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. In this article, I will take you through the task of Unemployment analysis with Python.

Unemployment Analysis with Python

The unemployment rate is calculated based on a particular region, so to analyze unemployment I will be using an unemployment dataset of India. The dataset I'm using here contains data on India's unemployment rate during Covid-19. So let's start the task of Unemployment analysis by importing the necessary Python libraries and the dataset:

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

data = pd.read_csv("unemployment.csv")
print(data.head())
```

```
            Region         Date  Frequency   Estimated Unemployment Rate (%)  \
0  Andhra Pradesh   31-01-2020          M                              5.48
1  Andhra Pradesh   29-02-2020          M                              5.83
2  Andhra Pradesh   31-03-2020          M                              5.79
3  Andhra Pradesh   30-04-2020          M                             20.51
4  Andhra Pradesh   31-05-2020          M                             17.43

    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
```

Let's see if this dataset contains missing values or not:

In [2]:
```
print(data.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
```

While analyzing the missing values, I found that the column names are not correct. So, for a better understanding of this data, I will rename all the columns:

In [3]:
```
data.columns= ["States","Date","Frequency",
               "Estimated Unemployment Rate",
               "Estimated Employed",
               "Estimated Labour Participation Rate",
               "Region","longitude","latitude"]
```

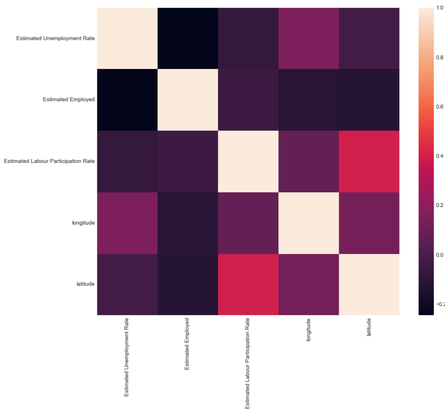
In [4]:
```
data.head()
```

Out[4]:

	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

Now let's have a look at the correlation between the features of this dataset:

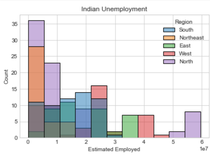
In [5]:
```
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(12, 10))
sns.heatmap(data.corr())
plt.show()
```



Unemployment Rate Analysis: Data Visualization

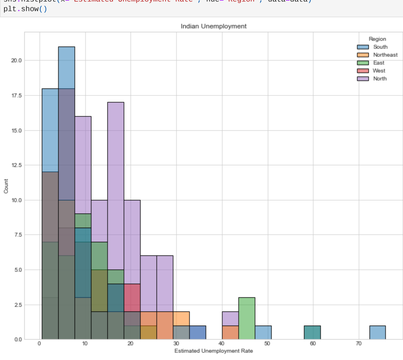
Now let's visualize the data to analyze the unemployment rate. I will first take a look at the estimated number of employees according to different regions of India:

In [6]:
```
data.columns= ["States","Date","Frequency",
               "Estimated Unemployment Rate","Estimated Employed",
               "Estimated Labour Participation Rate","Region",
               "longitude","latitude"]
plt.title("Indian Unemployment")
sns.histplot(x="Estimated Employed", hue="Region", data=data)
plt.show()
```



Now let's see the unemployment rate according to different regions of India:

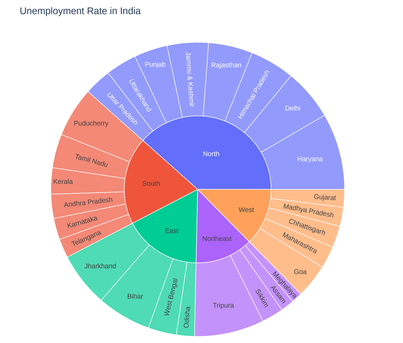
In [7]:
```
plt.figure(figsize=(12, 10))
plt.title("Indian Unemployment")
sns.histplot(x="Estimated Unemployment Rate", hue="Region", data=data)
plt.show()
```



Now let's create a dashboard to analyze the unemployment rate of each Indian state by region. For this, I'll use a sunburst plot:

In [8]:
```
unemployment = data[["States", "Region", "Estimated Unemployment Rate"]]
figure = px.sunburst(unemployment, path=["Region", "States"],
                     values="Estimated Unemployment Rate",
                     width=700, height=700, color_continuous_scale="RdYlGn",
                     title="Unemployment Rate in India")
figure.show()
```

```
D:\anaconda\lib\site-packages\plotly\express\_core.py:1637: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.
  df_all_trees = df_all_trees.append(df_tree, ignore_index=True)
D:\anaconda\lib\site-packages\plotly\express\_core.py:1637: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.
  df_all_trees = df_all_trees.append(df_tree, ignore_index=True)
```



Summary

So this is how you can analyze the unemployment rate by using the Python programming language. Unemployment is measured by the unemployment rate which is the number of people who are unemployed as a percentage of the total labour force. I hope you liked this article on unemployment rate analysis with Python.