Unemployment in India

Data Analysis Using Python

Created by
Aravind Udiyana

The COVID-19 pandemic caused a sharp rise in unemployment across India during 2019-2020. Lockdowns and restrictions hit businesses hard, especially in sectors like hospitality, retail, and manufacturing, leading to widespread job losses, particularly among daily wage workers and informal sector employees. The unemployment rate soared to unprecedented levels, with rural areas also impacted by reverse migration. While government interventions, including relief packages and employment schemes, sought to alleviate the crisis, the recovery remained gradual.

Now, we will analyze the different unemployment patterns observed across various states during this period.

I found this data in unemploymentinindia.cmie.com site which focuses mainly on the unemployment of Indian people across all states during the initial stages of covid 19.

The Aim of this analysis is

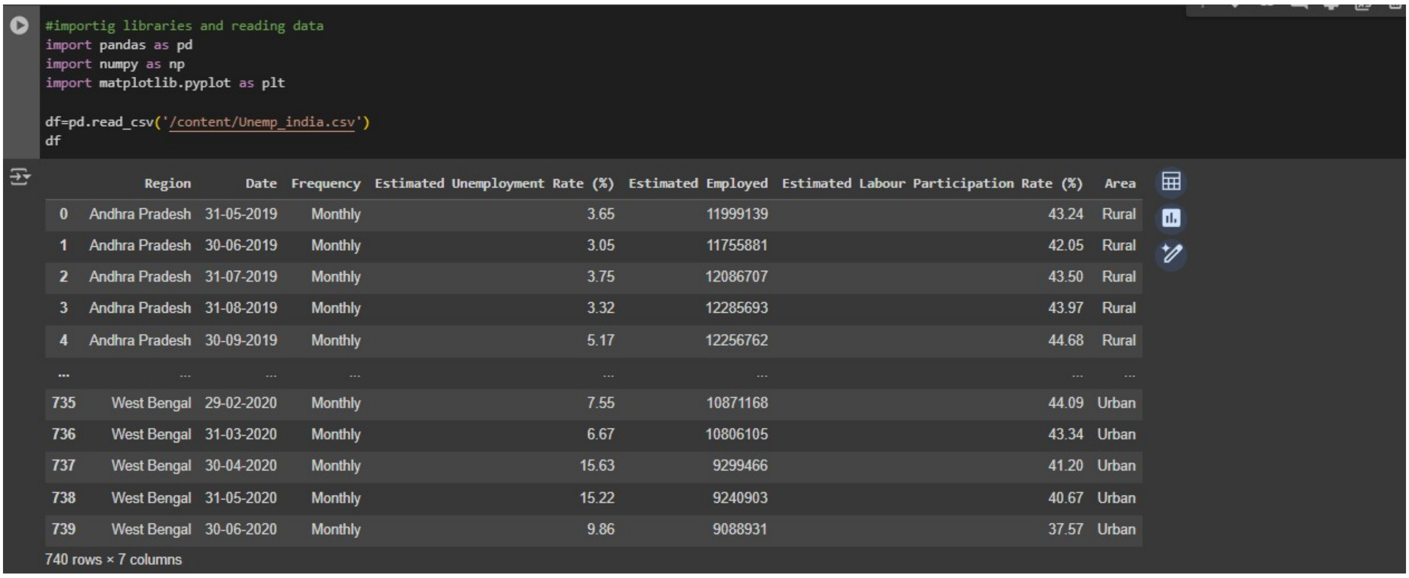
The main aim of this analysis is to analyze the different patterns among the states regarding unemployment and comparing the rate of employed people before and after covid 19 and visualizing them in chart like bar and line charts.

The workflow of this analysis is

1. Study the dataset for understanding different attributes
2. Cleaning the dataset for more and clarified data analysis
3. Formatting the dataset for our convenience
4. Extracting insights on unemployment rate, number of employed people, labour participation rare etc in different states before and after Covid 19
5. Visualizing them for better user experience
6. Finding any correlations between numerical typed attributes

Python Code and its following output

Importing libraries and reading data



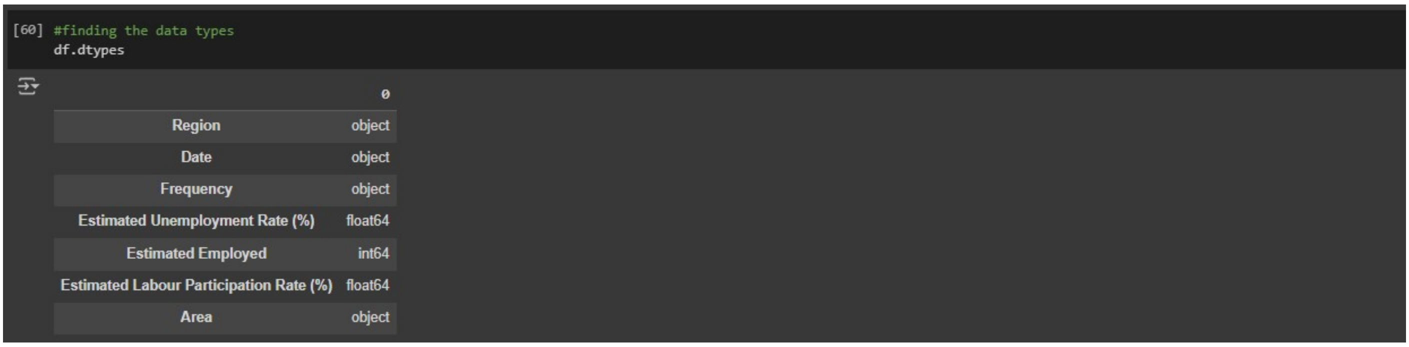
```
#importig libraries and reading data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df=pd.read_csv('/content/Unemp_india.csv')
df
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762	44.68	Rural
...	...	...	...	...	...	...	...
735	West Bengal	29-02-2020	Monthly	7.55	10871168	44.09	Urban
736	West Bengal	31-03-2020	Monthly	6.67	10806105	43.34	Urban
737	West Bengal	30-04-2020	Monthly	15.63	9299466	41.20	Urban
738	West Bengal	31-05-2020	Monthly	15.22	9240903	40.67	Urban
739	West Bengal	30-06-2020	Monthly	9.86	9088931	37.57	Urban

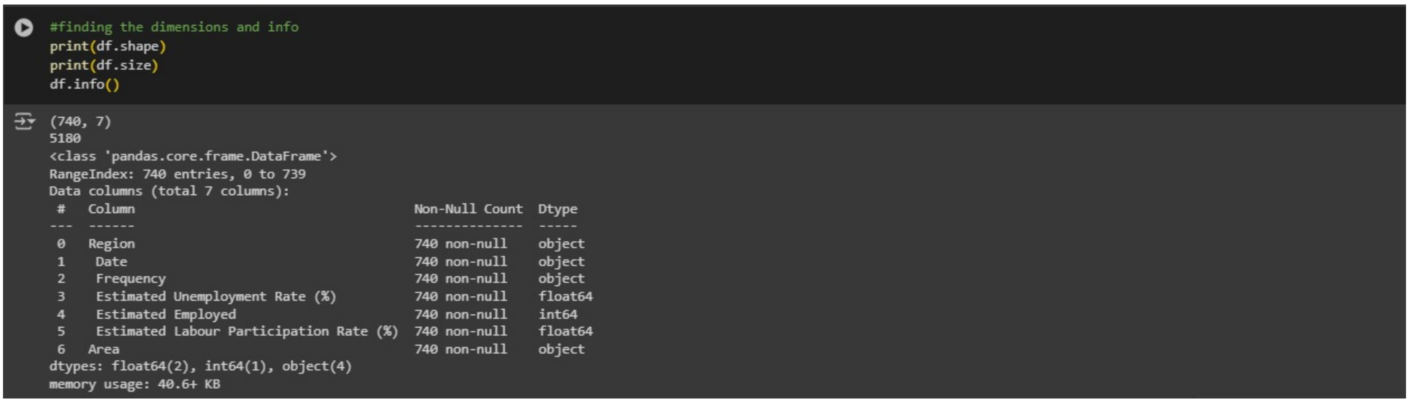
740 rows × 7 columns

Basic info about data



```
[60] #finding the data types
     df.dtypes
```

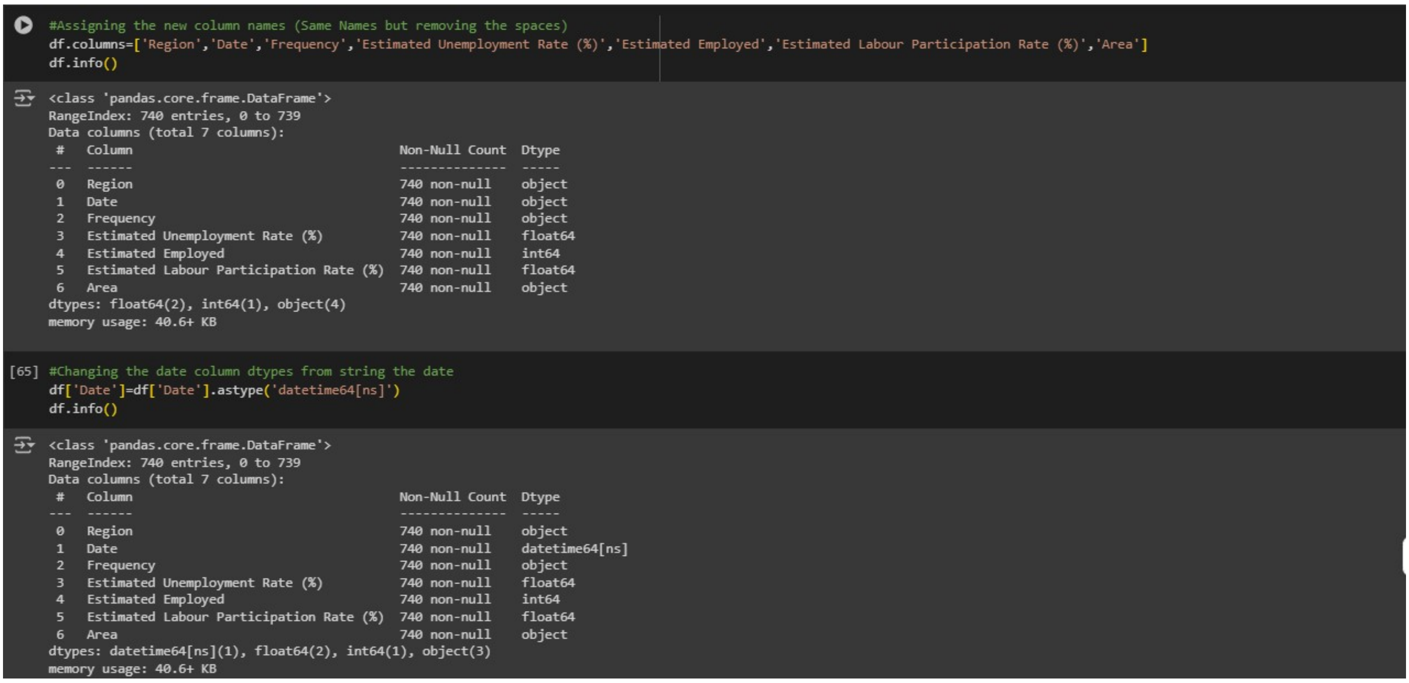
	0
Region	object
Date	object
Frequency	object
Estimated Unemployment Rate (%)	float64
Estimated Employed	int64
Estimated Labour Participation Rate (%)	float64
Area	object



```
#finding the dimensions and info
print(df.shape)
print(df.size)
df.info()
```

```
(740, 7)
5180
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 740 entries, 0 to 739
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    int64  
 5    Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                      740 non-null    object 
dtypes: float64(2), int64(1), object(4)
memory usage: 40.6+ KB
```

Cleaning the data



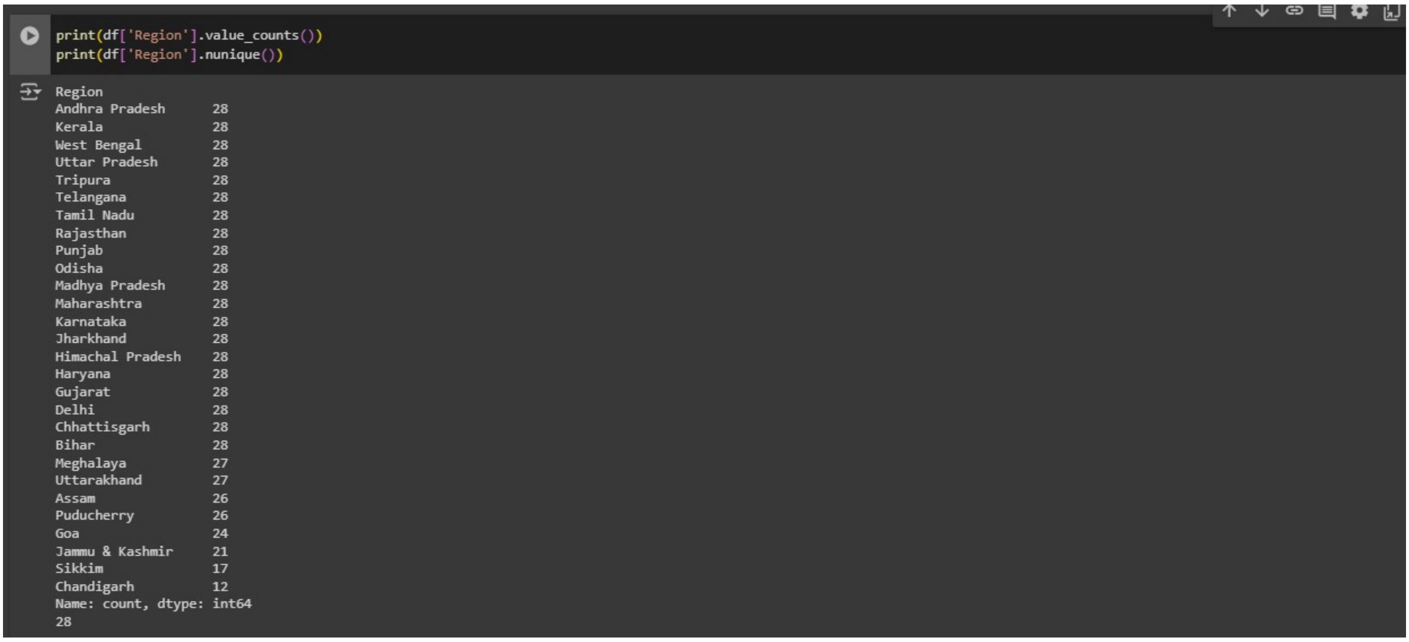
```
#Assigning the new column names (Same Names but removing the spaces)
df.columns=['Region','Date','Frequency','Estimated Unemployment Rate (%)','Estimated Employed','Estimated Labour Participation Rate (%)','Area']
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 740 entries, 0 to 739
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    int64
 5   Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                     740 non-null    object
dtypes: float64(2), int64(1), object(4)
memory usage: 40.6+ KB
```

```
[65] #Changing the date column dtypes from string the date
     df['Date']=df['Date'].astype('datetime64[ns]')
     df.info()
```

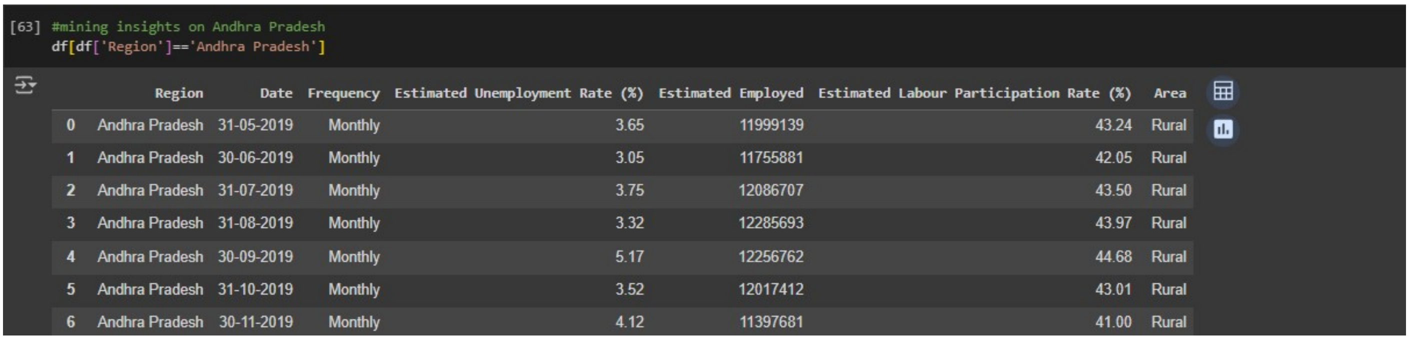
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 740 entries, 0 to 739
Data columns (total 7 columns):
 #   Column                                   Non-Null Count  Dtype
---  ------                                   --------------  -----
 0   Region                                   740 non-null    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    int64
 5   Estimated Labour Participation Rate (%)  740 non-null    float64
 6   Area                                     740 non-null    object
dtypes: datetime64[ns](1), float64(2), int64(1), object(3)
memory usage: 40.6+ KB
```

EDA : Exploratory Data Analysis



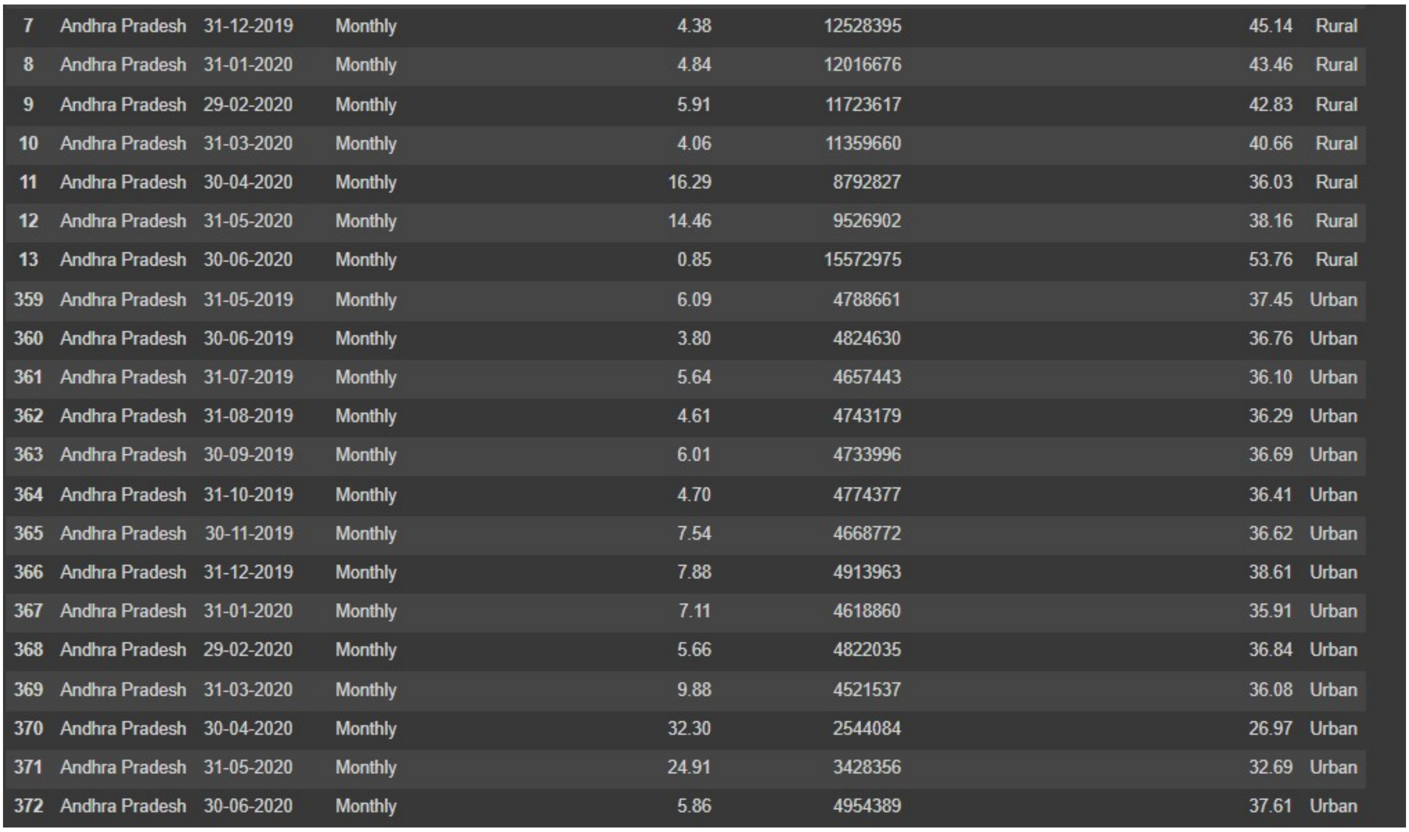
```
print(df['Region'].value_counts())
print(df['Region'].nunique())
```

```
Region
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: count, dtype: int64
28
```



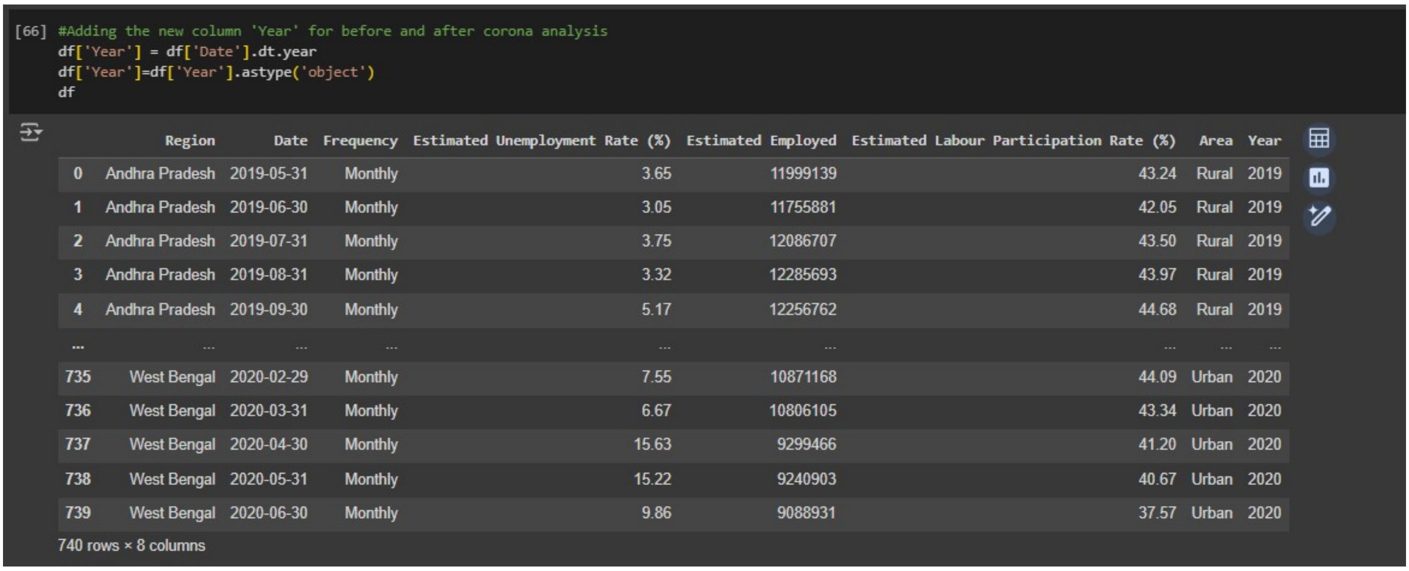
```
[63] #mining insights on Andhra Pradesh
     df[df['Region']=='Andhra Pradesh']
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139	43.24	Rural
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881	42.05	Rural
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707	43.50	Rural
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693	43.97	Rural
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762	44.68	Rural
5	Andhra Pradesh	31-10-2019	Monthly	3.52	12017412	43.01	Rural
6	Andhra Pradesh	30-11-2019	Monthly	4.12	11397681	41.00	Rural



7	Andhra Pradesh	31-12-2019	Monthly	4.38	12528395	45.14	Rural
8	Andhra Pradesh	31-01-2020	Monthly	4.84	12016676	43.46	Rural
9	Andhra Pradesh	29-02-2020	Monthly	5.91	11723617	42.83	Rural
10	Andhra Pradesh	31-03-2020	Monthly	4.06	11359660	40.66	Rural
11	Andhra Pradesh	30-04-2020	Monthly	16.29	8792827	36.03	Rural
12	Andhra Pradesh	31-05-2020	Monthly	14.46	9526902	38.16	Rural
13	Andhra Pradesh	30-06-2020	Monthly	0.85	15572975	53.76	Rural
359	Andhra Pradesh	31-05-2019	Monthly	6.09	4788661	37.45	Urban
360	Andhra Pradesh	30-06-2019	Monthly	3.80	4824630	36.76	Urban
361	Andhra Pradesh	31-07-2019	Monthly	5.64	4657443	36.10	Urban
362	Andhra Pradesh	31-08-2019	Monthly	4.61	4743179	36.29	Urban
363	Andhra Pradesh	30-09-2019	Monthly	6.01	4733996	36.69	Urban
364	Andhra Pradesh	31-10-2019	Monthly	4.70	4774377	36.41	Urban
365	Andhra Pradesh	30-11-2019	Monthly	7.54	4668772	36.62	Urban
366	Andhra Pradesh	31-12-2019	Monthly	7.88	4913963	38.61	Urban
367	Andhra Pradesh	31-01-2020	Monthly	7.11	4618860	35.91	Urban
368	Andhra Pradesh	29-02-2020	Monthly	5.66	4822035	36.84	Urban
369	Andhra Pradesh	31-03-2020	Monthly	9.88	4521537	36.08	Urban
370	Andhra Pradesh	30-04-2020	Monthly	32.30	2544084	26.97	Urban
371	Andhra Pradesh	31-05-2020	Monthly	24.91	3428356	32.69	Urban
372	Andhra Pradesh	30-06-2020	Monthly	5.86	4954389	37.61	Urban

Adding new column 'Year'

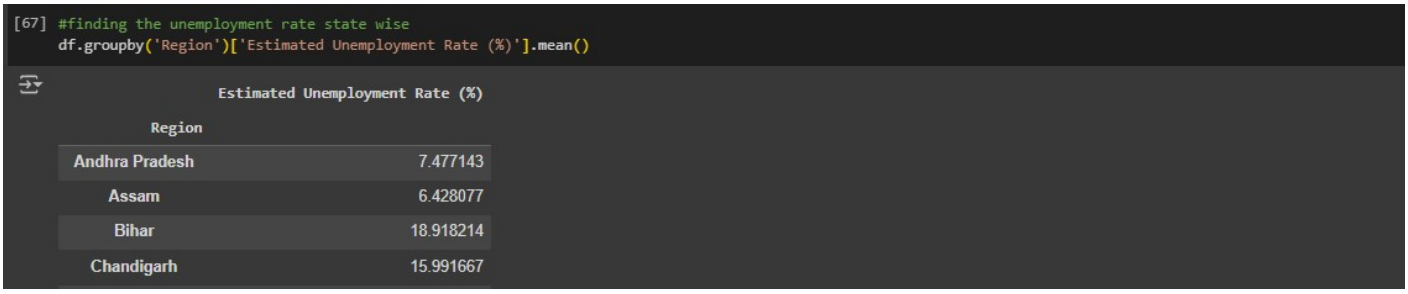


```
[66] #Adding the new column 'Year' for before and after corona analysis
     df['Year'] = df['Date'].dt.year
     df['Year']=df['Year'].astype('object')
     df
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Year
0	Andhra Pradesh	2019-05-31	Monthly	3.65	11999139	43.24	Rural	2019
1	Andhra Pradesh	2019-06-30	Monthly	3.05	11755881	42.05	Rural	2019
2	Andhra Pradesh	2019-07-31	Monthly	3.75	12086707	43.50	Rural	2019
3	Andhra Pradesh	2019-08-31	Monthly	3.32	12285693	43.97	Rural	2019
4	Andhra Pradesh	2019-09-30	Monthly	5.17	12256762	44.68	Rural	2019
...	...	...	...	...	...	...	...	...
735	West Bengal	2020-02-29	Monthly	7.55	10871168	44.09	Urban	2020
736	West Bengal	2020-03-31	Monthly	6.67	10806105	43.34	Urban	2020
737	West Bengal	2020-04-30	Monthly	15.63	9299466	41.20	Urban	2020
738	West Bengal	2020-05-31	Monthly	15.22	9240903	40.67	Urban	2020
739	West Bengal	2020-06-30	Monthly	9.86	9088931	37.57	Urban	2020

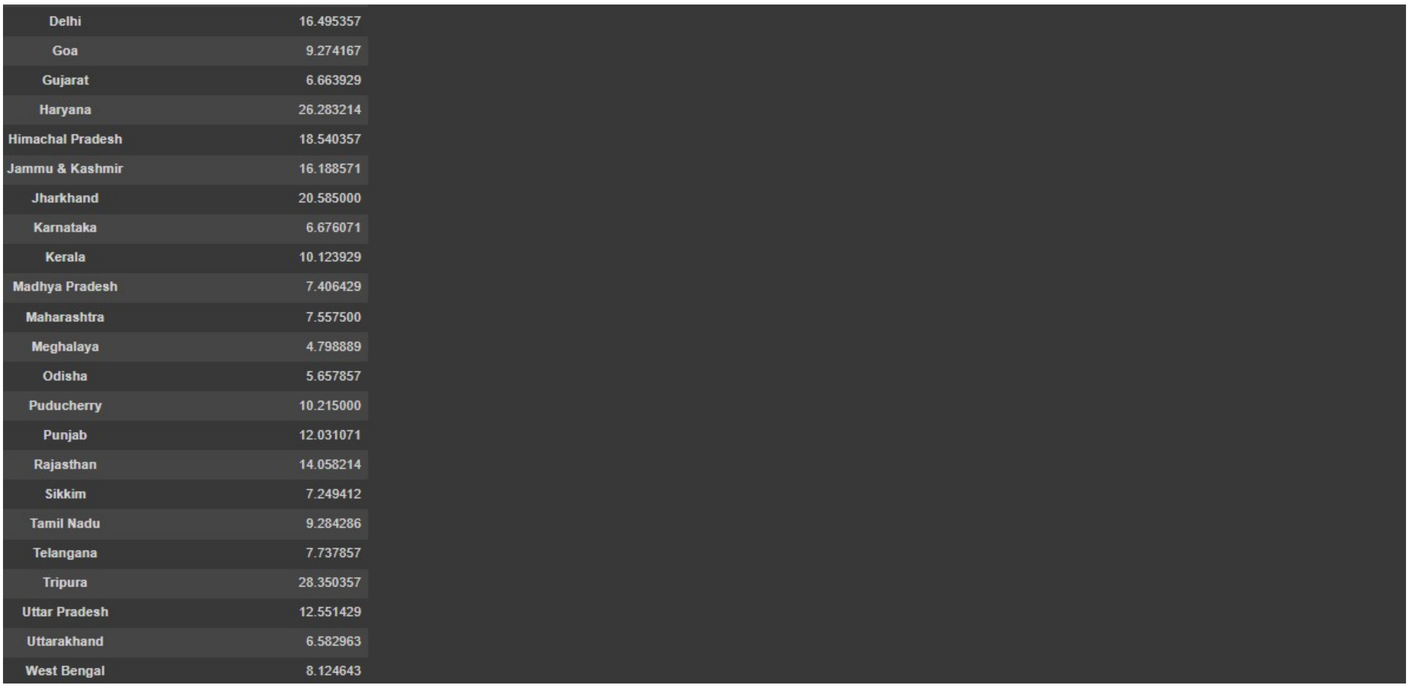
740 rows × 8 columns

Finding the Average unemployment rate during this period (2019-2020)



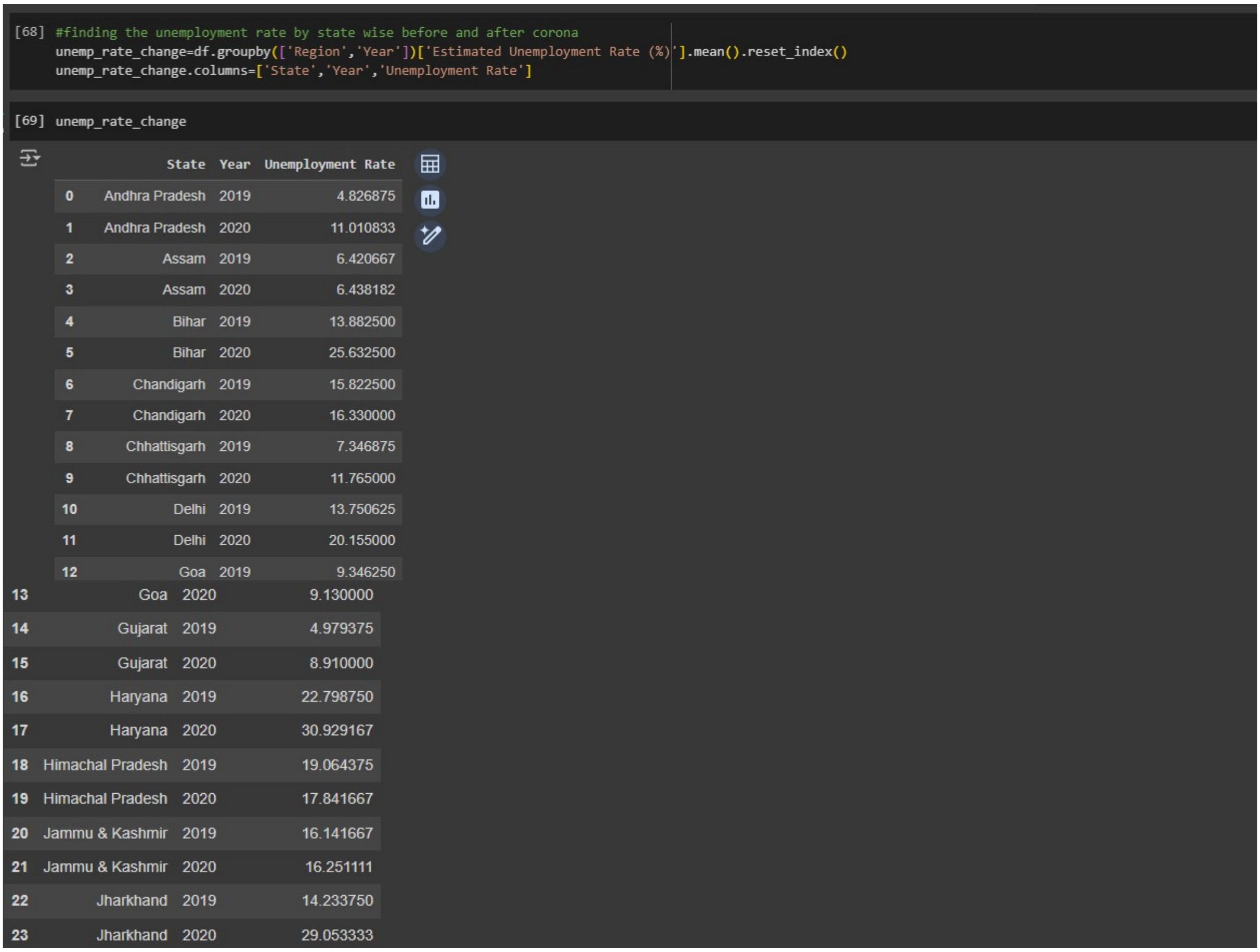
```
[67] #finding the unemployment rate state wise
     df.groupby('Region')['Estimated Unemployment Rate (%)'].mean()
```

	Estimated Unemployment Rate (%)
Region	
Andhra Pradesh	7.477143
Assam	6.428077
Bihar	18.918214
Chandigarh	15.991667



```
            Delhi   16.495357
              Goa    9.274167
          Gujarat    6.663929
          Haryana   26.283214
 Himachal Pradesh   18.540357
  Jammu & Kashmir   16.188571
        Jharkhand   20.585000
        Karnataka    6.676071
           Kerala   10.123929
   Madhya Pradesh    7.406429
      Maharashtra    7.557500
        Meghalaya    4.798889
           Odisha    5.657857
       Puducherry   10.215000
           Punjab   12.031071
        Rajasthan   14.058214
           Sikkim    7.249412
       Tamil Nadu    9.284286
        Telangana    7.737857
          Tripura   28.350357
    Uttar Pradesh   12.551429
      Uttarakhand    6.582963
      West Bengal    8.124643
```

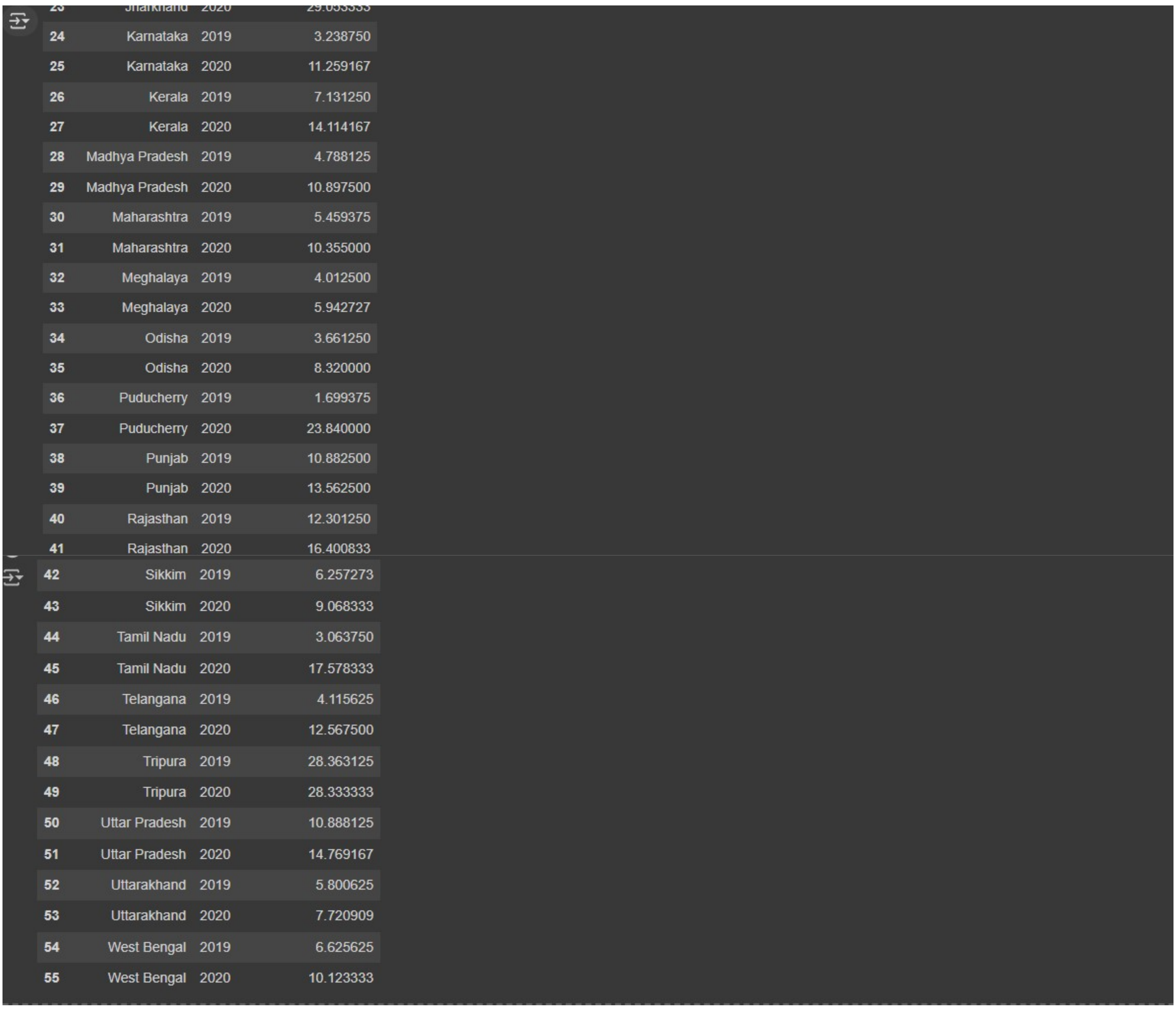
Finding the unemployment rate state wise before and after covid



```
[68] #finding the unemployment rate by state wise before and after corona
     unemp_rate_change=df.groupby(['Region','Year'])['Estimated Unemployment Rate (%)'].mean().reset_index()
     unemp_rate_change.columns=['State','Year','Unemployment Rate']
```

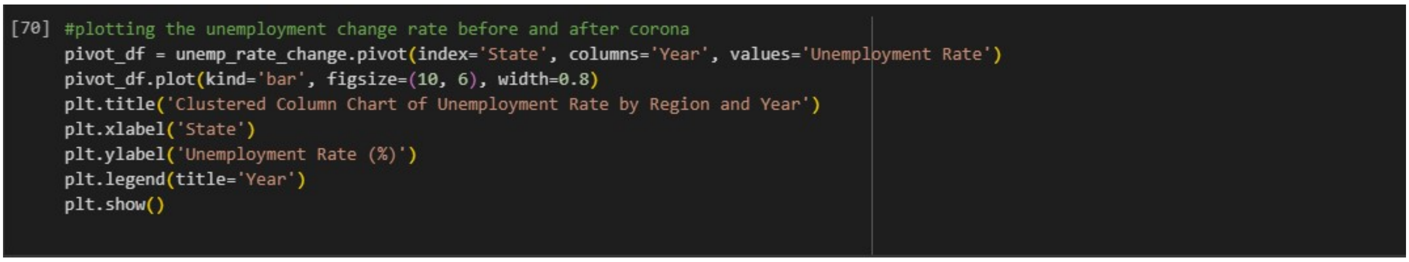
```
[69] unemp_rate_change
```

```
                State  Year  Unemployment Rate
 0     Andhra Pradesh  2019           4.826875
 1     Andhra Pradesh  2020          11.010833
 2              Assam  2019           6.420667
 3              Assam  2020           6.438182
 4              Bihar  2019          13.882500
 5              Bihar  2020          25.632500
 6         Chandigarh  2019          15.822500
 7         Chandigarh  2020          16.330000
 8       Chhattisgarh  2019           7.346875
 9       Chhattisgarh  2020          11.765000
10              Delhi  2019          13.750625
11              Delhi  2020          20.155000
12                Goa  2019           9.346250
13                Goa  2020           9.130000
14            Gujarat  2019           4.979375
15            Gujarat  2020           8.910000
16            Haryana  2019          22.798750
17            Haryana  2020          30.929167
18   Himachal Pradesh  2019          19.064375
19   Himachal Pradesh  2020          17.841667
20    Jammu & Kashmir  2019          16.141667
21    Jammu & Kashmir  2020          16.251111
22          Jharkhand  2019          14.233750
23          Jharkhand  2020          29.053333
```

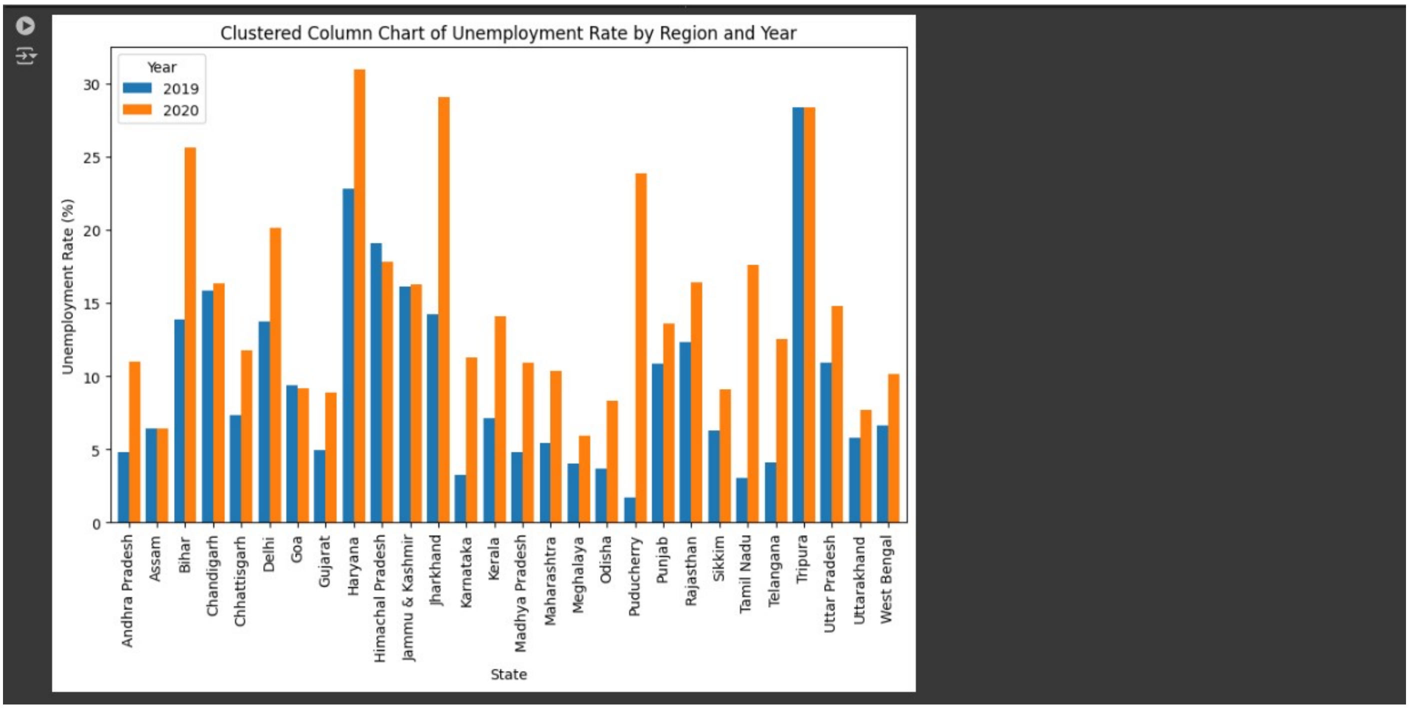


```
23       Jharkhand  2020   29.053333
24       Karnataka  2019    3.238750
25       Karnataka  2020   11.259167
26          Kerala  2019    7.131250
27          Kerala  2020   14.114167
28  Madhya Pradesh  2019    4.788125
29  Madhya Pradesh  2020   10.897500
30     Maharashtra  2019    5.459375
31     Maharashtra  2020   10.355000
32       Meghalaya  2019    4.012500
33       Meghalaya  2020    5.942727
34          Odisha  2019    3.661250
35          Odisha  2020    8.320000
36      Puducherry  2019    1.699375
37      Puducherry  2020   23.840000
38          Punjab  2019   10.882500
39          Punjab  2020   13.562500
40       Rajasthan  2019   12.301250
41       Rajasthan  2020   16.400833
42          Sikkim  2019    6.257273
43          Sikkim  2020    9.068333
44      Tamil Nadu  2019    3.063750
45      Tamil Nadu  2020   17.578333
46       Telangana  2019    4.115625
47       Telangana  2020   12.567500
48         Tripura  2019   28.363125
49         Tripura  2020   28.333333
50   Uttar Pradesh  2019   10.888125
51   Uttar Pradesh  2020   14.769167
52     Uttarakhand  2019    5.800625
53     Uttarakhand  2020    7.720909
54     West Bengal  2019    6.625625
55     West Bengal  2020   10.123333
```

Plotting the same table in clustered chart

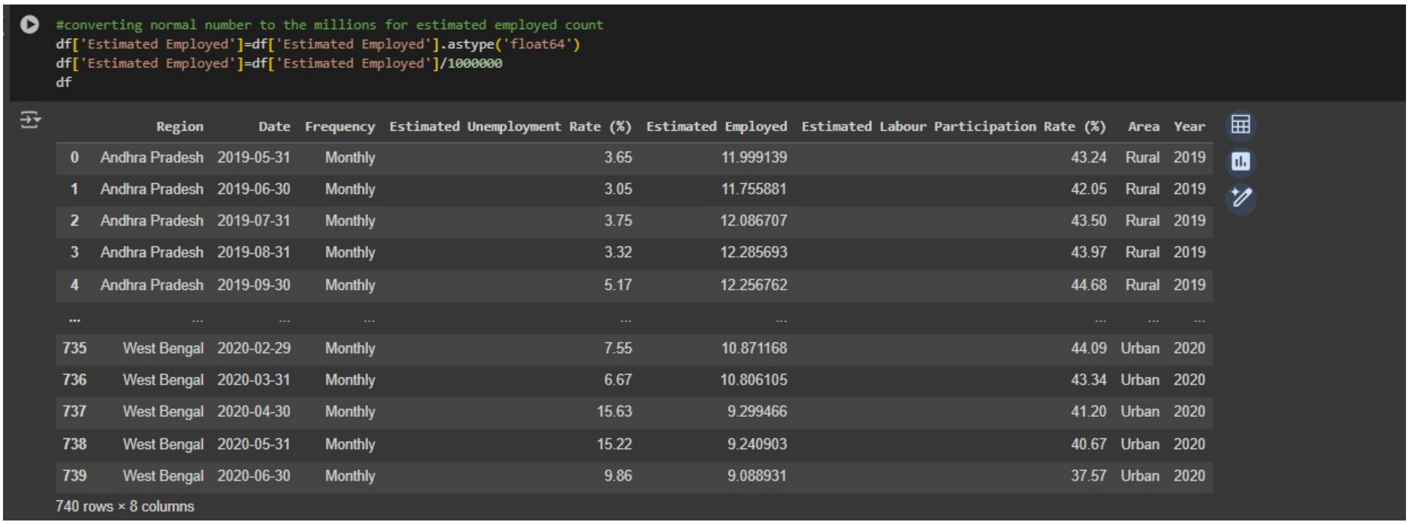


```
[70] #plotting the unemployment change rate before and after corona
     pivot_df = unemp_rate_change.pivot(index='State', columns='Year', values='Unemployment Rate')
     pivot_df.plot(kind='bar', figsize=(10, 6), width=0.8)
     plt.title('Clustered Column Chart of Unemployment Rate by Region and Year')
     plt.xlabel('State')
     plt.ylabel('Unemployment Rate (%)')
     plt.legend(title='Year')
     plt.show()
```



We can clearly see the increase in rate of unemployment in almost all states of India after the entrance of covid 19 into the country.

Converting the employed count to millions

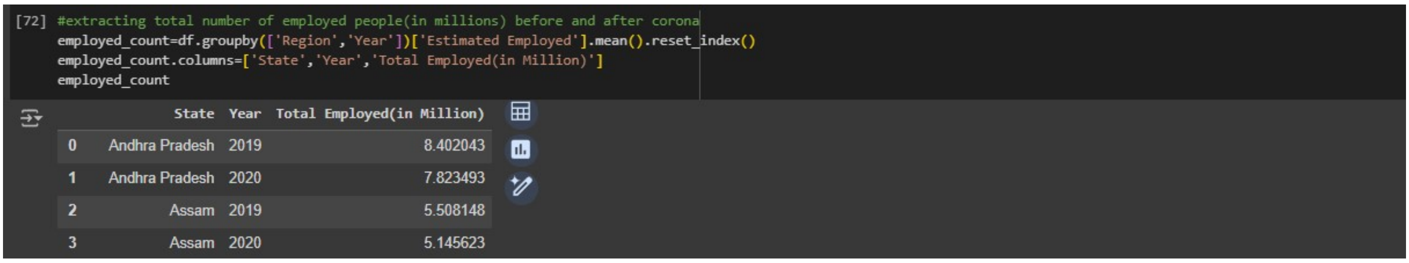


```
#converting normal number to the millions for estimated employed count
df['Estimated Employed']=df['Estimated Employed'].astype('float64')
df['Estimated Employed']=df['Estimated Employed']/1000000
df
```

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Year
0	Andhra Pradesh	2019-05-31	Monthly	3.65	11.999139	43.24	Rural	2019
1	Andhra Pradesh	2019-06-30	Monthly	3.05	11.755881	42.05	Rural	2019
2	Andhra Pradesh	2019-07-31	Monthly	3.75	12.086707	43.50	Rural	2019
3	Andhra Pradesh	2019-08-31	Monthly	3.32	12.285693	43.97	Rural	2019
4	Andhra Pradesh	2019-09-30	Monthly	5.17	12.256762	44.68	Rural	2019
...	...	...	...	...	...	...	...	...
735	West Bengal	2020-02-29	Monthly	7.55	10.871168	44.09	Urban	2020
736	West Bengal	2020-03-31	Monthly	6.67	10.806105	43.34	Urban	2020
737	West Bengal	2020-04-30	Monthly	15.63	9.299466	41.20	Urban	2020
738	West Bengal	2020-05-31	Monthly	15.22	9.240903	40.67	Urban	2020
739	West Bengal	2020-06-30	Monthly	9.86	9.088931	37.57	Urban	2020

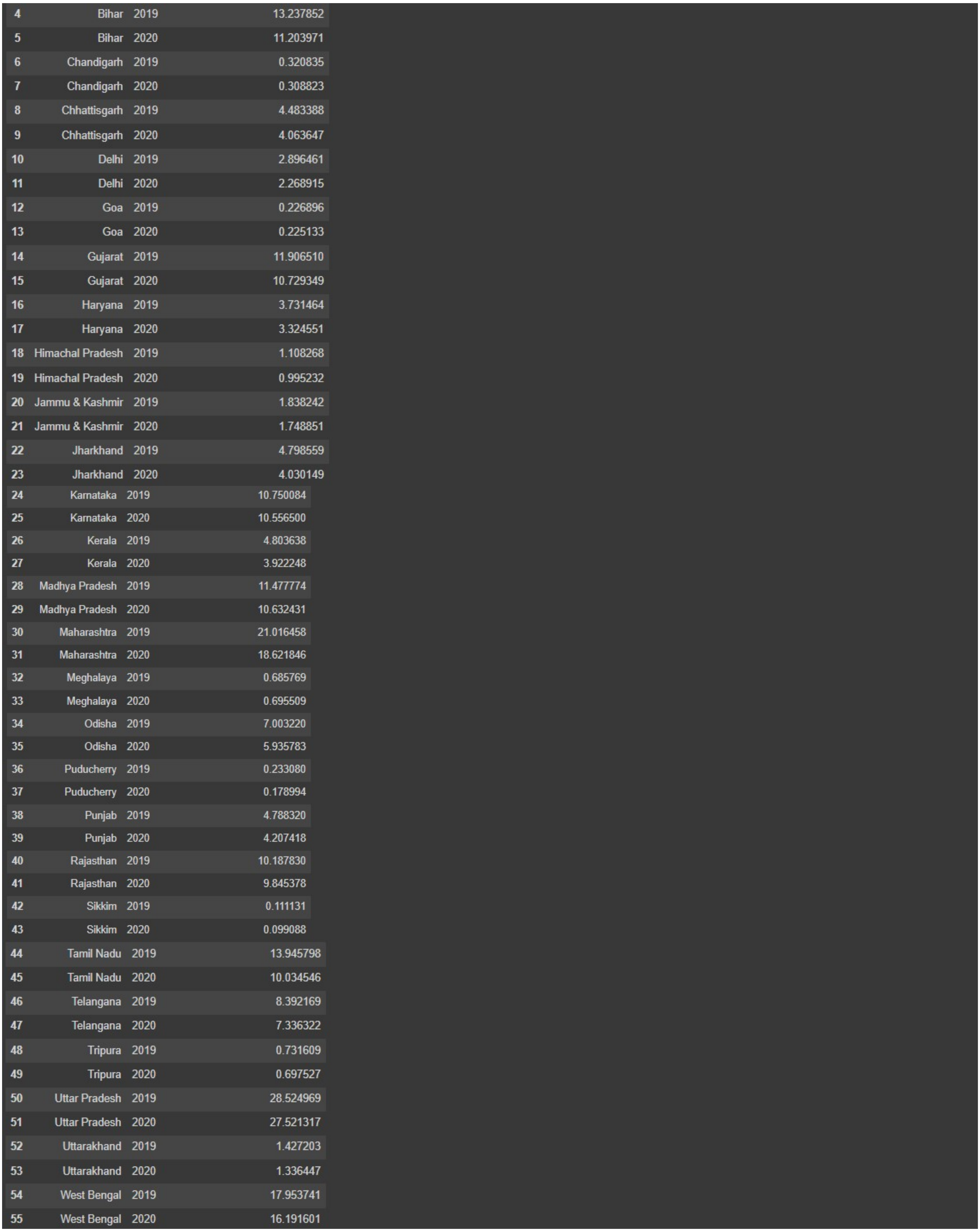
740 rows × 8 columns

Finding the no.of employed people before and after covid



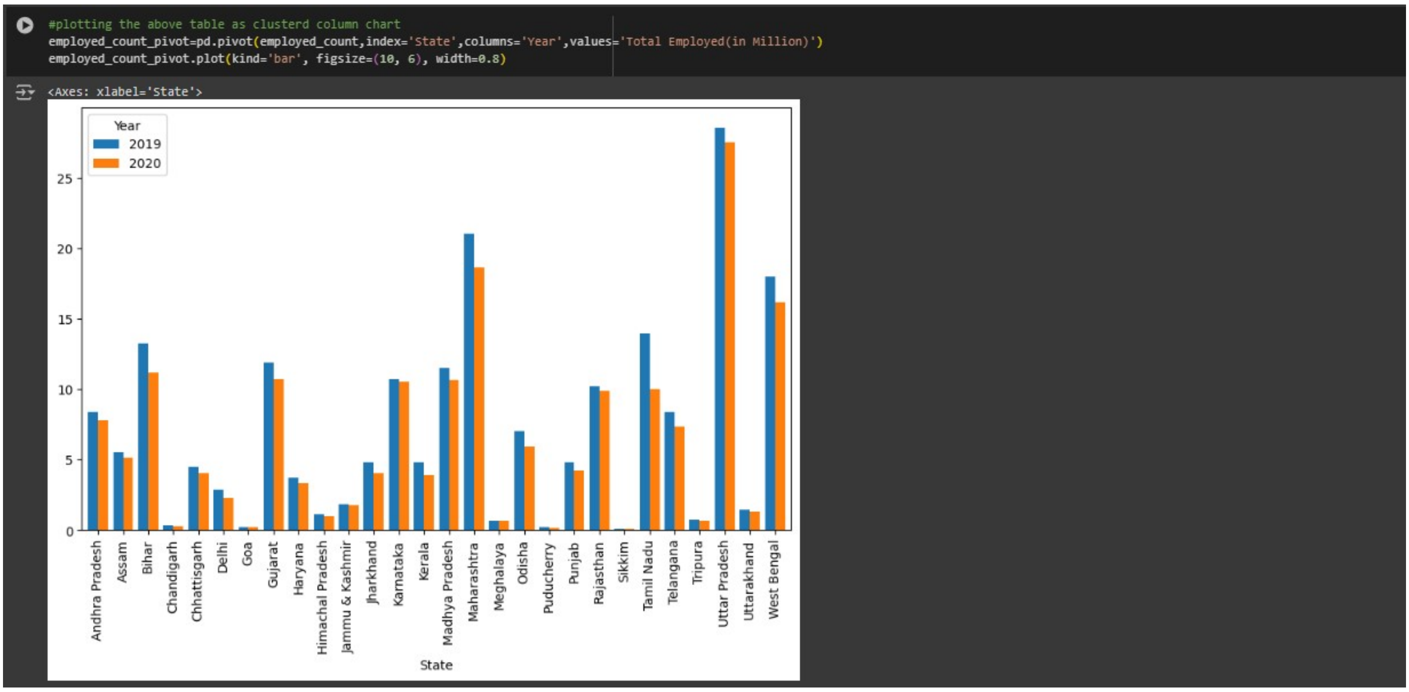
```
[72] #extracting total number of employed people(in millions) before and after corona
employed_count=df.groupby(['Region','Year'])['Estimated Employed'].mean().reset_index()
employed_count.columns=['State','Year','Total Employed(in Million)']
employed_count
```

	State	Year	Total Employed(in Million)
0	Andhra Pradesh	2019	8.402043
1	Andhra Pradesh	2020	7.823493
2	Assam	2019	5.508148
3	Assam	2020	5.145623



```
4            Bihar  2019  13.237852
5            Bihar  2020  11.203971
6       Chandigarh  2019   0.320835
7       Chandigarh  2020   0.308823
8     Chhattisgarh  2019   4.483388
9     Chhattisgarh  2020   4.063647
10           Delhi  2019   2.896461
11           Delhi  2020   2.268915
12             Goa  2019   0.226896
13             Goa  2020   0.225133
14         Gujarat  2019  11.906510
15         Gujarat  2020  10.729349
16         Haryana  2019   3.731464
17         Haryana  2020   3.324551
18 Himachal Pradesh 2019   1.108268
19 Himachal Pradesh 2020   0.995232
20 Jammu & Kashmir  2019   1.838242
21 Jammu & Kashmir  2020   1.748851
22       Jharkhand  2019   4.798559
23       Jharkhand  2020   4.030149
24       Karnataka  2019  10.750084
25       Karnataka  2020  10.556500
26          Kerala  2019   4.803638
27          Kerala  2020   3.922248
28  Madhya Pradesh  2019  11.477774
29  Madhya Pradesh  2020  10.632431
30     Maharashtra  2019  21.016458
31     Maharashtra  2020  18.621846
32       Meghalaya  2019   0.685769
33       Meghalaya  2020   0.695509
34          Odisha  2019   7.003220
35          Odisha  2020   5.935783
36      Puducherry  2019   0.233080
37      Puducherry  2020   0.178994
38          Punjab  2019   4.788320
39          Punjab  2020   4.207418
40       Rajasthan  2019  10.187830
41       Rajasthan  2020   9.845378
42          Sikkim  2019   0.111131
43          Sikkim  2020   0.099088
44      Tamil Nadu  2019  13.945798
45      Tamil Nadu  2020  10.034546
46       Telangana  2019   8.392169
47       Telangana  2020   7.336322
48         Tripura  2019   0.731609
49         Tripura  2020   0.697527
50   Uttar Pradesh  2019  28.524969
51   Uttar Pradesh  2020  27.521317
52     Uttarakhand  2019   1.427203
53     Uttarakhand  2020   1.336447
54     West Bengal  2019  17.953741
55     West Bengal  2020  16.191601
```

Plotting the same in Clustered column chart

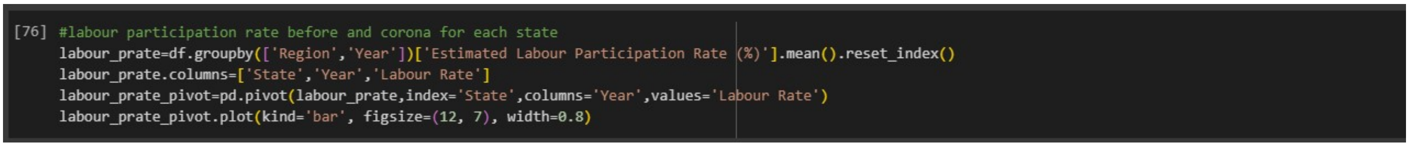


```
#plotting the above table as clusterd column chart
employed_count_pivot=pd.pivot(employed_count,index='State',columns='Year',values='Total Employed(in Million)')
employed_count_pivot.plot(kind='bar', figsize=(10, 6), width=0.8)
```

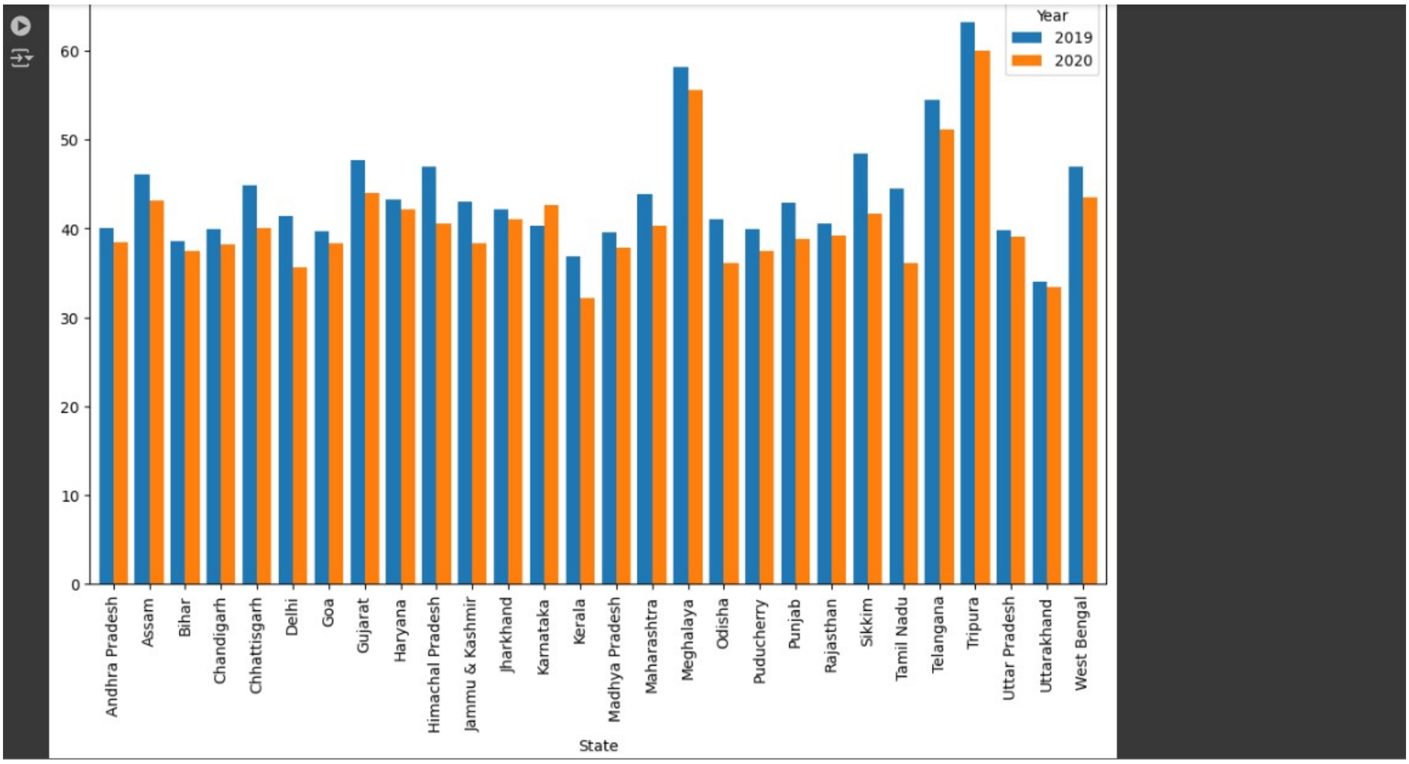
```
<Axes: xlabel='State'>
```

We can clearly see that the no.of employed people decreased and unemployment rate increased(by observing previous chart) in all states of India after the during the covid19 pandemic.

Labour Participation rate in India before and after covid19



```
[76] #labour participation rate before and corona for each state
     labour_prate=df.groupby(['Region','Year'])['Estimated Labour Participation Rate (%)'].mean().reset_index()
     labour_prate.columns=['State','Year','Labour Rate']
     labour_prate_pivot=pd.pivot(labour_prate,index='State',columns='Year',values='Labour Rate')
     labour_prate_pivot.plot(kind='bar', figsize=(12, 7), width=0.8)
```

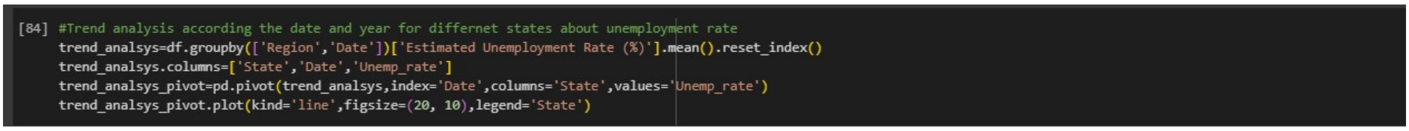


Labour participation means, the sum of population who are employed and unemployed (but still looking for employment) out of every 100 working age people

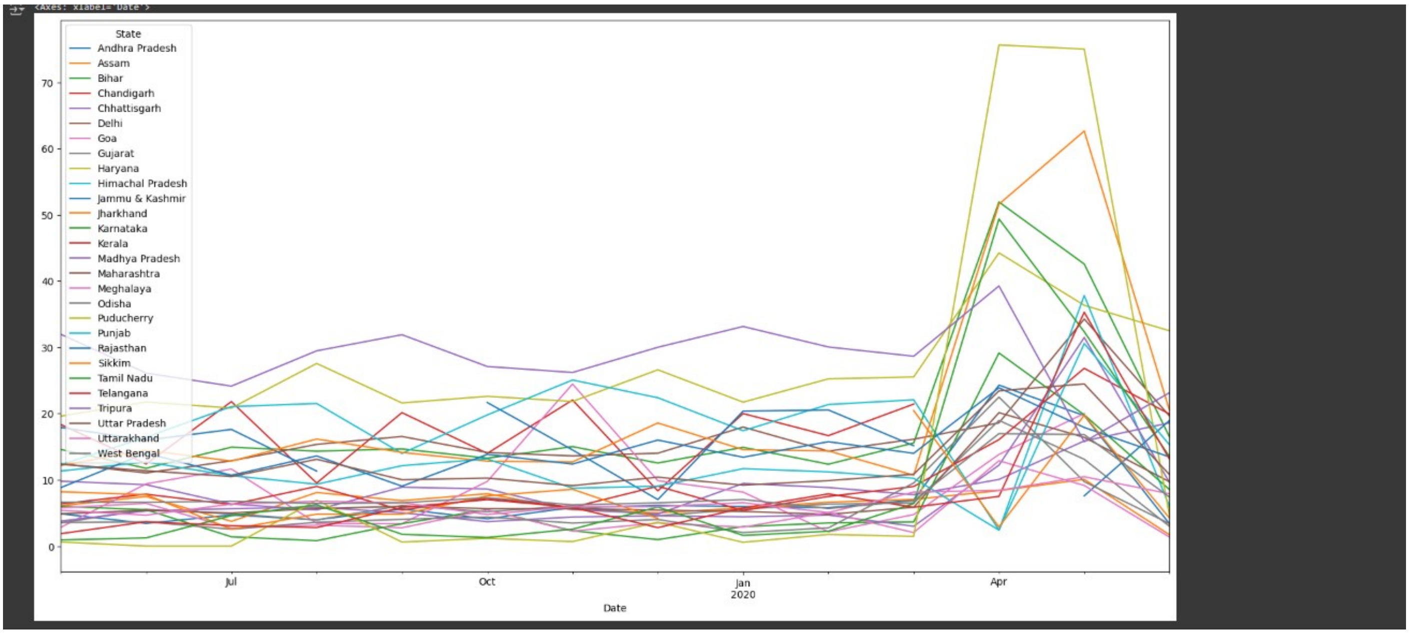
Labour participation rate = (labour Force/Working age population)*100

So according the output of that chart, the percentage of labour participation is also reduced significantly during the period of covid19

Trend Analysis of different states during each month of covid19



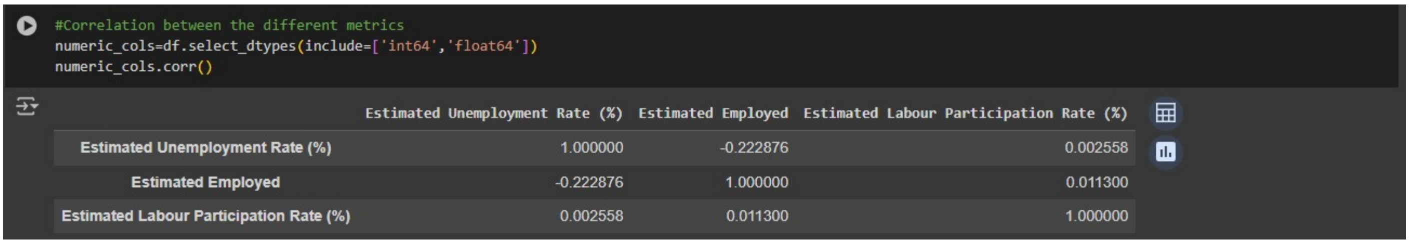
```
[84] #Trend analysis according the date and year for differnet states about unemployment rate
     trend_analsys=df.groupby(['Region','Date'])['Estimated Unemployment Rate (%)'].mean().reset_index()
     trend_analsys.columns=['State','Date','Unemp_rate']
     trend_analsys_pivot=pd.pivot(trend_analsys,index='Date',columns='State',values='Unemp_rate')
     trend_analsys_pivot.plot(kind='line',figsize=(20, 10),legend='State')
```



```
<Axes: xlabel='Date'>
```

We can see the huge rise of unemployment rate in India in all states during April 2020 which indicates us that covid19 was at its peak during this period and job market was very weak at that period.

Correlation of different metrics



```
#Correlation between the different metrics
numeric_cols=df.select_dtypes(include=['int64','float64'])
numeric_cols.corr()
```

	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)
Estimated Unemployment Rate (%)	1.000000	-0.222876	0.002558
Estimated Employed	-0.222876	1.000000	0.011300
Estimated Labour Participation Rate (%)	0.002558	0.011300	1.000000

We can see there is negative correlation between total employed and unemployment rate as we proved using the charts.

So, That’s it for this document. Meet you again in next project. Thank You.