Unempolyment Analysis

> About This Project

Here's a summary of the code and the conclusions you can Find It:

1. **Importing Libraries:** You've imported the necessary libraries such as NumPy, Pandas, Matplotlib, Seaborn, and Plotly Express for data analysis and visualization.
2. **Loading Data:** You've loaded the unemployment data from a CSV file and displayed the first few rows using `data.head()`.
3. **Data Information:** You've printed information about the dataset using `data.info()`. It shows that the dataset contains 267 entries with various columns such as Region, Date, Frequency, Estimated Unemployment Rate (%), Estimated Employed, Estimated Labour Participation Rate (%), Region.1, longitude, and latitude.
4. **Data Description:** You've provided a statistical description of the dataset using `data.describe()`, giving insights into the central tendency and spread of numerical columns.
5. **Checking for Missing Values:** You've checked for missing values in the dataset using `data.isnull().sum()` and found that there are no missing values in any column.
6. **Correlation Heatmap:** You've created a heatmap using Seaborn to visualize the correlation between numerical columns, showing how they are related to each other.
7. **Histograms:** You've plotted histograms to visualize the distribution of Estimated Employed and Estimated Unemployment Rate, grouped by the Region.1 column.
8. **Sunburst Plot:** You've created a sunburst plot using Plotly Express to visualize the Estimated Unemployment Rate across different states and regions in India.

Conclusions: Based on the analysis and visualizations performed in the code:

- **Unemployment Rate Distribution:** The histogram of Estimated Unemployment Rate shows how the unemployment rate is distributed across different regions. This gives you an idea of which regions have higher or lower unemployment rates.
- **Employment Distribution:** The histogram of Estimated Employed showcases the distribution of employed individuals across different regions. This provides insights into the employment distribution in various parts of India.
- **Unemployment by Region:** The sunburst plot visually presents the Estimated Unemployment Rate across different states and regions. This can help you understand the

variation in unemployment rates across India and how different regions contribute to the overall unemployment picture.

- **Correlation Analysis:** The correlation heatmap gives an idea of the relationships between different numerical variables. For instance, you can observe if there's any correlation between unemployment rate and labor participation rate, or between unemployment rate and employed population.

Remember that the conclusions will depend on the specific data in your dataset, the context of the analysis, and the goals you're trying to achieve. It's always a good practice to provide detailed context and interpretation along with your visualizations and analysis.

Import the Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly .express as px
```

```
data = pd.read_csv('/content/Unemployment.csv')
```

```
data.head()
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longit
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9

```
# 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
```

```
# describing the dataset
data.describe()
```

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	longitude
count	267.000000	2.670000e+02	267.000000	267.000000
mean	12.236929	1.396211e+07	41.681573	22.826048
std	10.803283	1.336632e+07	7.845419	6.270731
min	0.500000	1.175420e+05	16.770000	10.850500
25%	4.845000	2.838930e+06	37.265000	18.112400
50%	9.650000	9.732417e+06	40.390000	23.610200
75%	16.755000	2.187869e+07	44.055000	27.278400
max	75.850000	5.943376e+07	69.690000	33.778200

```
# check null/missing values
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
```

```
data.head()
```

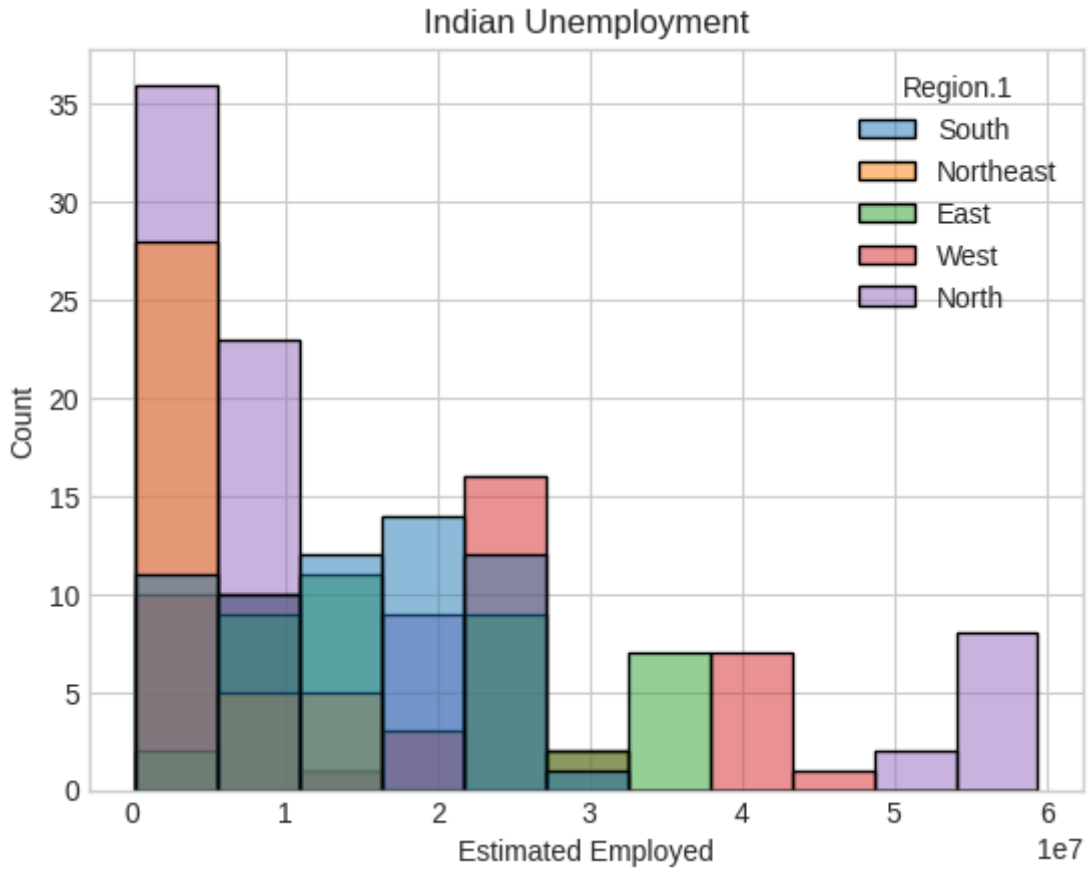
	States	Date	Frequency	Estimated Unemployment Rate	Estimated Employed	Estimated Labour Participation Rate	Region	Longitud
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.912
	Andhra	29-						

```
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(8,6))

# Compute the correlation matrix and plot the heatmap
ax = sns.heatmap(data.corr(), annot=True,linewidth=3)

plt.show()
```

```
<ipython-input-12-35ecd18313f7>:1: MatplotlibDeprecationWarning: The seaborn styles s
  plt.style.use('seaborn-whitegrid')
<ipython-input-12-35ecd18313f7>:5: FutureWarning: The default value of numeric_only i
  ax = sns.heatmap(data.corr(), annot=True,linewidth=3)
```


```
# plotting histplot

data.columns=['States','Date','Frequency','Estimated Unemployment Rate',
```

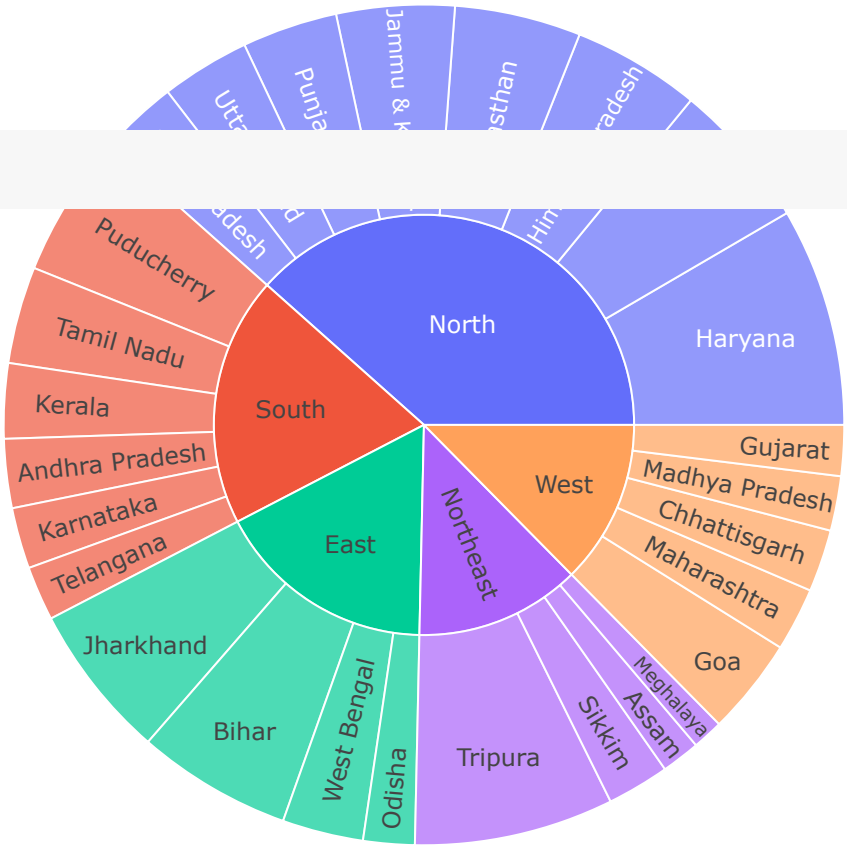
```
                'Estimated Employed','Estimated Labour Participation Rate',
                'Region.1','Longitude','Latitude']
plt.title('Indian Unemployment')
sns.histplot(x='Estimated Employed',hue='Region.1',data=data)
plt.show()
```




```
# plotting histplot

plt.figure(figsize=(10,8))
plt.title("Indian Unemployment")
sns.histplot(x="Estimated Unemployment Rate",hue='Region.1',data=data)
plt.show()
```


```
# plotting sunburst

unemployment = data[['States','Region.1','Estimated Unemployment Rate']]
figure = px.sunburst(unemployment,path=['Region.1','States'],
                     values='Estimated Unemployment Rate',
                     width=700,height=600, color_continuous_scale='RdY1Gn',
                     title="Unemployment Rate in India")
figure.show()
```

Unemployment Rate in India



✓ 1s completed at 11:27 AM