In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

In [2]:
```
df=pd.read_csv('C:\\Users\\anush\\OneDrive\\Desktop\\Oasis Infobyte\\Task 2\\Unemployment in India.csv')
```

In [3]:
```
df.head()
```

Out[3]:

	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 [4]:
```
df.tail()
```

Out[4]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
763	NaN	NaN	NaN	NaN	NaN	NaN	NaN
764	NaN	NaN	NaN	NaN	NaN	NaN	NaN
765	NaN	NaN	NaN	NaN	NaN	NaN	NaN
766	NaN	NaN	NaN	NaN	NaN	NaN	NaN
767	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [5]:
```
df.shape
```

Out[5]:
```
(768, 7)
```

In [6]:
```
df.size
```

Out[6]:
```
5376
```

In [7]:
```
df.isna().sum()
```

Out[7]:
```
Region                                   28
 Date                                    28
 Frequency                               28
 Estimated Unemployment Rate (%)         28
 Estimated Employed                      28
 Estimated Labour Participation Rate (%) 28
Area                                     28
dtype: int64
```

In [8]:
```
df.dropna(inplace=True)
```

In [9]:
```
df.shape
```

Out[9]:
```
(740, 7)
```

In [10]:
```
df.size
```

Out[10]:
```
5180
```

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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 [12]:
```
df.describe(include='all')
```

Out[12]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
count	740	740	740	740.000000	7.400000e+02	740.000000	740
unique	28	14	2	NaN	NaN	NaN	2
top	Andhra Pradesh	31-10-2019	Monthly	NaN	NaN	NaN	Urban
freq	28	55	381	NaN	NaN	NaN	381
mean	NaN	NaN	NaN	11.787946	7.204460e+06	42.630122	NaN
std	NaN	NaN	NaN	10.721298	8.087988e+06	8.111094	NaN
min	NaN	NaN	NaN	0.000000	4.942000e+04	13.330000	NaN
25%	NaN	NaN	NaN	4.657500	1.190404e+06	38.062500	NaN
50%	NaN	NaN	NaN	8.350000	4.744178e+06	41.160000	NaN
75%	NaN	NaN	NaN	15.887500	1.127549e+07	45.505000	NaN
max	NaN	NaN	NaN	76.740000	4.577751e+07	72.570000	NaN

In [13]:
```
df.columns
```

Out[13]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

In [14]:
```
df.columns=df.columns.str.strip()
```

In [15]:
```
df.columns=df.columns.str.replace(' ','_')
```

In [16]:
```
df.columns
```

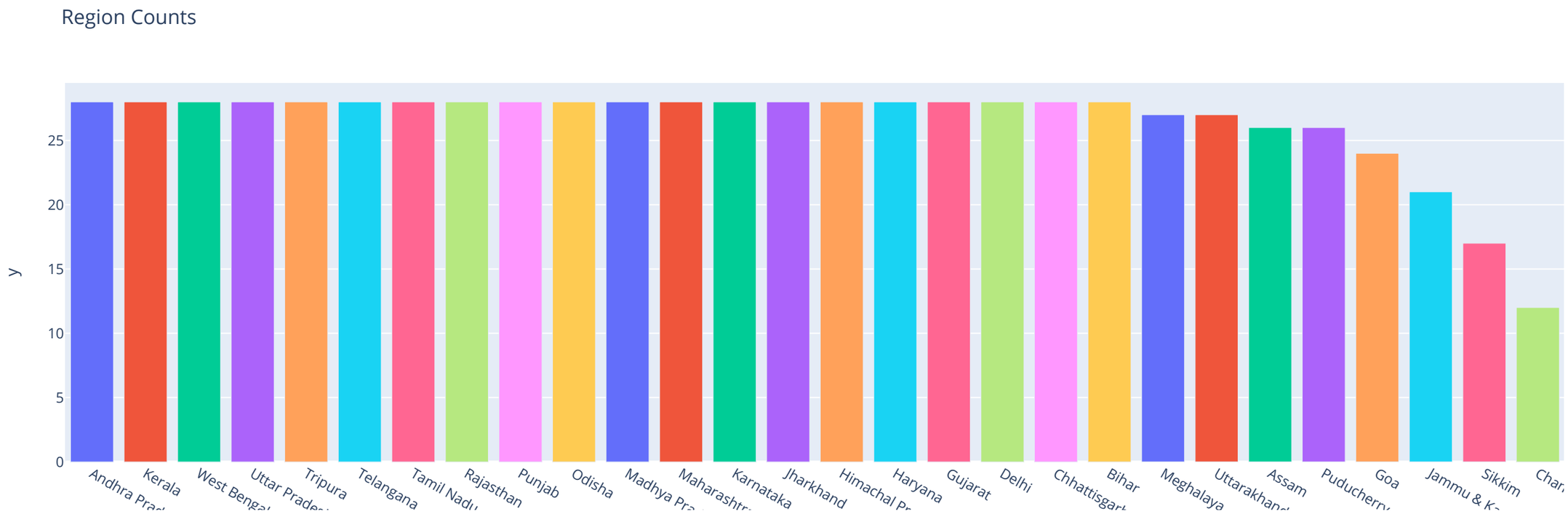
Out[16]:
```
Index(['Region', 'Date', 'Frequency', 'Estimated_Unemployment_Rate_(%)',
       'Estimated_Employed', 'Estimated_Labour_Participation_Rate_(%)',
       'Area'],
      dtype='object')
```

In [17]:
```
df['Date']=pd.to_datetime(df['Date'])
```

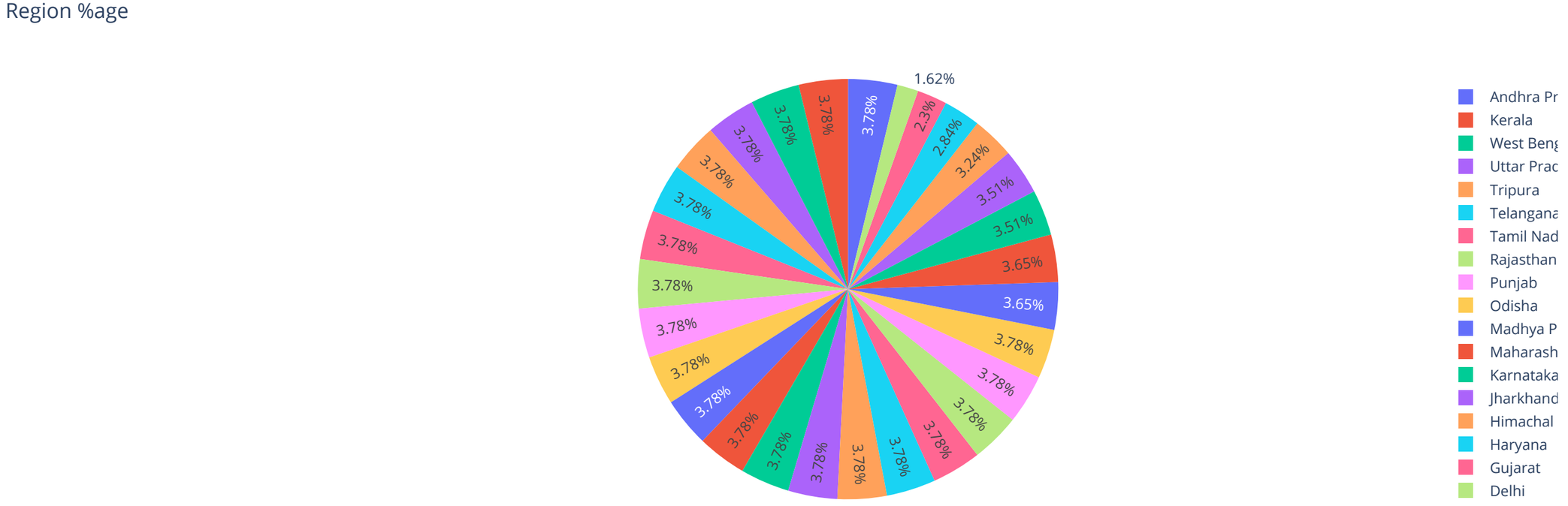
In [18]:
```
df['Region'].value_counts()
```

Out[18]:
```
Andhra Pradesh      28
Kerala              28
West Bengal         28
Uttar Pradesh       28
Tripura             28
Telangana           28
Tamil Nadu          28
Rajasthan           28
Punjab              28
Odisha              28
Madhya Pradesh      28
Maharashtra         28
Karnataka           28
Jharkhand           28
Himachal Pradesh    28
Haryana             28
Gujarat             28
Delhi               28
Chhattisgarh        28
Bihar               28
Meghalaya           27
Uttarakhand         27
Assam               26
Puducherry          26
Goa                 24
Jammu & Kashmir     21
Sikkim              17
Chandigarh          12
Name: Region, dtype: int64
```

In [19]:
```
px.bar(df,x=df['Region'].value_counts().keys(),y=df['Region'].value_counts(),color=df['Region'].value_counts().keys(),title='Region Counts')
```



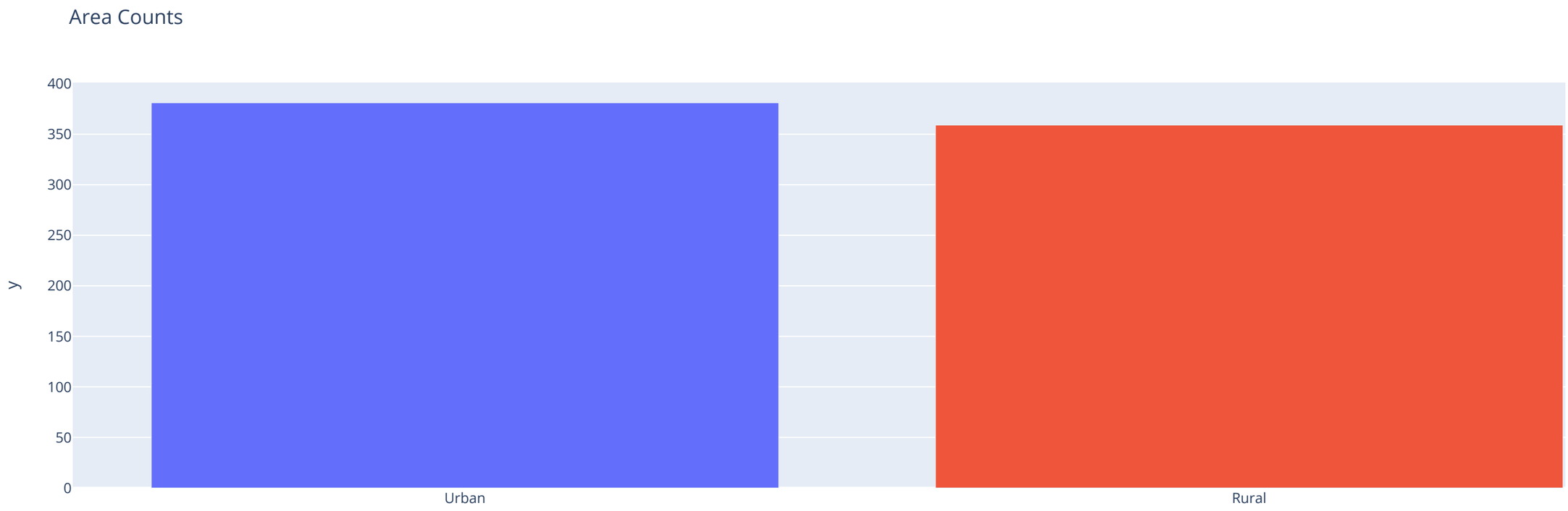
In [20]:
```
px.pie(df,names=df['Region'].value_counts().keys(),values=df['Region'].value_counts(),color=df['Region'].value_counts().keys(),title='Region %age')
```



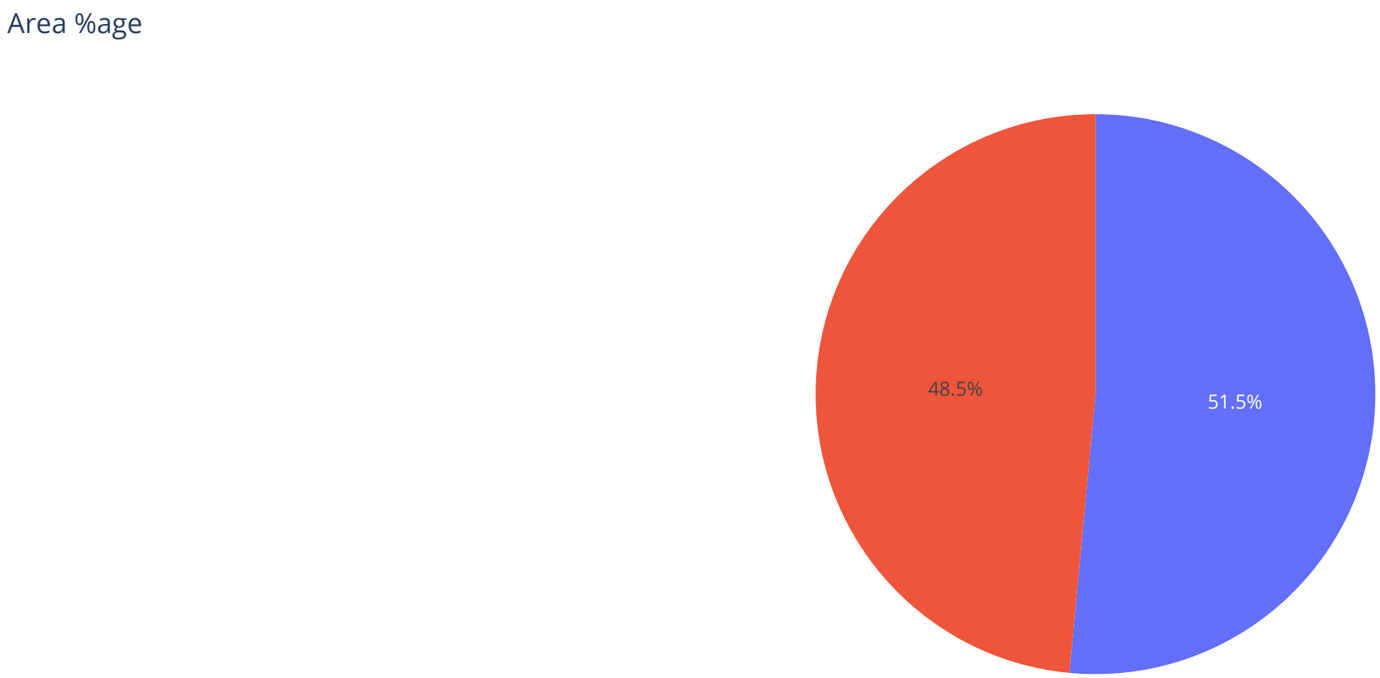
In [21]:
```
df['Area'].value_counts()
```

Out[21]:
```
Urban    381
Rural    359
Name: Area, dtype: int64
```

In [22]:
```
px.bar(df,x=df['Area'].value_counts().keys(),y=df['Area'].value_counts(),color=df['Area'].value_counts().keys(),title='Area Counts')
```



In [23]:
```
px.pie(df,names=df['Area'].value_counts().keys(),values=df['Area'].value_counts(),color=df['Area'].value_counts().keys(),title='Area %age')
```



In [24]:
```
import datetime as dt
```

In [25]:
```
df['Month']=df['Date'].dt.month_name()
```

In [26]:
```
df['Year']=df['Date'].dt.year
```

In [27]:
```
df.drop(['Frequency'],axis=1,inplace=True)
```

In [28]:
```
df.head()
```

Out[28]:

	Region	Date	Estimated_Unemployment_Rate_(%)	Estimated_Employed	Estimated_Labour_Participation_Rate_(%)	Area	Month	Year
0	Andhra Pradesh	2019-05-31	3.65	11999139.0	43.24	Rural	May	2019
1	Andhra Pradesh	2019-06-30	3.05	11755881.0	42.05	Rural	June	2019
2	Andhra Pradesh	2019-07-31	3.75	12086707.0	43.50	Rural	July	2019
3	Andhra Pradesh	2019-08-31	3.32	12285693.0	43.97	Rural	August	2019
4	Andhra Pradesh	2019-09-30	5.17	12256762.0	44.68	Rural	September	2019

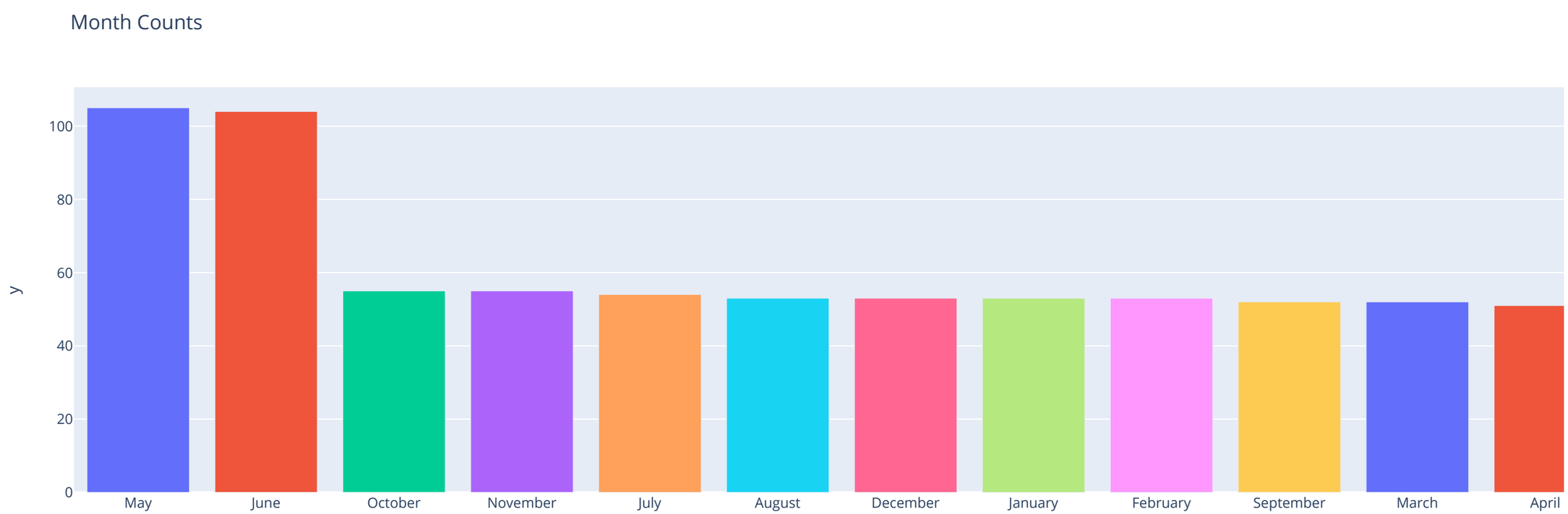
In [29]:
```
df['Year'].value_counts(normalize=True)*100
```

Out[29]:
```
2019    58.108108
2020    41.891892
Name: Year, dtype: float64
```

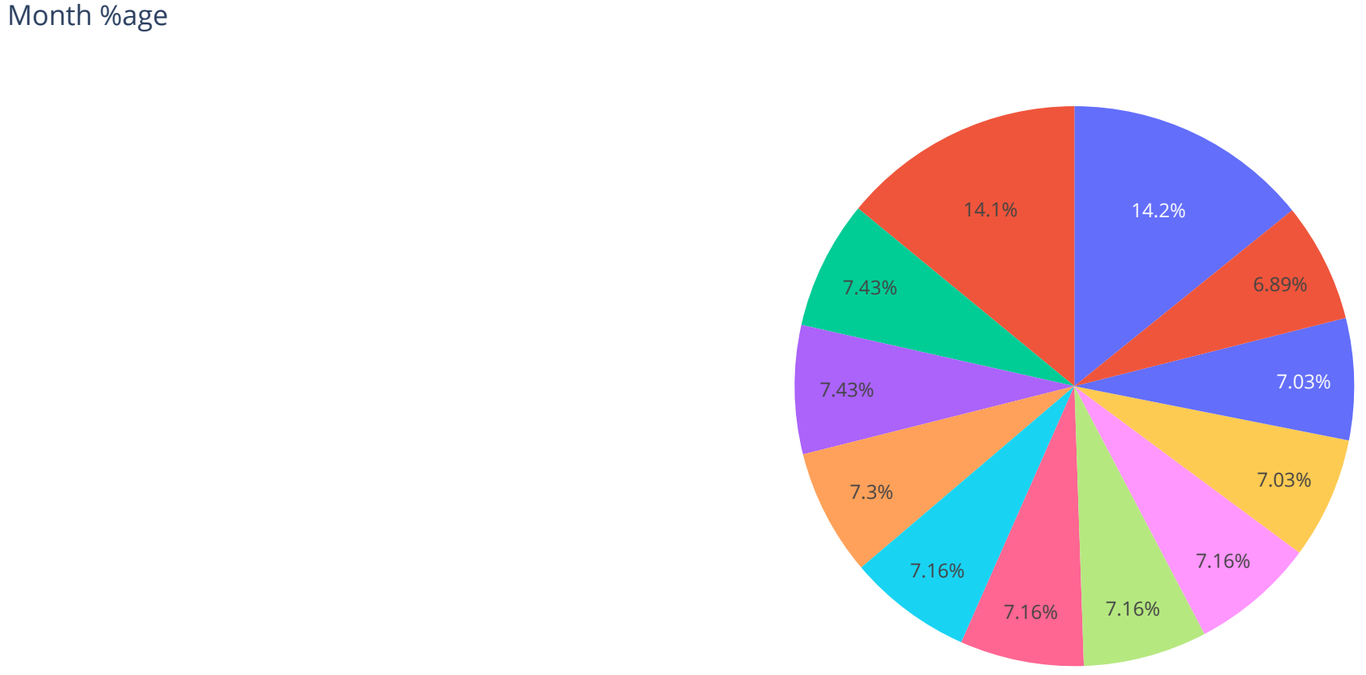
In [30]:
```
df['Month'].value_counts()
```

Out[30]:
```
May          105
June         104
October       55
November      55
July          54
August        53
December      53
January       53
February      53
September     52
March         52
April         51
Name: Month, dtype: int64
```

In [31]:
```
px.bar(df,x=df['Month'].value_counts().keys(),y=df['Month'].value_counts(),color=df['Month'].value_counts().keys(),title='Month Counts')
```



In [32]:
```
px.pie(df,names=df['Month'].value_counts().keys(),values=df['Month'].value_counts(),color=df['Month'].value_counts().keys(),title='Month %age')
```



In [33]:
```
df_region_rate=df.groupby(['Region','Year']).agg({"Estimated_Unemployment_Rate_(%)":'mean'}).reset_index()
df_region_rate.sort_values(by='Estimated_Unemployment_Rate_(%)',ascending=False)
```

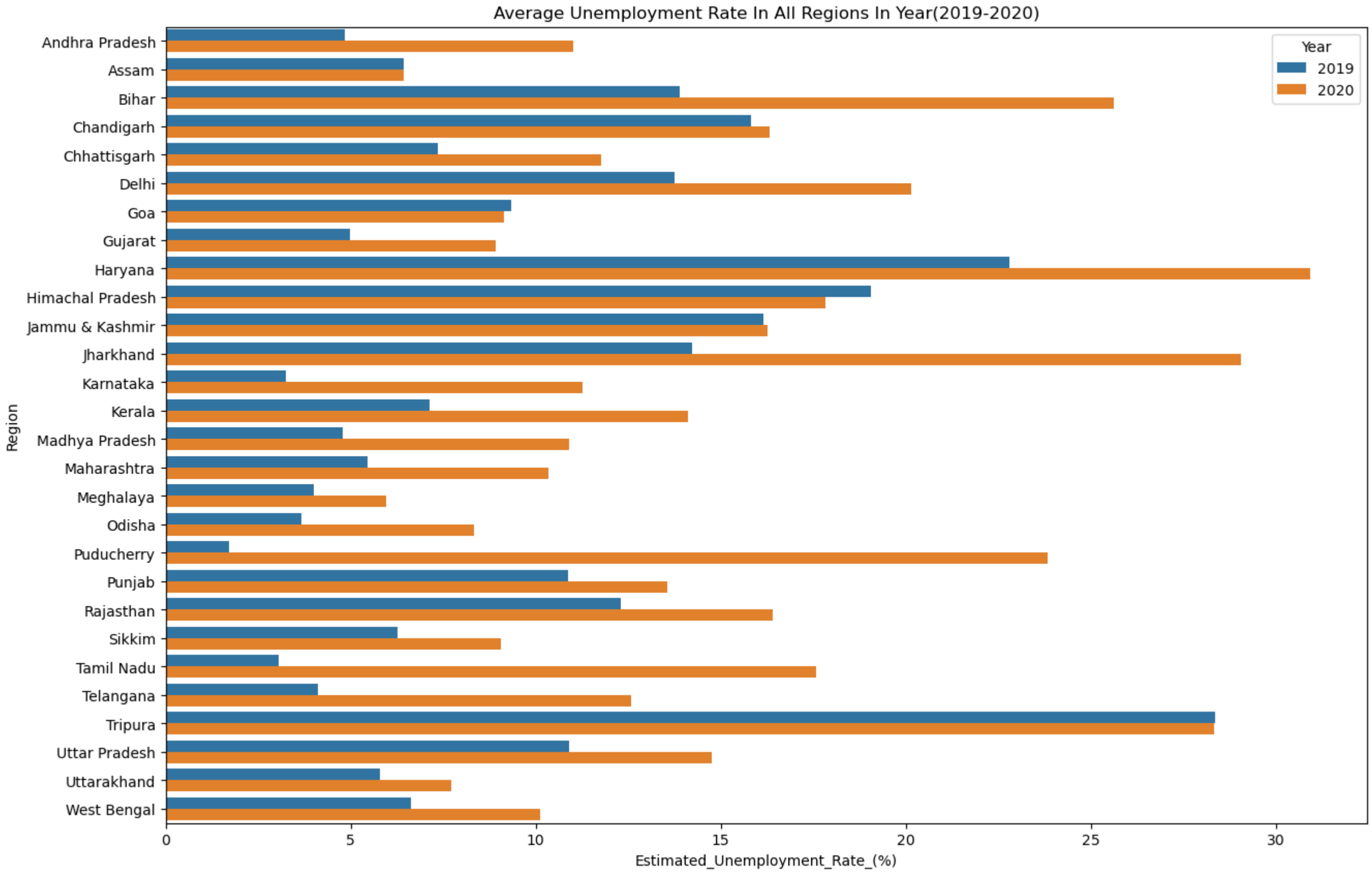
Out[33]:

	Region	Year	Estimated_Unemployment_Rate_(%)
17	Haryana	2020	30.929167
23	Jharkhand	2020	29.053333
48	Tripura	2019	28.363125
49	Tripura	2020	28.333333
5	Bihar	2020	25.632500
37	Puducherry	2020	23.840000
16	Haryana	2019	22.798750
11	Delhi	2020	20.155000
18	Himachal Pradesh	2019	19.064375
19	Himachal Pradesh	2020	17.841667
45	Tamil Nadu	2020	17.578333
41	Rajasthan	2020	16.400833
7	Chandigarh	2020	16.330000
21	Jammu & Kashmir	2020	16.251111
20	Jammu & Kashmir	2019	16.141667
6	Chandigarh	2019	15.822500
51	Uttar Pradesh	2020	14.769167
22	Jharkhand	2019	14.233750
27	Kerala	2020	14.114167
4	Bihar	2019	13.882500
10	Delhi	2019	13.750625
39	Punjab	2020	13.562500
47	Telangana	2020	12.567500
40	Rajasthan	2019	12.301250
9	Chhattisgarh	2020	11.765000
25	Karnataka	2020	11.259167
1	Andhra Pradesh	2020	11.010833
29	Madhya Pradesh	2020	10.897500
50	Uttar Pradesh	2019	10.888125
38	Punjab	2019	10.882500
31	Maharashtra	2020	10.355000
55	West Bengal	2020	10.123333
12	Goa	2019	9.346250
13	Goa	2020	9.130000
43	Sikkim	2020	9.068333

	Region	Year	Estimated_Unemployment_Rate_(%)
15	Gujarat	2020	8.910000
35	Odisha	2020	8.320000
53	Uttarakhand	2020	7.720909
8	Chhattisgarh	2019	7.346875
26	Kerala	2019	7.131250
54	West Bengal	2019	6.625625
3	Assam	2020	6.438182
2	Assam	2019	6.420667
42	Sikkim	2019	6.257273
33	Meghalaya	2020	5.942727
52	Uttarakhand	2019	5.800625
30	Maharashtra	2019	5.459375
14	Gujarat	2019	4.979375
0	Andhra Pradesh	2019	4.826875
28	Madhya Pradesh	2019	4.788125
46	Telangana	2019	4.115625
32	Meghalaya	2019	4.012500
34	Odisha	2019	3.661250
24	Karnataka	2019	3.238750
44	Tamil Nadu	2019	3.063750
36	Puducherry	2019	1.699375

In [34]:
```
plt.figure(figsize=(15,10))
plt.title("Average Unemployment Rate In All Regions In Year(2019-2020)")
sns.barplot(data=df_region_rate,y='Region',x='Estimated_Unemployment_Rate_(%)',hue='Year')
```

Out[34]:
```
<AxesSubplot:title={'center':'Average Unemployment Rate In All Regions In Year(2019-2020)'}, xlabel='Estimated_Unemployment_Rate_(%)', ylabel='Region'>
```


Average Unemployment Rate In All Regions In Year(2019-2020)
Year
2019
2020
Region
Andhra Pradesh
Assam
Bihar
Chandigarh
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
0
5
10
15
20
25
30
Estimated_Unemployment_Rate_(%)

In [35]:
```
df.head()
```

Out[35]:

	Region	Date	Estimated_Unemployment_Rate_(%)	Estimated_Employed	Estimated_Labour_Participation_Rate_(%)	Area	Month	Year
0	Andhra Pradesh	2019-05-31	3.65	11999139.0	43.24	Rural	May	2019
1	Andhra Pradesh	2019-06-30	3.05	11755881.0	42.05	Rural	June	2019
2	Andhra Pradesh	2019-07-31	3.75	12086707.0	43.50	Rural	July	2019
3	Andhra Pradesh	2019-08-31	3.32	12285693.0	43.97	Rural	August	2019
4	Andhra Pradesh	2019-09-30	5.17	12256762.0	44.68	Rural	September	2019

In [36]:
```
df_region_emp=df.groupby(['Region','Year']).agg({"Estimated_Employed":'mean'}).reset_index()
df_region_emp.sort_values(by='Estimated_Employed',ascending=False)
```

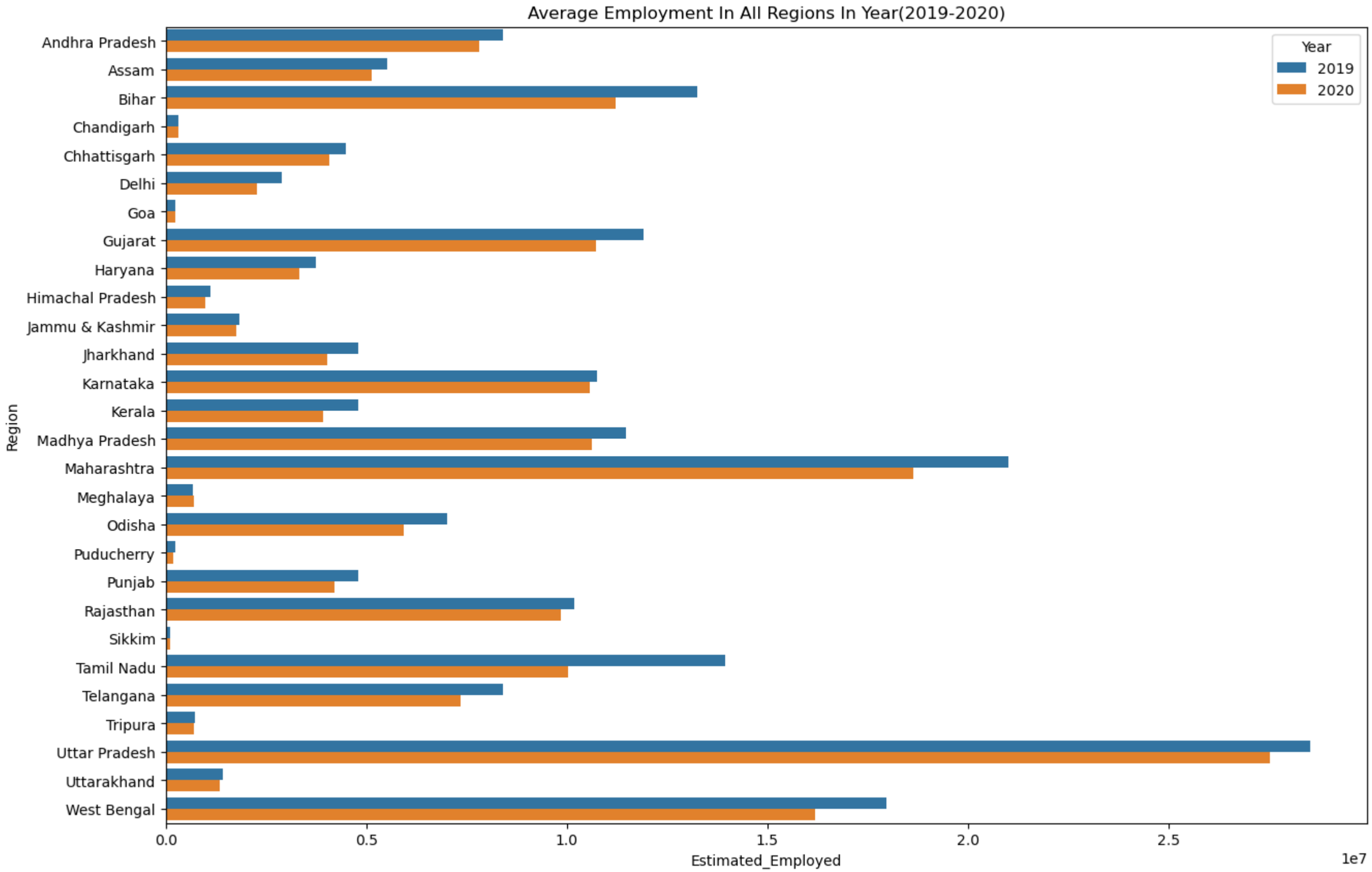
Out[36]:

	Region	Year	Estimated_Employed
50	Uttar Pradesh	2019	2.852497e+07
51	Uttar Pradesh	2020	2.752132e+07
30	Maharashtra	2019	2.101646e+07
31	Maharashtra	2020	1.862185e+07
54	West Bengal	2019	1.795374e+07
55	West Bengal	2020	1.619160e+07
44	Tamil Nadu	2019	1.394580e+07
4	Bihar	2019	1.323785e+07
14	Gujarat	2019	1.190651e+07
28	Madhya Pradesh	2019	1.147777e+07
5	Bihar	2020	1.120397e+07
24	Karnataka	2019	1.075008e+07
15	Gujarat	2020	1.072935e+07
29	Madhya Pradesh	2020	1.063243e+07
25	Karnataka	2020	1.055650e+07
40	Rajasthan	2019	1.018783e+07
45	Tamil Nadu	2020	1.003455e+07
41	Rajasthan	2020	9.845378e+06
0	Andhra Pradesh	2019	8.402043e+06
46	Telangana	2019	8.392168e+06
1	Andhra Pradesh	2020	7.823493e+06
47	Telangana	2020	7.336322e+06
34	Odisha	2019	7.003220e+06
35	Odisha	2020	5.935783e+06
2	Assam	2019	5.508148e+06
3	Assam	2020	5.145623e+06
26	Kerala	2019	4.803638e+06
22	Jharkhand	2019	4.798559e+06
38	Punjab	2019	4.788320e+06
8	Chhattisgarh	2019	4.483388e+06
39	Punjab	2020	4.207418e+06
9	Chhattisgarh	2020	4.063647e+06
23	Jharkhand	2020	4.030149e+06
27	Kerala	2020	3.922248e+06
16	Haryana	2019	3.731464e+06

	Region	Year	Estimated_Employed
17	Haryana	2020	3.324551e+06
10	Delhi	2019	2.896461e+06
11	Delhi	2020	2.268915e+06
20	Jammu & Kashmir	2019	1.838242e+06
21	Jammu & Kashmir	2020	1.748851e+06
52	Uttarakhand	2019	1.427203e+06
53	Uttarakhand	2020	1.336447e+06
18	Himachal Pradesh	2019	1.108268e+06
19	Himachal Pradesh	2020	9.952317e+05
48	Tripura	2019	7.316091e+05
49	Tripura	2020	6.975273e+05
33	Meghalaya	2020	6.955085e+05
32	Meghalaya	2019	6.857688e+05
6	Chandigarh	2019	3.208352e+05
7	Chandigarh	2020	3.088232e+05
36	Puducherry	2019	2.330804e+05
12	Goa	2019	2.268961e+05
13	Goa	2020	2.251329e+05
37	Puducherry	2020	1.789944e+05
42	Sikkim	2019	1.111313e+05
43	Sikkim	2020	9.908800e+04

In [37]:
```
plt.figure(figsize=(15,10))
plt.title("Average Employment In All Regions In Year(2019-2020)")
sns.barplot(data=df_region_emp,y='Region',x='Estimated_Employed',hue='Year')
```

Out[37]:
```
<AxesSubplot:title={'center':'Average Employment In All Regions In Year(2019-2020)'}, xlabel='Estimated_Employed', ylabel='Region'>
```



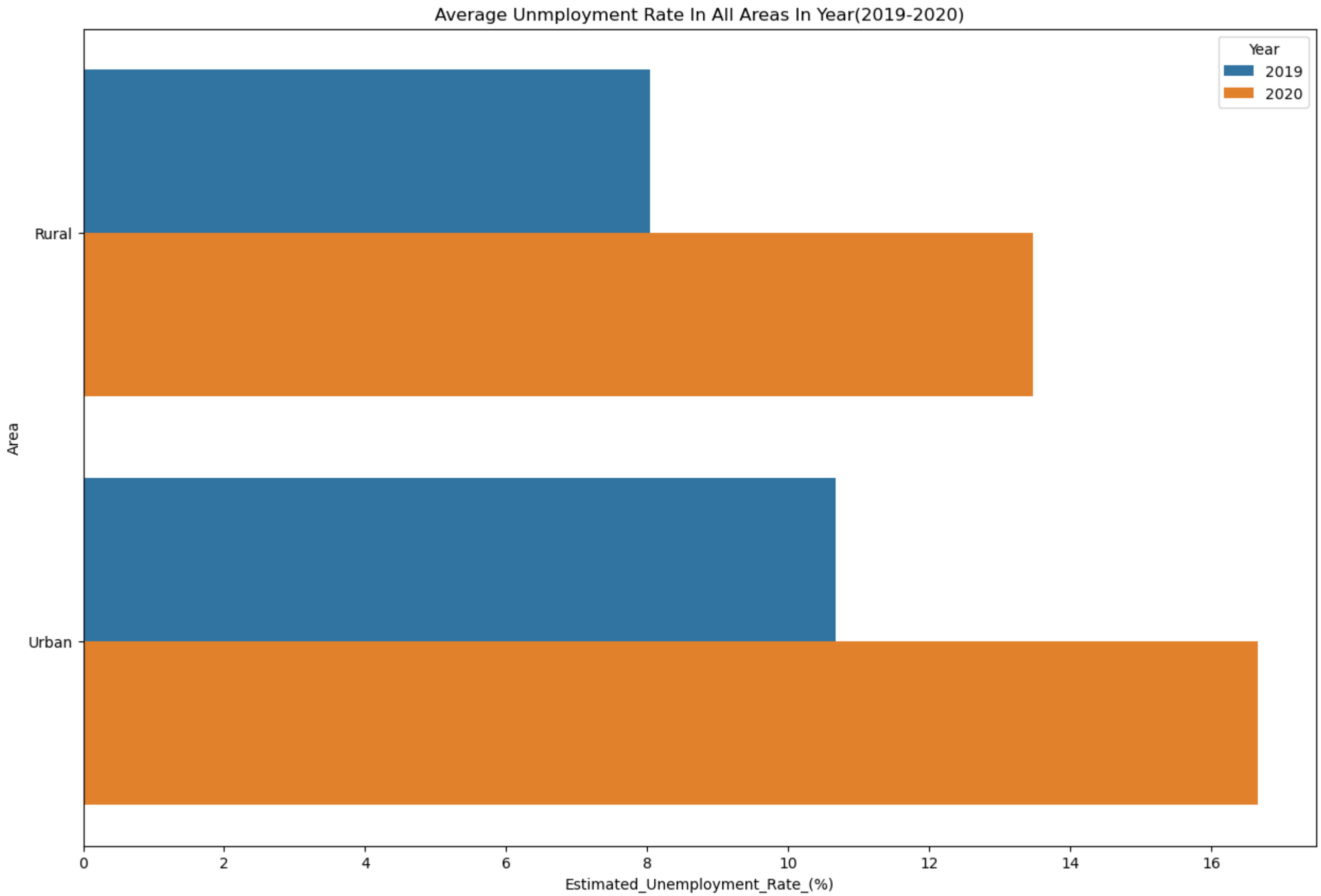
In [38]:
```
df_region_area=df.groupby(['Area','Year']).agg({"Estimated_Unemployment_Rate_(%)":'mean'}).reset_index()
df_region_area.sort_values(by='Estimated_Unemployment_Rate_(%)',ascending=False)
```

Out[38]:

	Area	Year	Estimated_Unemployment_Rate_(%)
3	Urban	2020	16.654528
1	Rural	2020	13.466358
2	Urban	2019	10.668514
0	Rural	2019	8.044135

In [39]:
```
plt.figure(figsize=(15,10))
plt.title("Average Unmployment Rate In All Areas In Year(2019-2020)")
sns.barplot(data=df_region_area,y='Area',x='Estimated_Unemployment_Rate_(%)',hue='Year')
```

Out[39]:
```
<AxesSubplot:title={'center':'Average Unmployment Rate In All Areas In Year(2019-2020)'}, xlabel='Estimated_Unemployment_Rate_(%)', ylabel='Area'>
```



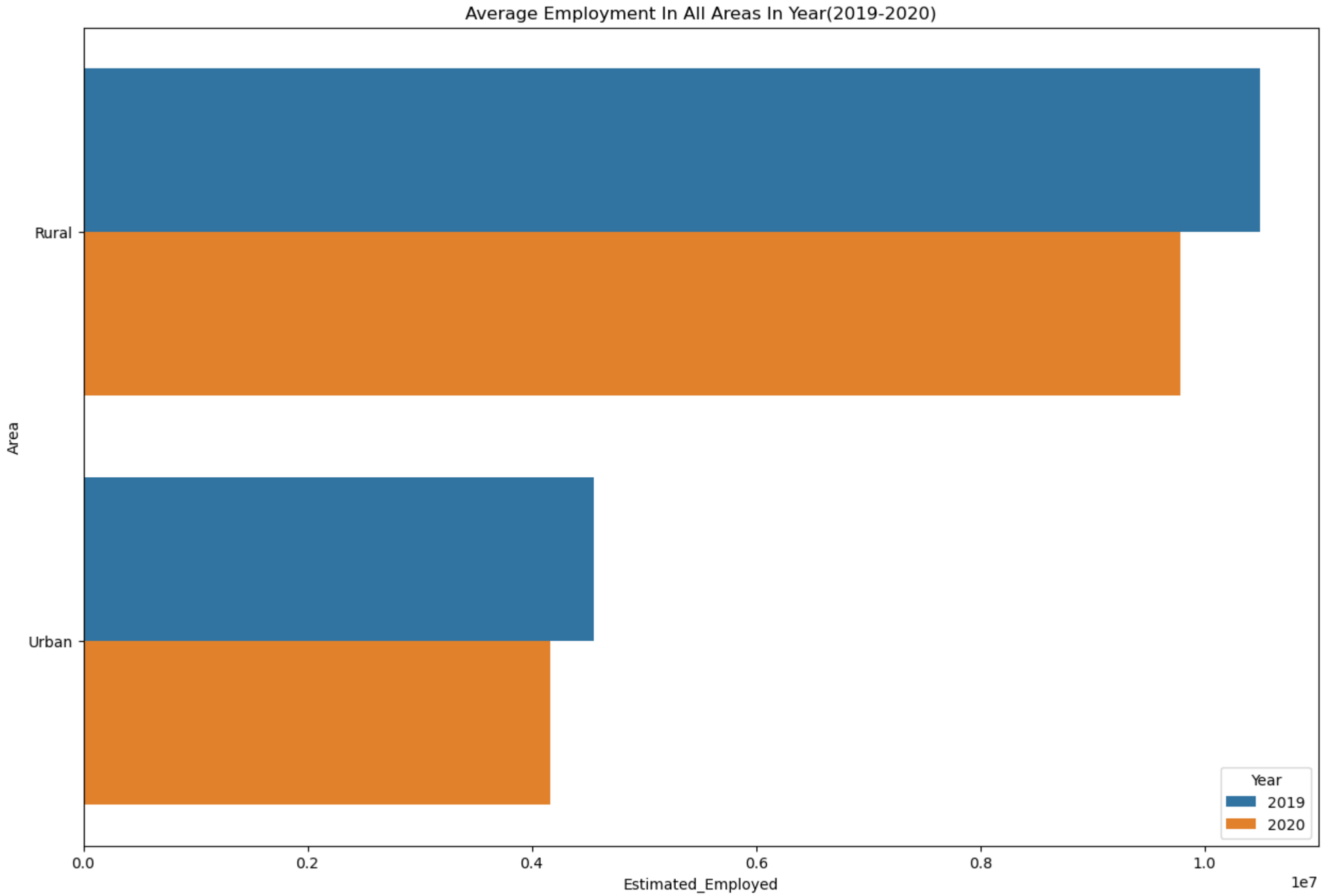
In [40]:
```
df_region_ar_em=df.groupby(['Area','Year']).agg({"Estimated_Employed":'mean'}).reset_index()
df_region_ar_em.sort_values(by='Estimated_Employed',ascending=False)
```

Out[40]:

	Area	Year	Estimated_Employed
0	Rural	2019	1.049050e+07
1	Rural	2020	9.782849e+06
2	Urban	2019	4.548901e+06
3	Urban	2020	4.164845e+06

In [41]:
```
plt.figure(figsize=(15,10))
plt.title("Average Employment In All Areas In Year(2019-2020)")
sns.barplot(data=df_region_ar_em,y='Area',x='Estimated_Employed',hue='Year')
```

Out[41]:
```
<AxesSubplot:title={'center':'Average Employment In All Areas In Year(2019-2020)'}, xlabel='Estimated_Employed', ylabel='Area'>
```


Average Employment In All Areas In Year(2019-2020)
Area
Rural
Urban
0.0
0.2
0.4
0.6
0.8
1.0
1e7
Estimated_Employed
Year
2019
2020