OASIS INFOBYTE :Data Science

Author: Sujata Gaikwad

Task1: Unemployment Analysis With Python

In [1]:
```python
import pandas as pd
import numpy as np
```

In [2]:
```
data=pd.read_csv(r"C:\Users\HP\Downloads\Unemployment_Rate_upto_11_2020.csv")
data
```

Out[2]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	lat
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	7
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	7
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	7
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	7
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	7
...	...	...	...	...	...	...	...	...	
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	8
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	8
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	8
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	8
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	8

267 rows × 9 columns

```
In [3]:  data.isnull().sum()
```

```
Out[3]:  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 [4]:  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 [5]:  data["Region"].unique()
```

```
Out[5]:  array(['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', 'Goa',
                'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',
                'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh',
                'Maharashtra', 'Meghalaya', 'Odisha', 'Puducherry', 'Punjab',
                'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana', 'Tripura',
                'Uttar Pradesh', 'Uttarakhand', 'West Bengal'], dtype=object)
```

In [6]: `data.corr()`

Out[6]:

|	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

In [7]: `data.describe()`

Out[7]:

	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

```python
import seaborn as sns
sns.pairplot(data,hue='Region.1')
```

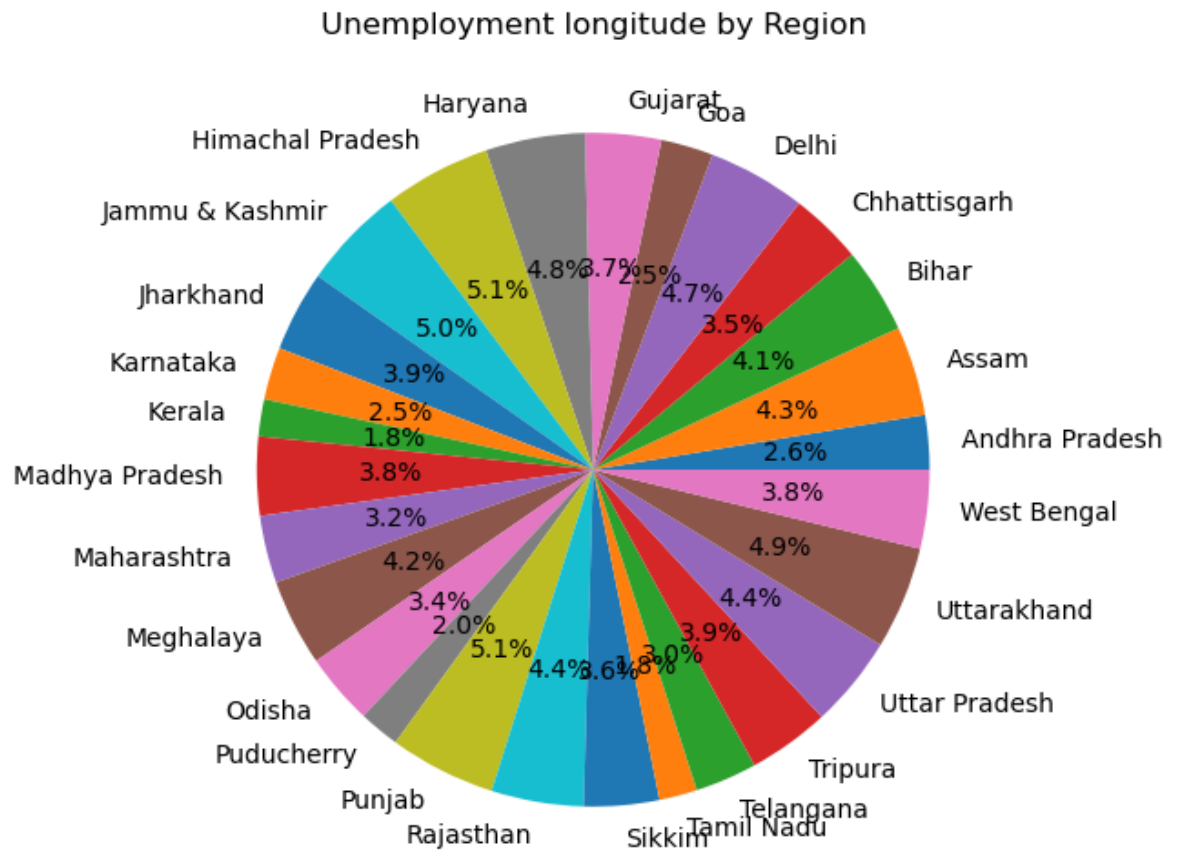Out[8]: <seaborn.axisgrid.PairGrid at 0x1c48d20ecd0>
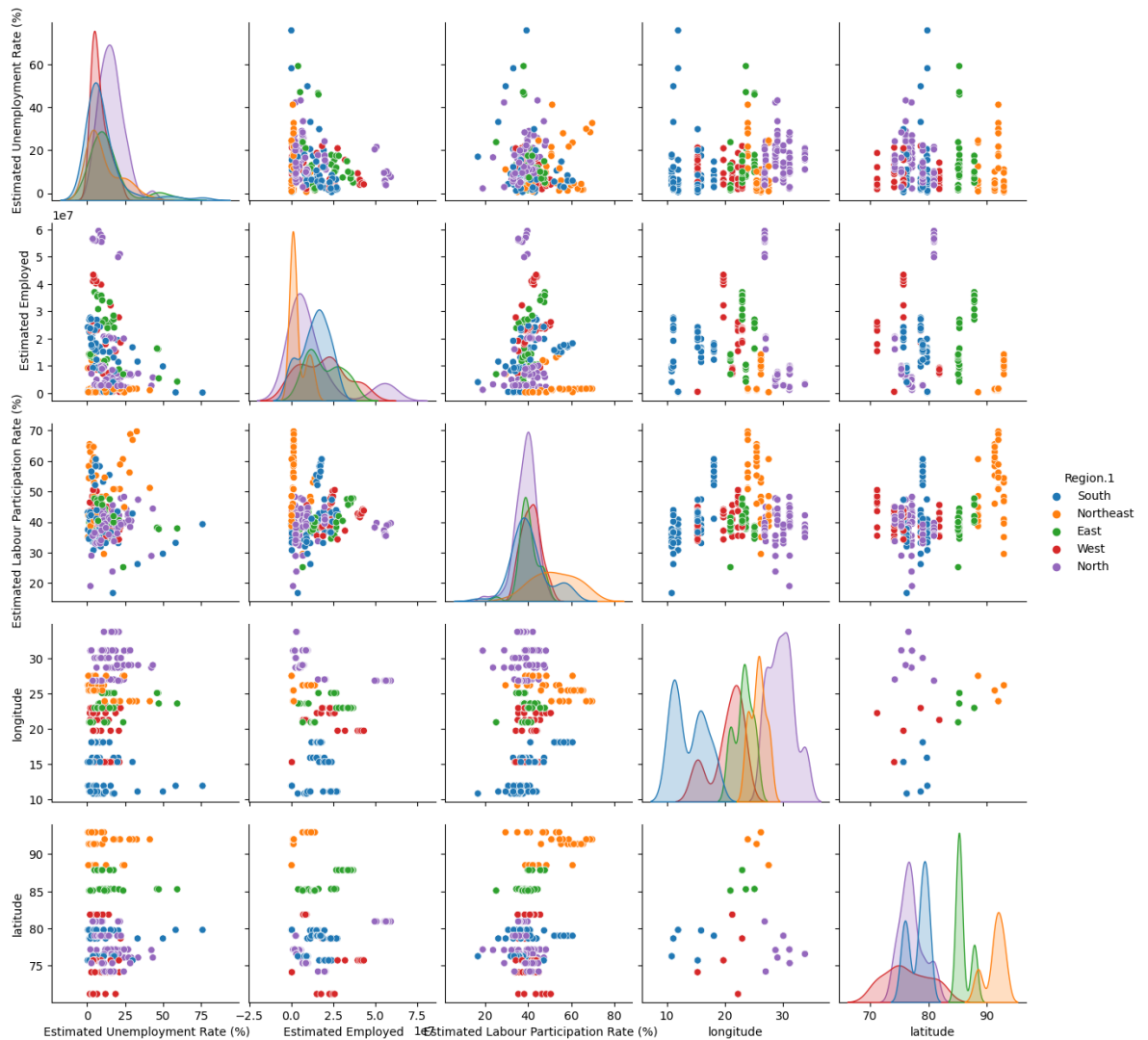
```
In [9]: import matplotlib.pyplot as plt
        plt.figure(figsize=(10,8))
        sns.heatmap(data.corr(),annot=True)
        plt.title("Correlation Heatmap")
        plt.show()
```


Correlation Heatmap

```python
# pie chart of unemployment rate distribution by region
region_unemployment= data.groupby("Region")["longitude"].sum().reset_index()
plt.figure(figsize=(8,6))
plt.pie(region_unemployment["longitude"],labels=region_unemployment["Region"],
plt.title("Unemployment longitude by Region")
plt.show()
```

Unemployment longitude by Region

In [ ]: