ANALYSIS OF UNEMPLOYMENT IN INDIA DURING AND BEFORE PANDEMIC YEAR 2020

In [75]:
```
import pandas as pd
```

In [76]:
```
data = pd.read_csv("D:\\Data Science Project\\Umemployment in India//Unemployment_in_India.csv")
```

In [77]:
```
data
```

Out[77]:

	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
...	...	...	...	...	...	...	...
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

768 rows × 7 columns

In [78]:
```
data1 = pd.read_csv('D:\\Data Science Project\\Umemployment in India//Unemployment_Rate_upto_11_2020.csv'
```

In [79]:
```
data1
```

Out[79]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.740
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.740
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.740
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.740
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.740
...	...	...	...	...	...	...	...	...	...
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

267 rows × 9 columns

In [80]:
```
print('Unemployment in India during 2020')
data.info()
print('\nUnemployment in India before 2020')
data1.info()
```

```
Unemployment in India during 2020
<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

Unemployment in India before 2020
<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 [81]:
```
data.columns
```

Out[81]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Area'],
      dtype='object')
```

In [82]:
```
data1.columns
```

Out[82]:
```
Index(['Region', ' Date', ' Frequency', ' Estimated Unemployment Rate (%)',
       ' Estimated Employed', ' Estimated Labour Participation Rate (%)',
       'Region.1', 'longitude', 'latitude'],
      dtype='object')
```

DATA PREPROCESSING

In [83]:
```
data.head(20)
```

Out[83]:

	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
5	Andhra Pradesh	31-10-2019	Monthly	3.52	12017412.0	43.01	Rural
6	Andhra Pradesh	30-11-2019	Monthly	4.12	11397681.0	41.00	Rural
7	Andhra Pradesh	31-12-2019	Monthly	4.38	12528395.0	45.14	Rural
8	Andhra Pradesh	31-01-2020	Monthly	4.84	12016676.0	43.46	Rural
9	Andhra Pradesh	29-02-2020	Monthly	5.91	11723617.0	42.83	Rural
10	Andhra Pradesh	31-03-2020	Monthly	4.06	11359660.0	40.66	Rural
11	Andhra Pradesh	30-04-2020	Monthly	16.29	8792827.0	36.03	Rural
12	Andhra Pradesh	31-05-2020	Monthly	14.46	9526902.0	38.16	Rural
13	Andhra Pradesh	30-06-2020	Monthly	0.85	15572975.0	53.76	Rural
14	Assam	31-05-2019	Monthly	4.29	11749334.0	57.39	Rural
15	Assam	30-06-2019	Monthly	5.08	8923222.0	43.87	Rural
16	Assam	31-07-2019	Monthly	4.26	9911534.0	48.21	Rural
17	Assam	31-08-2019	Monthly	5.79	9292039.0	45.83	Rural
18	Assam	30-09-2019	Monthly	4.46	11468349.0	55.67	Rural
19	Assam	31-10-2019	Monthly	4.65	8395906.0	40.76	Rural

In [84]:
```
data1.head(20)
```

Out[84]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
0	Andhra Pradesh	31-01-2020	M	5.48	16635535	41.02	South	15.9129	79.7400
1	Andhra Pradesh	29-02-2020	M	5.83	16545652	40.90	South	15.9129	79.7400
2	Andhra Pradesh	31-03-2020	M	5.79	15881197	39.18	South	15.9129	79.7400
3	Andhra Pradesh	30-04-2020	M	20.51	11336911	33.10	South	15.9129	79.7400
4	Andhra Pradesh	31-05-2020	M	17.43	12988845	36.46	South	15.9129	79.7400
5	Andhra Pradesh	30-06-2020	M	3.31	19805400	47.41	South	15.9129	79.7400
6	Andhra Pradesh	31-07-2020	M	8.34	15431615	38.91	South	15.9129	79.7400
7	Andhra Pradesh	31-08-2020	M	6.96	15251776	37.83	South	15.9129	79.7400
8	Andhra Pradesh	30-09-2020	M	6.40	15220312	37.47	South	15.9129	79.7400
9	Andhra Pradesh	31-10-2020	M	6.59	15157557	37.34	South	15.9129	79.7400
10	Assam	31-01-2020	M	4.66	13051904	52.98	Northeast	26.2006	92.9376
11	Assam	29-02-2020	M	4.41	10088268	40.77	Northeast	26.2006	92.9376
12	Assam	31-03-2020	M	4.77	11542888	46.73	Northeast	26.2006	92.9376
13	Assam	30-04-2020	M	11.06	6830817	29.55	Northeast	26.2006	92.9376
14	Assam	31-05-2020	M	9.55	11367897	48.26	Northeast	26.2006	92.9376
15	Assam	30-06-2020	M	0.60	9095944	35.07	Northeast	26.2006	92.9376
16	Assam	31-07-2020	M	3.77	10286757	40.88	Northeast	26.2006	92.9376
17	Assam	31-08-2020	M	5.53	9781310	39.52	Northeast	26.2006	92.9376
18	Assam	30-09-2020	M	1.19	14107641	54.38	Northeast	26.2006	92.9376
19	Assam	31-10-2020	M	3.02	11949329	46.84	Northeast	26.2006	92.9376

In [85]:
```
data.tail(10)
```

Out[85]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
758	NaN	NaN	NaN	NaN	NaN	NaN	NaN
759	NaN	NaN	NaN	NaN	NaN	NaN	NaN
760	NaN	NaN	NaN	NaN	NaN	NaN	NaN
761	NaN	NaN	NaN	NaN	NaN	NaN	NaN
762	NaN	NaN	NaN	NaN	NaN	NaN	NaN
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 [86]:
```
data1.tail(10)
```

Out[86]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1	longitude	latitude
257	West Bengal	31-01-2020	M	6.94	35820789	47.35	East	22.9868	87.855
258	West Bengal	29-02-2020	M	4.92	36964178	47.74	East	22.9868	87.855
259	West Bengal	31-03-2020	M	6.92	35903917	47.27	East	22.9868	87.855
260	West Bengal	30-04-2020	M	17.41	26938836	39.90	East	22.9868	87.855
261	West Bengal	31-05-2020	M	17.41	28356675	41.92	East	22.9868	87.855
262	West Bengal	30-06-2020	M	7.29	30726310	40.39	East	22.9868	87.855
263	West Bengal	31-07-2020	M	6.83	35372506	46.17	East	22.9868	87.855
264	West Bengal	31-08-2020	M	14.87	33298644	47.48	East	22.9868	87.855
265	West Bengal	30-09-2020	M	9.35	35707239	47.73	East	22.9868	87.855
266	West Bengal	31-10-2020	M	9.98	33962549	45.63	East	22.9868	87.855

In [87]:
```
data.isnull().sum()
```

Out[87]:
```
Region                                   28
 Date                                    28
 Frequency                               28
 Estimated Unemployment Rate (%)         28
 Estimated Employed                      28
 Estimated Labour Participation Rate (%) 28
Area                                     28
dtype: int64
```

In [88]:
```
data1.isnull().sum()
```

Out[88]:
```
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 [89]:
```
data.dropna(inplace=True)
```

In [90]:
```
data.isnull().sum()
```

Out[90]:
```
Region                                   0
 Date                                    0
 Frequency                               0
 Estimated Unemployment Rate (%)         0
 Estimated Employed                      0
 Estimated Labour Participation Rate (%) 0
Area                                     0
dtype: int64
```

In [91]:
```
data1.dropna(inplace=True)
```

In [92]:
```
data1.isnull().sum()
```

Out[92]:
```
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 [93]:
```
print('data')
print('Number of rows:',data.shape[0])
print('Number of columns:',data.shape[1])
```

```
data
Number of rows: 740
Number of columns: 7
```

In [94]:
```
print('data1')
print('Number of rows:',data1.shape[0])
print('Number of columns:',data1.shape[1])
```

```
data1
Number of rows: 267
Number of columns: 9
```

In [95]:
```
data1=data1.drop(columns=['longitude','latitude'])
```

In [96]:
```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 7 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 
dtypes: float64(2), int64(1), object(4)
memory usage: 14.7+ KB
```

In [97]:
```
data[' Date'] = pd.to_datetime(data[' Date'])
data.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    datetime64[ns]
 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: datetime64[ns](1), float64(3), object(3)
memory usage: 46.2+ KB
```

In [98]:
```
data1[' Date'] = pd.to_datetime(data1[' Date'])
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 267 entries, 0 to 266
Data columns (total 7 columns):
 #   Column                                   Non-Null Count  Dtype
---  ------                                   --------------  -----
 0   Region                                   267 non-null    object
 1    Date                                    267 non-null    datetime64[ns]
 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
dtypes: datetime64[ns](1), float64(2), int64(1), object(3)
memory usage: 14.7+ KB
```

In [104]:
```
data = data[data[' Date'].dt.year != 2020]
```

In [105]:
```
data.head(20)
```

Out[105]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	2019-05-31	Monthly	3.65	11999139.0	43.24	Rural
1	Andhra Pradesh	2019-06-30	Monthly	3.05	11755881.0	42.05	Rural
2	Andhra Pradesh	2019-07-31	Monthly	3.75	12086707.0	43.50	Rural
3	Andhra Pradesh	2019-08-31	Monthly	3.32	12285693.0	43.97	Rural
4	Andhra Pradesh	2019-09-30	Monthly	5.17	12256762.0	44.68	Rural
5	Andhra Pradesh	2019-10-31	Monthly	3.52	12017412.0	43.01	Rural
6	Andhra Pradesh	2019-11-30	Monthly	4.12	11397681.0	41.00	Rural
7	Andhra Pradesh	2019-12-31	Monthly	4.38	12528395.0	45.14	Rural
14	Assam	2019-05-31	Monthly	4.29	11749334.0	57.39	Rural
15	Assam	2019-06-30	Monthly	5.08	8923222.0	43.87	Rural
16	Assam	2019-07-31	Monthly	4.26	9911534.0	48.21	Rural
17	Assam	2019-08-31	Monthly	5.79	9292039.0	45.83	Rural
18	Assam	2019-09-30	Monthly	4.46	11468349.0	55.67	Rural
19	Assam	2019-10-31	Monthly	4.65	8395906.0	40.76	Rural
20	Assam	2019-11-30	Monthly	4.66	9625362.0	46.64	Rural
26	Bihar	2019-05-31	Monthly	9.27	24322330.0	39.75	Rural
27	Bihar	2019-06-30	Monthly	10.20	24097712.0	39.71	Rural
28	Bihar	2019-07-31	Monthly	13.44	23248875.0	39.66	Rural
29	Bihar	2019-08-31	Monthly	11.00	22260203.0	36.85	Rural
30	Bihar	2019-09-30	Monthly	8.87	23905700.0	38.57	Rural

In [109]:
```
data1.head(20)
```

Out[109]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1
0	Andhra Pradesh	2020-01-31	M	5.48	16635535	41.02	South
1	Andhra Pradesh	2020-02-29	M	5.83	16545652	40.90	South
2	Andhra Pradesh	2020-03-31	M	5.79	15881197	39.18	South
3	Andhra Pradesh	2020-04-30	M	20.51	11336911	33.10	South
4	Andhra Pradesh	2020-05-31	M	17.43	12988845	36.46	South
5	Andhra Pradesh	2020-06-30	M	3.31	19805400	47.41	South
6	Andhra Pradesh	2020-07-31	M	8.34	15431615	38.91	South
7	Andhra Pradesh	2020-08-31	M	6.96	15251776	37.83	South
8	Andhra Pradesh	2020-09-30	M	6.40	15220312	37.47	South
9	Andhra Pradesh	2020-10-31	M	6.59	15157557	37.34	South
10	Assam	2020-01-31	M	4.66	13051904	52.98	Northeast
11	Assam	2020-02-29	M	4.41	10088268	40.77	Northeast
12	Assam	2020-03-31	M	4.77	11542888	46.73	Northeast
13	Assam	2020-04-30	M	11.06	6830817	29.55	Northeast
14	Assam	2020-05-31	M	9.55	11367897	48.26	Northeast
15	Assam	2020-06-30	M	0.60	9095944	35.07	Northeast
16	Assam	2020-07-31	M	3.77	10286757	40.88	Northeast
17	Assam	2020-08-31	M	5.53	9781310	39.52	Northeast
18	Assam	2020-09-30	M	1.19	14107641	54.38	Northeast
19	Assam	2020-10-31	M	3.02	11949329	46.84	Northeast

In [117]:
```
data1[' Frequency'] = 'Monthly'
```

In [118]:
```
data1
```

Out[118]:

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

267 rows × 7 columns

In [125]:
```
data1
```

Out[125]:

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

267 rows × 7 columns

In [129]:
```
data_2020 = data1
```

In [130]:
```
data_2019 = data
```

In [131]:
```
data_2020.head()
```

Out[131]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Region.1
0	Andhra Pradesh	2020-01-31	Monthly	5.48	16635535	41.02	South
1	Andhra Pradesh	2020-02-29	Monthly	5.83	16545652	40.90	South
2	Andhra Pradesh	2020-03-31	Monthly	5.79	15881197	39.18	South
3	Andhra Pradesh	2020-04-30	Monthly	20.51	11336911	33.10	South
4	Andhra Pradesh	2020-05-31	Monthly	17.43	12988845	36.46	South

In [132]:
```
data_2019.head()
```

Out[132]:

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

In [134]:
```
data_merged = pd.concat([data_2019,data_2020],axis=0)
```

In [135]:
```
data_merged.shape
```

Out[135]: (697, 8)

In [150]:
```
data_merged.head()
```

Out[150]:

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

In [137]:
```
data_merged.tail(5)
```

Out[137]:

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Region.1
262	West Bengal	2020-06-30	Monthly	7.29	30726310.0	40.39	NaN	East
263	West Bengal	2020-07-31	Monthly	6.83	35372506.0	46.17	NaN	East
264	West Bengal	2020-08-31	Monthly	14.87	33298644.0	47.48	NaN	East
265	West Bengal	2020-09-30	Monthly	9.35	35707239.0	47.73	NaN	East
266	West Bengal	2020-10-31	Monthly	9.98	33962549.0	45.63	NaN	East

In [142]:
```
data_merged = data_merged.drop(columns=[' Frequency','Area','Region.1'])
```

In [145]:
```
data_merged.describe()
```

Out[145]:

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
count	697.000000	6.970000e+02	697.000000
mean	10.486155	9.927923e+06	43.041463
std	8.867251	1.097239e+07	7.625807
min	0.000000	8.797400e+04	16.770000
25%	4.460000	1.946957e+06	38.600000
50%	7.600000	6.437868e+06	41.440000
75%	14.230000	1.410764e+07	45.610000
max	75.850000	5.943376e+07	72.570000

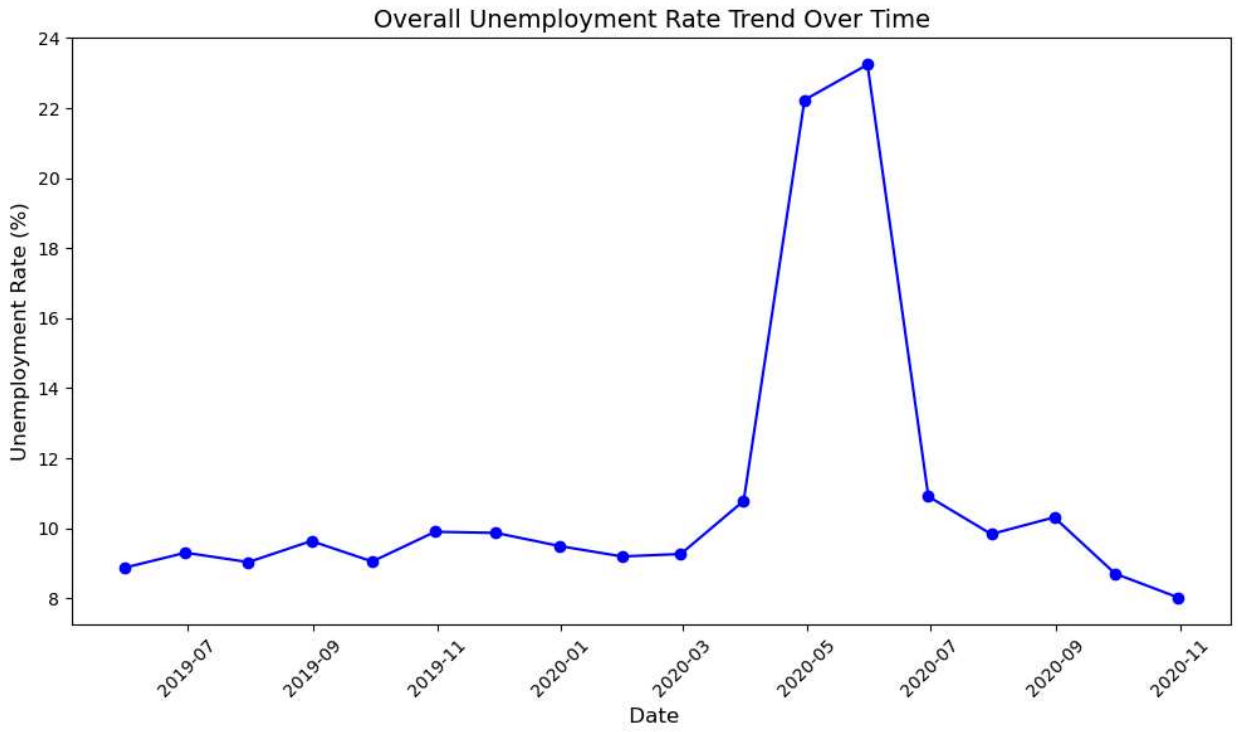
In [146]:
```
data_merged.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 697 entries, 0 to 266
Data columns (total 5 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   Region                                    697 non-null    object
 1    Date                                     697 non-null    datetime64[ns]
 2    Estimated Unemployment Rate (%)          697 non-null    float64
 3    Estimated Employed                       697 non-null    float64
 4    Estimated Labour Participation Rate (%)  697 non-null    float64
dtypes: datetime64[ns](1), float64(3), object(1)
memory usage: 32.7+ KB
```

EXPLORATORY DATA ANALYSIS

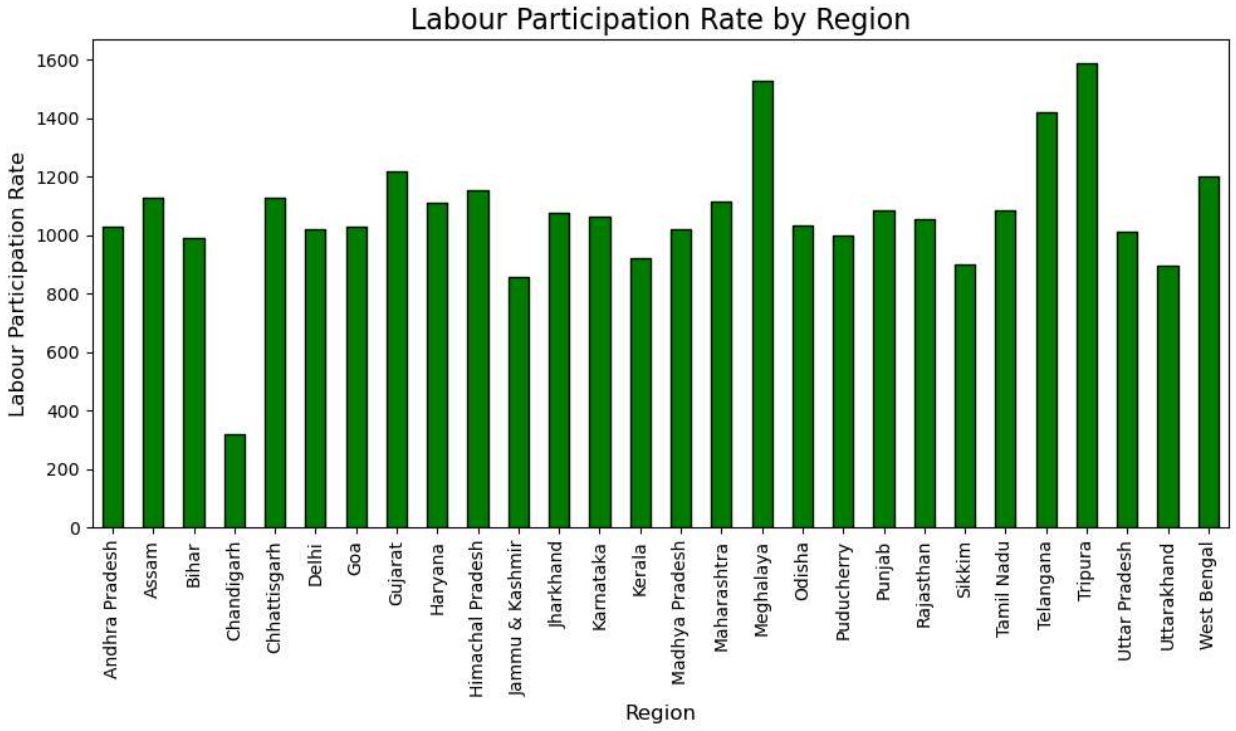
In [152]:
```
# Unemployment rate trend over time
```

In [163]:
```
unemployment_trend = data_merged.groupby(' Date')[' Estimated Unemployment Rate (%)'].mean()

import matplotlib.pyplot as plt
plt.figure(figsize=(10,6))
plt.plot(unemployment_trend,marker='o',color='blue')
plt.title('Overall Unemployment Rate Trend Over Time',fontsize=14)
plt.xlabel(' Date',fontsize=12)
plt.ylabel('Unemployment Rate (%)',fontsize=12)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



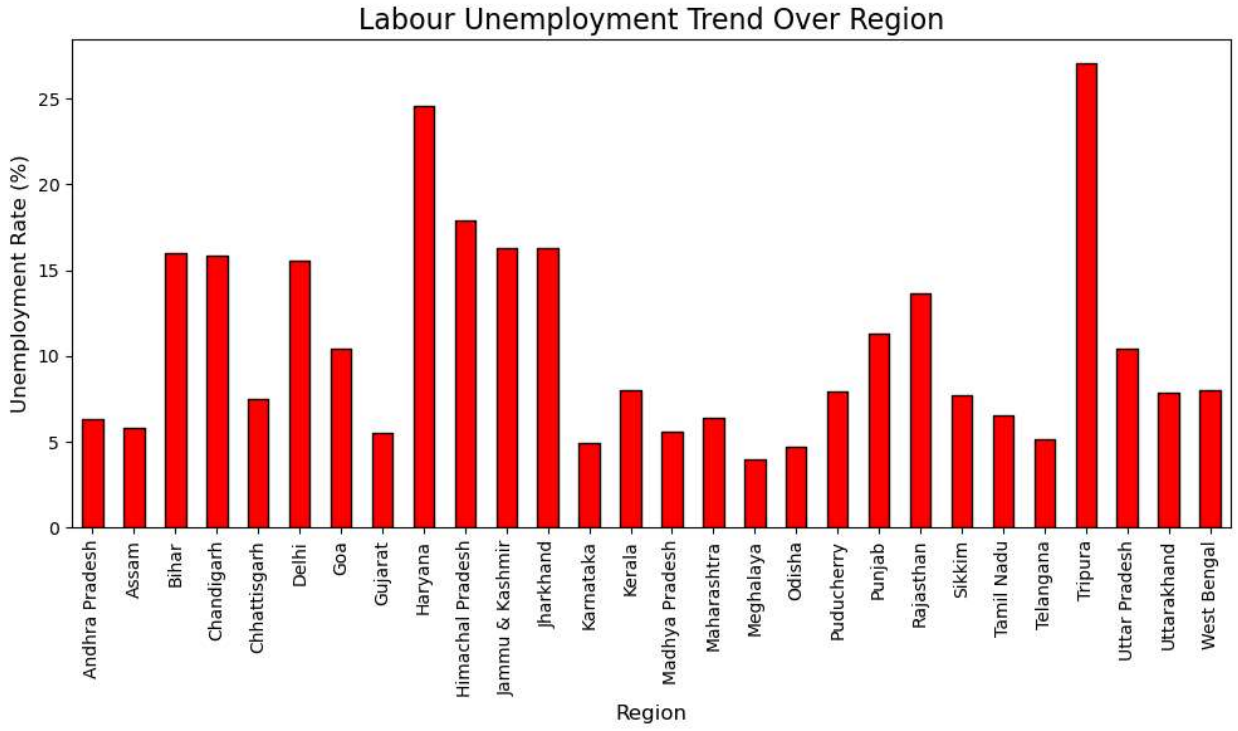
In [161]:
```
# Labour Participation Rate by Region
```

In [189]:
```
labour_participation_by_region = data_merged.groupby('Region')[' Estimated Labour Participation Rate (%)'

plt.figure(figsize=(10,6))
labour_participation_by_region.plot.bar(color='green',edgecolor='black')
plt.title("Labour Participation Rate by Region",fontsize=16)
plt.xlabel("Region",fontsize=12)
plt.ylabel("Labour Participation Rate",fontsize=12)
plt.tight_layout()
plt.show()
```



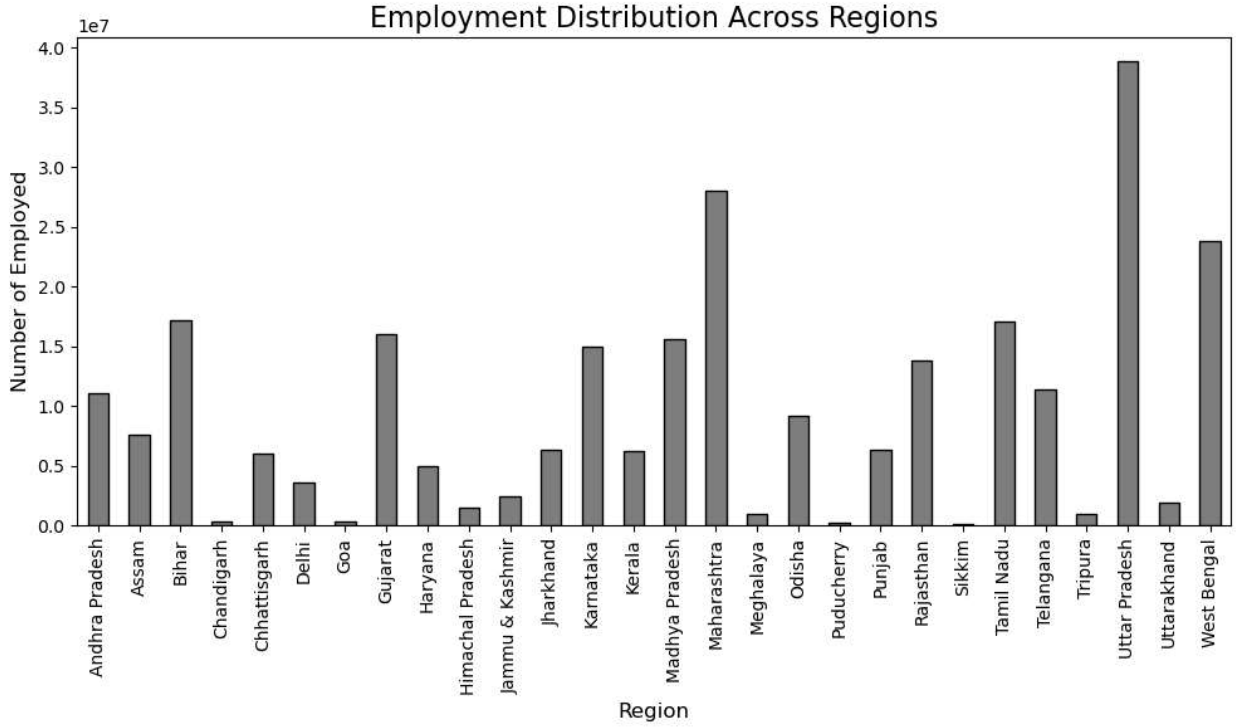
In [166]:
```
# Labor Unemployment Trend Over Region
```

In [188]:
```
unemployment_over_region = data_merged.groupby('Region')[' Estimated Unemployment Rate (%)'].mean()

plt.figure(figsize=(10,6))
unemployment_over_region.plot.bar(color='red',edgecolor='black')
plt.title("Labour Unemployment Trend Over Region",fontsize=16)
plt.xlabel("Region",fontsize=12)
plt.ylabel("Unemployment Rate (%)",fontsize=12)
plt.tight_layout()
plt.show()
```



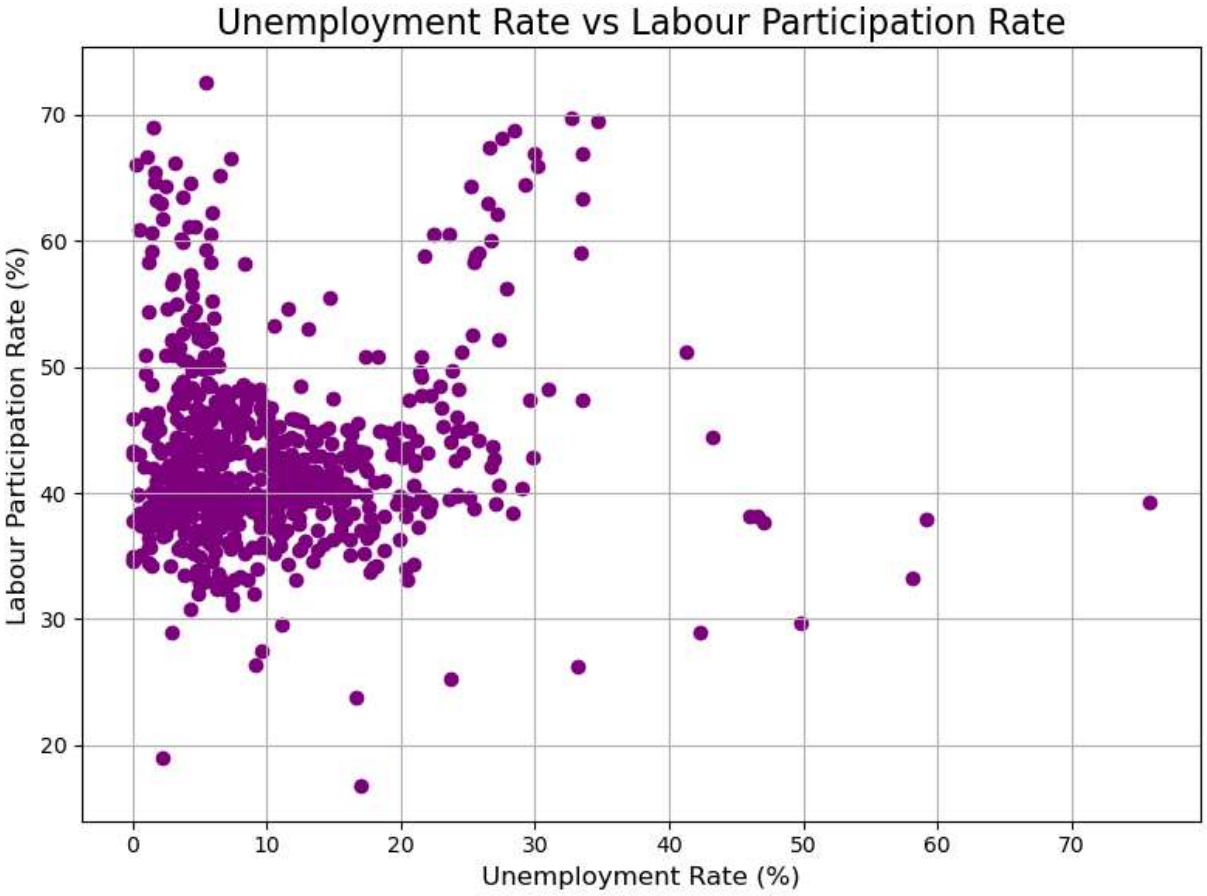
In [176]:
```
# Employment Distribution Across Regions
```

In [190]:
```
employment_by_region = data_merged.groupby('Region')[' Estimated Employed'].mean()

plt.figure(figsize=(10,6))
employment_by_region.plot.bar(color='grey',edgecolor='black')
plt.title('Employment Distribution Across Regions',fontsize=16)
plt.xlabel("Region",fontsize=12)
plt.ylabel("Number of Employed",fontsize=12)
plt.tight_layout()
plt.show()
```



In [191]:
```
# Relationship Between Unemployment Rate and Labour Participation Rate
```

In [193]:
```
plt.figure(figsize=(8,6))
plt.scatter(data_merged[' Estimated Unemployment Rate (%)'],
            data_merged[' Estimated Labour Participation Rate (%)'],color='purple')
plt.title("Unemployment Rate vs Labour Participation Rate",fontsize=16)
plt.xlabel('Unemployment Rate (%)',fontsize=12)
plt.ylabel("Labour Participation Rate (%)",fontsize=12)
plt.grid(True)
plt.tight_layout()
plt.show()
```

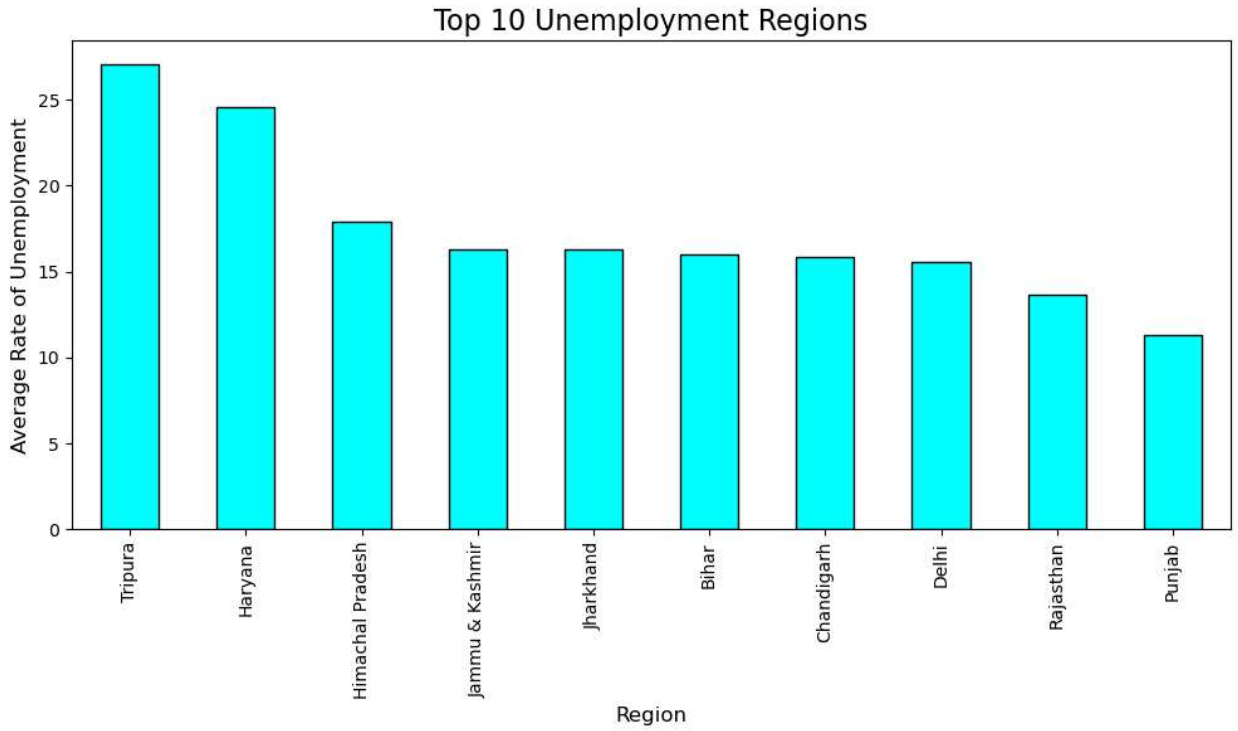


In [194]:
```
# Top 10 Unemployment Regions
```

In [206]:
```
top_unemployment = data_merged.groupby('Region')[' Estimated Unemployment Rate (%)'].mean().sort_values(a
top_unemployment
```

Out[206]:
```
Region
Tripura              27.090769
Haryana              24.598077
Himachal Pradesh     17.910769
Jammu & Kashmir      16.285714
Jharkhand            16.274231
Bihar                16.031923
Chandigarh           15.822500
Delhi                15.544231
Rajasthan            13.673077
Punjab               11.305000
Name:  Estimated Unemployment Rate (%), dtype: float64
```

In [207]:
```
plt.figure(figsize=(10,6))
top_unemployment.plot.bar(color='cyan',edgecolor='black')
plt.title("Top 10 Unemployment Regions",fontsize=16)
plt.xlabel('Region',fontsize=12)
plt.ylabel('Average Rate of Unemployment',fontsize=12)
plt.tight_layout()
plt.show()
```

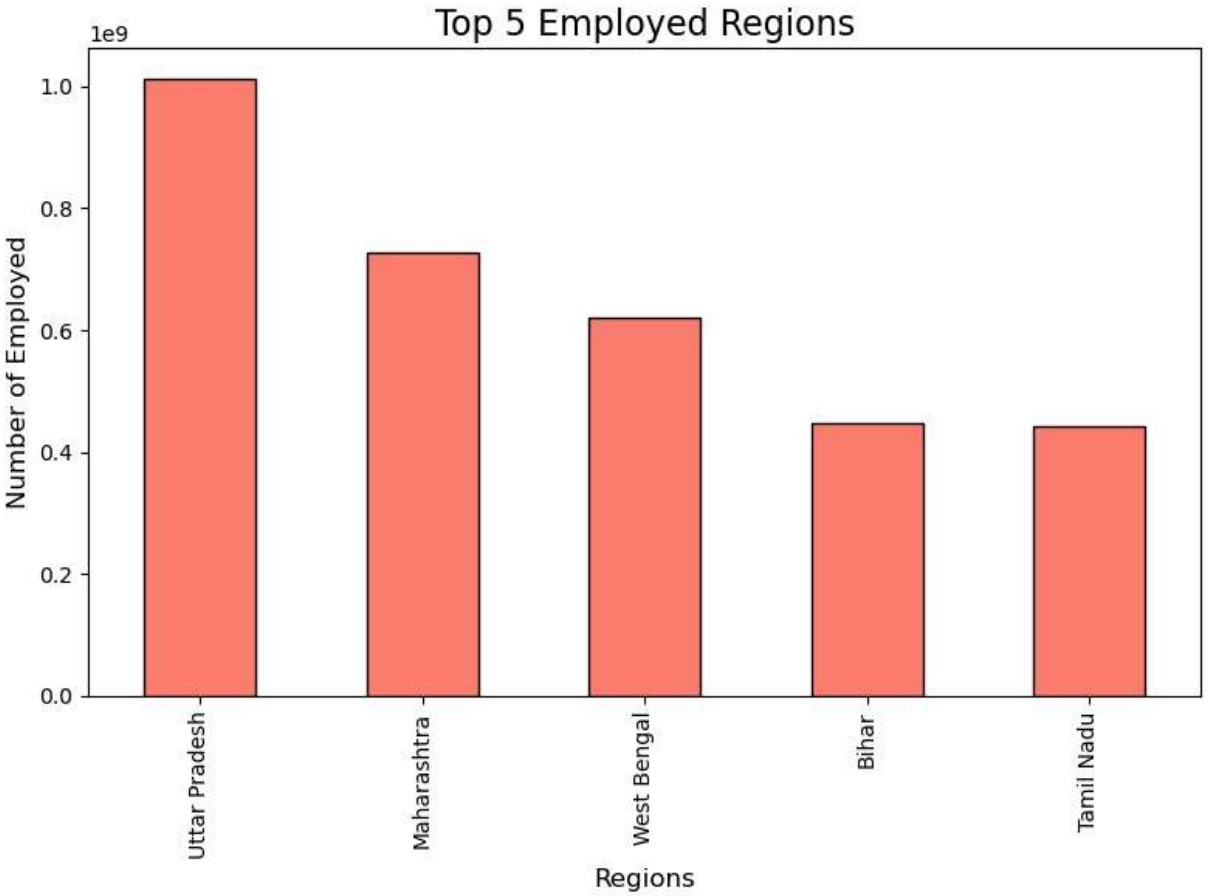


In [198]:
```
# Top 5 Employed Regions
```

In [211]:
```
top_employed = data_merged.groupby('Region')[' Estimated Employed'].sum().sort_values(ascending=False).he
top_employed
```

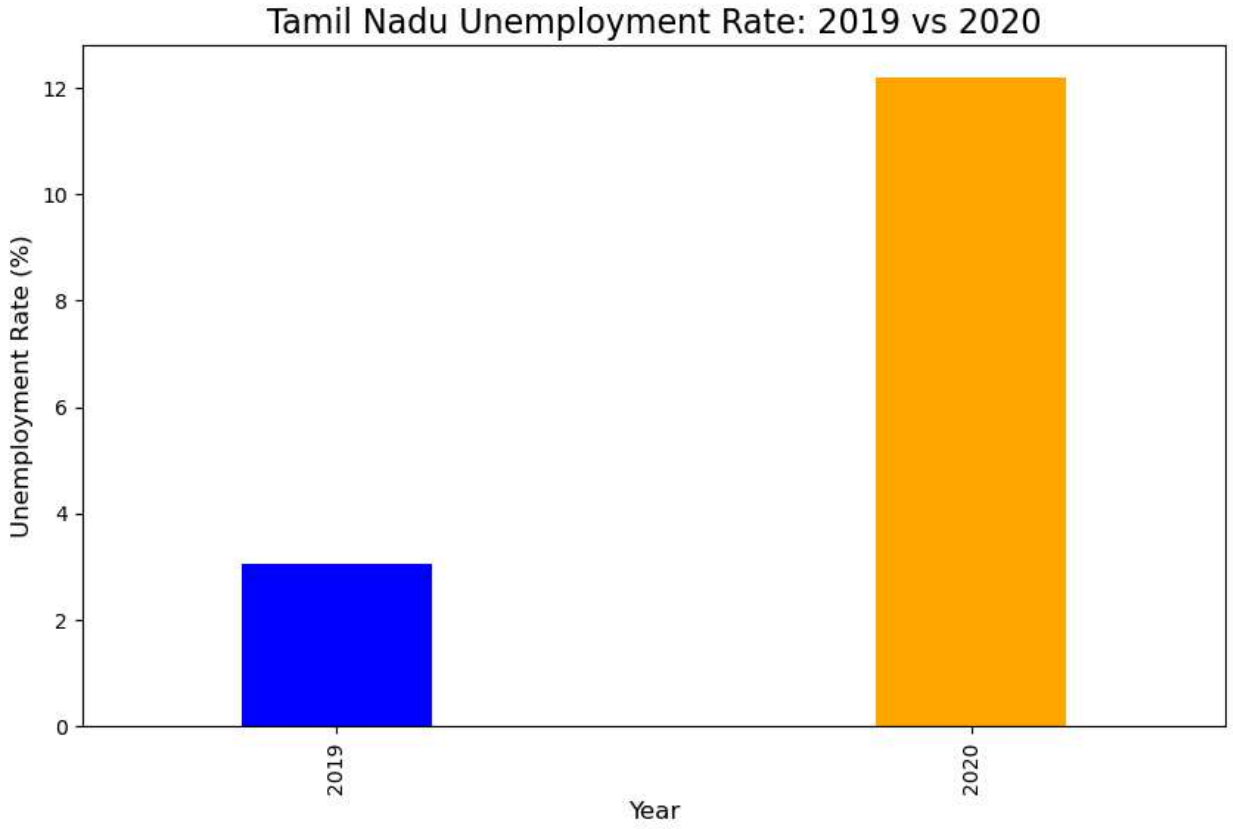
Out[211]:
```
Region
Uttar Pradesh    1.011647e+09
Maharashtra      7.283109e+08
West Bengal      6.203115e+08
Bihar            4.478739e+08
Tamil Nadu       4.430117e+08
Name:  Estimated Employed, dtype: float64
```

In [212]:
```
plt.figure(figsize=(8,6))
top_employed.plot.bar(color='salmon',edgecolor='black')
plt.title("Top 5 Employed Regions",fontsize=16)
plt.xlabel("Regions",fontsize=12)
plt.ylabel("Number of Employed",fontsize=12)
plt.tight_layout()
plt.show()
```



In [213]:
```
# Unemployment Rate of Tamil Nadu over 2019 vs 2020
```

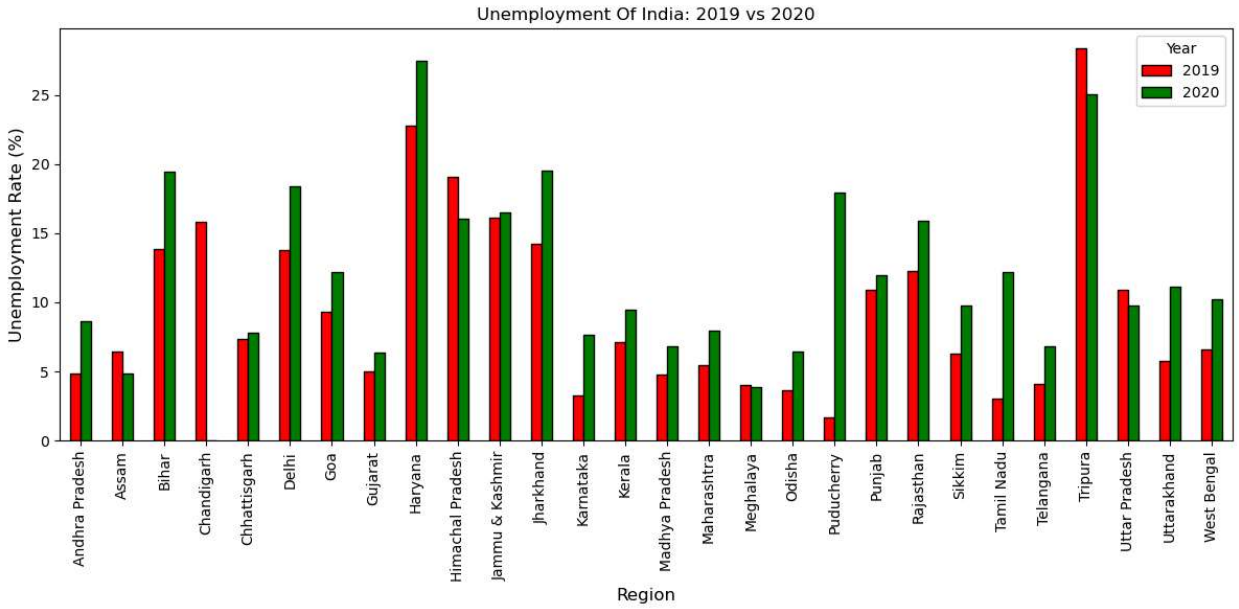
In [218]:
```
data_merged['Year'] = data_merged[' Date'].dt.year

tamil_nadu_data = data_merged[(data_merged['Region'] == 'Tamil Nadu') & (data_merged['Year'].isin([2019,2

tn_unemployment = tamil_nadu_data.groupby('Year')[' Estimated Unemployment Rate (%)'].mean()

plt.figure(figsize=(10,6))
tn_unemployment.plot.bar(color=['blue','orange'],width=0.3)
plt.title("Tamil Nadu Unemployment Rate: 2019 vs 2020",fontsize=16)
plt.xlabel("Year",fontsize=12)
plt.ylabel("Unemployment Rate (%)",fontsize=12)
plt.show()
```



In [219]:
```
# Unemployment Rate of all Regions: 2019 vs 2020
```

In [224]:
```
filtered_data = data_merged[data_merged['Year'].isin([2019,2020])]
unemployment_regions = filtered_data.groupby(['Region','Year'])[' Estimated Unemployment Rate (%)'].mean(
plt.figure(figsize=(12,6))
unemployment_regions.plot.bar(color=['red','green'],edgecolor='black',figsize=(12,6),width=0.5)
plt.title("Unemployment Of India: 2019 vs 2020",fontsize=12)
plt.xlabel("Region",fontsize=12)
plt.ylabel("Unemployment Rate (%)",fontsize=12)
plt.tight_layout()
plt.show()
```

```
<Figure size 1200x600 with 0 Axes>
```



In [ ]: