Data Loading and Cleaning

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

df=pd.read_csv('/content/Unemployment in India.csv')

df
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 768,\n  \"fields\": [\
n    {\n      \"column\": \"Region\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n
\"samples\": [\n          \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Date\",\n      \"properties\": {\
n        \"dtype\": \"object\",\n        \"num_unique_values\": 14,\n
\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-
2020\",\n          \" 31-05-2019\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Monthly\",\n          \" Monthly\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Estimated Unemployment Rate
(%)\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 10.721298373157783,\n        \"min\": 0.0,\n        \"max\":
76.74,\n        \"num_unique_values\": 624,\n        \"samples\": [\n
22.01,\n          5.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8087988.4294577865,\n        \"min\":
49420.0,\n        \"max\": 45777509.0,\n        \"num_unique_values\":
740,\n        \"samples\": [\n          1732050.0,\n
2392400.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Labour Participation Rate (%)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8.111094365865872,\n
\"min\": 13.33,\n        \"max\": 72.57,\n
\"num_unique_values\": 626,\n        \"samples\": [\n          47.8,\n
38.2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Area\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
```

\"num_unique_values\": 2,\n   \"samples\": [\n
\"Urban\",\n   \"Rural\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
df.head()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 768,\n \"fields\": [\
n   {\n   \"column\": \"Region\",\n   \"properties\": {\n
\"dtype\": \"category\",\n   \"num_unique_values\": 28,\n
\"samples\": [\n   \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n   \"Himachal Pradesh\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \" Date\",\n   \"properties\": {\
n   \"dtype\": \"object\",\n   \"num_unique_values\": 14,\n
\"samples\": [\n   \" 29-02-2020\",\n   \" 30-04-
2020\",\n   \" 31-05-2019\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \" Frequency\",\n
\"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n
\"Monthly\",\n   \" Monthly\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \" Estimated Unemployment Rate
(%)\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 10.721298373157783,\n   \"min\": 0.0,\n   \"max\":
76.74,\n   \"num_unique_values\": 624,\n   \"samples\": [\n
22.01,\n   5.8\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"
Estimated Employed\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 8087988.4294577865,\n   \"min\":
49420.0,\n   \"max\": 45777509.0,\n   \"num_unique_values\":
740,\n   \"samples\": [\n   1732050.0,\n
2392400.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"
Estimated Labour Participation Rate (%)\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 8.111094365865872,\n
\"min\": 13.33,\n   \"max\": 72.57,\n
\"num_unique_values\": 626,\n   \"samples\": [\n   47.8,\n
38.2\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Area\",\n   \"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n
\"Urban\",\n   \"Rural\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

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

```
array(['Andhra Pradesh', 'Assam', 'Bihar', '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', nan,
'Chandigarh'],
      dtype=object)
```

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

```
array(['Rural', nan, 'Urban'], dtype=object)
```

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

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

```
df.dtypes
```

```
Region                                       object
 Date                                        object
 Frequency                                   object
 Estimated Unemployment Rate (%)            float64
 Estimated Employed                         float64
 Estimated Labour Participation Rate (%)    float64
Area                                         object
dtype: object
```

```
df=df.dropna()
```

```
df
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 740,\n  \"fields\": [\
n    {\n      \"column\": \"Region\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n
\"samples\": [\n          \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Date\",\n      \"properties\": {\
n        \"dtype\": \"object\",\n        \"num_unique_values\": 14,\n
\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-
2020\",\n          \" 31-05-2019\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \" Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Monthly\",\n          \" Monthly\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Estimated Unemployment Rate
(%)\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 10.721298373157783,\n        \"min\": 0.0,\n        \"max\":
76.74,\n        \"num_unique_values\": 624,\n        \"samples\": [\n
22.01,\n          5.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8087988.4294577865,\n        \"min\":
49420.0,\n        \"max\": 45777509.0,\n        \"num_unique_values\":
740,\n        \"samples\": [\n          1732050.0,\n
2392400.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Labour Participation Rate (%)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8.111094365865872,\n
\"min\": 13.33,\n        \"max\": 72.57,\n
\"num_unique_values\": 626,\n        \"samples\": [\n          47.8,\n
38.2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Area\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Urban\",\n          \"Rural\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


```
df.head(10)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 740,\n  \"fields\": [\
n    {\n      \"column\": \"Region\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n
\"samples\": [\n          \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Date\",\n      \"properties\": {\
n        \"dtype\": \"object\",\n        \"num_unique_values\": 14,\n
\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-
2020\",\n          \" 31-05-2019\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Monthly\",\n          \" Monthly\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" Estimated Unemployment Rate
(%)\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 10.721298373157783,\n        \"min\": 0.0,\n        \"max\":
```

```
76.74,\n        \"num_unique_values\": 624,\n        \"samples\": [\n
22.01,\n          5.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8087988.4294577865,\n        \"min\":
49420.0,\n        \"max\": 45777509.0,\n        \"num_unique_values\":
740,\n        \"samples\": [\n          1732050.0,\n
2392400.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
Estimated Labour Participation Rate (%)\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8.111094365865872,\n
\"min\": 13.33,\n        \"max\": 72.57,\n
\"num_unique_values\": 626,\n        \"samples\": [\n          47.8,\n
38.2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Area\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Urban\",\n          \"Rural\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


Changing names of columns:

since column names contain some % sign i.e two columns so it might create problem while visualization

```
header=['Region','Date','Frequency','Estimated Unemployment
rate','Estimated employed','Estimated labour participation
rate','Area']

df.columns=header

df.head(10)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 740,\n  \"fields\": [\
n    {\n      \"column\": \"Region\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n
\"samples\": [\n          \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Date\",\n      \"properties\": {\n
\"dtype\": \"object\",\n        \"num_unique_values\": 14,\n
\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-
2020\",\n          \" 31-05-2019\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
```

\"Monthly\",\n   \" Monthly\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Estimated Unemployment rate\",\n
\"properties\": {\n   \"dtype\": \"number\",\n   \"std\":
10.721298373157783,\n   \"min\": 0.0,\n   \"max\": 76.74,\n
\"num_unique_values\": 624,\n   \"samples\": [\n   22.01,\
n   5.8\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Estimated employed\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 8087988.4294577865,\n   \"min\":
49420.0,\n   \"max\": 45777509.0,\n   \"num_unique_values\":
740,\n   \"samples\": [\n   1732050.0,\n
2392400.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Estimated labour participation rate\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 8.111094365865872,\n
\"min\": 13.33,\n   \"max\": 72.57,\n
\"num_unique_values\": 626,\n   \"samples\": [\n   47.8,\n
38.2\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Area\",\n   \"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n
\"Urban\",\n   \"Rural\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

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

```
Region                                0
Date                                  0
Frequency                             0
Estimated Unemployment rate           0
Estimated employed                    0
Estimated labour participation rate    0
Area                                  0
dtype: int64
```

```
df.head()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 740,\n \"fields\": [\
n   {\n   \"column\": \"Region\",\n   \"properties\": {\n
\"dtype\": \"category\",\n   \"num_unique_values\": 28,\n
\"samples\": [\n   \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n   \"Himachal Pradesh\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Date\",\n   \"properties\": {\n
\"dtype\": \"object\",\n   \"num_unique_values\": 14,\n
\"samples\": [\n   \" 29-02-2020\",\n   \" 30-04-
2020\",\n   \" 31-05-2019\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\

```
n    },\n    {\n      \"column\": \"Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Monthly\",\n          \" Monthly\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Estimated Unemployment rate\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
10.721298373157783,\n        \"min\": 0.0,\n        \"max\": 76.74,\n
\"num_unique_values\": 624,\n        \"samples\": [\n          22.01,\
n          5.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8087988.4294577865,\n        \"min\":
49420.0,\n        \"max\": 45777509.0,\n        \"num_unique_values\":
740,\n        \"samples\": [\n          1732050.0,\n
2392400.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated labour participation rate\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8.111094365865872,\n
\"min\": 13.33,\n        \"max\": 72.57,\n
\"num_unique_values\": 626,\n        \"samples\": [\n          47.8,\n
38.2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Area\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Urban\",\n          \"Rural\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


Observing data

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

```
array(['Rural', 'Urban'], dtype=object)
```

```
df['Area'].value_counts()
```

```
Area
Urban    381
Rural    359
Name: count, dtype: int64
```

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

```
array(['Andhra Pradesh', 'Assam', 'Bihar', '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', 'Chandigarh'],
      dtype=object)
```

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

```
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
```

```
df.head(10)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 740,\n  \"fields\": [\
n    {\n      \"column\": \"Region\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n
\"samples\": [\n          \"Jammu & Kashmir\",\n
\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Date\",\n      \"properties\": {\n
\"dtype\": \"object\",\n        \"num_unique_values\": 14,\n
\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-
2020\",\n          \" 31-05-2019\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \"Frequency\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Monthly\",\n          \" Monthly\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Estimated Unemployment rate\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
10.721298373157783,\n        \"min\": 0.0,\n        \"max\": 76.74,\n
\"num_unique_values\": 624,\n        \"samples\": [\n          22.01,\
n          5.8\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 8087988.4294577865,\n        \"min\":
49420.0,\n        \"max\": 45777509.0,\n        \"num_unique_values\":
740,\n        \"samples\": [\n          1732050.0,\n
2392400.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated labour participation rate\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 8.111094365865872,\n
\"min\": 13.33,\n        \"max\": 72.57,\n
\"num_unique_values\": 626,\n        \"samples\": [\n          47.8,\n
38.2\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Area\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Urban\",\n          \"Rural\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```


Descriptive statistics

```
df.describe()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"Estimated Unemployment rate\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
256.313829221936,\n        \"min\": 0.0,\n        \"max\": 740.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
11.787945945945946,\n          8.35,\n          740.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Estimated employed\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
15104451.557182284,\n        \"min\": 740.0,\n        \"max\":
45777509.0,\n        \"num_unique_values\": 8,\n        \"samples\":
[\n          7204460.025675676,\n          4744178.5,\n
740.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated labour participation rate\",\n      \"properties\": {\n
```

```
\"dtype\": \"number\",\n        \"std\": 249.22691963141102,\n
\"min\": 8.111094365865872,\n        \"max\": 740.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
42.630121621621626,\n          41.16,\n          740.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```

```
np.median(df['Estimated Unemployment rate'])
```

```
8.35
```

```
np.median(df['Estimated employed'])
```

```
4744178.5
```

```
np.median(df['Estimated labour participation rate'])
```

```
41.16
```

```
from scipy import stats
```

so the mode is defined as the value that occurs more frequently within the dataset and here in Estimated Unemployment rate it is 0.00 so it is possible that most of the states or region may have 0 percent of unemployment

```
stats.mode(df['Estimated Unemployment rate'])
```

```
ModeResult(mode=0.0, count=11)
```

Here are the region with 0 rate of unemployement at somepoint

```
for index in df.index:
  if df.loc[index, 'Estimated Unemployment rate'] == 0.00:
    print(df.loc[index, 'Region'])
```

```
Assam
Chhattisgarh
Goa
Puducherry
Puducherry
Puducherry
Puducherry
Puducherry
Sikkim
Puducherry
Puducherry
```

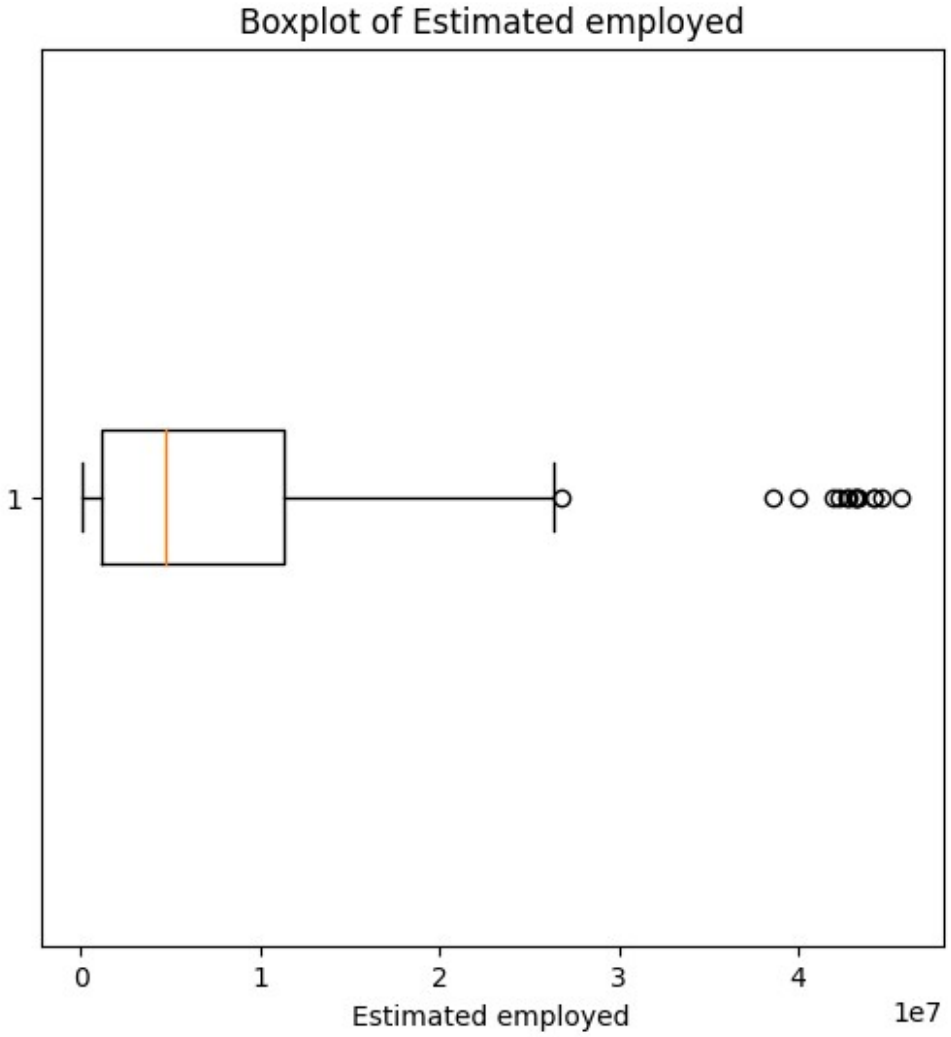
```
stats.mode(df['Estimated employed'])
```

```
ModeResult(mode=49420.0, count=1)
```

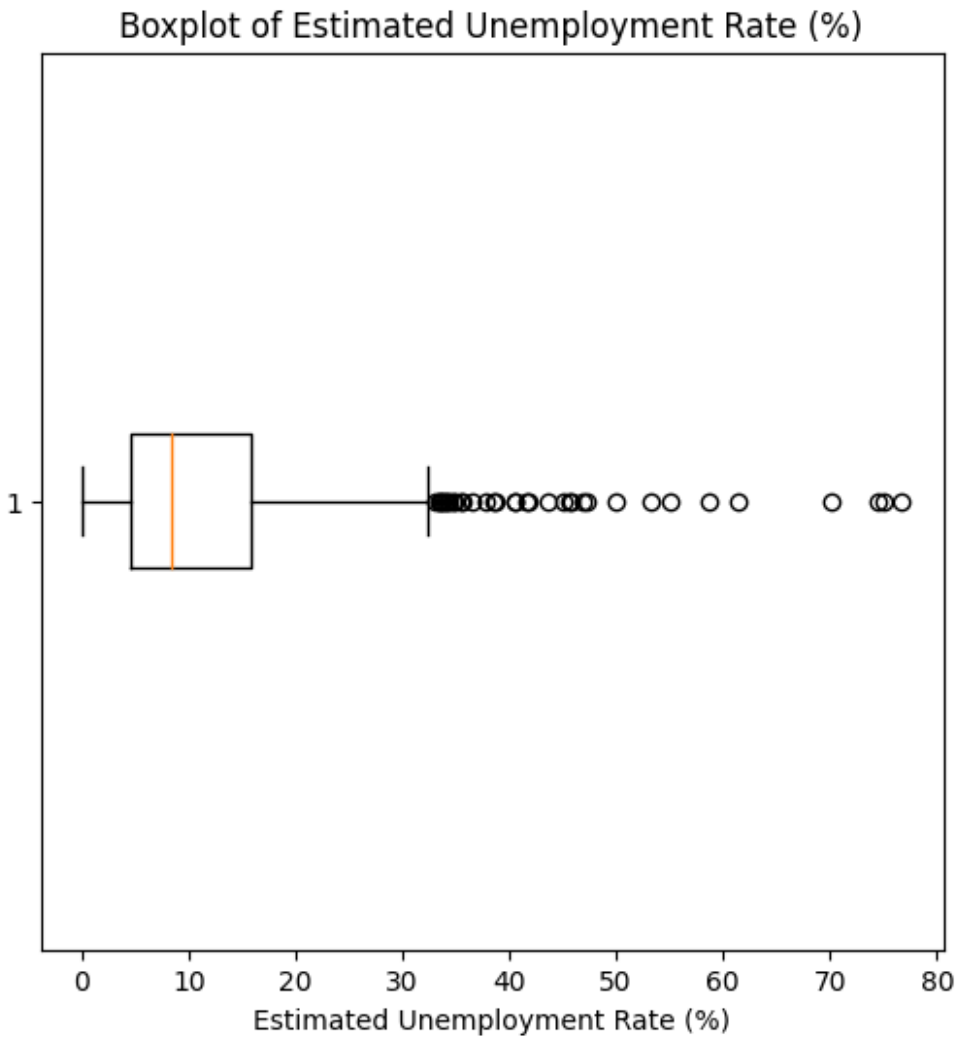
```
stats.mode(df['Estimated labour participation rate'])
```

```
ModeResult(mode=39.04, count=3)
```

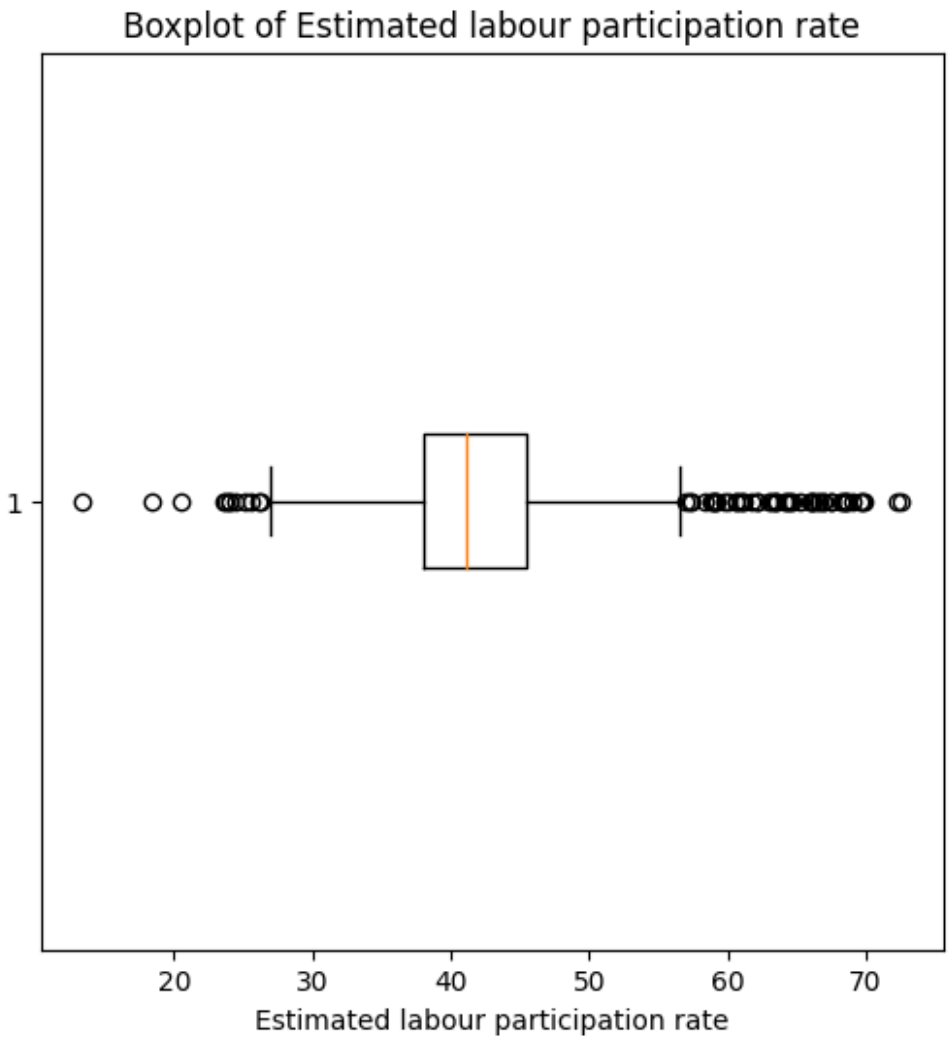
```
# Create the boxplot
plt.figure(figsize=(6, 6))
plt.boxplot(df['Estimated employed'], vert=False)
plt.xlabel('Estimated employed')
plt.title('Boxplot of Estimated employed')
plt.show()
```




```
plt.figure(figsize=(6, 6))
plt.boxplot(df['Estimated Unemployment rate'], vert=False)
plt.xlabel('Estimated Unemployment Rate (%)')
plt.title('Boxplot of Estimated Unemployment Rate (%)')
plt.show()
```




```
plt.figure(figsize=(6, 6))
plt.boxplot(df['Estimated labour participation rate'], vert=False)
plt.xlabel('Estimated labour participation rate')
plt.title('Boxplot of Estimated labour participation rate')
plt.show()
```



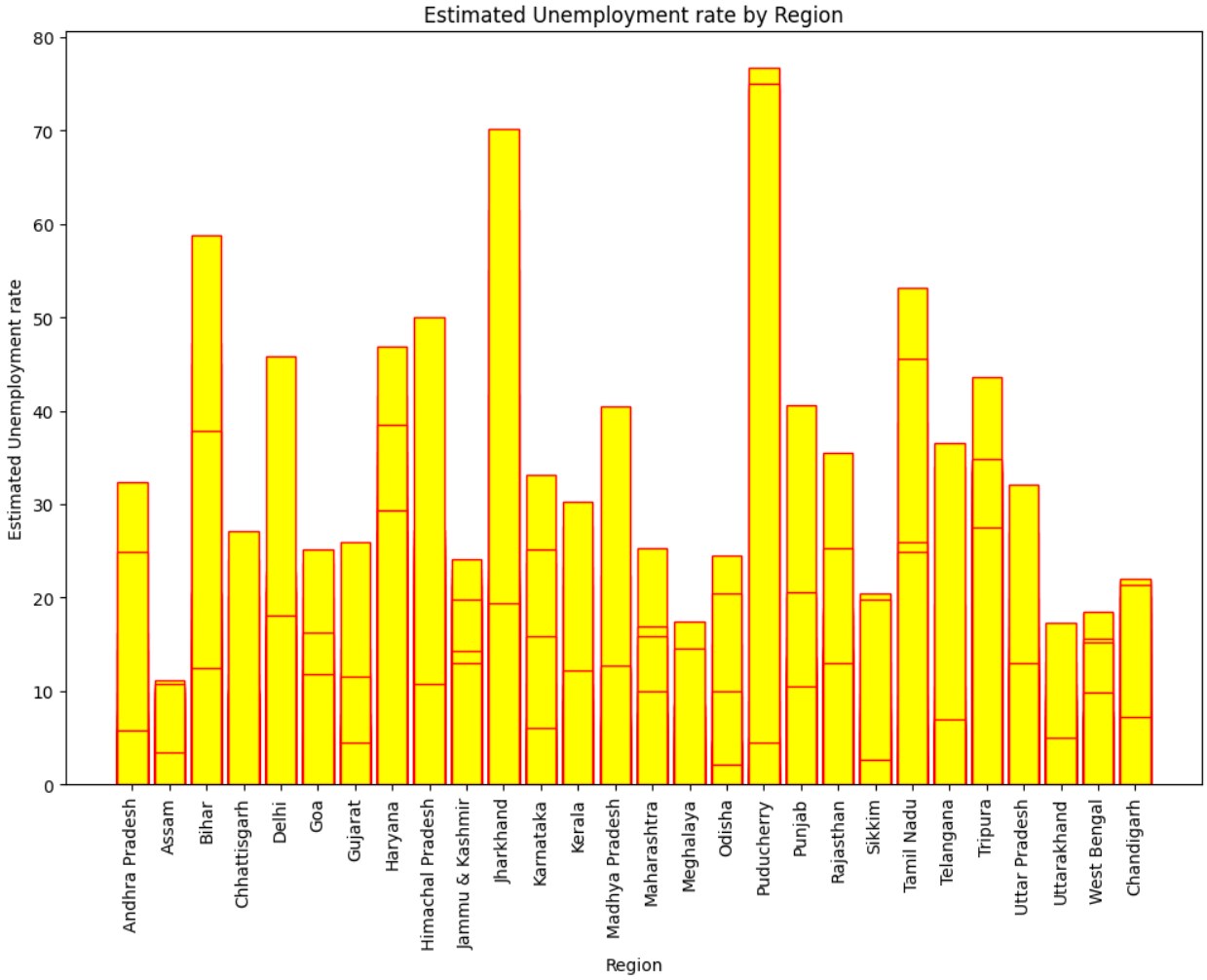
Visualization

```
df.head(10)
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 740,\n  \"fields\": [\n    {\n      \"column\": \"Region\",\n      \"properties\": {\n\"dtype\": \"category\",\n        \"num_unique_values\": 28,\n\"samples\": [\n          \"Jammu & Kashmir\",\n\"Uttarakhand\",\n          \"Himachal Pradesh\"\n        ],\n\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Date\",\n      \"properties\": {\n\"dtype\": \"object\",\n        \"num_unique_values\": 14,\n\"samples\": [\n          \" 29-02-2020\",\n          \" 30-04-2020\",\n          \" 31-05-2019\"\n        ],\n\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Frequency\",\n\"properties\": {\n        \"dtype\": \"category\",\n
```

\"num_unique_values\": 2,\n   \"samples\": [\n
\"Monthly\",\n   \" Monthly\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"Estimated Unemployment rate\",\n
\"properties\": {\n   \"dtype\": \"number\",\n   \"std\":
10.721298373157783,\n   \"min\": 0.0,\n   \"max\": 76.74,\n
\"num_unique_values\": 624,\n   \"samples\": [\n   22.01,\
n   5.8\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Estimated employed\",\n   \"properties\": {\n   \"dtype\":
\"number\",\n   \"std\": 8087988.4294577865,\n   \"min\":
49420.0,\n   \"max\": 45777509.0,\n   \"num_unique_values\":
740,\n   \"samples\": [\n   1732050.0,\n
2392400.0\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Estimated labour participation rate\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 8.111094365865872,\n
\"min\": 13.33,\n   \"max\": 72.57,\n
\"num_unique_values\": 626,\n   \"samples\": [\n   47.8,\n
38.2\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"Area\",\n   \"properties\": {\n   \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n
\"Urban\",\n   \"Rural\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df"}

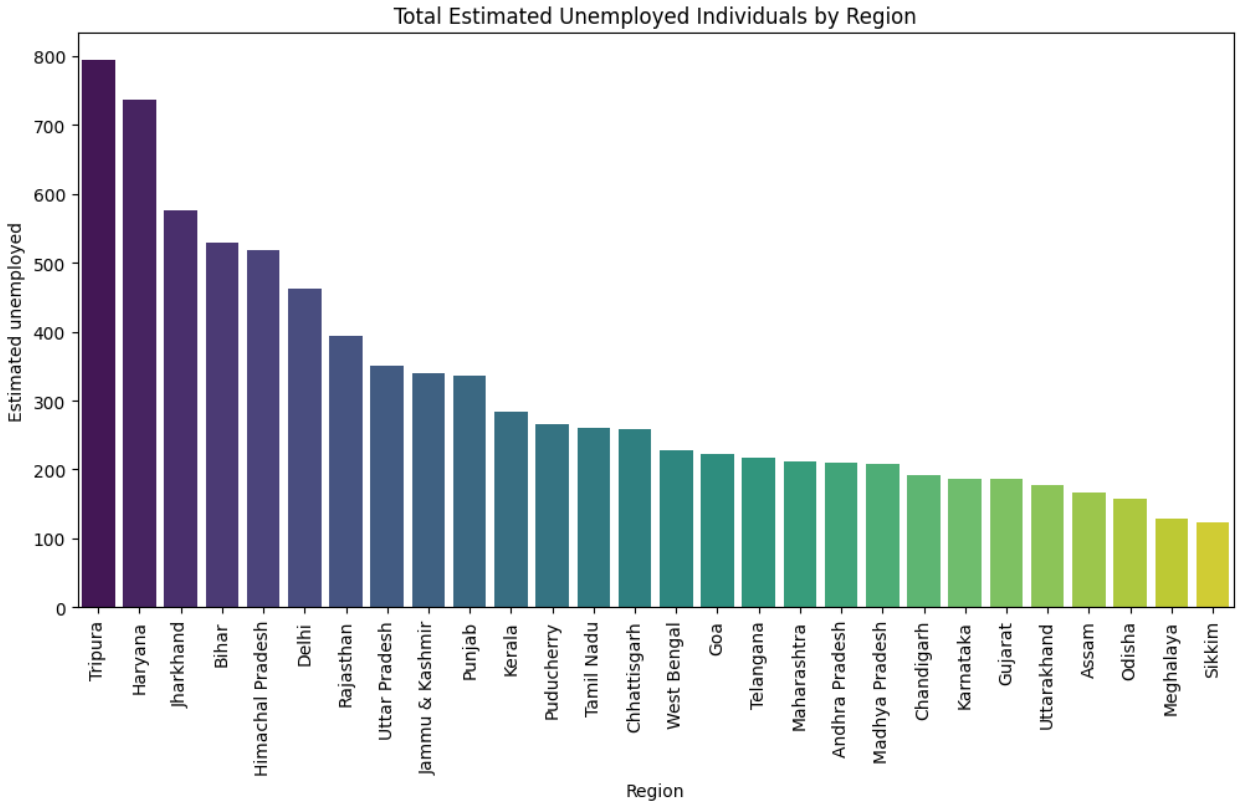
```
plt.figure(figsize=(12,8))
plt.bar(df['Region'],df['Estimated Unemployment
rate'],color='yellow',edgecolor='red')
plt.xlabel('Region')
plt.ylabel('Estimated Unemployment rate')
plt.title('Estimated Unemployment rate by Region')
plt.xticks(rotation=90)
plt.show()
```



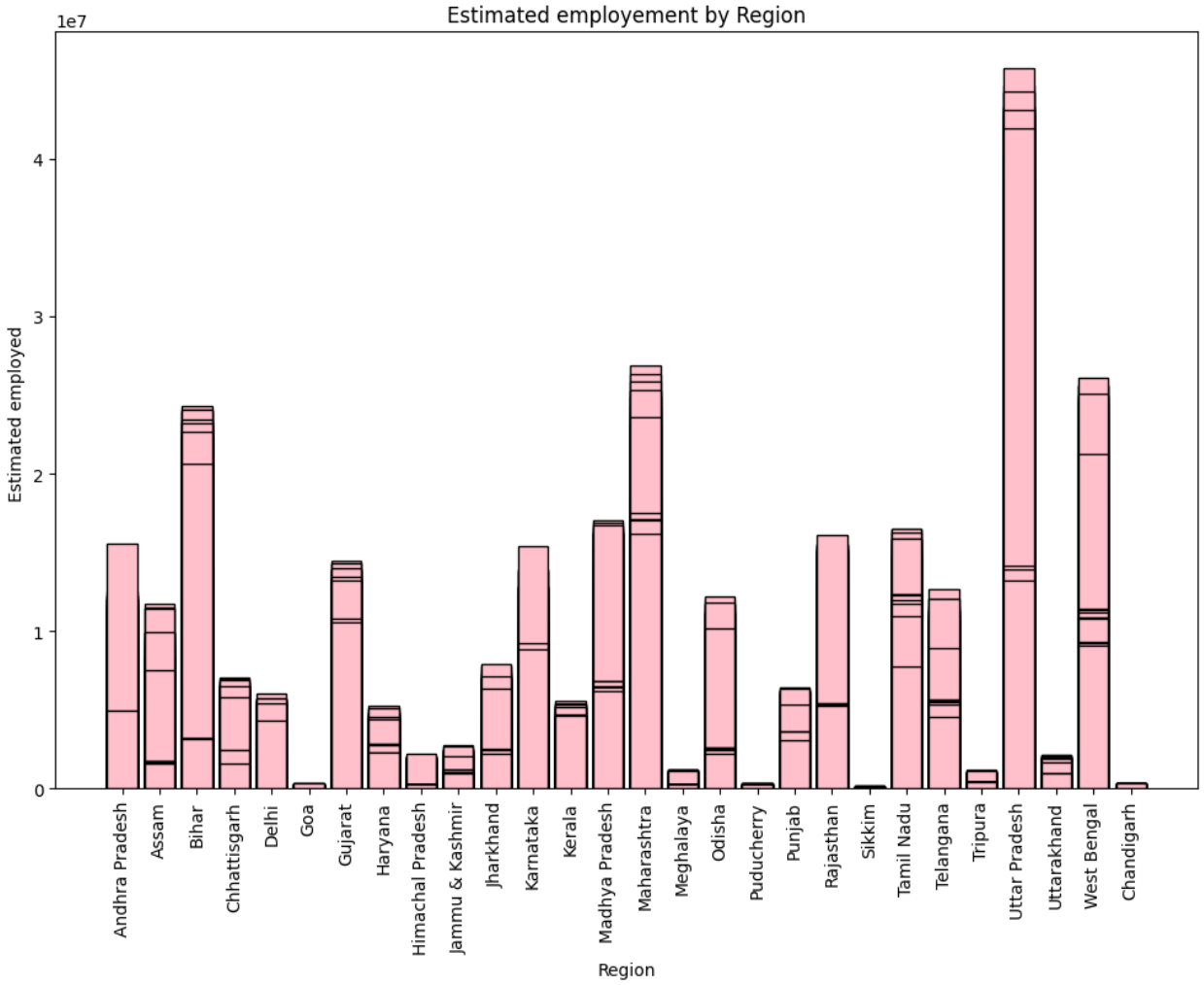

```
unemployed_by_region = df.groupby('Region')['Estimated Unemployment
rate'].sum().reset_index()

unemployed_by_region
```


```
{"summary":"{\n  \"name\": \"unemployed_by_region\",\n  \"rows\": 28,\
n  \"fields\": [\n    {\n      \"column\": \"Region\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 28,\n        \"samples\": [\n
\"Himachal Pradesh\",\n          \"Uttar Pradesh\",\n
\"Haryana\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated Unemployment rate\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 176.98440596120085,\n
\"min\": 123.24,\n        \"max\": 793.81,\n
\"num_unique_values\": 28,\n        \"samples\": [\n          519.13,\
n          351.44,\n          735.93\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    }\n  ]\
n}","type":"dataframe","variable_name":"unemployed_by_region"}

plt.figure(figsize=(12, 6))
unemployed_by_region = unemployed_by_region.sort_values(by=['Estimated
Unemployment rate'], ascending=False)
sns.barplot(data=unemployed_by_region, x='Region', y='Estimated
Unemployment rate', palette='viridis')
plt.title('Total Estimated Unemployed Individuals by Region')
plt.xlabel('Region')
plt.ylabel('Estimated unemployed')
plt.xticks(rotation=90)
plt.show()
```



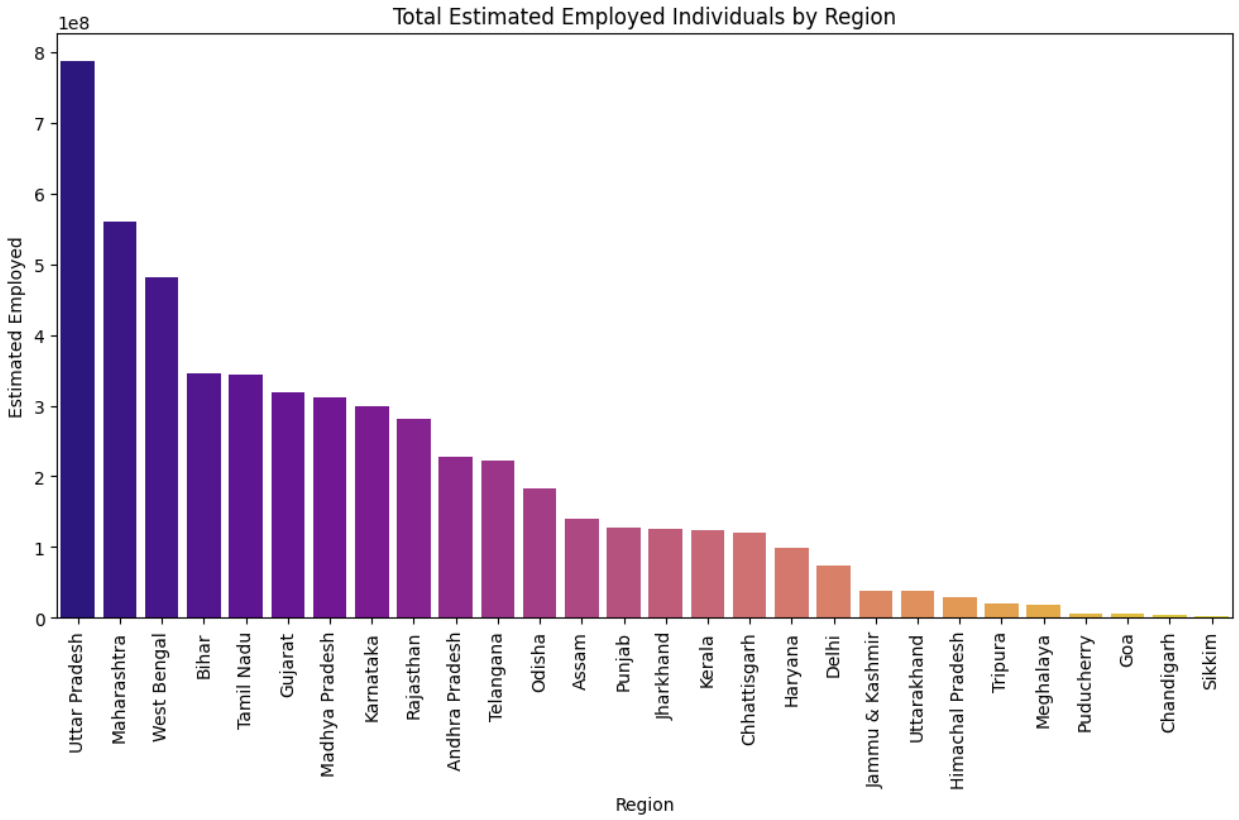

```
plt.figure(figsize=(12,8))
plt.bar(df['Region'],df['Estimated
employed'],color='pink',edgecolor='black')
plt.xlabel('Region')
plt.ylabel('Estimated employed ')
plt.title('Estimated employement by Region')
plt.xticks(rotation=90)
plt.show()
```



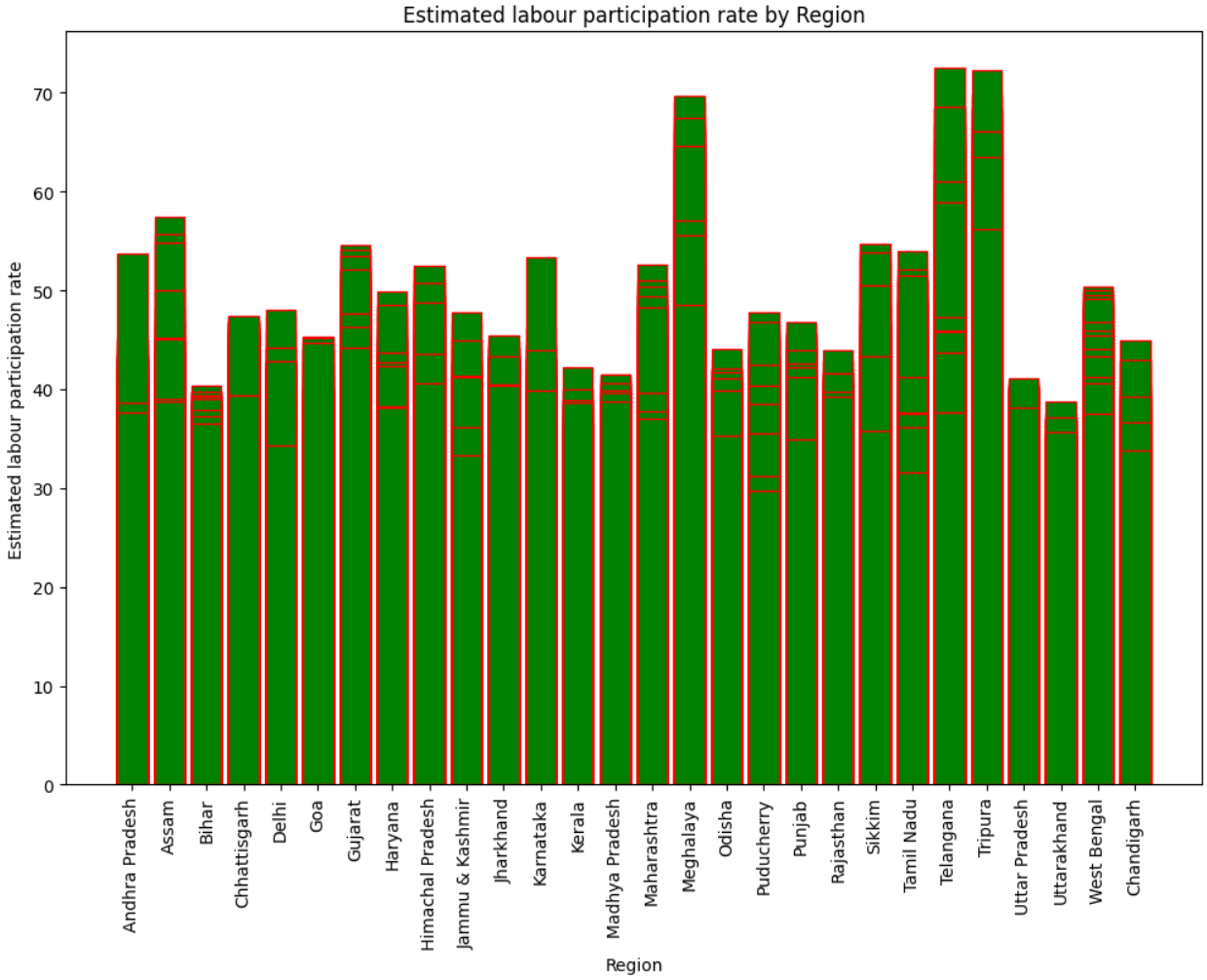

```
employed_by_region = df.groupby('Region')['Estimated
employed'].sum().reset_index()

employed_by_region
```


```
{"summary":"{\n  \"name\": \"employed_by_region\",\n  \"rows\": 28,\n
\"fields\": [\n    {\n      \"column\": \"Region\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 28,\n        \"samples\": [\n
\"Himachal Pradesh\",\n          \"Uttar Pradesh\",\n
\"Haryana\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated employed\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 191483110.51178244,\n        \"min\":
1816972.0,\n        \"max\": 786655301.0,\n
\"num_unique_values\": 28,\n        \"samples\": [\n
29675064.0,\n          786655301.0,\n          99598029.0\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
```

```
}\n    }\n  ]\
n}","type":"dataframe","variable_name":"employed_by_region"}

plt.figure(figsize=(12, 6))
employed_by_region=employed_by_region.sort_values(by=['Estimated
employed'], ascending=False)
sns.barplot(data=employed_by_region, x='Region', y='Estimated
employed', palette='plasma')
plt.title('Total Estimated Employed Individuals by Region')
plt.xlabel('Region')
plt.ylabel('Estimated Employed')
plt.xticks(rotation=90)
plt.show()
```



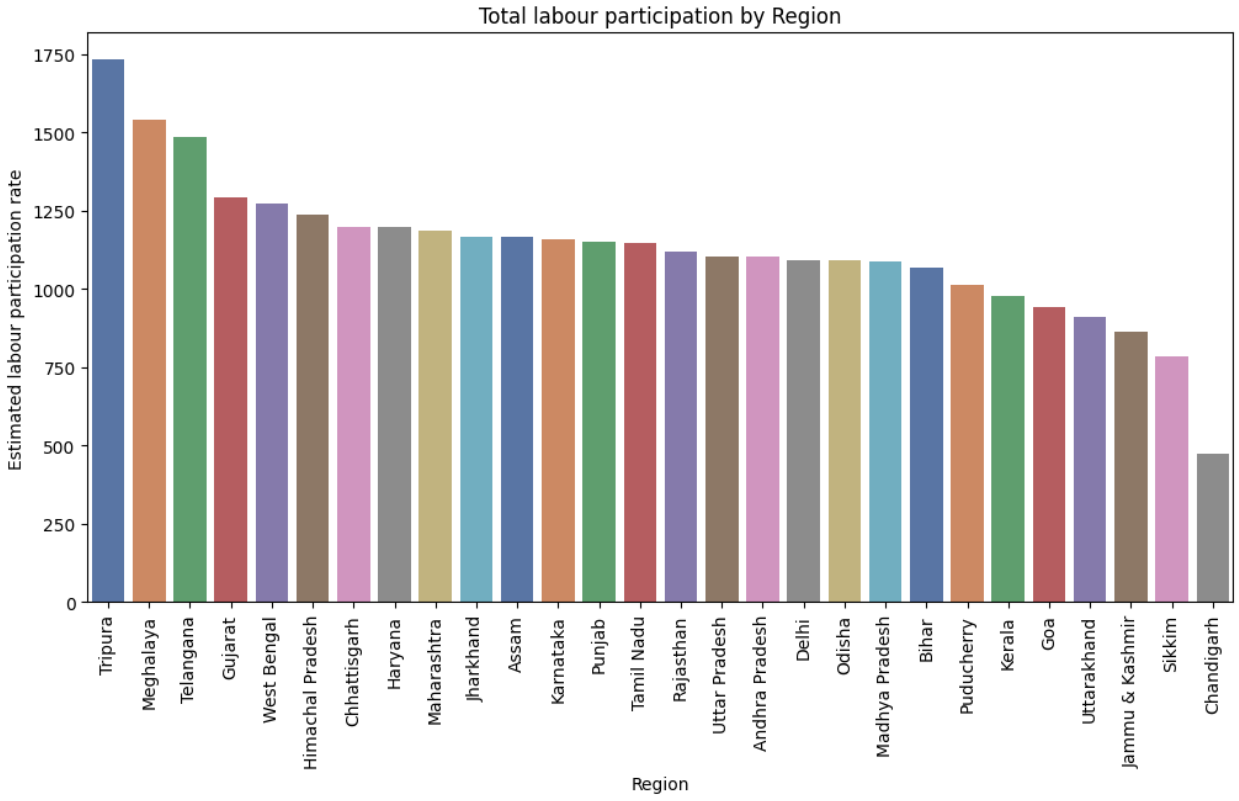

```
plt.figure(figsize=(12,8))
plt.bar(df['Region'],df['Estimated labour participation
rate'],color='green',edgecolor='red')
plt.xlabel('Region')
plt.ylabel('Estimated labour participation rate')
plt.title('Estimated labour participation rate by Region')
plt.xticks(rotation=90)
plt.show()
```




```
labour_parti_by_region = df.groupby('Region')['Estimated labour
participation rate'].sum().reset_index()

labour_parti_by_region
```


```
{"summary":"{\n  \"name\": \"labour_parti_by_region\",\n  \"rows\":
28,\n  \"fields\": [\n    {\n      \"column\": \"Region\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 28,\n        \"samples\": [\n
\"Himachal Pradesh\",\n          \"Uttar Pradesh\",\n
\"Haryana\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Estimated labour participation rate\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 234.51703804570397,\n
\"min\": 472.04,\n        \"max\": 1731.07,\n
\"num_unique_values\": 28,\n        \"samples\": [\n
1238.22,\n          1104.1100000000001,\n          1196.64\
n        ],\n        \"semantic_type\": \"\",\n
```

```
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"labour_parti_by_region"}

plt.figure(figsize=(12, 6))
labour_parti_by_region=labour_parti_by_region.sort_values(by=['Estimat
ed labour participation rate'],ascending=False)
sns.barplot(data=labour_parti_by_region, x='Region', y='Estimated
labour participation rate', palette='deep')
plt.title('Total labour participation by Region')
plt.xlabel('Region')
plt.ylabel('Estimated labour participation rate')
plt.xticks(rotation=90)
plt.show()
```

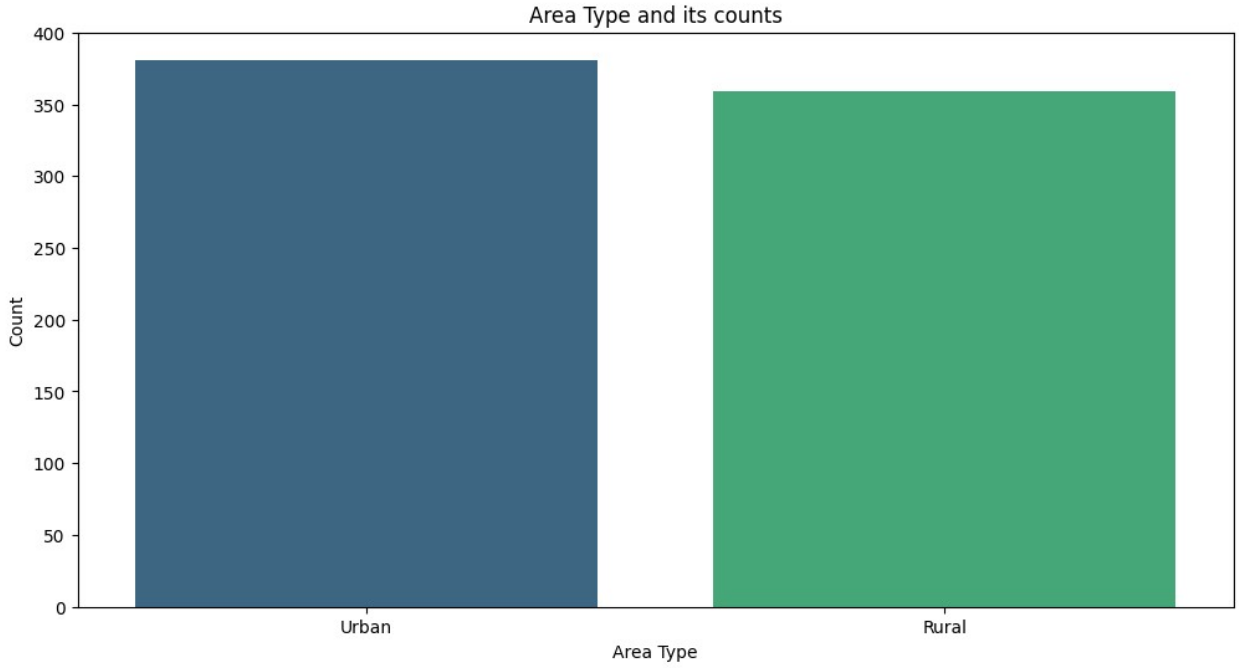



```
area_counts=df['Area'].value_counts()
area_counts
```

```
Area
Urban    381
Rural    359
Name: count, dtype: int64
```

```
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='Area',
order=area_counts.index,palette='viridis')
```

```
plt.title('Area Type and its counts')
plt.xlabel('Area Type')
plt.ylabel('Count')
plt.show()
```



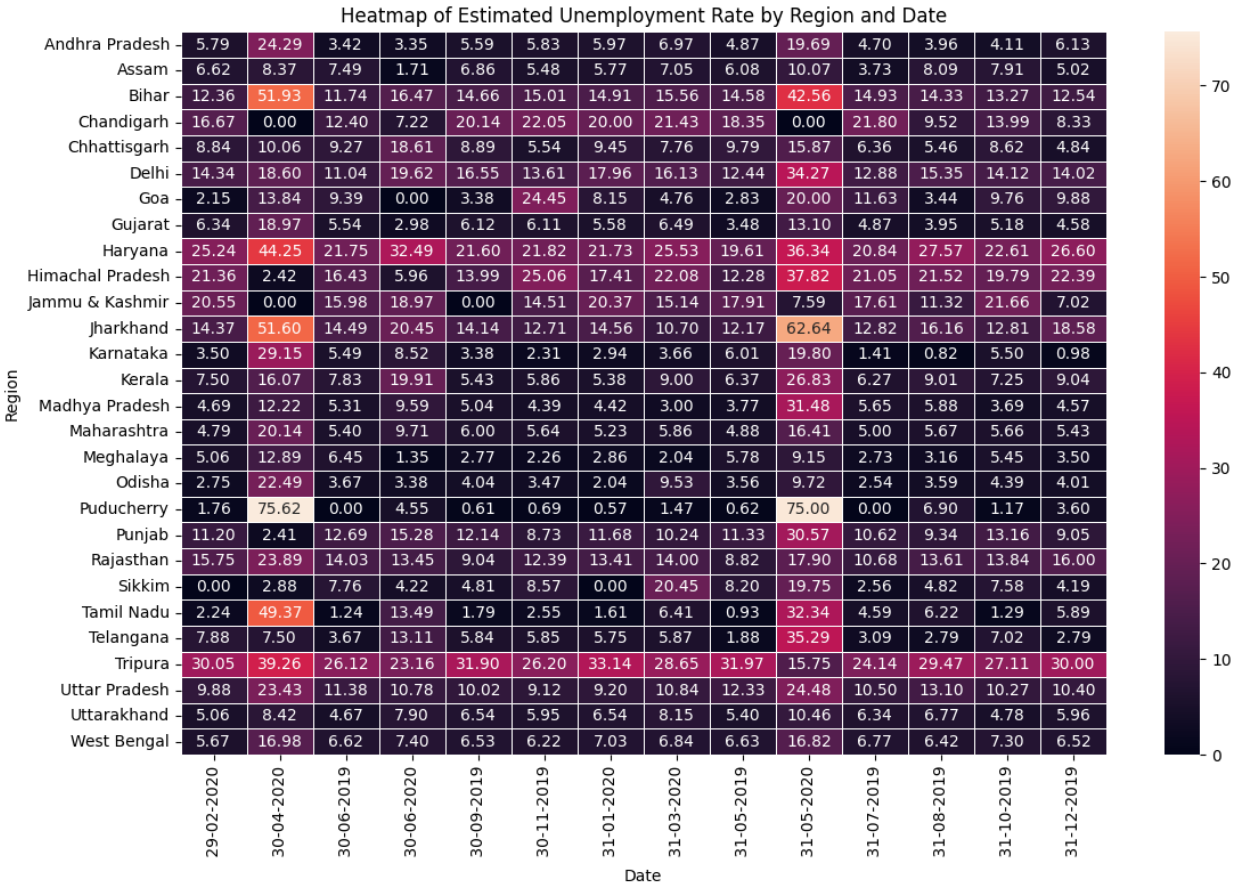

```
heatmap_data = df.pivot_table(values='Estimated Unemployment
rate',index='Region',columns='Date')
heatmap_data.fillna(0, inplace=True)

heatmap_data
```


```
{"summary":"{\n  \"name\": \"heatmap_data\",\n  \"rows\": 28,\n
\"fields\": [\n    {\n      \"column\": \"Region\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 28,\n        \"samples\": [\n
\"Himachal Pradesh\",\n          \"Uttar Pradesh\",\n
\"Haryana\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
29-02-2020\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 7.593993832499624,\n        \"min\":
0.0,\n        \"max\": 30.05,\n        \"num_unique_values\": 27,\n
\"samples\": [\n          25.240000000000002,\n          7.495,\n
21.365\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
30-04-2020\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 18.661212768841374,\n        \"min\":
0.0,\n        \"max\": 75.625,\n        \"num_unique_values\": 27,\n
\"samples\": [\n          44.25,\n          12.219999999999999,\n
2.415\n        ],\n        \"semantic_type\": \"\",\n
```

```
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
30-06-2019\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 5.995686467530738,\n        \"min\":
0.0,\n        \"max\": 26.115000000000002,\n
\"num_unique_values\": 28,\n        \"samples\": [\n          16.43,\n
11.375,\n          21.75\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \" 30-06-2020\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 7.885441299682646,\n
\"min\": 0.0,\n        \"max\": 32.49,\n        \"num_unique_values\":
28,\n        \"samples\": [\n          5.965,\n
10.780000000000001,\n          32.49\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" 30-09-2019\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
7.173291991088749,\n        \"min\": 0.0,\n        \"max\": 31.9,\n
\"num_unique_values\": 27,\n        \"samples\": [\n          21.6,\n
5.04,\n          13.99\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
30-11-2019\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 7.57960837129263,\n        \"min\":
0.685,\n        \"max\": 26.2,\n        \"num_unique_values\": 28,\n
\"samples\": [\n          25.064999999999998,\n          9.125,\n
21.825\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"
31-01-2020\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 7.821747043901579,\n        \"min\":
0.0,\n        \"max\": 33.14,\n        \"num_unique_values\": 28,\n
\"samples\": [\n          17.41,\n          9.2,\n          21.725\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \" 31-03-2020\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
7.235427619062181,\n        \"min\": 1.465,\n        \"max\": 28.65,\n
\"num_unique_values\": 28,\n        \"samples\": [\n
22.075000000000003,\n          10.844999999999999,\n          25.53\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \" 31-05-2019\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
6.904985833807291,\n        \"min\": 0.625,\n        \"max\": 31.97,\n
\"num_unique_values\": 28,\n        \"samples\": [\n
12.280000000000001,\n          12.33,\n          19.605\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \" 31-05-2020\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
16.435876524732212,\n        \"min\": 0.0,\n        \"max\": 75.0,\n
\"num_unique_values\": 28,\n        \"samples\": [\n          37.82,\n
24.475,\n          36.34\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \" 31-07-2019\",\n      \"properties\": {\n
```

\"dtype\": \"number\",\n   \"std\": 6.8415856166534965,\n   \"min\": 0.0,\n   \"max\": 24.14,\n   \"num_unique_values\": 28,\n   \"samples\": [\n   21.05,\n   10.495000000000001,\n   20.835\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \" 31-08-2019\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 7.202565705684953,\n   \"min\": 0.8200000000000001,\n   \"max\": 29.47,\n   \"num_unique_values\": 28,\n   \"samples\": [\n   21.52,\n   13.100000000000001,\n   27.565\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \" 31-10-2019\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 6.570272326102887,\n   \"min\": 1.1749999999999998,\n   \"max\": 27.105,\n   \"num_unique_values\": 28,\n   \"samples\": [\n   19.79,\n   10.265,\n   22.605\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \" 31-12-2019\",\n   \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 7.30591724691589,\n   \"min\": 0.985,\n   \"max\": 30.005,\n   \"num_unique_values\": 28,\n   \"samples\": [\n   22.39,\n   10.399999999999999,\n   26.604999999999997\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe","variable_name":"heatmap_data"}

```
plt.figure(figsize=(12, 8))
sns.heatmap(heatmap_data,annot=True, fmt=".2f", linewidths=.5)
plt.title('Heatmap of Estimated Unemployment Rate by Region and Date')
plt.xlabel('Date')
plt.ylabel('Region')
plt.tight_layout()
plt.show()
```



Unemployment rates vary significantly across regions. Some regions, such as Sikkim, Puducherry, and Kerala, consistently have lower unemployment rates, while others, like Bihar, Jharkhand, and Rajasthan, often experience higher rates.