

MACHINE LEARNING

GROUP ASSIGNMENT

UNEMPLOYEMENT ANALYSIS

SUBMITTED BY :-

ADITYA ANAND 2106005

ARPIT KUMAR 2106016

MOHD. SAUD KHAN 2106039

ISTIAK AHAMED TURJYA 2106305

SECTION : IT-1(ML)

SUBMITTED TO :-

PROF. MANOJ KUMAR BEURIA

#  Import Necessary Libraries

import numpy as np #linear algebra import pandas as pd #data processing

 Reading CSV File

 Code  Text

df=pd.read\_csv("Unemployment\_Rate\_upto\_11\_2020.csv") #read dataset

df.head() #returns first 5 entries

### Region Date Frequency

### Estimated Unemployment Rate

### Estimated Employed

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **(%)** |  | | | | | |
| **0** Andhra 31-01- M | 5.48 | 16635535 | 41.02 | South | 15.9129 | 79.74 |  |
| **1** Andhra 29-02- M | 5.83 | 16545652 | 40.90 | South | 15.9129 | 79.74 |  |
|  | **2** Andhra 31-03- M  Pradesh 2020 | 5.79 | 15881197 | 39.18 | South | 15.9129 | 79.74 |  |

### Estimated Labour Region.1 longitude latitude Participation Rate (%)

Pradesh 2020

Pradesh 2020

Next steps:

Generate code with df

 View recommended plots

df.tail() #returns last 5 entries

Region Date Frequency

Estimated Unemployment Rate

(%)

[Estimated Employed](#_TOC_250001)

[Estimated Labour Region.1 longitude latitude Participation Rate (%)](#_TOC_250000)

### 262

West Bengal

30-06- M 7.29 30726310 40.39 East 22.9868 87.855

2020

### 263

West Bengal

31-07- M 6.83 35372506 46.17 East 22.9868 87.855

2020

### 264

West Bengal

31-08- M 14.87 33298644 47.48 East 22.9868 87.855

2020

(267, 9)

#returns tuple of shape (Rows, columns) of dataframe df.shape

<class 'pandas.core.frame.DataFrame'> RangeIndex: 267 entries, 0 to 266 Data columns (total 9 columns):

#prints information about the dataframe df.info()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # |  | 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

#returns numerical description of the data in the dataframe df.describe()

### Estimated Unemployment

### Estimated Employed

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Rate (%)** |  | **Rate (%)** |  | |
| **count** | 267.000000 | 2.670000e+02 | 267.000000 | 267.000000 | 267.000000 |
| **mean** | 12.236929 | 1.396211e+07 | 41.681573 | 22.826048 | 80.532425 |
| **std** | 10.803283 | 1.336632e+07 | 7.845419 | 6.270731 | 5.831738 |
| **min** | 0.500000 | 1.175420e+05 | 16.770000 | 10.850500 | 71.192400 |
| **25%** | 4.845000 | 2.838930e+06 | 37.265000 | 18.112400 | 76.085600 |
| **50%** | 9.650000 | 9.732417e+06 | 40.390000 | 23.610200 | 79.019300 |
| **75%** | 16.755000 | 2.187869e+07 | 44.055000 | 27.278400 | 85.279900 |

**Estimated Labour Participation**

### longitude latitude

x = df['Region'] #plotting column 'Region' on x-axis

x #print x

1. Andhra Pradesh
2. Andhra Pradesh
3. Andhra Pradesh
4. Andhra Pradesh
5. Andhra Pradesh

...

1. West Bengal
2. West Bengal
3. West Bengal
4. West Bengal
5. West Bengal

Name: Region, Length: 267, dtype: object

y=df[' Estimated Unemployment Rate (%)'] #plotting column 'Estimated Unemployment Rate (%)' on y-axis

y #print y

|  |  |  |
| --- | --- | --- |
| 0 | | 5.48 |
| 1 | | 5.83 |
| 2 | | 5.79 |
| 3 | | 20.51 |
| 4  262 | | 17.43  ...  7.29 |
| 263 | | 6.83 |
| 264 | | 14.87 |
| 265 | | 9.35 |
| 266 | | 9.98 |
| Name: | | Estimated Unemployment Rate (%), Length: 267, dtype: float64 |
| df2=df.iloc[:,3] | | |
| df2 |  |  |
|  | 0 | 5.48 |
|  | 1 | 5.83 |
|  | 2 | 5.79 |
|  | 3 | 20.51 |
|  | 4 | 17.43  ... |
|  | 262 | 7.29 |
|  | 263 | 6.83 |
|  | 264 | 14.87 |
|  | 265 | 9.35 |
|  | 266 | 9.98 |
|  | Name: | Estimated Unemployment Rate (%), Length: 267, dtype: float64 |

#  Import Necessary Libraries

import plotly.express as px import matplotlib.pyplot as plt

 Aanalyzing Data By Bar Graphs

fg = px.box(df,x='Region',y=' Estimated Unemployment Rate (%)',color='Region', title='Unemploymeny Rate (Statewise) by Box Plot',template='plotly')

fg.update\_layout(xaxis={'categoryorder':'total descending'}) fg.show()

fg = px.bar(df,x='Region',y=' Estimated Unemployment Rate (%)',color='Region', title='Unemploymeny Rate (State Wise) by Bar Graph',template='plotly')

fg.update\_layout(xaxis={'categoryorder':'total descending'}) fg.show()

## Unemploymeny Rate (State Wise) by Bar Graph

250

Estimated Unemployment Rate (%)

200

150

100

50

0

Tamil N

Jammu & Kashmir

Rajasthan

Himachal Pradesh

Puducherry

Delhi

Bihar

Jharkhand

Tripura

Haryana

fg = px.bar(df,x='Region.1',y=' Estimated Unemployment Rate (%)',color='Region', title='Unemploymeny Rate (Region Wise) by Bar Graph',template='plotly')

fg.update\_layout(xaxis={'categoryorder':'total descending'}) fg.show()

## Unemploymeny Rate (Region Wise) by Bar Graph

1200

1000

Estimated Unemployment Rate (%)

800

600

400

200

0 North South

#  Aanalyzing Data By Box Plot

fg = px.histogram(df,x='Region',y=' Estimated Unemployment Rate (%)',color='Region', title='Unemploymeny Rate (Statewise) by Histogram',template='plotly')

fg.update\_layout(xaxis={'categoryorder':'total descending'}) fg.show()

## Unemploymeny Rate (Statewise) by Box Plot

80

70

Estimated Unemployment Rate (%)

60

50

40

30

20

10

0

Tamil N

Jammu & Kashmir

Rajasthan

Himachal Pradesh

Puducherry

Delhi

Bihar

Jharkhand

Tripura

Haryana

#  Aanalyzing Data By Scatter Plot

fg = px.scatter(df,x='Region',y=' Estimated Unemployment Rate (%)',color='Region', title='Unemploymeny Rate (Statewise) by Scatter Plot',template='plotly')

fg.update\_layout(xaxis={'categoryorder':'total descending'}) fg.show()

## Unemploymeny Rate (Statewise) by Scatter Plot

80



70

Estimated Unemployment Rate (%)

60

50

40

30

20

10

0

Tam

Jammu & Kas

Rajasthan

Himachal Pradesh

Puducherry

Delhi

Bihar

Jharkhand

Tripura

Haryana

#  Aanalyzing Data By Histogram

## Unemploymeny Rate (Statewise) by Histogram

250

sum of Estimated Unemployment Rate (%)

200

150

100

50

0

Tamil N

Jammu & Kashmir

Rajasthan

Himachal Pradesh

Puducherry

Delhi

Bihar

Jharkhand

Tripura

Haryana