```python
In [1]:  #load require liabraries of python
         !pip install plotly
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         import plotly.express as px
```

Requirement already satisfied: plotly in c:\users\admin\documents\arduino\lib\site-pa
ckages (5.9.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\admin\documents\arduino\li
b\site-packages (from plotly) (8.2.2)

```python
In [3]:  df = pd.read_csv('E:\\Unemployment in India.csv') #Read dataset
         df.sample(5)   #fetch five sample of dataset
```

Out[3]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
190	Maharashtra	30-06-2019	Monthly	4.34	23056511.0	45.69	Rural
183	Madhya Pradesh	31-01-2020	Monthly	3.66	16183702.0	39.35	Rural
235	Puducherry	30-09-2019	Monthly	0.00	175718.0	43.34	Rural
303	Telangana	30-06-2020	Monthly	19.30	8891181.0	58.97	Rural
438	Chhattisgarh	30-04-2020	Monthly	20.13	1066126.0	24.06	Urban

```python
In [4]:  df.columns
```

Out[4]: Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
               ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
               'Area'],
              dtype='object')

```python
In [5]:  df[' Frequency'].value_counts()
```

Out[5]:   Frequency
         Monthly   381
          Monthly   359
         Name: count, dtype: int64

```python
In [12]:  df.isnull().sum()
```

```
Out[12]: Region                                       28
          Date                                        28
          Frequency                                   28
          Estimated Unemployment Rate (%)             28
          Estimated Employed                          28
          Estimated Labour Participation Rate (%)     28
         Area                                         28
         day                                          28
         month                                        28
         year                                         28
         dtype: int64
```

In [13]: `df.duplicated().sum()`

Out[13]: 27

In [14]:
```python
print('row count--->',df.shape[0])
print('column count--->',df.shape[1])
```

```
row count---> 768
column count---> 10
```

In [15]: `df.dtypes`

```
Out[15]: Region                                       object
          Date                                        object
          Frequency                                   object
          Estimated Unemployment Rate (%)            float64
          Estimated Employed                         float64
          Estimated Labour Participation Rate (%)    float64
         Area                                         object
         day                                          object
         month                                        object
         year                                         object
         dtype: object
```

In [16]:
```python
df[["day", "month", "year"]] = df[' Date'].str.split("-", expand = True)
df
```

Out[16]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	day	month	year
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural	31	05	2019
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural	30	06	2019
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural	31	07	2019
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural	31	08	2019
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural	30	09	2019
...	...	...	...	...	...	...	...	...	...	...
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

768 rows × 10 columns

```python
In [17]: df.drop(columns=[' Frequency'],axis=1,inplace=True)
```

```python
In [18]: df[:5]
```

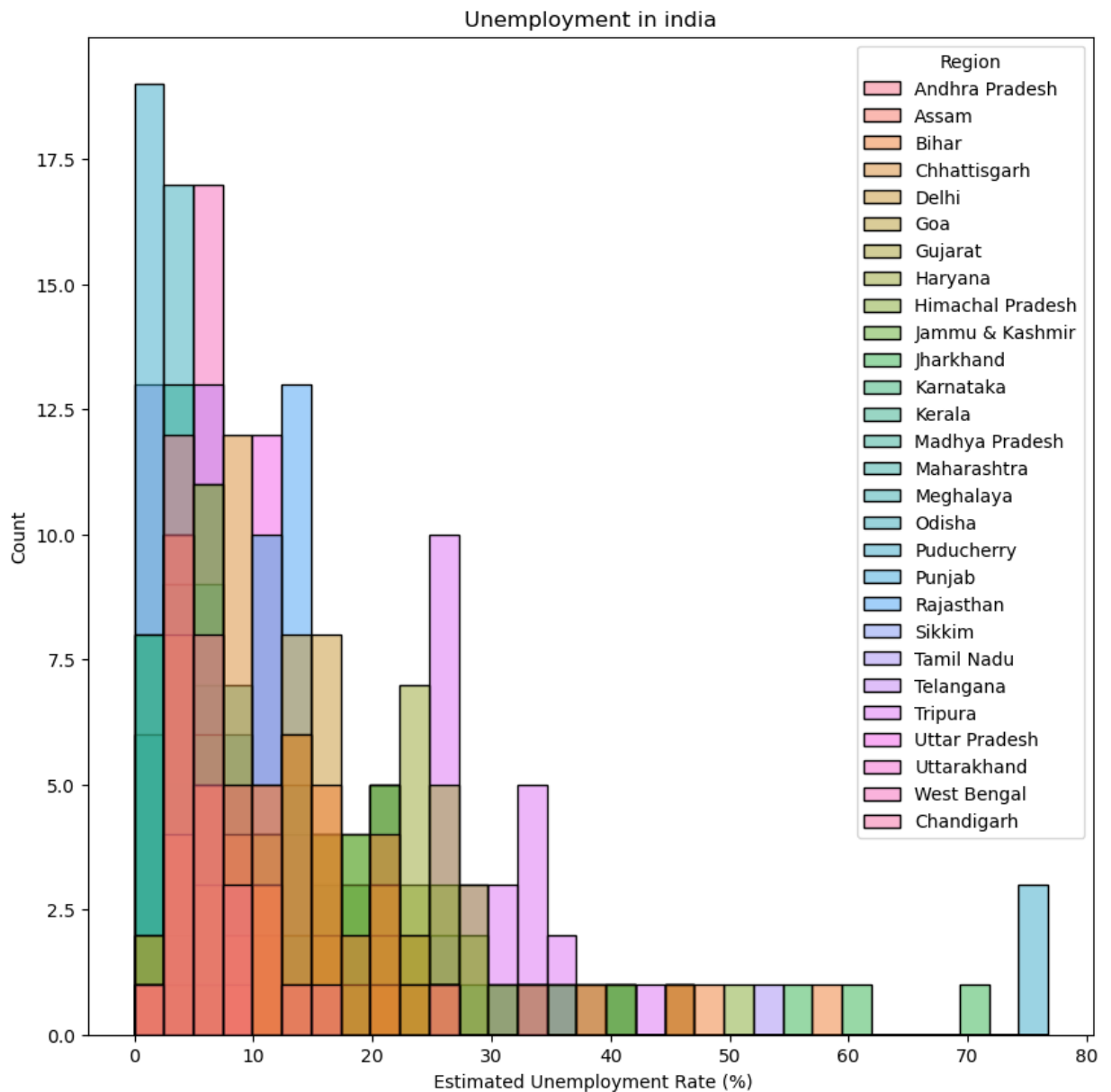|	Region	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	day	month	year
0	Andhra Pradesh	31-05-2019	3.65	11999139.0	43.24	Rural	31	05	2019
1	Andhra Pradesh	30-06-2019	3.05	11755881.0	42.05	Rural	30	06	2019
2	Andhra Pradesh	31-07-2019	3.75	12086707.0	43.50	Rural	31	07	2019
3	Andhra Pradesh	31-08-2019	3.32	12285693.0	43.97	Rural	31	08	2019
4	Andhra Pradesh	30-09-2019	5.17	12256762.0	44.68	Rural	30	09	2019

In [19]:
```python
import matplotlib.pyplot as plt
```

In [22]:
```python
df.columns
```

Out[22]:
```
Index(['Region', ' Date', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area', 'day', 'month', 'year'],
      dtype='object')
```

In [23]:
```python
plt.figure(figsize=(10,10))
plt.title("Unemployment in india")
sns.histplot(x=' Estimated Unemployment Rate (%)',hue= "Region", data=df,kde=False)
plt.show()
```

Unemployment in india

```
In [24]: df.columns
```

```
Out[24]: Index(['Region', ' Date', ' Estimated Unemployment Rate (%)',
                ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
                'Area', 'day', 'month', 'year'],
               dtype='object')
```

```
In [25]: df.month.unique()
```

```
Out[25]: array(['05', '06', '07', '08', '09', '10', '11', '12', '01', '02', '03',
                '04', nan], dtype=object)
```

```
In [27]: sns.barplot(x='month',y=' Estimated Unemployment Rate (%)',hue='year',data=df)
```

```
Out[27]: <Axes: xlabel='month', ylabel=' Estimated Unemployment Rate (%)'>
```
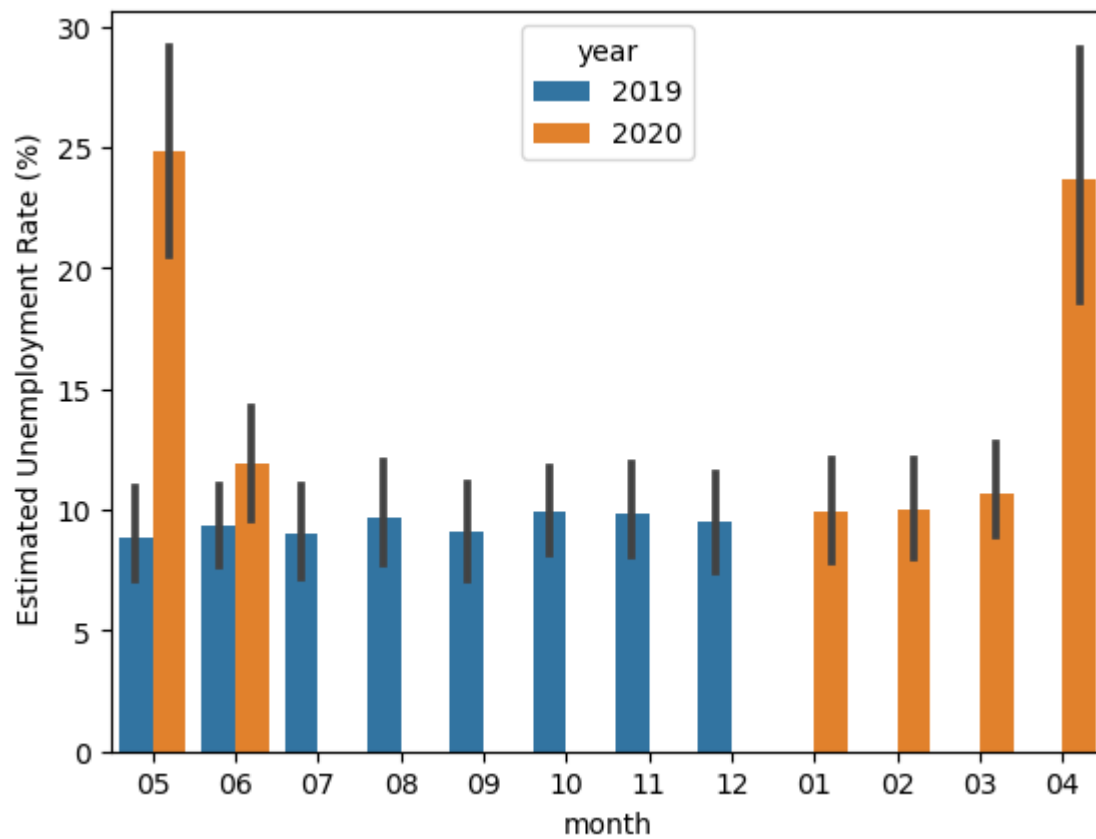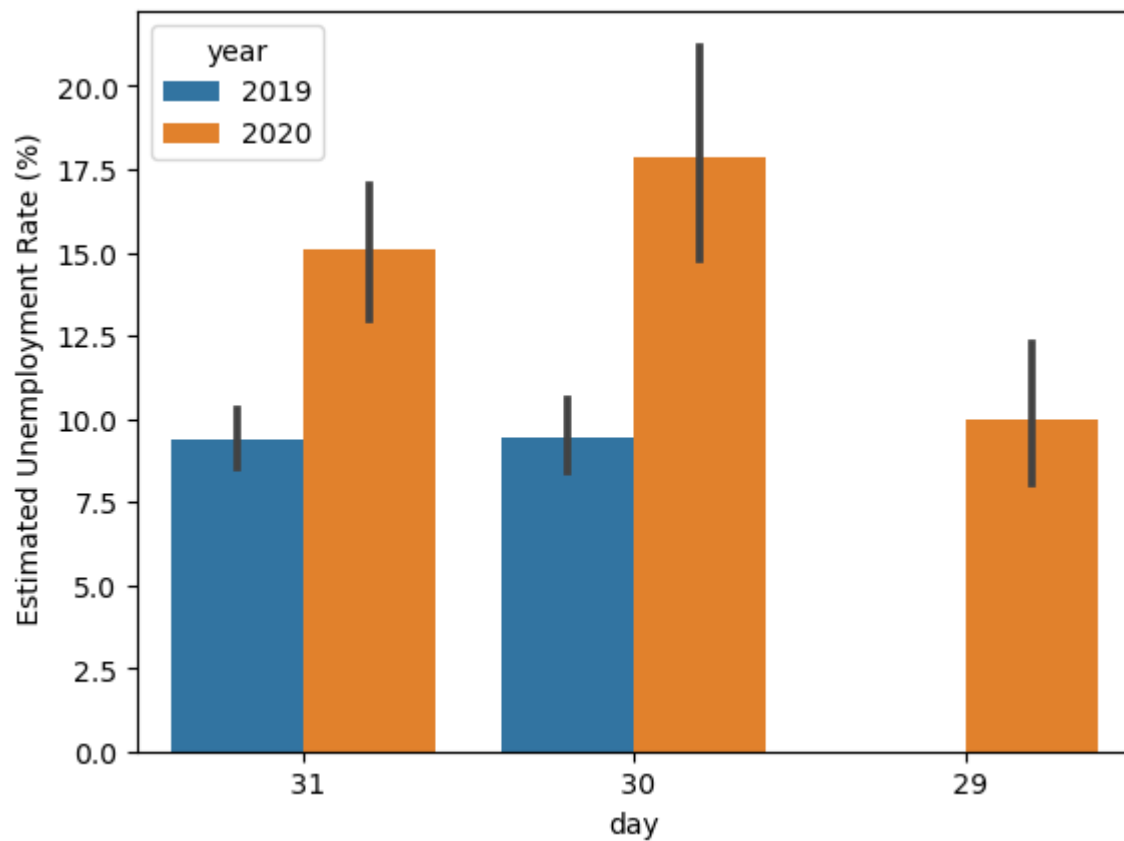
In [28]: `df.day.unique()`

Out[28]: `array([' 31', ' 30', ' 29', nan], dtype=object)`
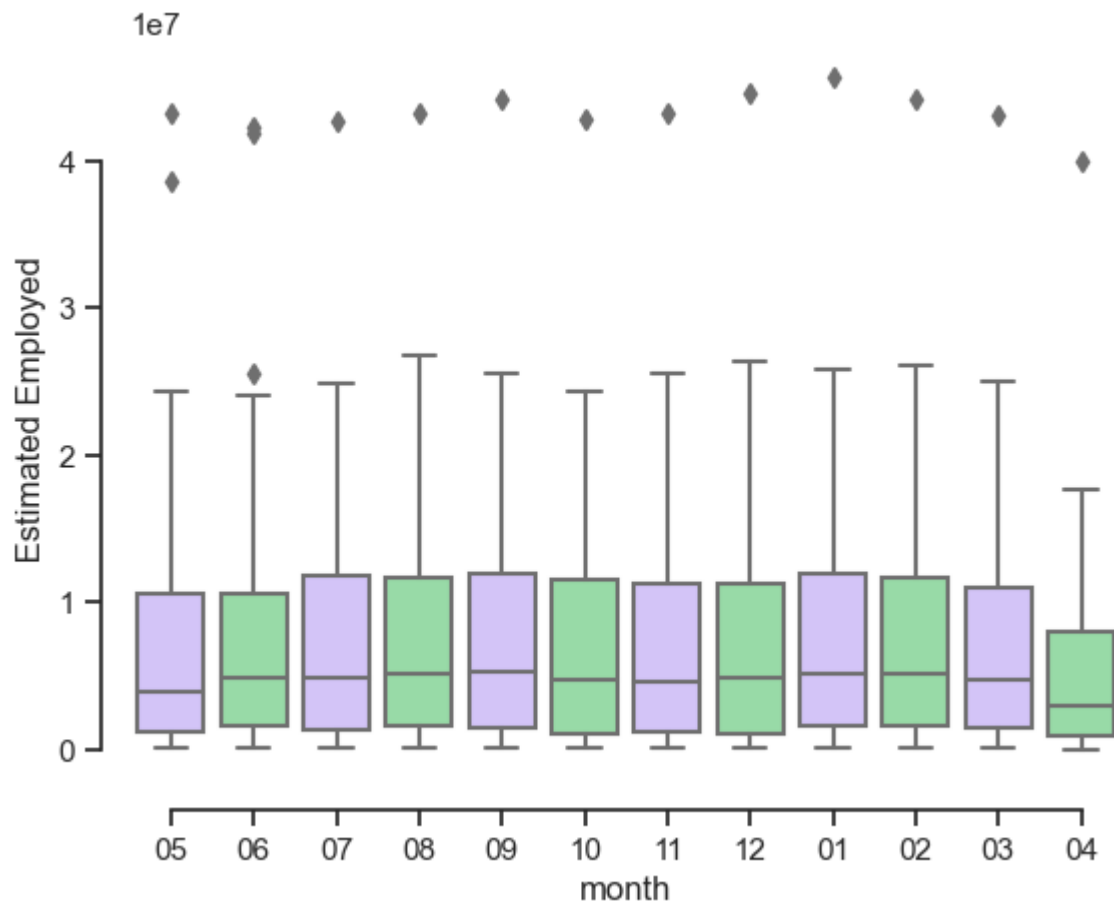
In [29]: `sns.barplot(x='day',y=' Estimated Unemployment Rate (%)',hue='year',data=df)`

Out[29]: `<Axes: xlabel='day', ylabel=' Estimated Unemployment Rate (%)'>`

```
In [30]:  import seaborn as sns
          sns.set_theme(style="ticks", palette="pastel")

          # Draw a nested boxplot to show bills by day and time
          sns.boxplot(x="month", y=' Estimated Employed', palette=["m", "g"],
                      data=df)
          sns.despine(offset=10, trim=True)
```

`df[:5]`

Out[31]:

	Region	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	day	month	year
0	Andhra Pradesh	31-05-2019	3.65	11999139.0	43.24	Rural	31	05	2019
1	Andhra Pradesh	30-06-2019	3.05	11755881.0	42.05	Rural	30	06	2019
2	Andhra Pradesh	31-07-2019	3.75	12086707.0	43.50	Rural	31	07	2019
3	Andhra Pradesh	31-08-2019	3.32	12285693.0	43.97	Rural	31	08	2019
4	Andhra Pradesh	30-09-2019	5.17	12256762.0	44.68	Rural	30	09	2019

In [32]: `df.drop('year',axis=1)`

	Region	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	day	month
0	Andhra Pradesh	31-05-2019	3.65	11999139.0	43.24	Rural	31	05
1	Andhra Pradesh	30-06-2019	3.05	11755881.0	42.05	Rural	30	06
2	Andhra Pradesh	31-07-2019	3.75	12086707.0	43.50	Rural	31	07
3	Andhra Pradesh	31-08-2019	3.32	12285693.0	43.97	Rural	31	08
4	Andhra Pradesh	30-09-2019	5.17	12256762.0	44.68	Rural	30	09
...	...	...	...	...	...	...	...	...
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

768 rows × 8 columns

In [37]:
```python
df.columns
```

Out[37]:
```
Index(['Region', ' Date', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area', 'day', 'month', 'year'],
      dtype='object')
```