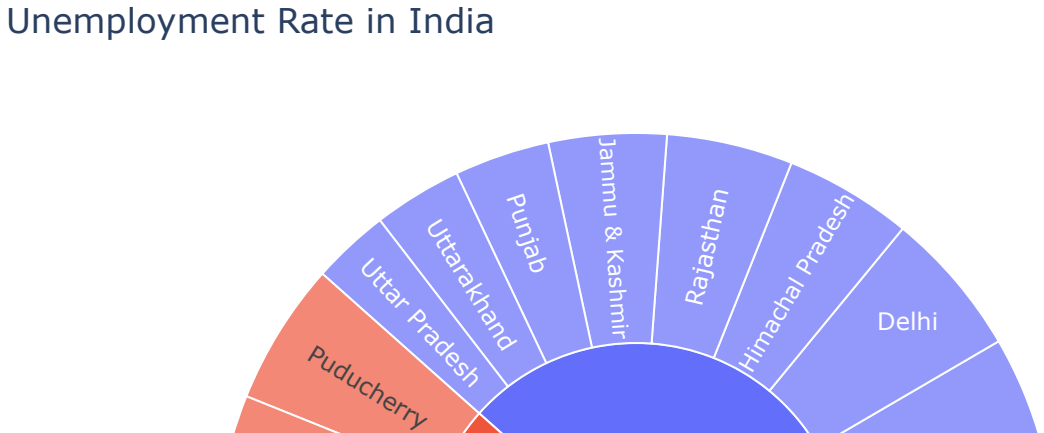
In [10]:
```
# Set the figure size
plt.figure(figsize=(10, 8))

# Rename the columns
data.columns = ['States', 'Date', 'Frequency', 'Estimated Unemployment Rate',
                'Estimated Employed', 'Estimated Labour Participation Rate',
                'Region', 'Longitude', 'Latitude']

# Specify a custom color palette (e.g., 'Set3')
custom_palette = sns.color_palette("Set3")

# Create a histogram with separate bars for each region and set the color palette
sns.histplot(data=data, x="Estimated Unemployment Rate", hue="Region", element="step", c

# Set plot title
plt.title("Distribution of Estimated Unemployment Rate in Different Regions")

# Show the plot
plt.show()
```

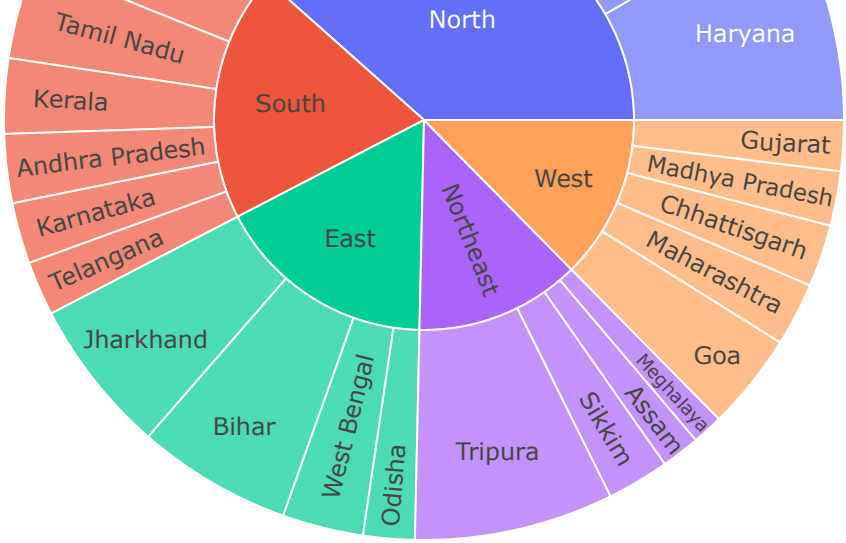
```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_16948\1085341482.py:13: UserWarning:
The palette list has more values (12) than needed (5), which may not be intended.
  sns.histplot(data=data, x="Estimated Unemployment Rate", hue="Region", element="step",
common_norm=False, kde=True, palette=custom_palette)
```



Dashboard to analyze the unemployment rate of each Indian state

In [11]:
```
# plotting sunburst
unemployment = data[['States', 'Region', 'Estimated Unemployment Rate']]
figure = px.sunburst(unemployment, path=['Region', 'States'],
                     values='Estimated Unemployment Rate',
                     width=700, height=600, color_continuous_scale='YlOrRd',  # Change t
                     title="Unemployment Rate in India")
figure.show()
```




North
Haryana
South
Tamil Nadu
Kerala
Andhra Pradesh
Karnataka
Telangana
West
Gujarat
Madhya Pradesh
Chhattisgarh
Maharashtra
Goa
East
Jharkhand
Bihar
West Bengal
Odisha
Northeast
Tripura
Sikkim
Assam
Meghalaya

In [ ]: