Oasis Infobyte Internship

Intern Name-Akshay Anandkar

Task 2-UNEMPLOYMENT ANALYSIS WITH PYTHON

Problem Statement-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 [1]:
```
#importing liabraries
import pandas as pd
import numpy as np
```

In [2]:
```
#import dataset
data=pd.read_csv(r"D:\Data-Science-Internship\Unemployment_Rate_upto_11_2020.csv")
data.head()
```

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	79.74
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 [3]:
```
data.shape
```

Out[3]:
```
(267, 9)
```

In [4]:
```
data.duplicated().sum()
```

Out[4]:
```
0
```

In [5]:
```
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]:
```
data['Region'].value_counts()
```

Out[6]:
```
Andhra Pradesh      10
Assam               10
Uttarakhand         10
Uttar Pradesh       10
Tripura             10
Telangana           10
Tamil Nadu          10
Rajasthan           10
Punjab              10
Puducherry          10
Odisha              10
Meghalaya           10
Maharashtra         10
Madhya Pradesh      10
Kerala              10
Karnataka           10
Jharkhand           10
Himachal Pradesh    10
Haryana             10
Gujarat             10
Goa                 10
Delhi               10
Chhattisgarh        10
Bihar               10
West Bengal         10
Jammu & Kashmir      9
Sikkim               8
Name: Region, dtype: int64
```

In [7]:
```
data['Region.1'].value_counts()
```

Out[7]:
```
North        79
South        60
West         50
East         40
Northeast    38
Name: Region.1, dtype: int64
```

In [8]:
```
data.dtypes
```

Out[8]:
```
Region                                       object
 Date                                        object
 Frequency                                   object
 Estimated Unemployment Rate (%)            float64
 Estimated Employed                           int64
 Estimated Labour Participation Rate (%)    float64
Region.1                                     object
longitude                                   float64
latitude                                    float64
dtype: object
```

In [9]:
```
data[["day", "month", "year"]]=data[' Date'].str.split("-", expand=True)
data
```

Out[9]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude	day	month	year
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.740	31	01	2020
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740	29	02	2020
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740	31	03	2020
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740	30	04	2020
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740	31	05	2020
...	...	...	...	...	...	...	...	...	...	...	...	...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855	30	06	2020
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855	31	07	2020
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855	31	08	2020
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855	30	09	2020
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855	31	10	2020

267 rows × 12 columns

In [10]:
```
data.drop(columns=[' Frequency'],axis=1,inplace=True)
```

In [11]:
```
data.head()
```

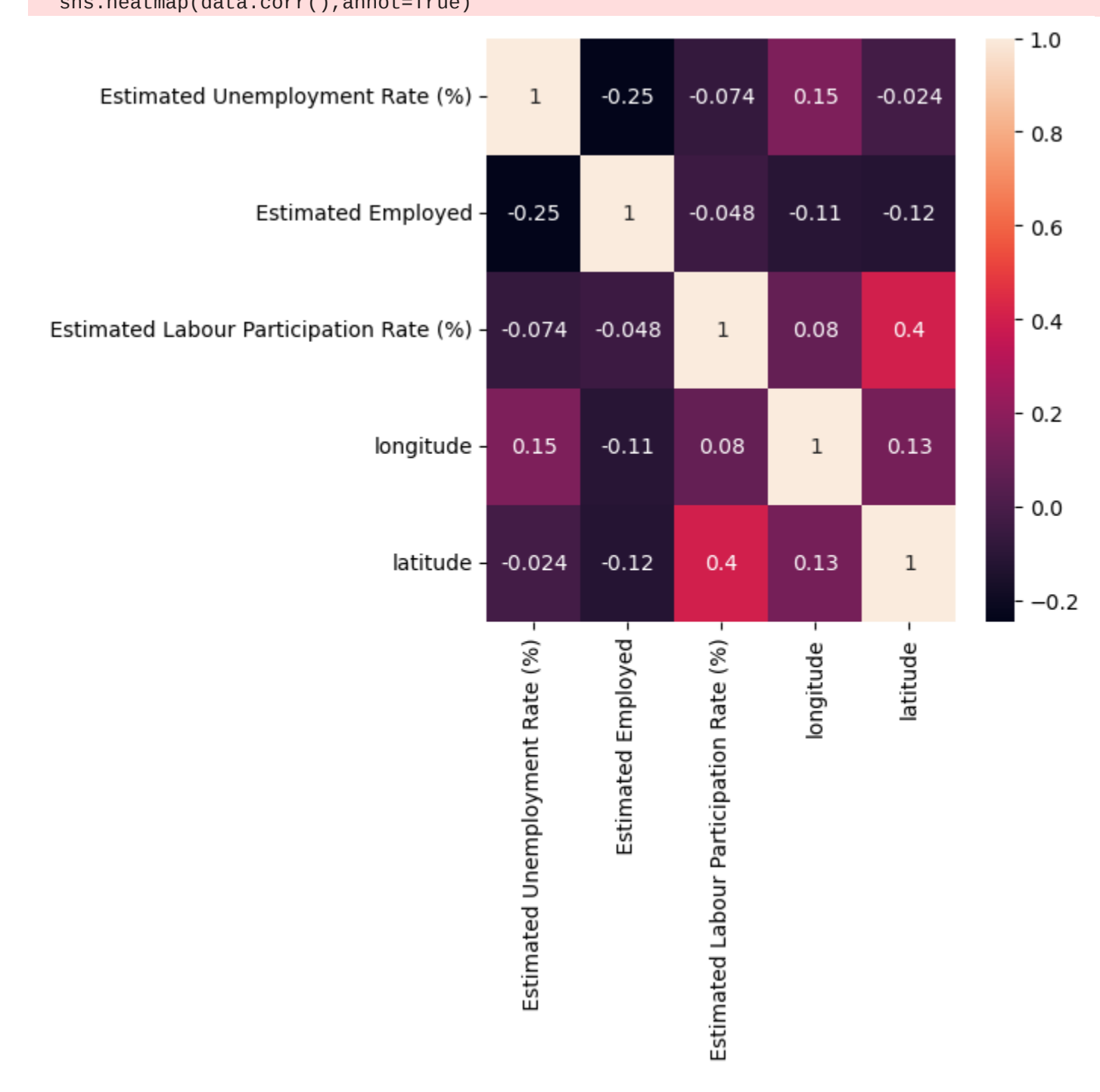
Out[11]:

	Region	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude	day	month	year
0	Andhra Pradesh	31-01-2020	5.48	16635535	41.02	South	15.9129	79.74	31	01	2020
1	Andhra Pradesh	29-02-2020	5.83	16545652	40.90	South	15.9129	79.74	29	02	2020
2	Andhra Pradesh	31-03-2020	5.79	15881197	39.18	South	15.9129	79.74	31	03	2020
3	Andhra Pradesh	30-04-2020	20.51	11336911	33.10	South	15.9129	79.74	30	04	2020
4	Andhra Pradesh	31-05-2020	17.43	12988845	36.46	South	15.9129	79.74	31	05	2020

In [12]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [13]:
```
#Heatmap
plt.figure(figsize=(5,5))
sns.heatmap(data.corr(),annot=True)
plt.show()
```

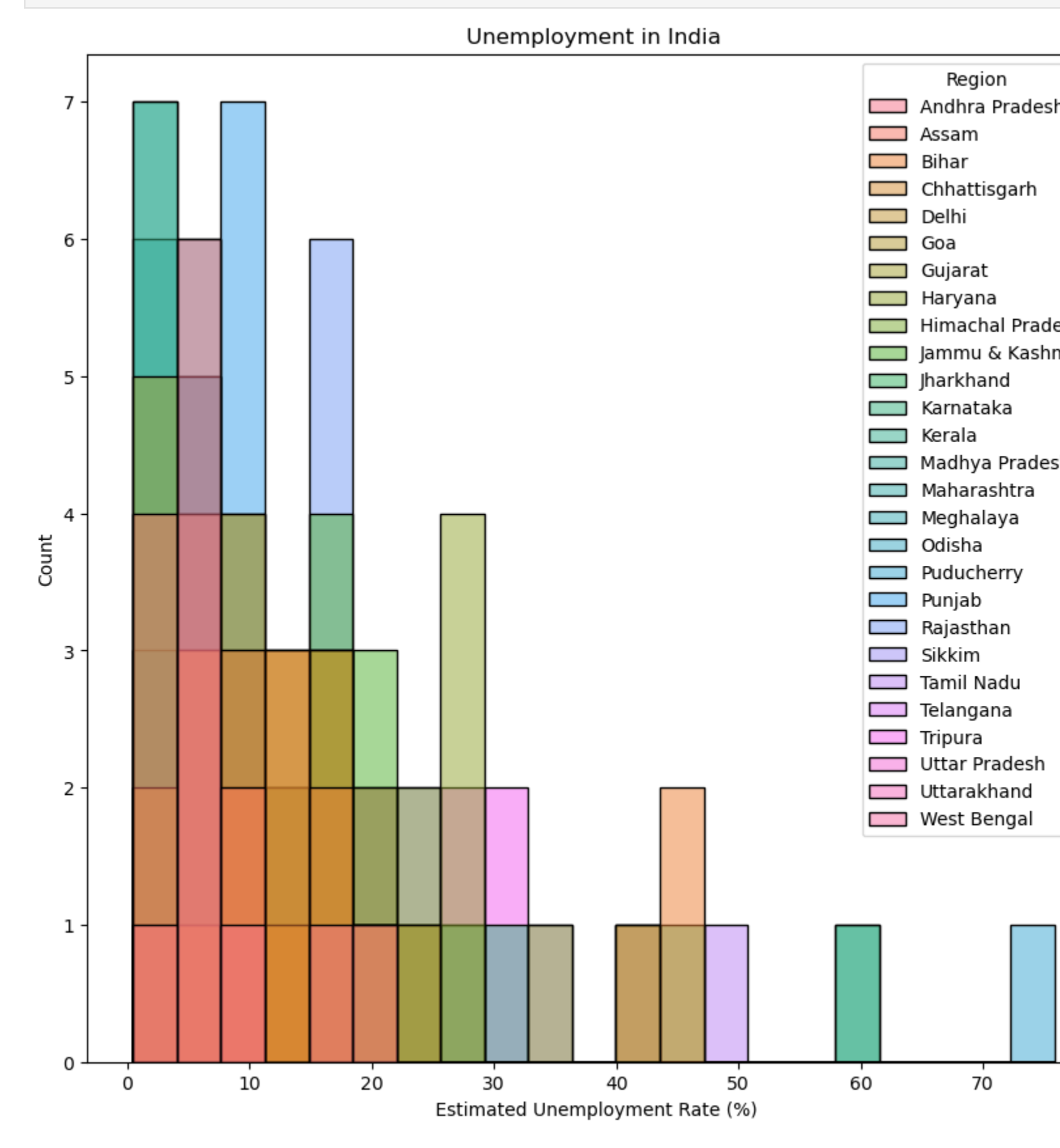
```
C:\Users\Akshay\AppData\Local\Temp\ipykernel_15440\1205428686.py:3: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  sns.heatmap(data.corr(),annot=True)
```



In [14]:
```
data.columns
```

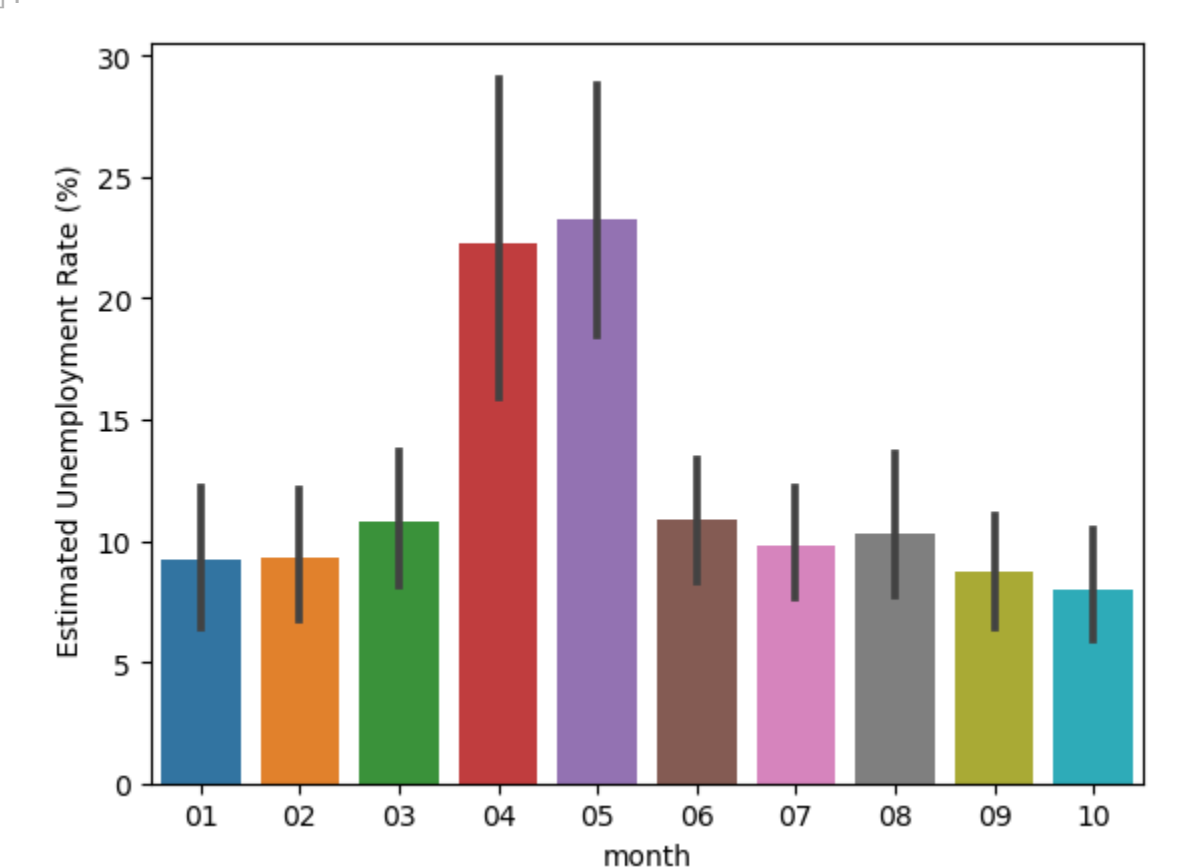
Out[14]:
```
Index(['Region', ' Date', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude', 'day', 'month', 'year'],
      dtype='object')
```

In [15]:
```
plt.figure(figsize=(10,10))
plt.title("Unemployment in India")
sns.histplot(x=' Estimated Unemployment Rate (%)',hue='Region',data=data,kde=False)
plt.show()
```



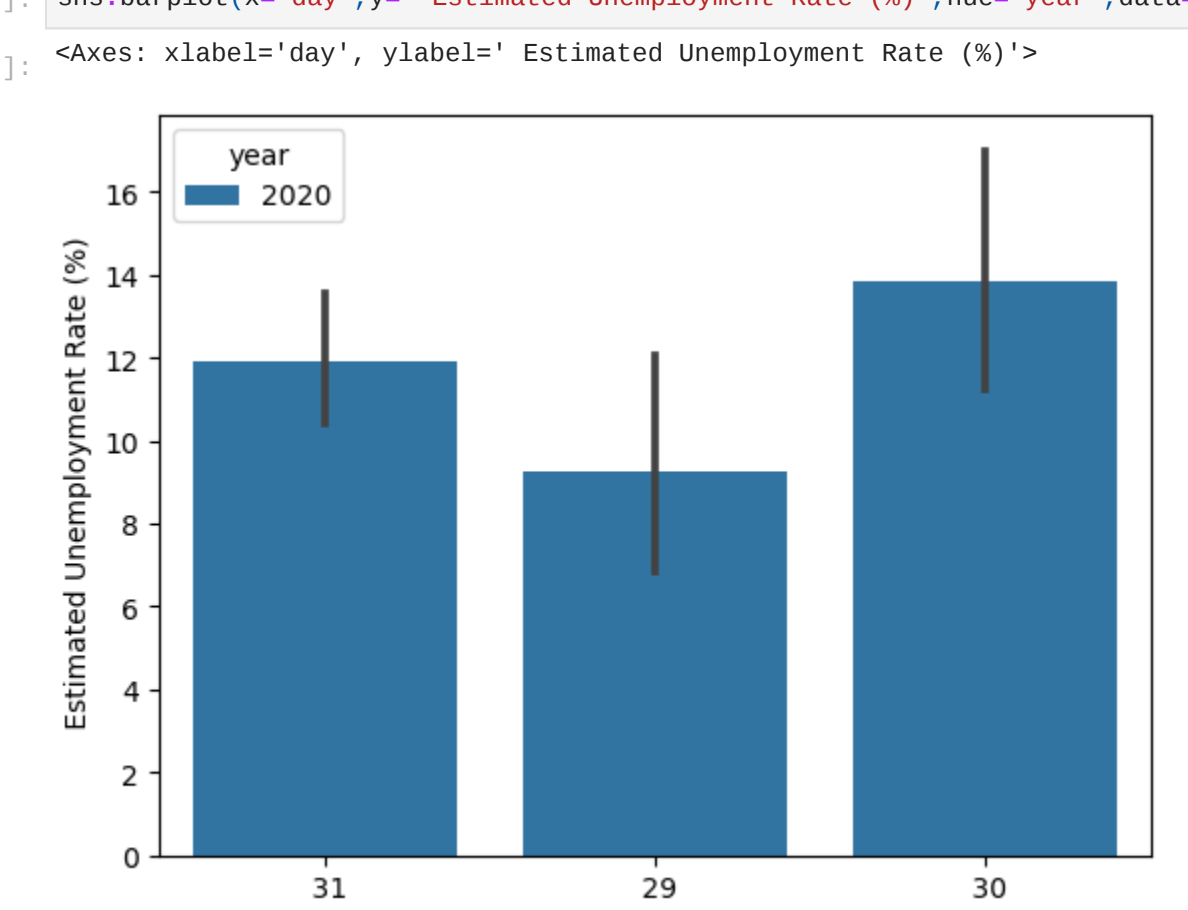
In [16]:
```
sns.barplot(x='month',y=' Estimated Unemployment Rate (%)',data=data)
```

Out[16]:
```
<Axes: xlabel='month', ylabel=' Estimated Unemployment Rate (%)'>
```

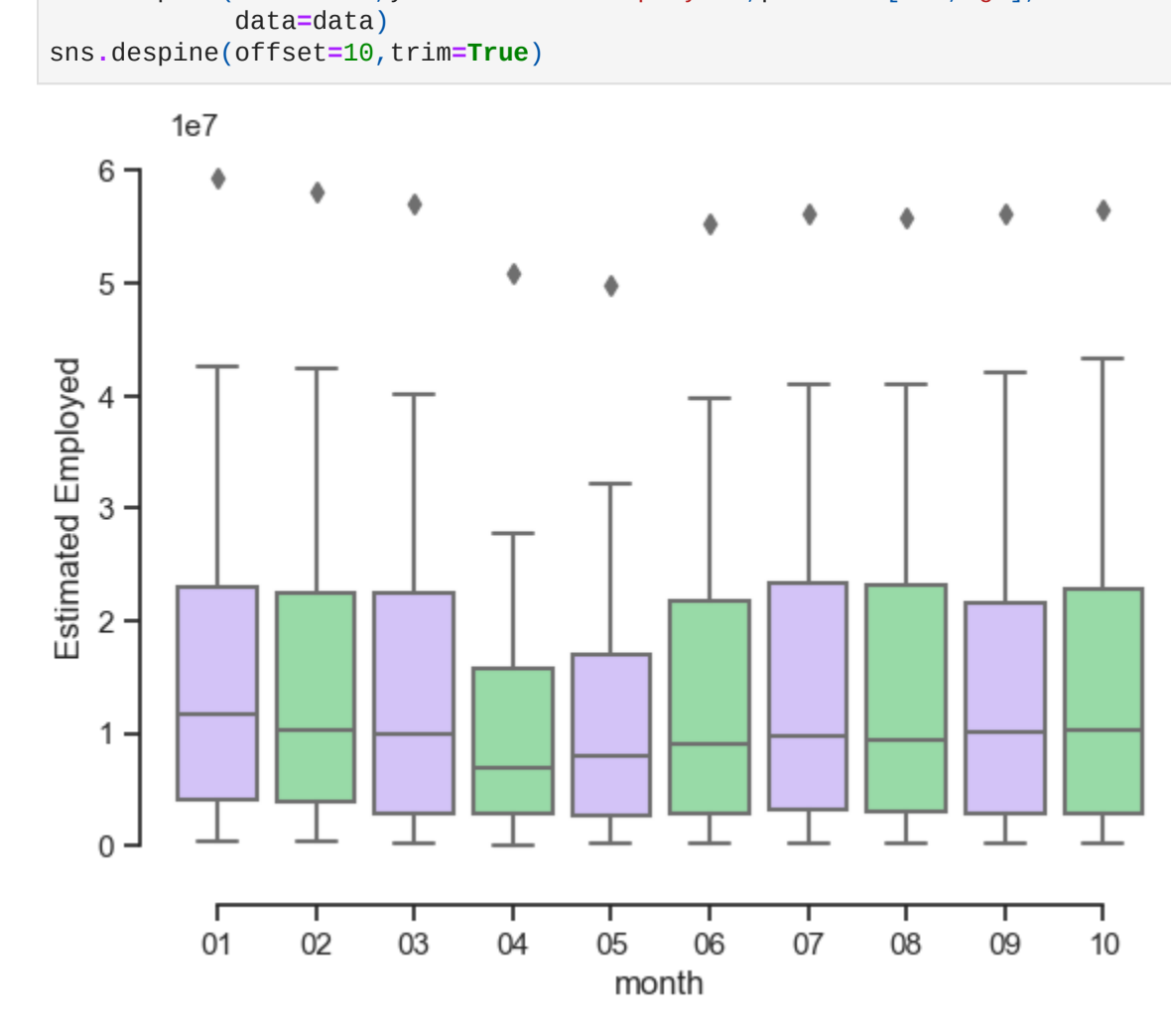


In [17]:
```
sns.barplot(x='day',y=' Estimated Unemployment Rate (%)',hue='year',data=data)
```

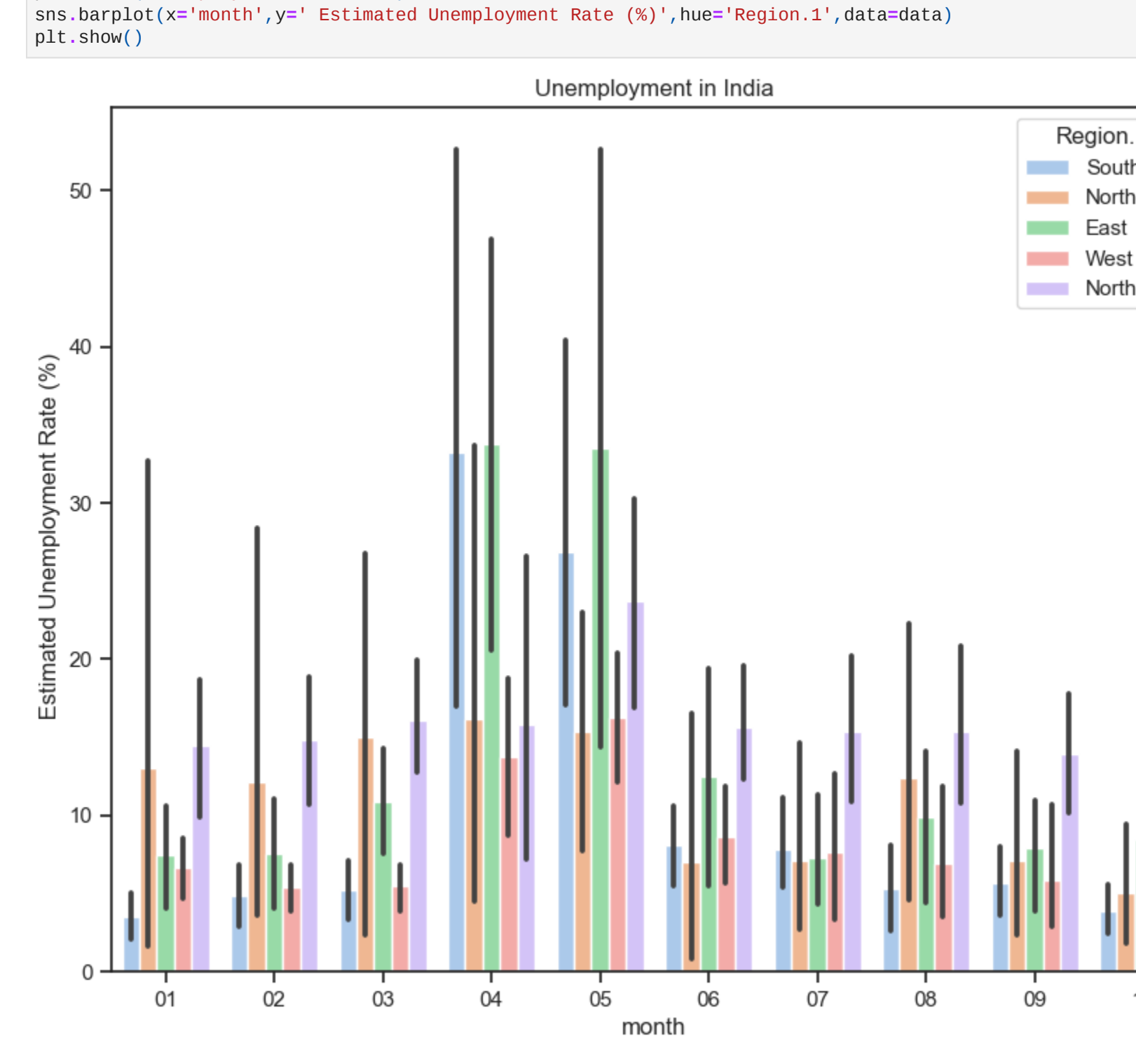
Out[17]:
```
<Axes: xlabel='day', ylabel=' Estimated Unemployment Rate (%)'>
```



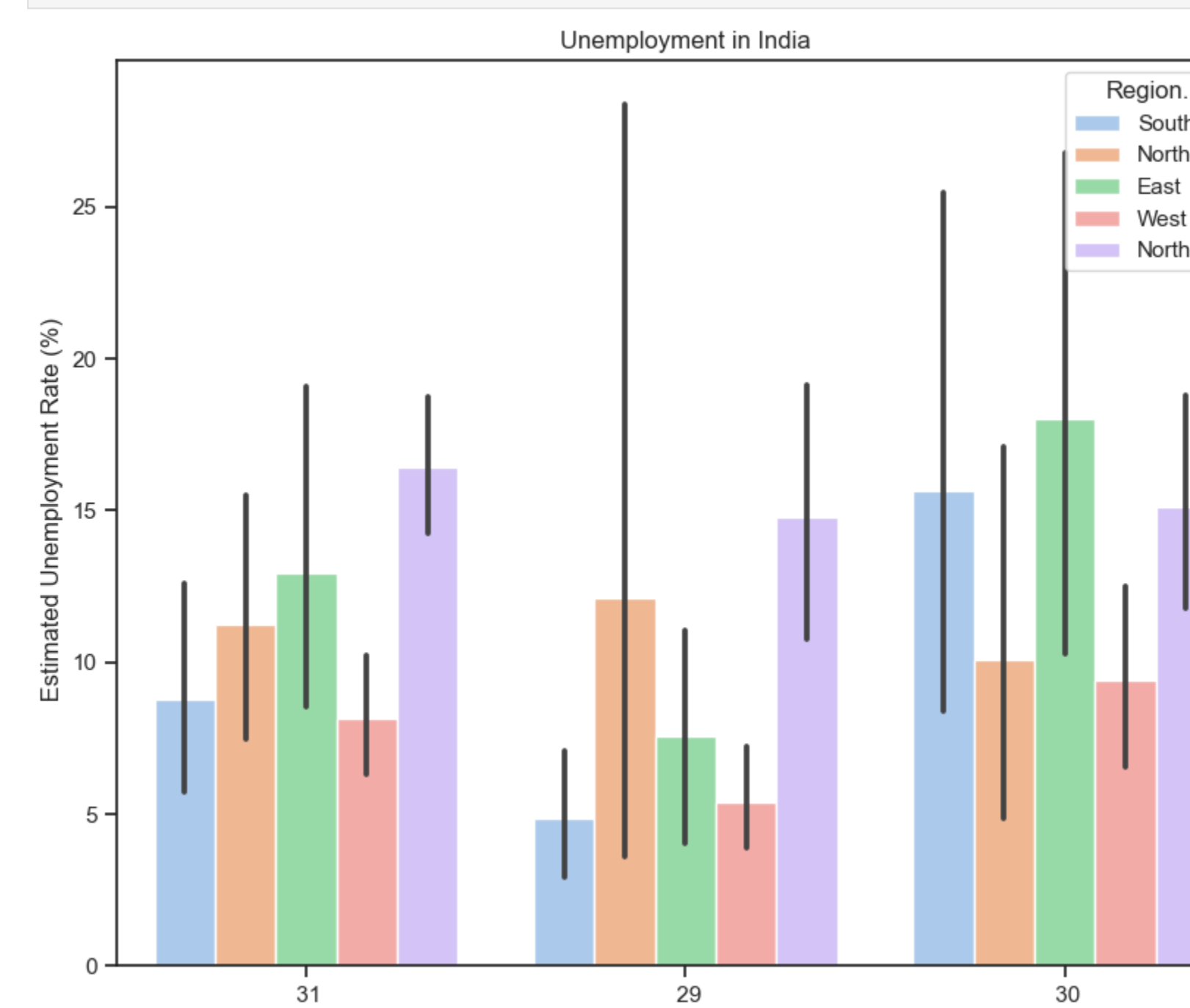
In [18]:
```
sns.set_theme(style="ticks", palette="pastel")

sns.boxplot(x='month',y=' Estimated Employed',palette=["m","g"],
            data=data)
sns.despine(offset=10,trim=True)
```



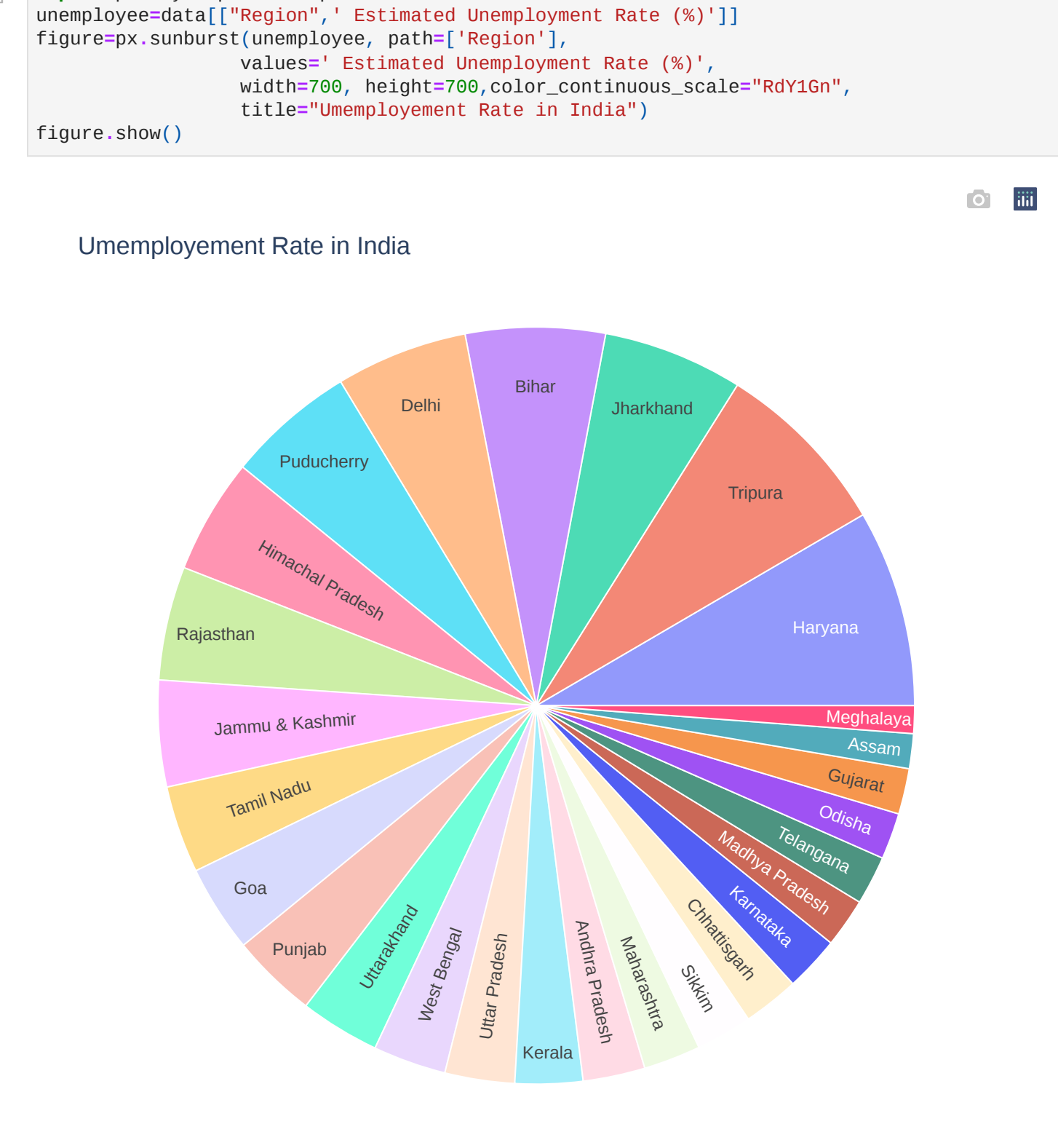
In [19]:
```
plt.figure(figsize=(10,8))
plt.title("Unemployment in India")
sns.barplot(x='month',y=' Estimated Unemployment Rate (%)',hue='Region.1',data=data)
plt.show()
```



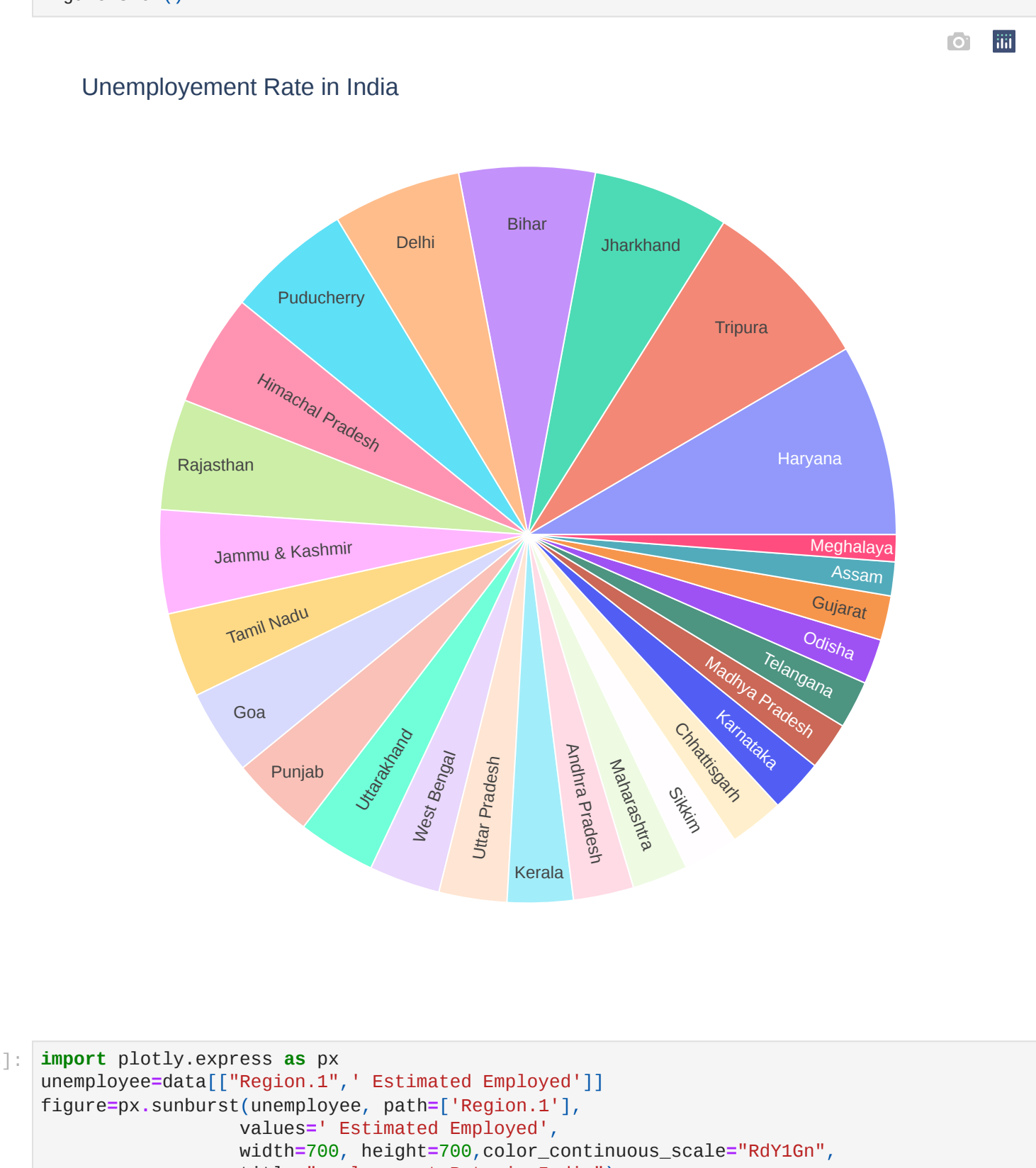
In [20]:
```
plt.figure(figsize=(10,8))
plt.title("Unemployment in India")
sns.barplot(x='day',y=' Estimated Unemployment Rate (%)',hue='Region.1',data=data)
plt.show()
```



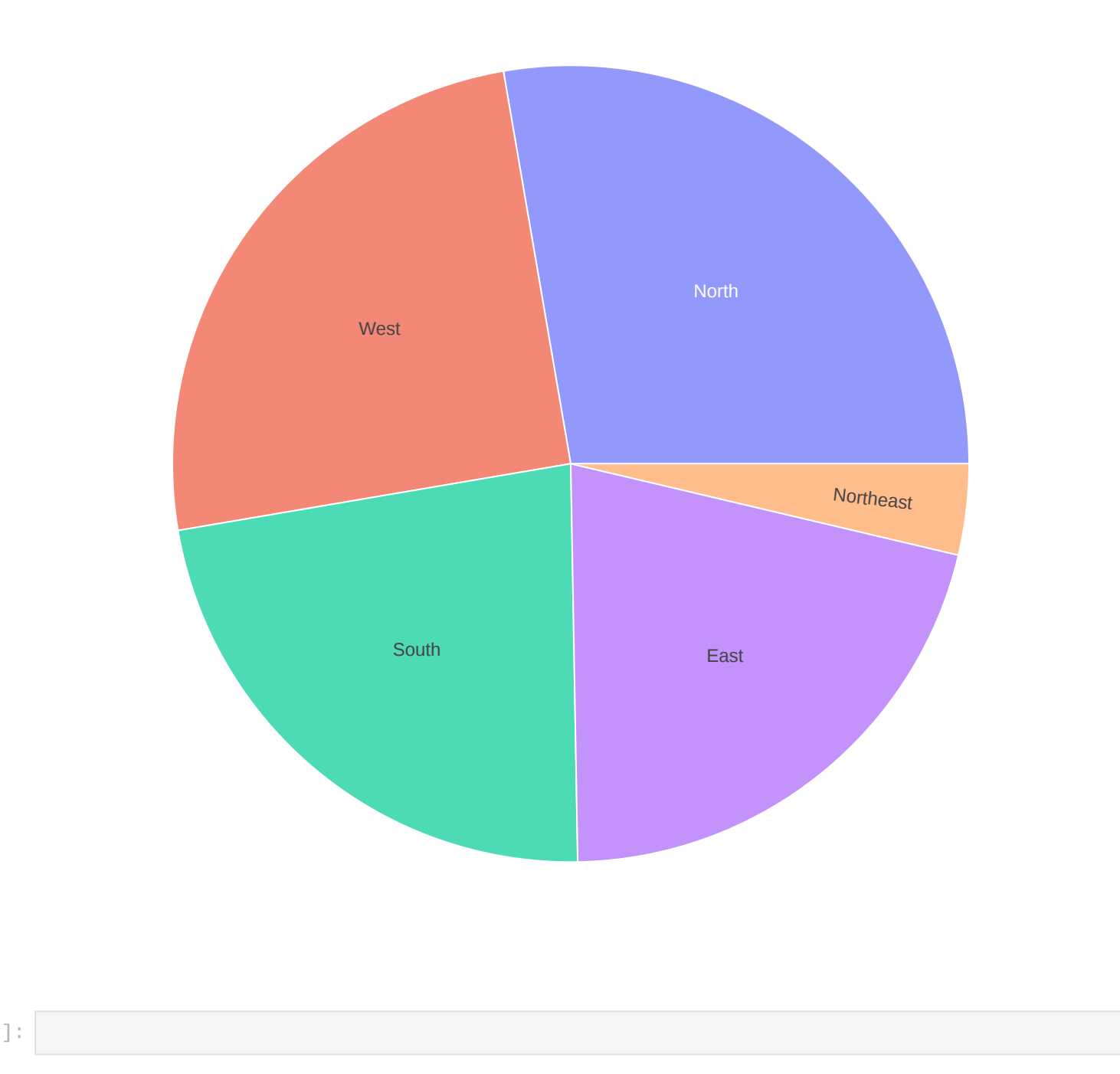
In [21]:
```
import plotly.express as px
unemployee=data[["Region",' Estimated Unemployment Rate (%)']]
figure=px.sunburst(unemployee, path=['Region'],
              values=' Estimated Unemployment Rate (%)',
              width=700, height=700,color_continuous_scale="RdY1Gn",
              title="Umemployement Rate in India")
figure.show()
```



In [22]:
```
import plotly.express as px
unemployee=data[["Region",' Estimated Unemployment Rate (%)']]
figure=px.sunburst(unemployee, path=['Region'],
              values=' Estimated Unemployment Rate (%)',
              width=700, height=700,color_continuous_scale="RdY1Gn",
              title="Unemployement Rate in India")
figure.show()
```



In [23]:
```
import plotly.express as px
unemployee=data[["Region.1",' Estimated Employed']]
figure=px.sunburst(unemployee, path=['Region.1'],
              values=' Estimated Employed',
              width=700, height=700,color_continuous_scale="RdY1Gn",
              title="employement Rate in India")
figure.show()
```



In [ ]: