Unemployment Analysis in india using python (During Covid pandemics) [Oasis infobyte Internship Task 2]   **PURAM HAREESH**

In [ ]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [ ]:

```
df=pd.read_csv("Unemployment in India.csv")
```

In [ ]:

```
df.head()
```

Out[ ]:

	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

In [ ]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 7 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Region                                    740 non-null    object 
 1    Date                                     740 non-null    object 
 2    Frequency                                740 non-null    object 
 3    Estimated Unemployment Rate (%)          740 non-null    float64
 4    Estimated Employed                       740 non-null    float64
 5    Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                      740 non-null    object 
dtypes: float64(3), object(4)
memory usage: 42.1+ KB
```

In [ ]:

```
#checking for null values
df.isnull().sum()
```

Out[ ]:

```
Region                                      28
 Date                                       28
 Frequency                                  28
 Estimated Unemployment Rate (%)            28
 Estimated Employed                         28
 Estimated Labour Participation Rate (%)    28
Area                                        28
dtype: int64
```

In [ ]:

```
df[df['Region'].isnull()].head()
```

Out[ ]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
359	NaN	NaN	NaN	NaN	NaN	NaN	NaN
360	NaN	NaN	NaN	NaN	NaN	NaN	NaN
361	NaN	NaN	NaN	NaN	NaN	NaN	NaN
362	NaN	NaN	NaN	NaN	NaN	NaN	NaN
363	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [ ]:

```
df.dropna(inplace=True)
```

In [ ]:

```
df.isnull().sum()
```

Out[ ]:

```
Region                                      0
 Date                                       0
 Frequency                                  0
 Estimated Unemployment Rate (%)            0
 Estimated Employed                         0
 Estimated Labour Participation Rate (%)    0
Area                                        0
dtype: int64
```

In [ ]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 740 entries, 0 to 753
Data columns (total 7 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Region                                    740 non-null    object 
 1    Date                                     740 non-null    object 
 2    Frequency                                740 non-null    object 
 3    Estimated Unemployment Rate (%)          740 non-null    float64
 4    Estimated Employed                       740 non-null    float64
 5    Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                      740 non-null    object 
dtypes: float64(3), object(4)
memory usage: 46.2+ KB
```

In [ ]:
```
df[' Date'].unique()
```

Out[ ]:
```
array([' 31-05-2019', ' 30-06-2019', ' 31-07-2019', ' 31-08-2019',
       ' 30-09-2019', ' 31-10-2019', ' 30-11-2019', ' 31-12-2019',
       ' 31-01-2020', ' 29-02-2020', ' 31-03-2020', ' 30-04-2020',
       ' 31-05-2020', ' 30-06-2020'], dtype=object)
```

So this dataset contains data from May 2019 to June 2020

In [ ]:
```
df[" Date"]=pd.to_datetime(df[' Date'])
df.sort_values(by=['Region',' Date'],inplace=True)
```

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

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

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

In [ ]:
```
#checking for duplicates
df.duplicated().sum()
```

Out[ ]:
```
0
```

In [ ]:
```
df['month']=df[' Date'].dt.strftime('%m-%y')
df.to_csv('final_unemployementdata.csv')
```

In [ ]:
```
df.head()
```

Out[ ]:

	Region	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	month
0	Andhra Pradesh	2019-05-31	3.65	11999139.0	43.24	Rural	05-19
373	Andhra Pradesh	2019-05-31	6.09	4788661.0	37.45	Urban	05-19
1	Andhra Pradesh	2019-06-30	3.05	11755881.0	42.05	Rural	06-19
374	Andhra Pradesh	2019-06-30	3.80	4824630.0	36.76	Urban	06-19
2	Andhra Pradesh	2019-07-31	3.75	12086707.0	43.50	Rural	07-19

In [ ]:

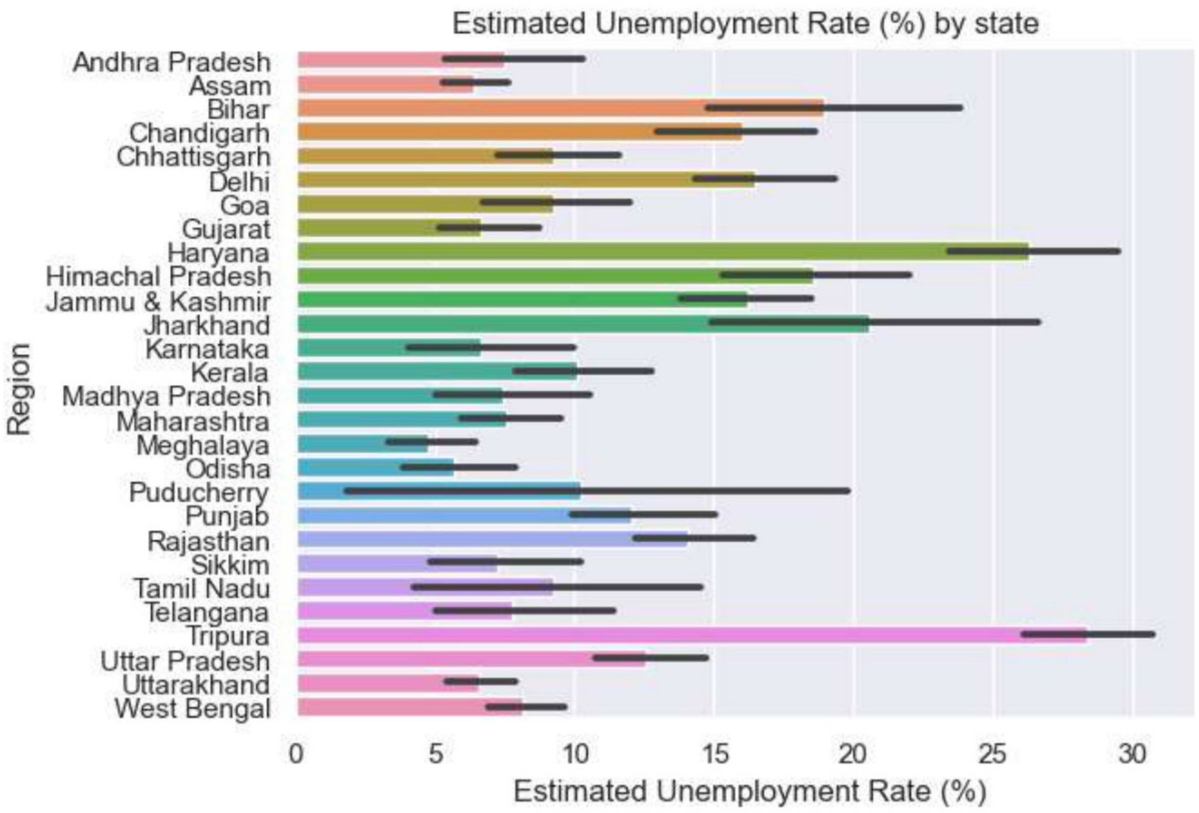
```
df.describe()
```

Out[ ]:

	Date	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
count	740	740.000000	7.400000e+02	740.000000
mean	2019-12-12 18:36:58.378378496	11.787946	7.204460e+06	42.630122
min	2019-05-31 00:00:00	0.000000	4.942000e+04	13.330000
25%	2019-08-31 00:00:00	4.657500	1.190404e+06	38.062500
50%	2019-11-30 00:00:00	8.350000	4.744178e+06	41.160000
75%	2020-03-31 00:00:00	15.887500	1.127549e+07	45.505000
max	2020-06-30 00:00:00	76.740000	4.577751e+07	72.570000
std	NaN	10.721298	8.087988e+06	8.111094

In [ ]:

```
#Estimated Unemployment Rate (%) by state
sns.barplot(y='Region',x=' Estimated Unemployment Rate (%)',data=df)
plt.title('Estimated Unemployment Rate (%) by state')
plt.show()
```



In [ ]:

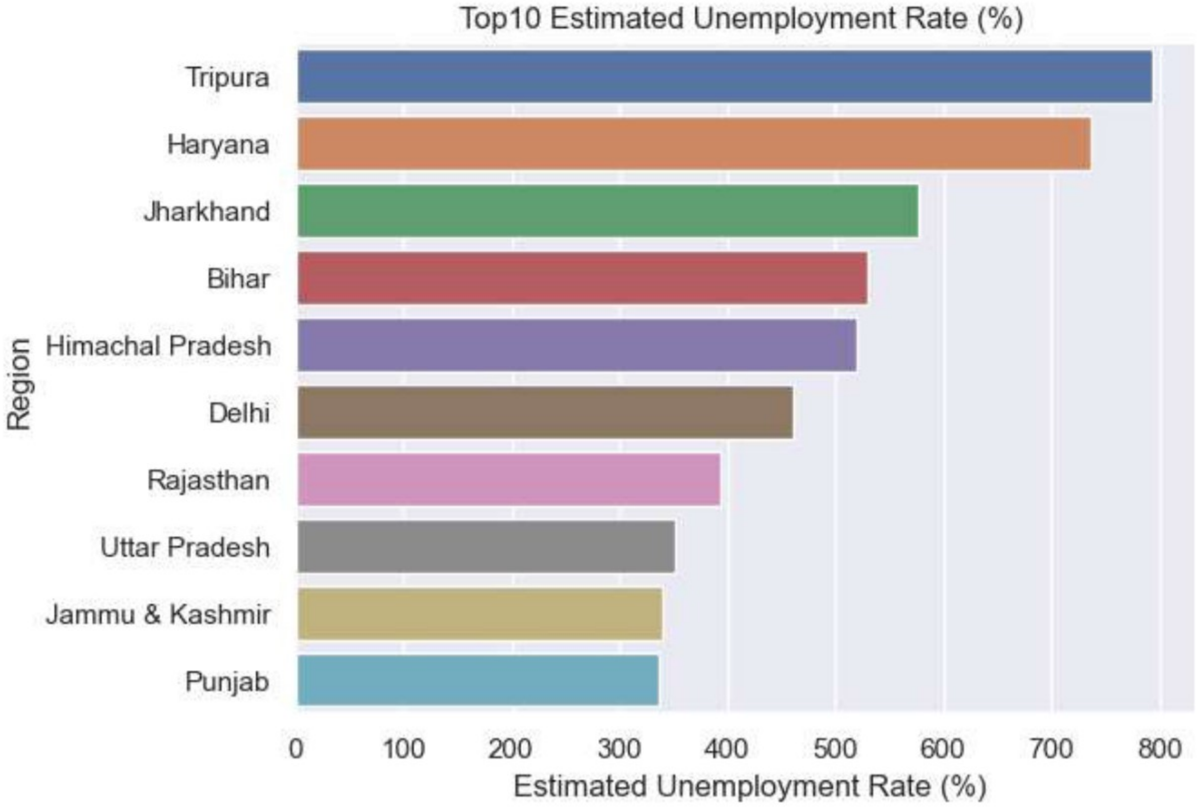
```
#Top10 Estimated Unemployment Rate (%)
top10=df.groupby(['Region'])[' Estimated Unemployment Rate (%)'].sum().reset_index(
top10=top10.sort_values(by=[' Estimated Unemployment Rate (%)'],ascending=False)
top10.head(10)
```

Out[ ]:

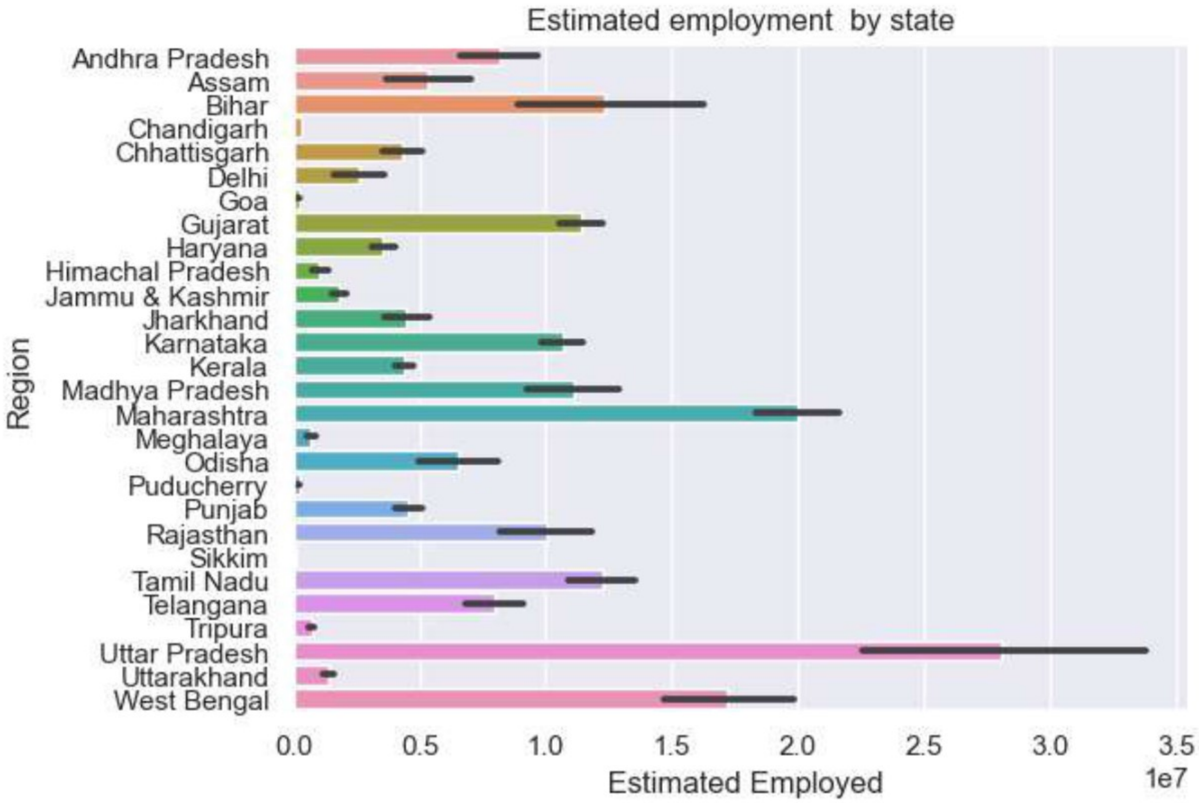
	Region	Estimated Unemployment Rate (%)
24	Tripura	793.81
8	Haryana	735.93
11	Jharkhand	576.38
2	Bihar	529.71
9	Himachal Pradesh	519.13
5	Delhi	461.87
20	Rajasthan	393.63
25	Uttar Pradesh	351.44
10	Jammu & Kashmir	339.96
19	Punjab	336.87

In [ ]:

```
sns.barplot(y='Region',x=' Estimated Unemployment Rate (%)',data=top10.head(10))
plt.title('Top10 Estimated Unemployment Rate (%) ')
plt.show()
```

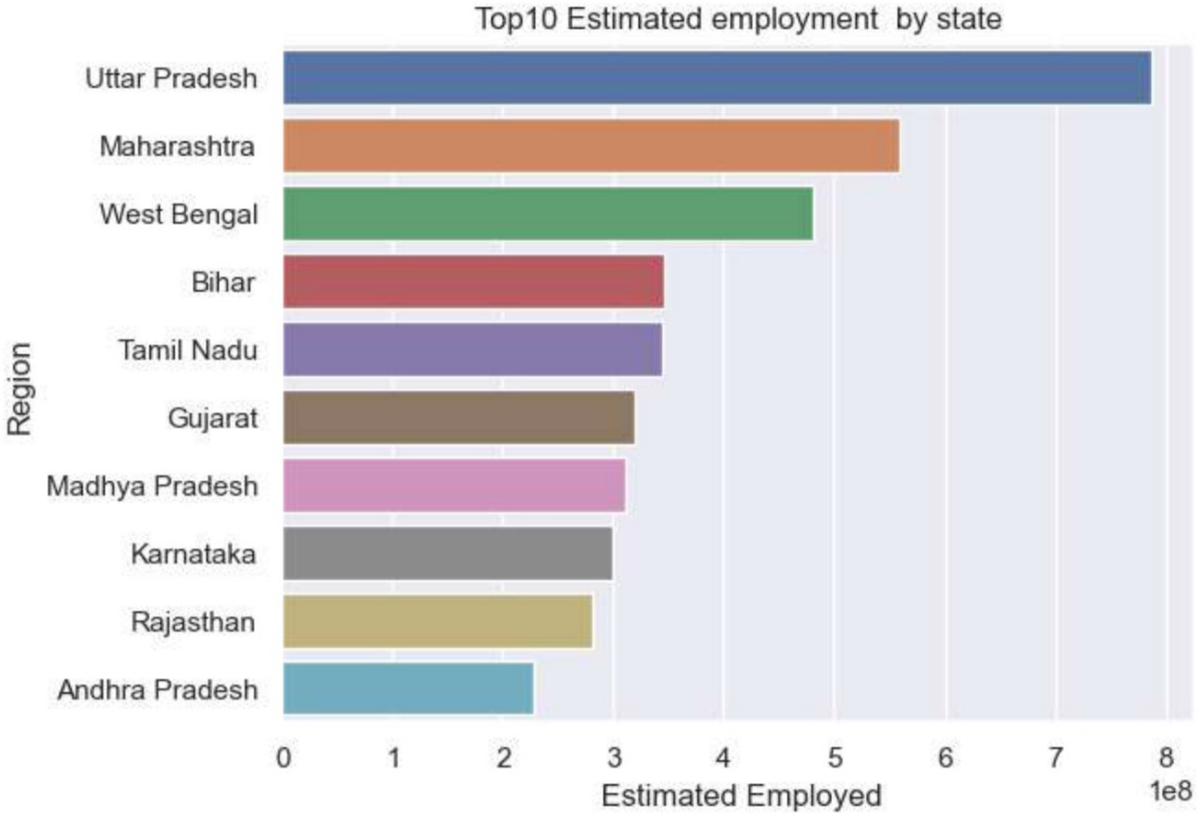


```
In [ ]: #Estimated employment  by state
        sns.barplot(y='Region',x=' Estimated Employed',data=df)
        plt.title('Estimated employment  by state')
        plt.show()
```



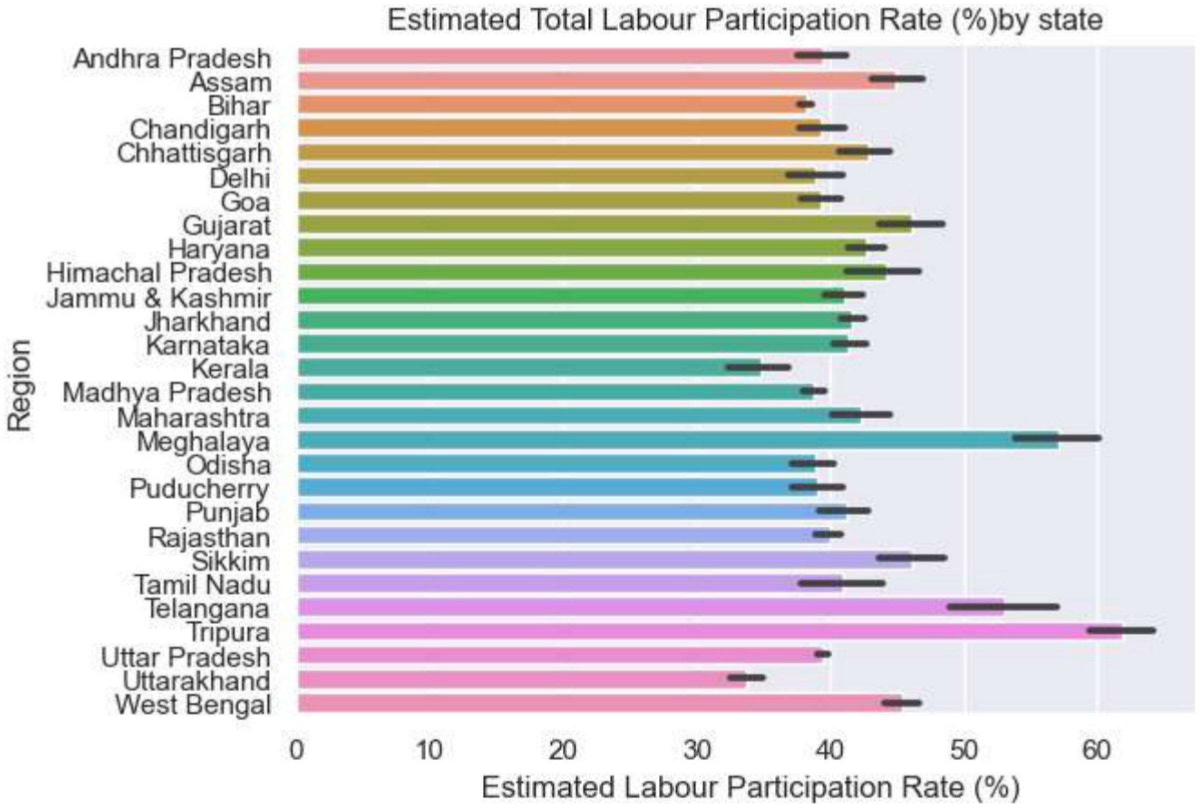
```
In [ ]: #Top10 Estimated employment  by state
        top10e=df.groupby(['Region'])[' Estimated Employed'].sum().reset_index()
        top10e=top10e.sort_values(by=[' Estimated Employed'],ascending=False)
        top10e.head(10)

Out[ ]:
```

```
In [ ]: sns.barplot(y='Region',x=' Estimated Employed',data=top10e.head(10))
        plt.title('Top10 Estimated employment  by state')
        plt.show()
```



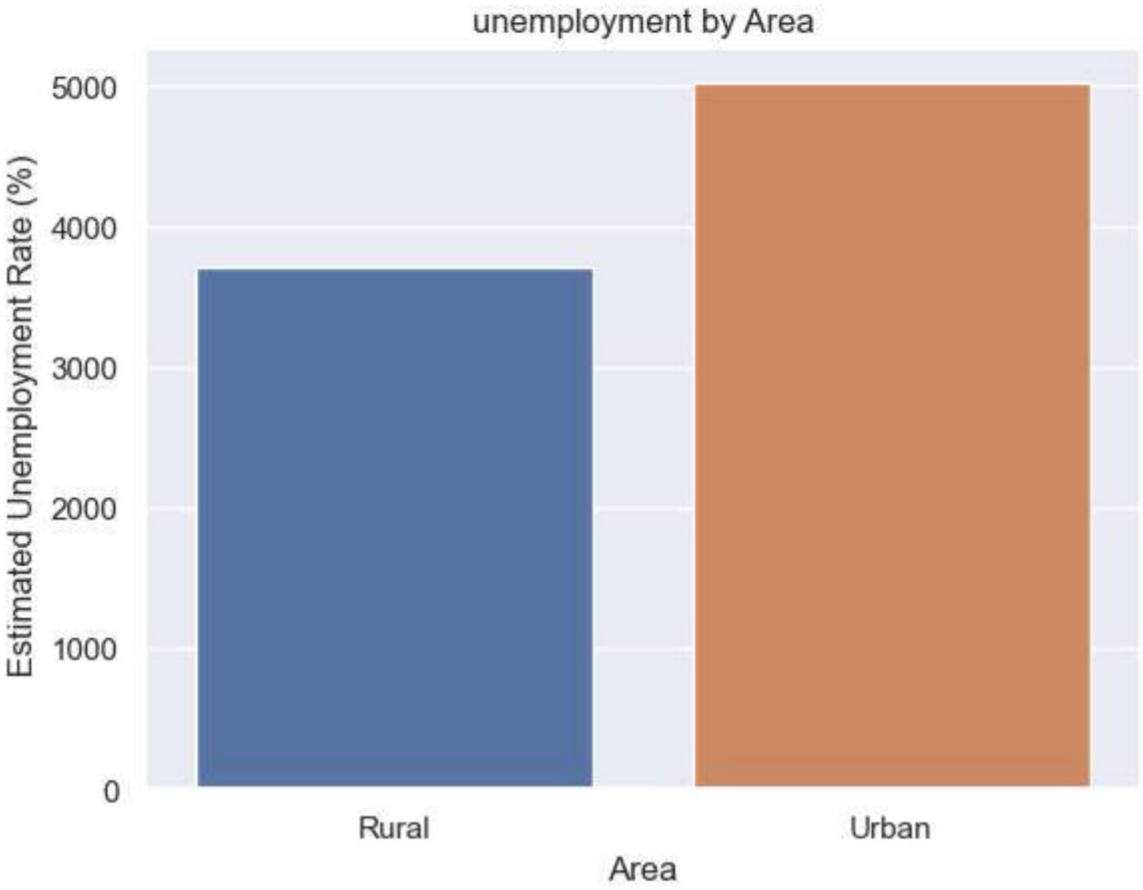
In [ ]:

```
#Estimated Total Labour Participation Rate (%)by state
sns.barplot(y='Region',x=' Estimated Labour Participation Rate (%)',data=df)
plt.title('Estimated Total Labour Participation Rate (%)by state')
plt.show()
```



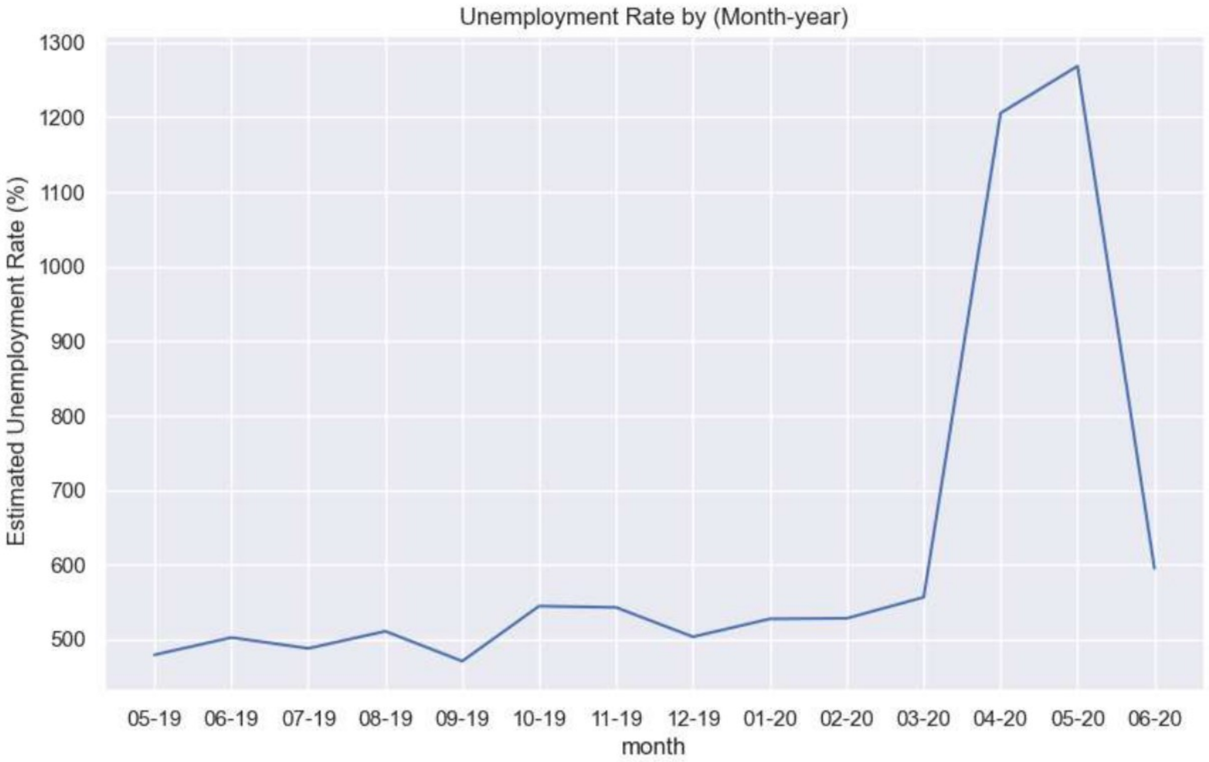
In [ ]:

```
#unemployment by Area
ar=df.groupby(['Area'])[' Estimated Unemployment Rate (%)'].sum().reset_index()
sns.barplot(x='Area',y=' Estimated Unemployment Rate (%)',data=ar)
plt.title('unemployment by Area')
plt.show()
```



In [ ]:

```
#Unemployment Rate by (Month-year)
month=df.groupby('month')[' Estimated Unemployment Rate (%)'].sum().reset_index()
month=month.sort_values(by=['month'],key=lambda x:pd.to_datetime(x,format='%m-%y'))
plt.figure(figsize=(10,6))
sns.lineplot(x=month['month'],y=month[' Estimated Unemployment Rate (%)'],data=mont
plt.title('Unemployment Rate by (Month-year)')
plt.show()
```


Unemployment Rate by (Month-year)
Estimated Unemployment Rate (%)
1300
1200
1100
1000
900
800
700
600
500
05-19
06-19
07-19
08-19
09-19
10-19
11-19
12-19
01-20
02-20
03-20
04-20
05-20
06-20
month