

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

```
df = pd.read_csv("Unemployment.csv")
df.head()
```

↗

	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19	Unnamed: 20	Unnamed: 21	Unnamed: 2
0	Andhra Pradesh	31-05-2019	Monthly	3.65	11999139.0	43.24	Rural	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	Na
1	Andhra Pradesh	30-06-2019	Monthly	3.05	11755881.0	42.05	Rural	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	Na
2	Andhra Pradesh	31-07-2019	Monthly	3.75	12086707.0	43.50	Rural	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	Na
3	Andhra Pradesh	31-08-2019	Monthly	3.32	12285693.0	43.97	Rural	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	Na
4	Andhra Pradesh	30-09-2019	Monthly	5.17	12256762.0	44.68	Rural	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	Na

5 rows × 26 columns

```
df.info()
```

↗

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 754 entries, 0 to 753
Data columns (total 26 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
```

```
7 Unnamed: 7      0 non-null    float64
8 Unnamed: 8      0 non-null    float64
9 Unnamed: 9      0 non-null    float64
10 Unnamed: 10    0 non-null    float64
11 Unnamed: 11    0 non-null    float64
12 Unnamed: 12    0 non-null    float64
13 Unnamed: 13    0 non-null    float64
14 Unnamed: 14    0 non-null    float64
15 Unnamed: 15    0 non-null    float64
16 Unnamed: 16    0 non-null    float64
17 Unnamed: 17    0 non-null    float64
18 Unnamed: 18    0 non-null    float64
19 Unnamed: 19    0 non-null    float64
20 Unnamed: 20    0 non-null    float64
21 Unnamed: 21    0 non-null    float64
22 Unnamed: 22    0 non-null    float64
23 Unnamed: 23    0 non-null    float64
24 Unnamed: 24    0 non-null    float64
25 Unnamed: 25    0 non-null    float64
```

```
dtypes: float64(22), object(4)
memory usage: 153.3+ KB
```


df.tail()



	Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19	Unnamed: 20	Unnamed: 21	Unnam
749	West Bengal	29-02-2020	Monthly	7.55	10871168.0	44.09	Urban	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	↑
750	West Bengal	31-03-2020	Monthly	6.67	10806105.0	43.34	Urban	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	↑
751	West Bengal	30-04-2020	Monthly	15.63	9299466.0	41.20	Urban	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	↑
752	West Bengal	31-05-2020	Monthly	15.22	9240903.0	40.67	Urban	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	↑
753	West Bengal	30-06-2020	Monthly	9.86	9088931.0	37.57	Urban	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	↑

5 rows × 26 columns

df.shape

 (754, 26)

df.describe()



	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Unnamed: 7	Unnamed: 8	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12	Unnamed: 13	...	Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19	Unnamed: 20	Unn
count	740.000000	7.400000e+02	740.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
mean	11.787946	7.204460e+06	42.630122	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
std	10.721298	8.087988e+06	8.111094	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
min	0.000000	4.942000e+04	13.330000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
25%	4.657500	1.190404e+06	38.062500	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
50%	8.350000	4.744178e+06	41.160000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
75%	15.887500	1.127549e+07	45.505000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	
max	76.740000	4.577751e+07	72.570000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	

8 rows × 22 columns



x = df['Region']

x



```
0    Andhra Pradesh
1    Andhra Pradesh
2    Andhra Pradesh
3    Andhra Pradesh
4    Andhra Pradesh
...
749  West Bengal
750  West Bengal
751  West Bengal
752  West Bengal
753  West Bengal
Name: Region, Length: 754, dtype: category
Categories (28, object): ['Andhra Pradesh', 'Assam', 'Bihar', 'Chandigarh', ..., 'Tripura', 'Uttar Pradesh',
                        'Uttarakhand', 'West Bengal']
```

df2 = df.iloc[:,3]

df2

```
0      3.65
1      3.05
2      3.75
3      3.32
4      5.17
...
749     7.55
750     6.67
751    15.63
752    15.22
753     9.86
Name: Estimated Unemployment Rate (%), Length: 754, dtype: float64
```

```
df_g=df.groupby('Region')['Area'].apply(lambda x: x.mode().iloc[0]).reset_index()
```

```
df_g.columns = ['Region', 'Mode_Area']
df_g
```



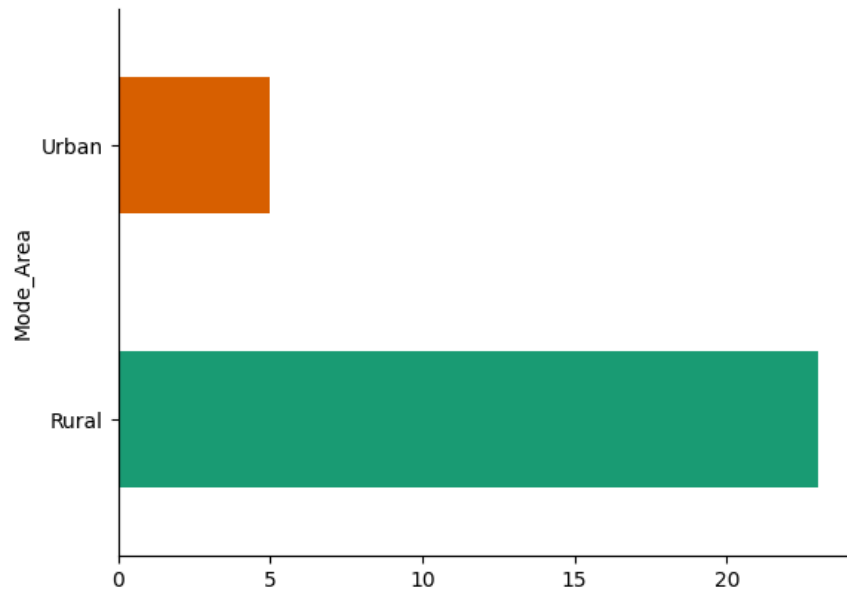
	Region	Mode_Area
0	Andhra Pradesh	Rural
1	Assam	Urban
2	Bihar	Rural
3	Chandigarh	Urban
4	Chhattisgarh	Rural
5	Delhi	Rural
6	Goa	Rural
7	Gujarat	Rural
8	Haryana	Rural
9	Himachal Pradesh	Rural
10	Jammu & Kashmir	Rural
11	Jharkhand	Rural
12	Karnataka	Rural
13	Kerala	Rural
14	Madhya Pradesh	Rural
15	Maharashtra	Rural
16	Meghalaya	Rural
17	Odisha	Rural
18	Puducherry	Urban
19	Punjab	Rural
20	Rajasthan	Rural
21	Sikkim	Urban
22	Tamil Nadu	Rural
23	Telangana	Rural
24	Tripura	Rural
25	Uttar Pradesh	Rural
26	Uttarakhand	Urban
27	West Bengal	Rural

```
df.isna().sum().reset_index().rename(columns={0: 'count'})
```



	index	count
0	Region	14
1	Date	14
2	Frequency	14
3	Estimated Unemployment Rate (%)	14
4	Estimated Employed	14
5	Estimated Labour Participation Rate (%)	14
6	Area	14
7	Unnamed: 7	754
8	Unnamed: 8	754
9	Unnamed: 9	754
10	Unnamed: 10	754
11	Unnamed: 11	754
12	Unnamed: 12	754
13	Unnamed: 13	754
14	Unnamed: 14	754
15	Unnamed: 15	754
16	Unnamed: 16	754
17	Unnamed: 17	754
18	Unnamed: 18	754
19	Unnamed: 19	754
20	Unnamed: 20	754
21	Unnamed: 21	754
22	Unnamed: 22	754
23	Unnamed: 23	754
24	Unnamed: 24	754
25	Unnamed: 25	754

```
from matplotlib import pyplot as plt
import seaborn as sns
df_g.groupby('Mode_Area').size().plot(kind='barh', color=sns.palettes.mpl_palette('Dark2'))
plt.gca().spines[['top', 'right']].set_visible(False)
```



```
df['Region'].unique()
```



```
['Andhra Pradesh', 'Assam', 'Bihar', 'Chhattisgarh', 'Delhi', ..., 'Uttar Pradesh', 'Uttarakhand', 'West Bengal', NaN, 'Chandigarh']
Length: 29
Categories (28, object): ['Andhra Pradesh', 'Assam', 'Bihar', 'Chandigarh', ..., 'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal']
```

```
df['Region'].value_counts().reset_index()
```



	Region	count
0	Andhra Pradesh	28
1	Karnataka	28
2	Uttar Pradesh	28
3	Tripura	28
4	Telangana	28
5	Tamil Nadu	28
6	Rajasthan	28
7	Punjab	28
8	Odisha	28
9	Maharashtra	28
10	Kerala	28
11	Madhya Pradesh	28
12	Jharkhand	28
13	Himachal Pradesh	28
14	Haryana	28
15	Gujarat	28
16	Delhi	28
17	Chhattisgarh	28
18	Bihar	28
19	West Bengal	28
20	Meghalaya	27
21	Uttarakhand	27
22	Assam	26
23	Puducherry	26
24	Goa	24
25	Jammu & Kashmir	21
26	Sikkim	17
27	Chandigarh	12


```
df.loc[df['Region'].isna()]
```



quency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 16	Unnamed: 17	Unnamed: 18	Unnamed: 19	Un
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	

```
df.dropna(how='all', inplace=True)
```

```
df.isna().sum().reset_index().rename(columns={0: 'count'})
```



	index	count
0	Region	0
1	Date	0
2	Frequency	0
3	Estimated Unemployment Rate (%)	0
4	Estimated Employed	0
5	Estimated Labour Participation Rate (%)	0
6	Area	0
7	Unnamed: 7	740
8	Unnamed: 8	740
9	Unnamed: 9	740
10	Unnamed: 10	740
11	Unnamed: 11	740
12	Unnamed: 12	740
13	Unnamed: 13	740
14	Unnamed: 14	740
15	Unnamed: 15	740
16	Unnamed: 16	740
17	Unnamed: 17	740
18	Unnamed: 18	740
19	Unnamed: 19	740
20	Unnamed: 20	740
21	Unnamed: 21	740
22	Unnamed: 22	740
23	Unnamed: 23	740
24	Unnamed: 24	740
25	Unnamed: 25	740

```
df_c['Region'].value_counts().reset_index()
```



	Region	count
0	Andhra Pradesh	0
1	Assam	0
2	Uttarakhand	0
3	Uttar Pradesh	0
4	Tripura	0
5	Telangana	0
6	Tamil Nadu	0
7	Sikkim	0
8	Rajasthan	0
9	Punjab	0
10	Puducherry	0
11	Odisha	0
12	Meghalaya	0
13	Maharashtra	0
14	Madhya Pradesh	0
15	Kerala	0
16	Karnataka	0
17	Jharkhand	0
18	Jammu & Kashmir	0
19	Himachal Pradesh	0
20	Haryana	0
21	Gujarat	0
22	Goa	0
23	Delhi	0
24	Chhattisgarh	0
25	Chandigarh	0
26	Bihar	0
27	West Bengal	0

df.dtypes

```
↔ Region category
   Date object
   Frequency category
   Estimated Unemployment Rate (%) float64
   Estimated Employed float64
   Estimated Labour Participation Rate (%) float64
   Area object
   Unnamed: 7 float64
   Unnamed: 8 float64
   Unnamed: 9 float64
   Unnamed: 10 float64
   Unnamed: 11 float64
   Unnamed: 12 float64
   Unnamed: 13 float64
   Unnamed: 14 float64
   Unnamed: 15 float64
   Unnamed: 16 float64
   Unnamed: 17 float64
   Unnamed: 18 float64
   Unnamed: 19 float64
   Unnamed: 20 float64
   Unnamed: 21 float64
   Unnamed: 22 float64
   Unnamed: 23 float64
   Unnamed: 24 float64
   Unnamed: 25 float64
dtype: object
```

df.dropna()

```
↔
```

Region	Date	Frequency	Estimated Unemployment Rate (%)	Estimated Employed	Estimated Labour Participation Rate (%)	Area	Unnamed: 7	Unnamed: 8	Unnamed: 9	...	Unnamed: 16	Unnamed: 17	U
--------	------	-----------	---------------------------------------	-----------------------	--	------	---------------	---------------	---------------	-----	----------------	----------------	---

0 rows × 26 columns

df.drop_duplicates()