Cognorise Infotech Internship

Name: Anuttama Mondal

Task 1: Unemployment in India

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.

Import Libraries:

In [4]:
```
import pandas as pd         #For data manipulation and analysis
import numpy as mp          # For numerical computation and handling arrays
import matplotlib.pyplot as plt #For data visualization
import seaborn as sns           # For enhanced data visualization
from sklearn.model_selection import train_test_split      #For Data splitting
from datetime import datetime               #For Working with dates and times
import warnings
warnings.filterwarnings("ignore")           #disable warning
%matplotlib inline
```

In [6]:
```
df1= pd.read_csv("Unemployment in India.csv")
```

In [7]:
```
df1
```

Out[7]:

	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
...	...	...	...	...	...	...	...
749	West Bengal	29-02-2020	Monthly	7.55	10871168.0	44.09	Urban
750	West Bengal	31-03-2020	Monthly	6.67	10806105.0	43.34	Urban
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban

754 rows × 7 columns

In [8]:
```
df2= pd.read_csv("Unemployment_Rate_upto_11_2020.csv")
```

In [9]:
```
df2
```

Out[9]:

	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

Checking with the Rows for storing the Length

In [10]:
```
df1_len=len(df1)
df1_len
```

Out[10]: 754

In [11]:
```
df2_len=len(df2)
df2_len
```

Out[11]: 267

Displaying first and last rows and columns of the dataset

In [12]:
```
df2.head()
```

Out[12]:

	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 [14]:
```
df2.tail()
```

Out[14]:

	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

View the information

In [15]:
```
df2.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 [17]:
```
df2.shape
```

Out[17]:
```
(267, 9)
```

checking for any null or missing values:

In [18]:
```
df2.isnull().sum()
```

Out[18]:
```
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 [19]:
```
df2.describe()
```

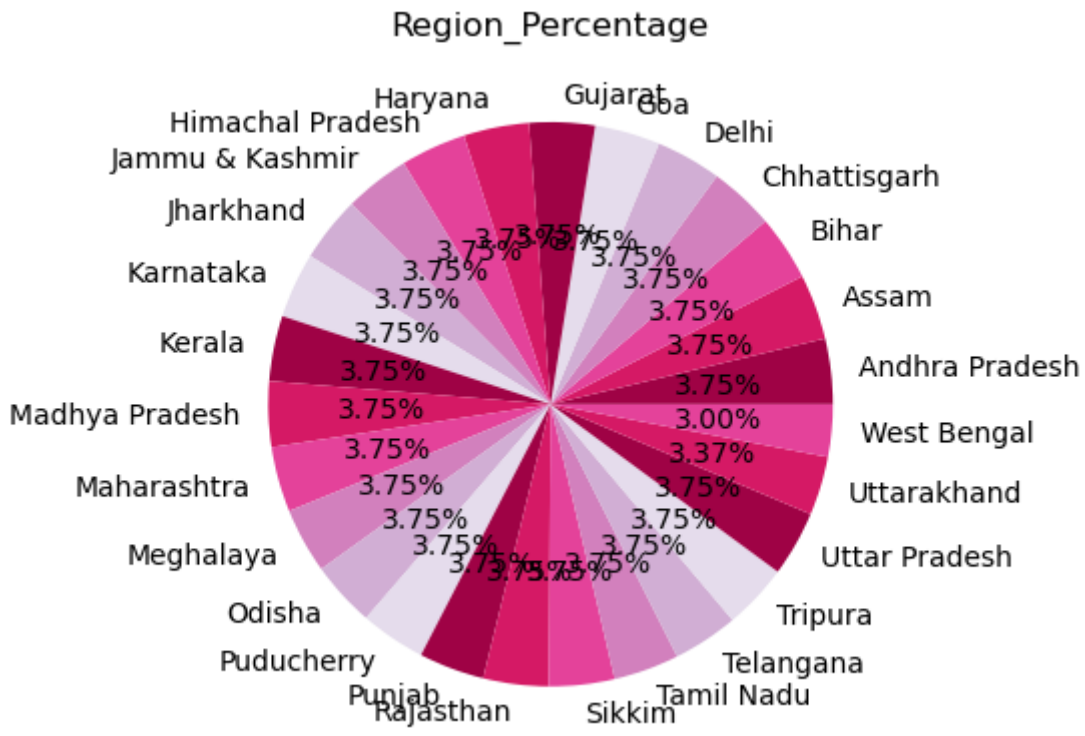
Out[19]:

	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

Using EDA(Exploratory Data Analysis)

In [41]:
```
colors= sns.color_palette('PuRd_r')
labels=df2['Region'].dropna().unique()
plt.figure(figsize=(10,7))
plt.subplot(1,2,1)
plt.title('Region_Percentage')
plt.pie(df2['Region'].value_counts(),labels=labels,colors=colors, autopct='%.2f%%')
```

Out[41]:
```
([<matplotlib.patches.Wedge at 0x205ba8f8640>,
  <matplotlib.patches.Wedge at 0x205ba8f8eb0>,
  <matplotlib.patches.Wedge at 0x205ba903610>,
  <matplotlib.patches.Wedge at 0x205ba903d30>,
  <matplotlib.patches.Wedge at 0x205ba90f490>,
  <matplotlib.patches.Wedge at 0x205ba90fbb0>,
  <matplotlib.patches.Wedge at 0x205ba8e04f0>,
  <matplotlib.patches.Wedge at 0x205ba91ba00>,
  <matplotlib.patches.Wedge at 0x205ba92a160>,
  <matplotlib.patches.Wedge at 0x205ba92a880>,
  <matplotlib.patches.Wedge at 0x205ba92afa0>,
  <matplotlib.patches.Wedge at 0x205ba936700>,
  <matplotlib.patches.Wedge at 0x205ba936e20>,
  <matplotlib.patches.Wedge at 0x205ba945580>,
  <matplotlib.patches.Wedge at 0x205ba945ca0>,
  <matplotlib.patches.Wedge at 0x205ba953400>,
  <matplotlib.patches.Wedge at 0x205ba953b20>,
  <matplotlib.patches.Wedge at 0x205ba961280>,
  <matplotlib.patches.Wedge at 0x205ba9619a0>,
  <matplotlib.patches.Wedge at 0x205ba96e100>,
  <matplotlib.patches.Wedge at 0x205ba96e820>,
  <matplotlib.patches.Wedge at 0x205ba96ef40>,
  <matplotlib.patches.Wedge at 0x205ba97a6a0>,
  <matplotlib.patches.Wedge at 0x205ba97adc0>,
  <matplotlib.patches.Wedge at 0x205ba989520>,
  <matplotlib.patches.Wedge at 0x205ba989c40>,
  <matplotlib.patches.Wedge at 0x205baca43a0>],
 [Text(1.092394307356846, 0.1291304660192809, 'Andhra Pradesh'),
  Text(1.0321783651761125, 0.38027335229064874, 'Assam'),
  Text(0.9150657581747471, 0.6104544685855576, 'Bihar'),
  Text(0.7475120728789119, 0.8069855642452798, 'Chhattisgarh'),
  Text(0.5387533543302139, 0.9590332753340434, 'Delhi'),
  Text(0.3002969884983478, 1.058216291076084, 'Goa'),
  Text(0.04528738179486117, 1.099067356011526, 'Gujarat'),
  Text(-0.2122185967272179, 1.0793346409724516, 'Haryana'),
  Text(-0.4580264707607053, 1.0001058704369719, 'Himachal Pradesh'),
  Text(-0.678586597442501, 0.8657483640015782, 'Jammu & Kashmir'),
  Text(-0.8617410629623605, 0.6836682970596933, 'Jharkhand'),
  Text(-0.997393862030241, 0.46390245093597043, 'Karnataka'),
  Text(-1.0780674189664503, 0.21856495639286794, 'Kerala'),
  Text(-1.0993147733581872, -0.03882047238813081, 'Madhya Pradesh'),
  Text(-1.0599647094137454, -0.2940660041511672, 'Maharashtra'),
  Text(-0.962186316773495, -0.5331017649697436, 'Meghalaya'),
  Text(-0.8113694239985132, -0.7427514105003915, 'Odisha'),
  Text(-0.6158274955897184, -0.9114584442944701, 'Puducherry'),
  Text(-0.3863393697154118, -1.029923245396422, 'Punjab'),
  Text(-0.13555509739265537, -1.0916156904198784, 'Rajasthan'),
  Text(0.12270136501011895, -1.0931351128861673, 'Sikkim'),
  Text(0.37419417233780977, -1.0343977578225998, 'Tamil Nadu'),
  Text(0.6050603117177515, -0.9186413985794553, 'Telangana'),
  Text(0.8025737720321927, -0.7522468613733249, 'Tripura'),
  Text(0.9558470385567217, -0.5443862956415647, 'Uttar Pradesh'),
  Text(1.0527516335591014, -0.31892632070533733, 'Uttarakhand'),
  Text(1.095130309519196, -0.1033905468231516, 'West Bengal')],
 [Text(0.5958514403764614, 0.07043479964688049, '3.75%'),
  Text(0.5630063810051522, 0.207421828522172, '3.75%'),
  Text(0.49912677718622567, 0.3329751646830314, '3.75%'),
  Text(0.4077338579339519, 0.4401739441337889, '3.75%'),
  Text(0.29386546599829844, 0.5231090592731146, '3.75%'),
  Text(0.16379835736273513, 0.5772088860415003, '3.75%'),
  Text(0.024702208251742453, 0.599491285097196, '3.75%'),
  Text(-0.11575559821484611, 0.5887279859849734, '3.75%'),
  Text(-0.24983262041493012, 0.5455122929656209, '3.75%'),
  Text(-0.370138144059546, 0.47222638036449716, '3.75%'),
  Text(-0.4700405797976511, 0.37290998021437816, '3.75%'),
  Text(-0.5440330156528587, 0.2530377005105293, '3.75%'),
  Text(-0.5880367739817002, 0.11921724894156431, '3.75%'),
  Text(-0.5996262400135566, -0.02117480312079862, '3.75%'),
  Text(-0.5781625687711338, -0.16039963862790937, '3.75%'),
  Text(-0.5248289000582699, -0.29078278089258736, '3.75%'),
  Text(-0.44256514036282535, -0.4051371330002135, '3.75%'),
  Text(-0.3359059066853009, -0.4971591514333472, '3.75%'),
  Text(-0.2107305652993155, -0.5617763156707756, '3.75%'),
  Text(-0.07393914403235746, -0.5954267402290245, '3.75%'),
  Text(0.0669280172782467, -0.5962555161197276, '3.75%'),
  Text(0.20410591218425986, -0.5642169588123271, '3.75%'),
  Text(0.3300328973005917, -0.5010771264978846, '3.75%'),
  Text(0.4377675120175596, -0.4103164698399953, '3.75%'),
  Text(0.5213711119400299, -0.2969379794408535, '3.75%'),
  Text(0.5742281637595098, -0.17395981129382035, '3.37%'),
  Text(0.5973438051922886, -0.056394843721719046, '3.00%')])
```

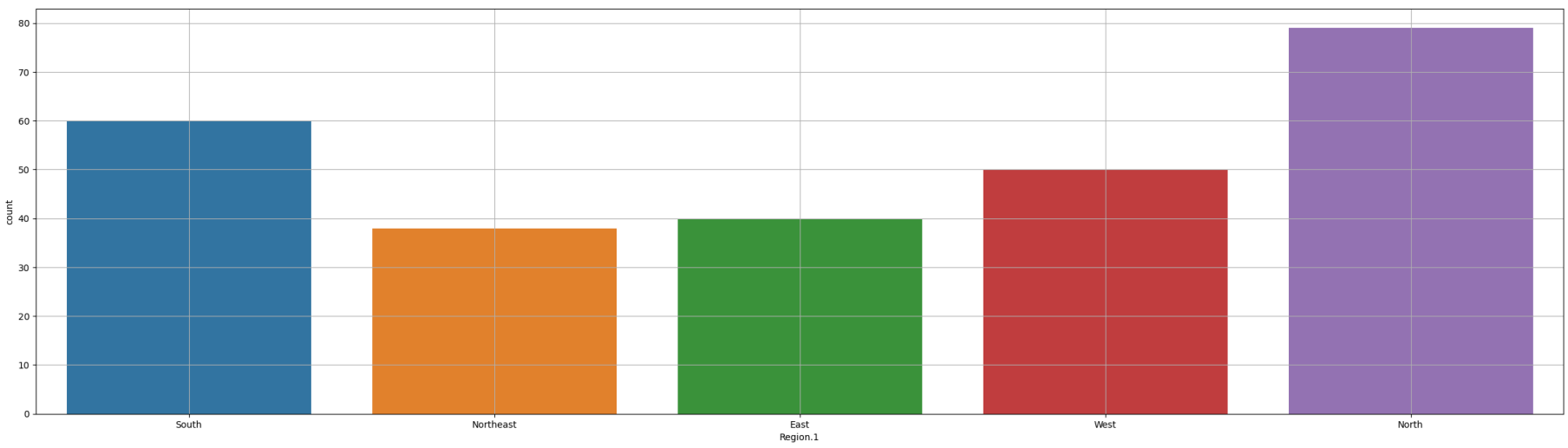


so from above chart we can clearly see the percentage of the regions

Region 1

In [42]:
```
plt.figure(figsize=(30,8))
sns.countplot(x='Region.1',data= df2)
plt.grid(True)
plt.show
```

Out[42]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [43]:
```
df2.corr()
```

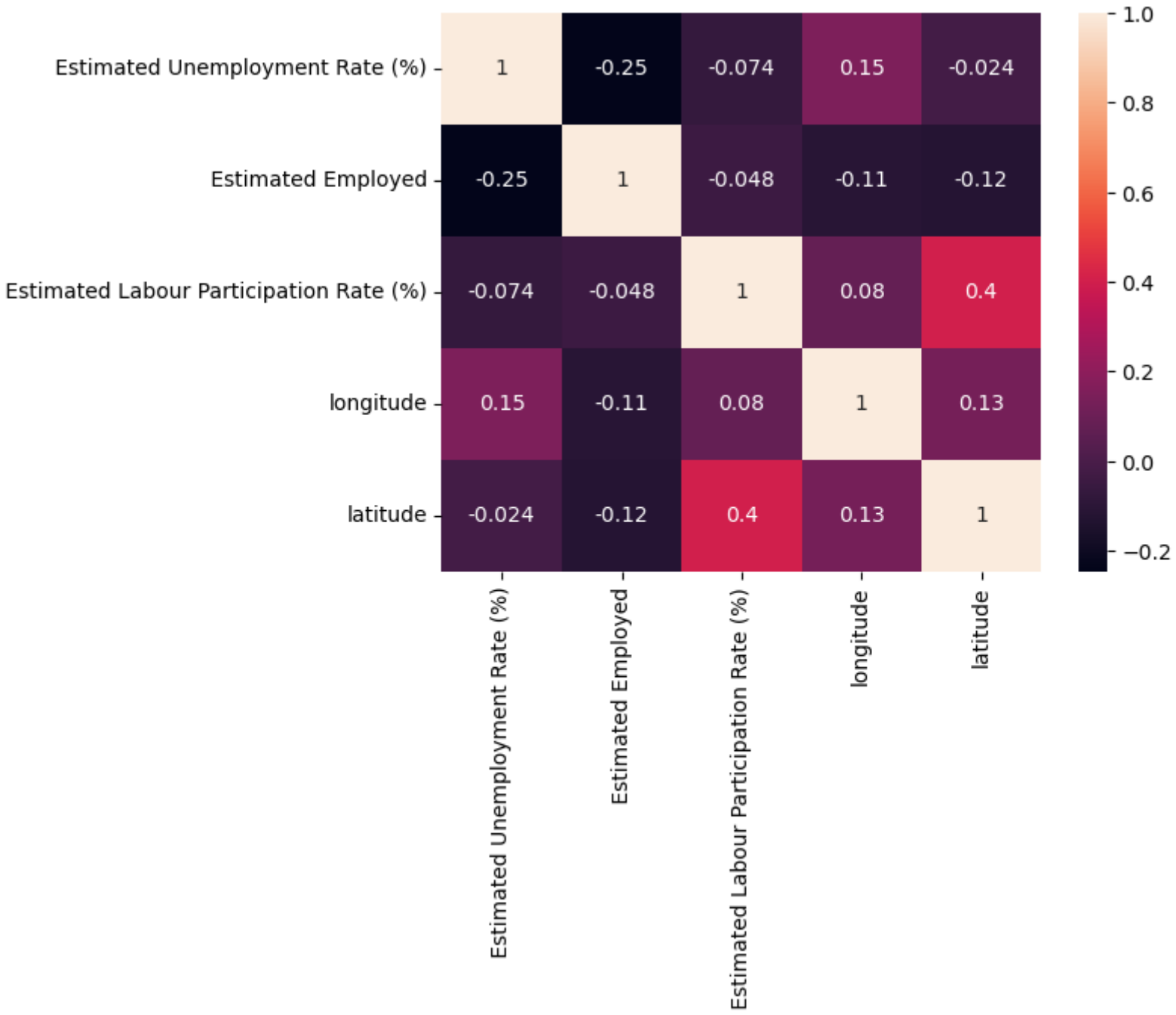
Out[43]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	longitude	latitude
Estimated Unemployment Rate (%)	1.000000	-0.245176	-0.073540	0.149976	-0.023976
Estimated Employed	-0.245176	1.000000	-0.047948	-0.113664	-0.119321
Estimated Labour Participation Rate (%)	-0.073540	-0.047948	1.000000	0.080372	0.397836
longitude	0.149976	-0.113664	0.080372	1.000000	0.125895
latitude	-0.023976	-0.119321	0.397836	0.125895	1.000000

Displaying The Data Using Heatmap

In [44]:
```
sns.heatmap(df2.corr(),annot=True,cmap="rocket")
```

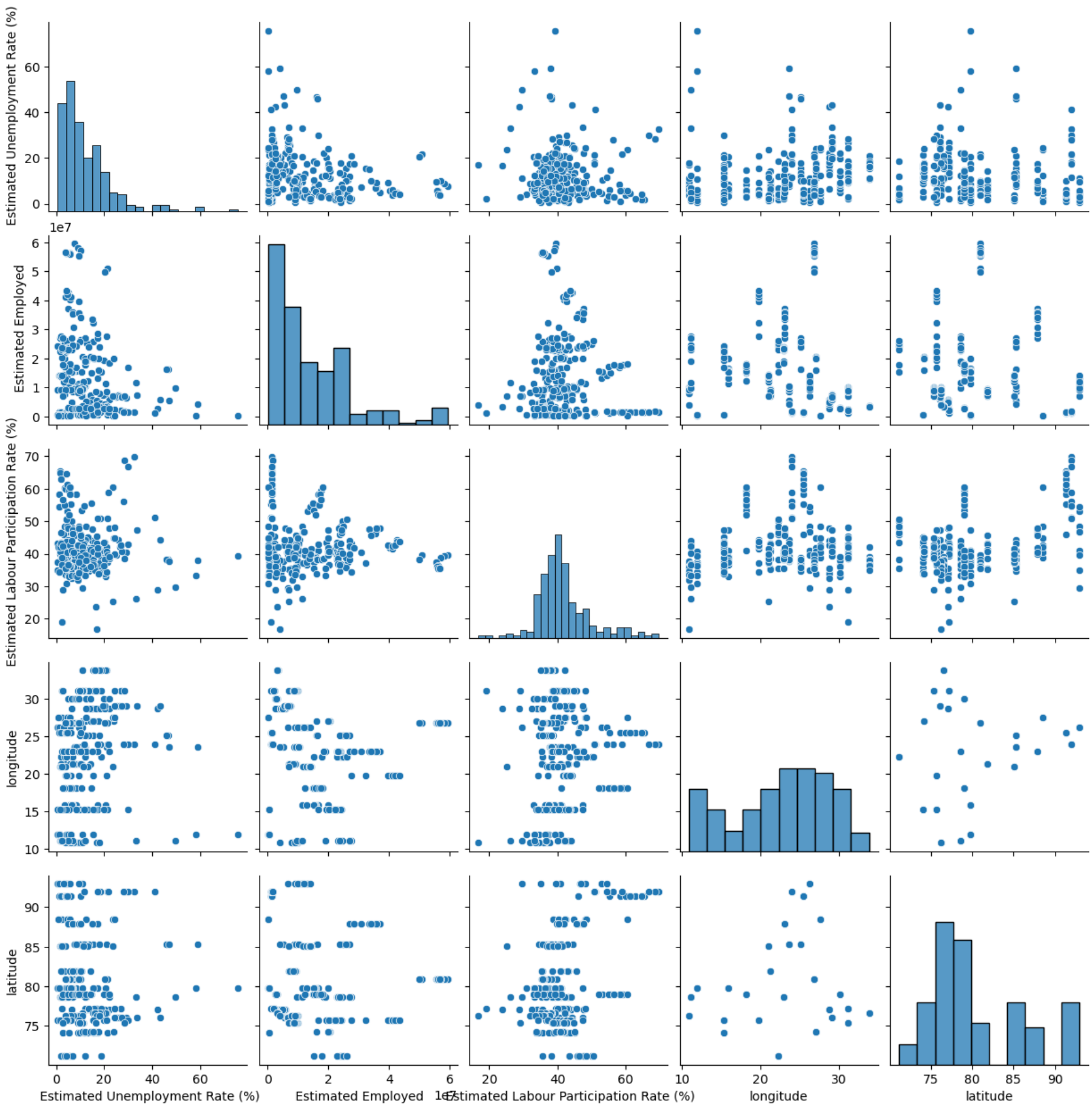
Out[44]:
```
<AxesSubplot:>
```



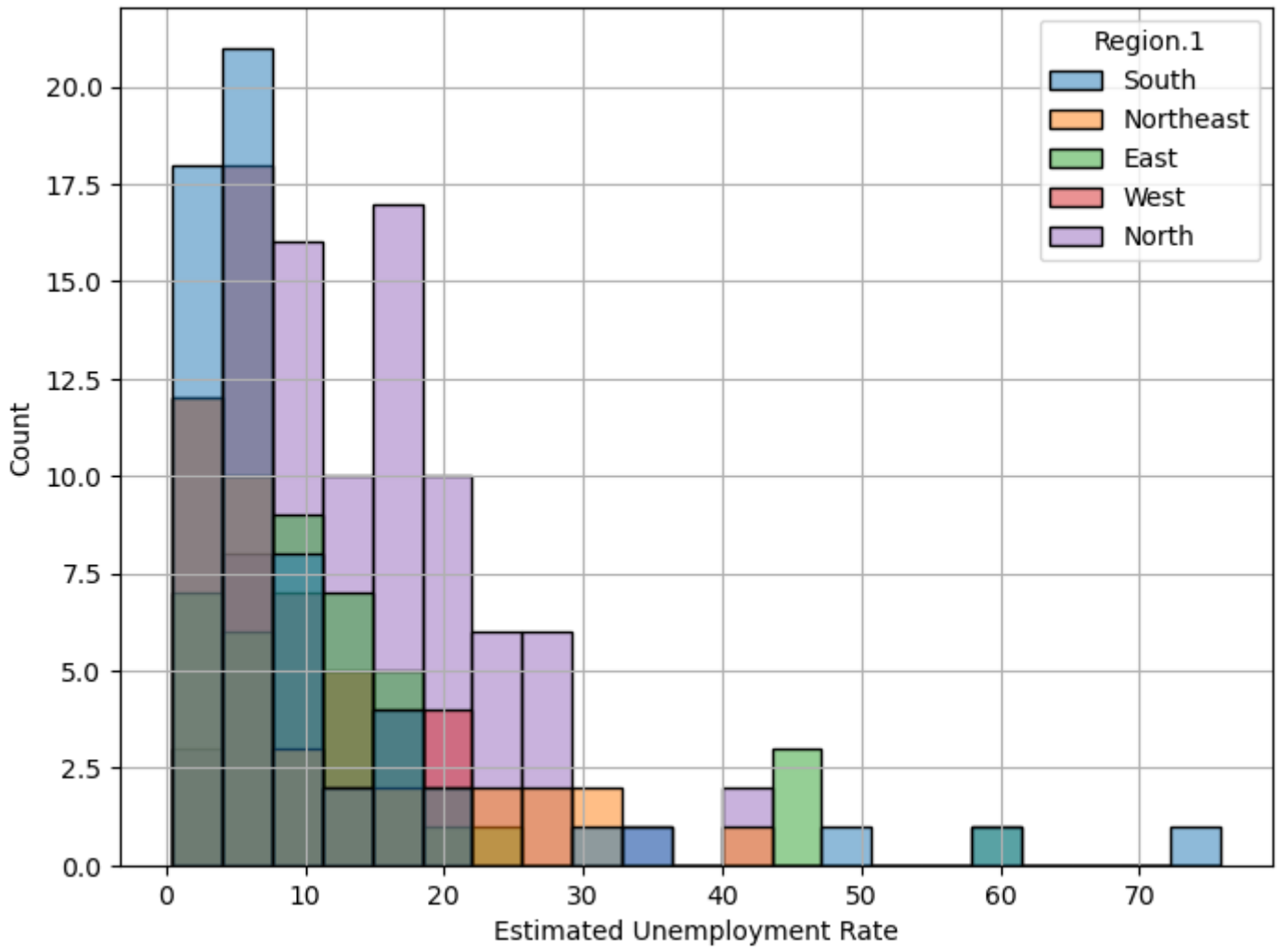
Data is presenting by using Pairplot

In [46]:
```
sns.pairplot(df2, palette='hls')
```

Out[46]:
```
<seaborn.axisgrid.PairGrid at 0x205ba8f8730>
```



In [57]:
```
df2.columns=['States','Date','Frequency','Estimated Unemployment Rate','Estimated Employed','Estimated Labour Participation Ra
plt.figure(figsize=(8,6))
sns.histplot(x="Estimated Unemployment Rate",hue='Region.1',data=df2)
plt.grid(True)
plt.show()
```



So from the above chart we can analyze the unployment rate in India

THANK YOU