

ASSIGNMENT 4

NAME : HANSINI JYOTHIRMAYI VADDEY

REDG NO: 21BDS0345

Question: 1.Download the Employee Attrition Dataset <https://www.kaggle.com/datasets/patelprashant/employee-attrition> 2.Perfrom Data Preprocessing
3.Model Building using Logistic Regression and Decision Tree and Random forest 4.Calculate Performance metrics

2.DATA PREPROCESSING

Steps 1.import necessary libraries

2.import dataset

3.Handling null values

4.Data Visualization

5.Outlier detection

6.Seperate dependent and independent variables

7.Encoding

8.Splitting into training set and testing set

9.Feature Scaling

```
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
In [ ]:
```

1.Import necessary libraries

```
In [1]: import numpy as np
In [2]: import pandas as pd
In [3]: import matplotlib.pyplot as plt
In [4]: import seaborn as sns
```

2.Load the Dataset

```
In [5]: dataset=pd.read_csv("HR-Employee-Attrition.csv")
In [6]: dataset
```

Out[6]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisf
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	
...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	

1470 rows × 35 columns



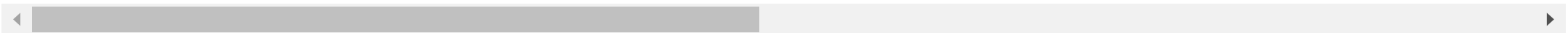
In [7]:

dataset.head()

Out[7]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisf
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	

5 rows × 35 columns



In [8]:

dataset.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   1470 non-null   int64
1   Attrition                           1470 non-null   object
2   BusinessTravel                      1470 non-null   object
3   DailyRate                           1470 non-null   int64
4   Department                          1470 non-null   object
5   DistanceFromHome                   1470 non-null   int64
6   Education                           1470 non-null   int64
7   EducationField                     1470 non-null   object
8   EmployeeCount                      1470 non-null   int64
9   EmployeeNumber                     1470 non-null   int64
10  EnvironmentSatisfaction             1470 non-null   int64
11  Gender                              1470 non-null   object
12  HourlyRate                          1470 non-null   int64
13  JobInvolvement                     1470 non-null   int64
14  JobLevel                           1470 non-null   int64
15  JobRole                             1470 non-null   object
16  JobSatisfaction                     1470 non-null   int64
17  MaritalStatus                       1470 non-null   object
18  MonthlyIncome                      1470 non-null   int64
19  MonthlyRate                         1470 non-null   int64
20  NumCompaniesWorked                 1470 non-null   int64
21  Over18                             1470 non-null   object
22  OverTime                           1470 non-null   object
23  PercentSalaryHike                  1470 non-null   int64
24  PerformanceRating                  1470 non-null   int64
25  RelationshipSatisfaction            1470 non-null   int64
26  StandardHours                      1470 non-null   int64
27  StockOptionLevel                   1470 non-null   int64
28  TotalWorkingYears                  1470 non-null   int64
29  TrainingTimesLastYear              1470 non-null   int64
30  WorkLifeBalance                    1470 non-null   int64
31  YearsAtCompany                     1470 non-null   int64
32  YearsInCurrentRole                 1470 non-null   int64
33  YearsSinceLastPromotion             1470 non-null   int64
34  YearsWithCurrManager                1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

In [9]: dataset.shape

Out[9]: (1470, 35)

In [10]: dataset.describe()

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLevel
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470.000000	1470.000000	1470.00
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306	2.721769	65.891156	2.729932	2.06
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335	1.093082	20.329428	0.711561	1.10
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	1.000000	30.000000	1.000000	1.00
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000	2.000000	48.000000	2.000000	1.00
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000	3.000000	66.000000	3.000000	2.00
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000	4.000000	83.750000	3.000000	3.00
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000	4.000000	100.000000	4.000000	5.00

8 rows × 26 columns

In [11]: dataset.Age.value_counts()

```
Out[11]: 35    78
          34    77
          36    69
          31    69
          29    68
          32    61
          30    60
          33    58
          38    58
          40    57
          37    50
          27    48
          28    48
          42    46
          39    42
          45    41
          41    40
          26    39
          44    33
          46    33
          43    32
          50    30
          25    26
          24    26
          49    24
          47    24
          55    22
          51    19
          53    19
          48    19
          54    18
          52    18
          22    16
          56    14
          23    14
          58    14
          21    13
          20    11
          59    10
          19     9
          18     8
          60     5
          57     4
Name: Age, dtype: int64
```

```
In [12]: dataset.DailyRate.value_counts()
```

```
Out[12]: 691     6
          408     5
          530     5
          1329    5
          1082     5
          ..
          650     1
          279     1
          316     1
          314     1
          628     1
Name: DailyRate, Length: 886, dtype: int64
```

3.Handling Null values

```
In [13]: dataset.isnull().any()
```

```
Out[13]: Age                False
Attrition                False
BusinessTravel           False
DailyRate               False
Department              False
DistanceFromHome        False
Education               False
EducationField           False
EmployeeCount           False
EmployeeNumber          False
EnvironmentSatisfaction  False
Gender                  False
HourlyRate              False
JobInvolvement          False
JobLevel                False
JobRole                 False
JobSatisfaction          False
MaritalStatus           False
MonthlyIncome           False
MonthlyRate             False
NumCompaniesWorked      False
Over18                  False
OverTime                False
PercentSalaryHike       False
PerformanceRating       False
RelationshipSatisfaction False
StandardHours           False
StockOptionLevel        False
TotalWorkingYears       False
TrainingTimesLastYear   False
WorkLifeBalance         False
YearsAtCompany          False
YearsInCurrentRole      False
YearsSinceLastPromotion False
YearsWithCurrManager    False
dtype: bool
```

No null values present

```
In [14]: dataset.isnull().sum()
```

```
Out[14]: Age                0
Attrition                0
BusinessTravel           0
DailyRate               0
Department              0
DistanceFromHome        0
Education               0
EducationField           0
EmployeeCount           0
EmployeeNumber          0
EnvironmentSatisfaction  0
Gender                  0
HourlyRate              0
JobInvolvement          0
JobLevel                0
JobRole                 0
JobSatisfaction          0
MaritalStatus           0
MonthlyIncome           0
MonthlyRate             0
NumCompaniesWorked      0
Over18                  0
OverTime                0
PercentSalaryHike       0
PerformanceRating       0
RelationshipSatisfaction 0
StandardHours           0
StockOptionLevel        0
TotalWorkingYears       0
TrainingTimesLastYear   0
WorkLifeBalance         0
YearsAtCompany          0
YearsInCurrentRole      0
YearsSinceLastPromotion 0
YearsWithCurrManager    0
dtype: int64
```

4.Data Visualization

```
In [115... dataset
```

Out[115]:

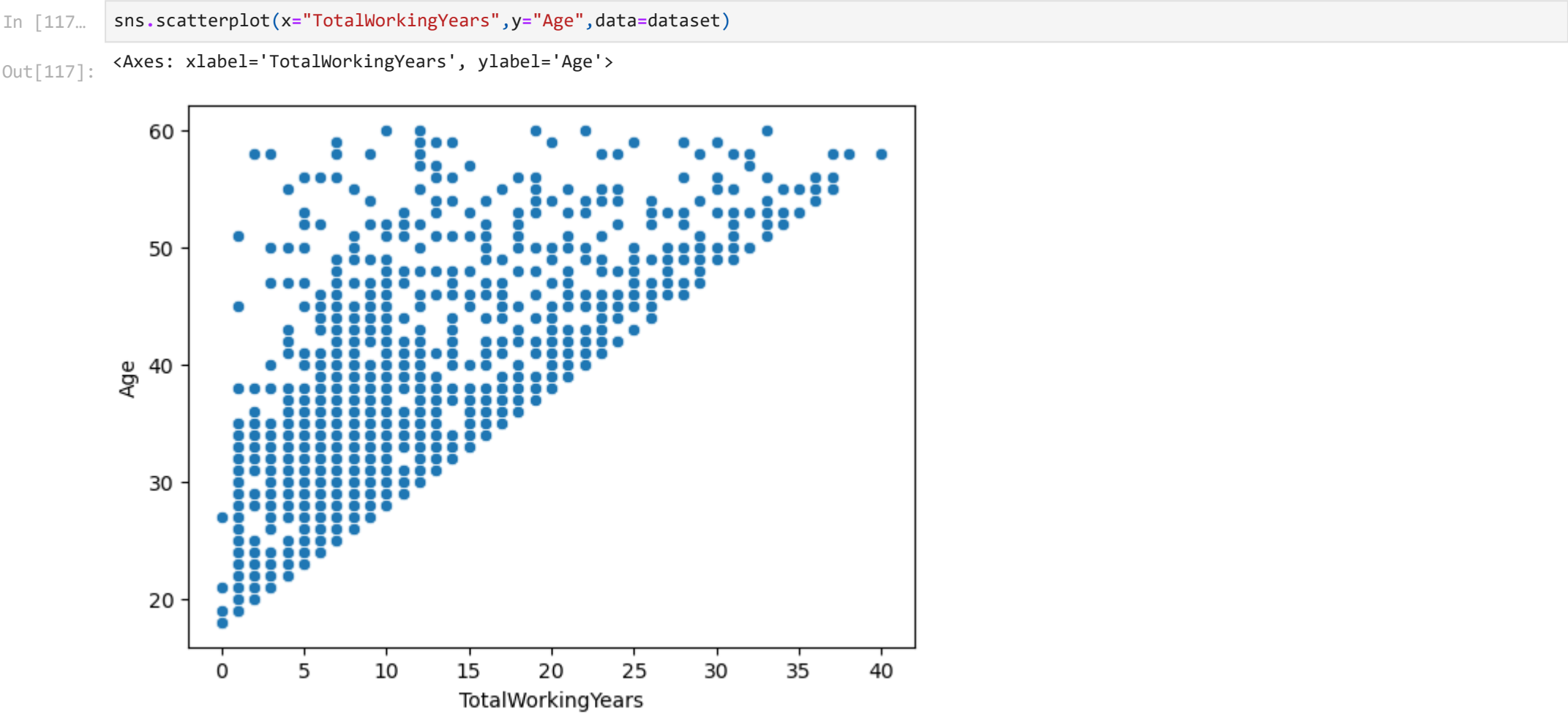
	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSa
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	
...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	

1470 rows × 35 columns

◀

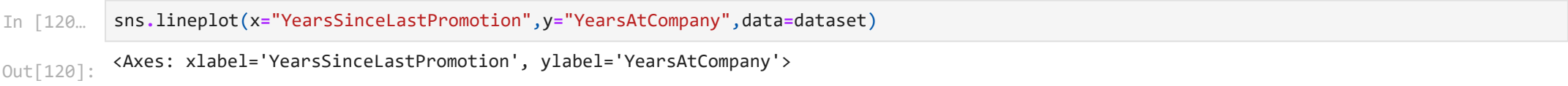
▶

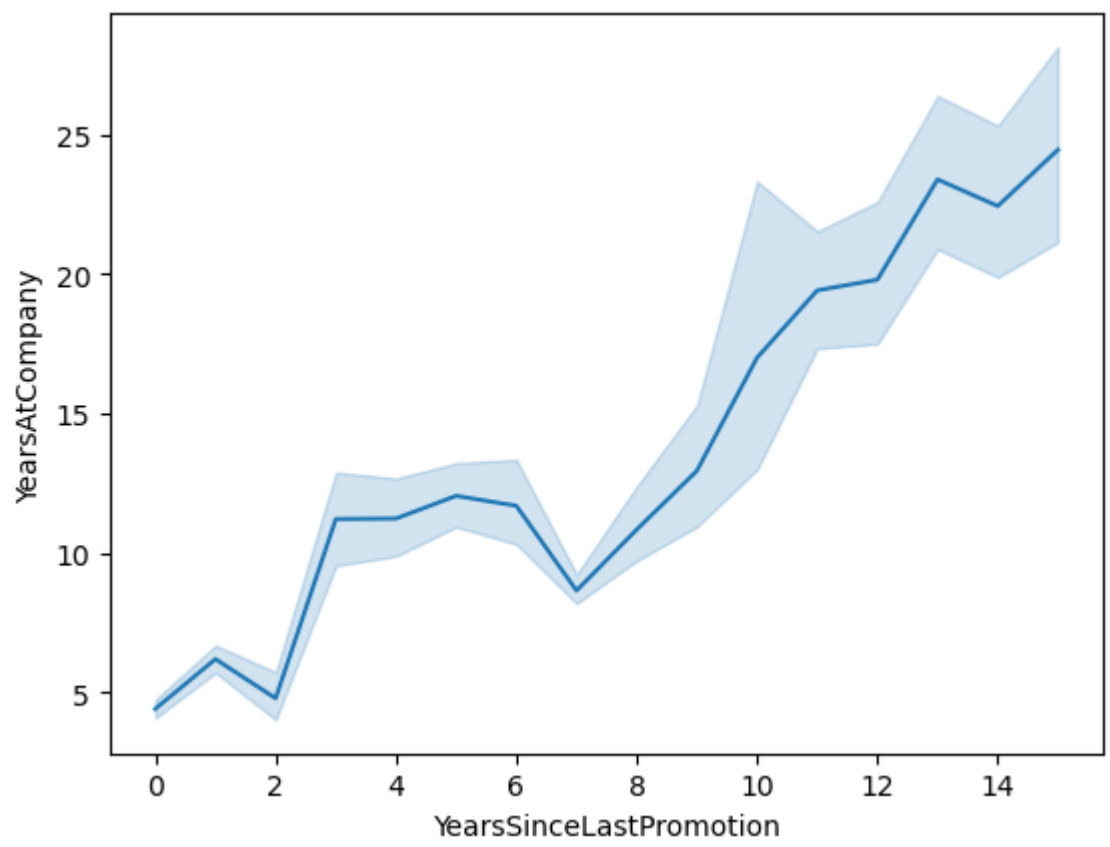
i) Scatter plot



Inference: The scatterplot comparing "TotalWorkingYears" on the x-axis and "Age" on the y-axis using the data from the dataset.this scatterplot confirms the expected positive correlation between total working years and age, but it also highlights the presence of outliers and individual variations within the dataset.

ii)Line Plot

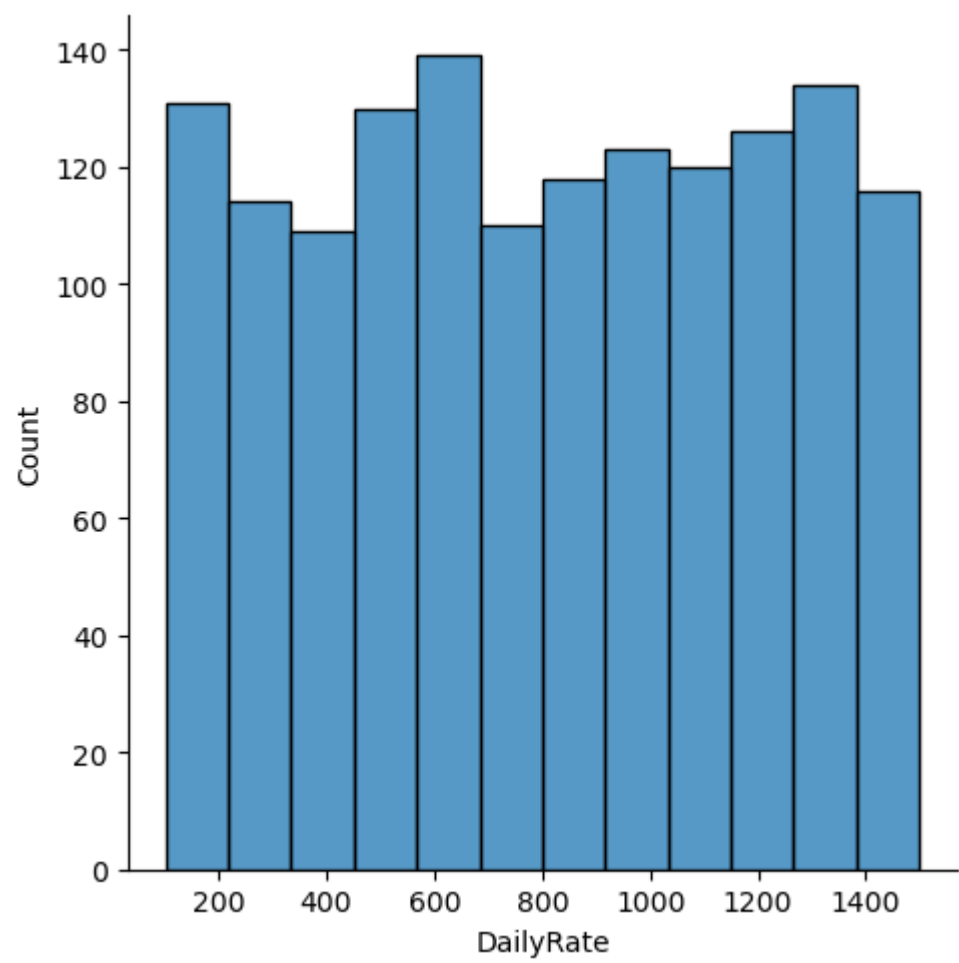




Inference :The line plot displays the relationship between "YearsSinceLastPromotion" on the x-axis and "YearsAtCompany" on the y-axis using the data from the dataset.this line plot shows a negative trend between "YearsSinceLastPromotion" and "YearsAtCompany," indicating that promotions tend to occur more frequently for employees who have been with the company for a shorter period.

iii)Distribution plot

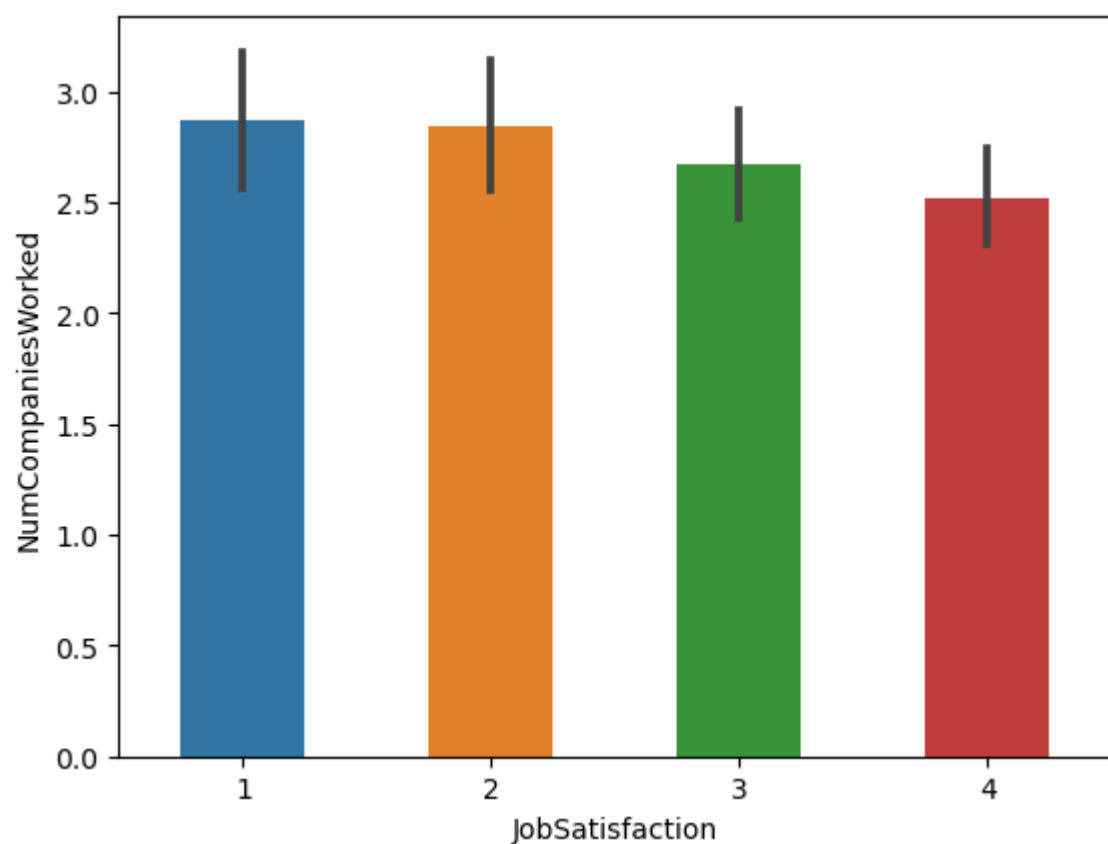
```
In [121]: sns.displot(dataset["DailyRate"])
Out[121]: <seaborn.axisgrid.FacetGrid at 0x1fb8ac054d0>
```



Inference : This is a distribution plot (histogram) of the "DailyRate" variable from your dataset using Seaborn.seful for understanding the distribution of daily rates and identifying any patterns or irregularities in the data

iv)Relational Plot

```
In [133]: sns.barplot(data=dataset, x="JobSatisfaction", y="NumCompaniesWorked", width=0.5)
Out[133]: <Axes: xlabel='JobSatisfaction', ylabel='NumCompaniesWorked'>
```

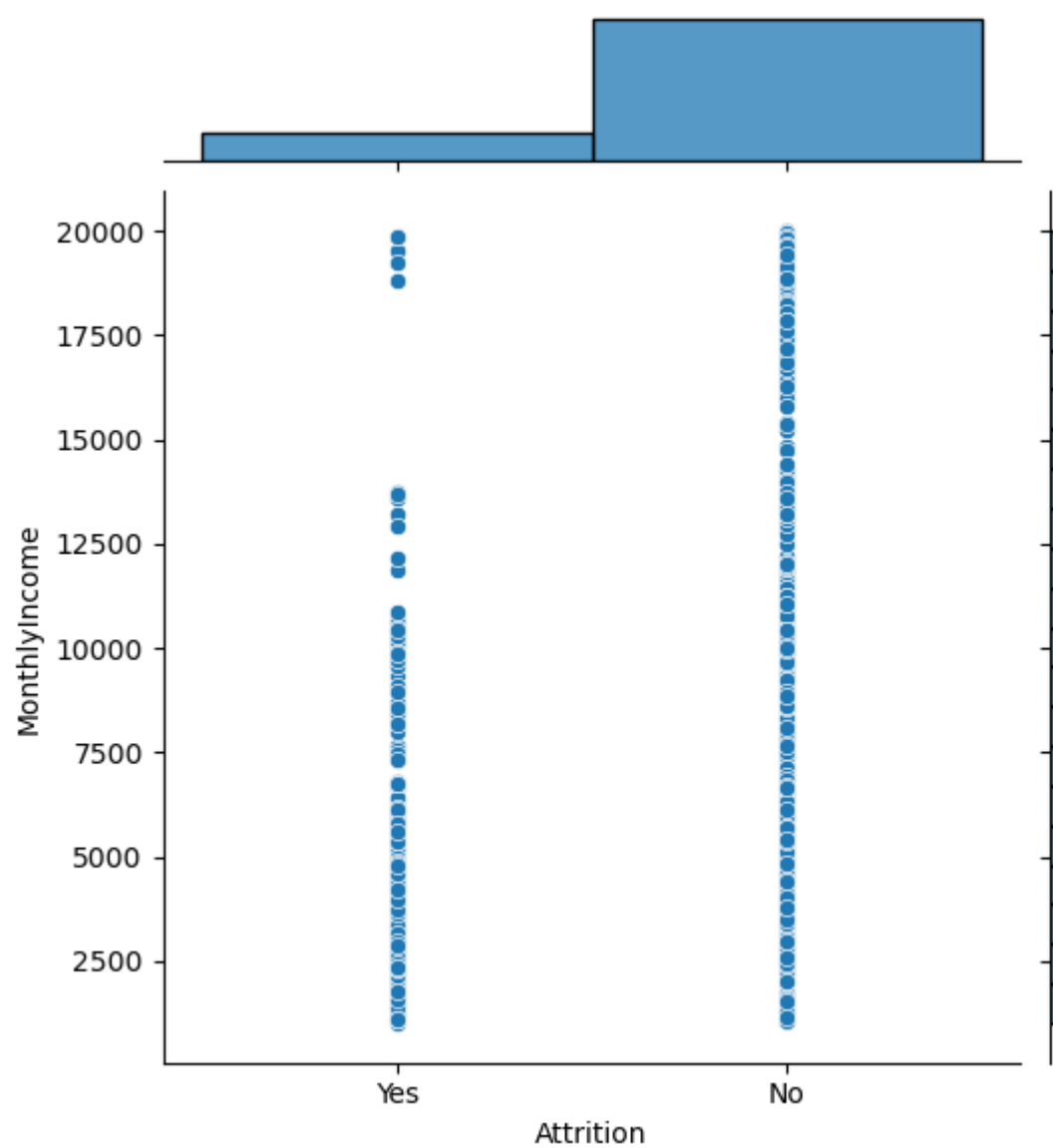


Inference : This Bar plot shows the relationship between job satisfaction and the number of companies worked for in your dataset.: The x-axis represents different levels of job satisfaction, and the y-axis represents the average or aggregated number of companies employees have worked for ("NumCompaniesWorked") for each level of job satisfaction.

v)Joint Plot

```
sns.jointplot(x="Attrition",y="MonthlyIncome",data=dataset)
```

<seaborn.axisgrid.JointGrid at 0x1fb92580150>

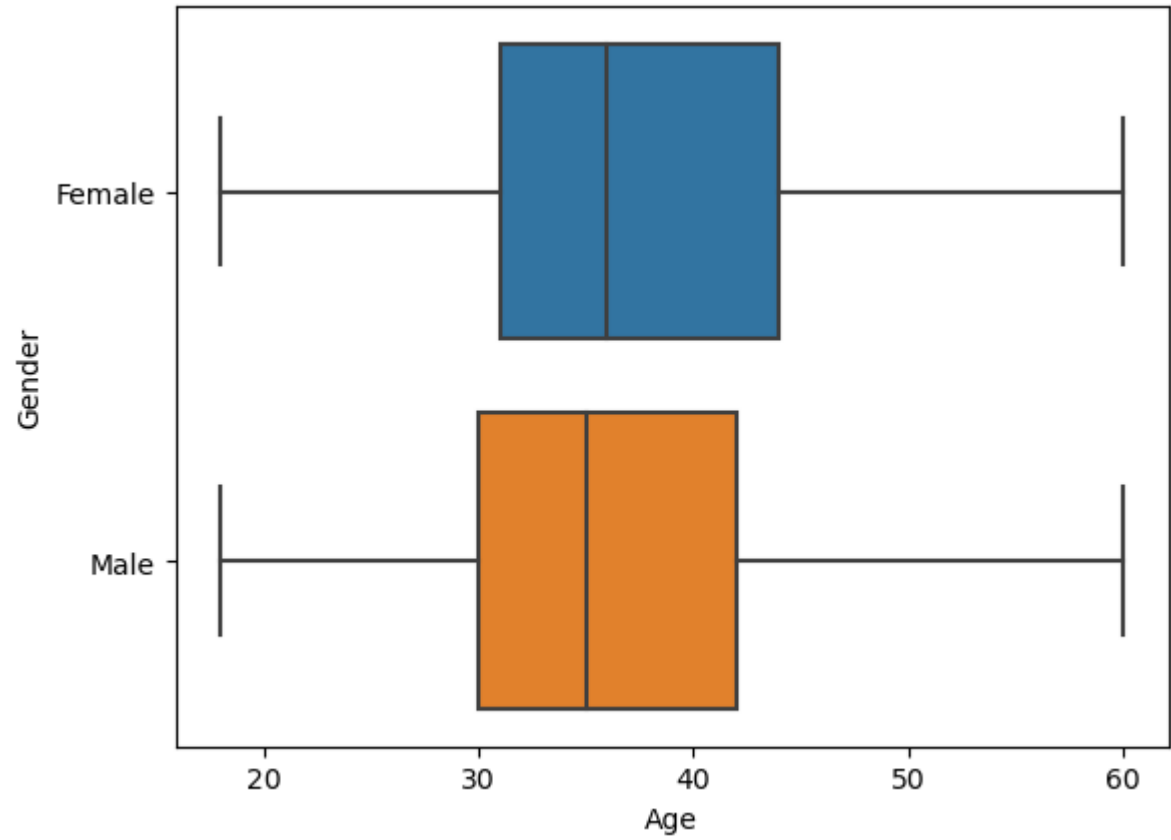


Inference : This joint plot provides a visual representation of the relationship between attrition and monthly income in your dataset.The central part of the joint plot is a scatter plot. In this case, "Attrition" is on the x-axis (usually a categorical variable), and "MonthlyIncome" is on the y-axis (a continuous variable). Each point on the scatter plot represents an employee's monthly income, and its position on the x-axis is determined by their attrition status. he marginal histograms on the top and right sides provide insights into the individual distributions of each variable.It helps assess whether there's a correlation between the two variables

vi)Box Plot

```
sns.boxplot(x="Age",y="Gender",data=dataset)
```

<Axes: xlabel='Age', ylabel='Gender'>



Inference : This box plot visualizes and compare the distribution of ages within different gender categories in your dataset. It provides insights into the central tendency, spread, and potential outliers in the age data for each gender group, facilitating a better understanding of how age varies by gender in your dataset.

Heatmap

Correlation before Encoding

In [137...

```
corr=dataset.corr() #As Data visualization step in the question is given before encoding we are getting warning message.
```

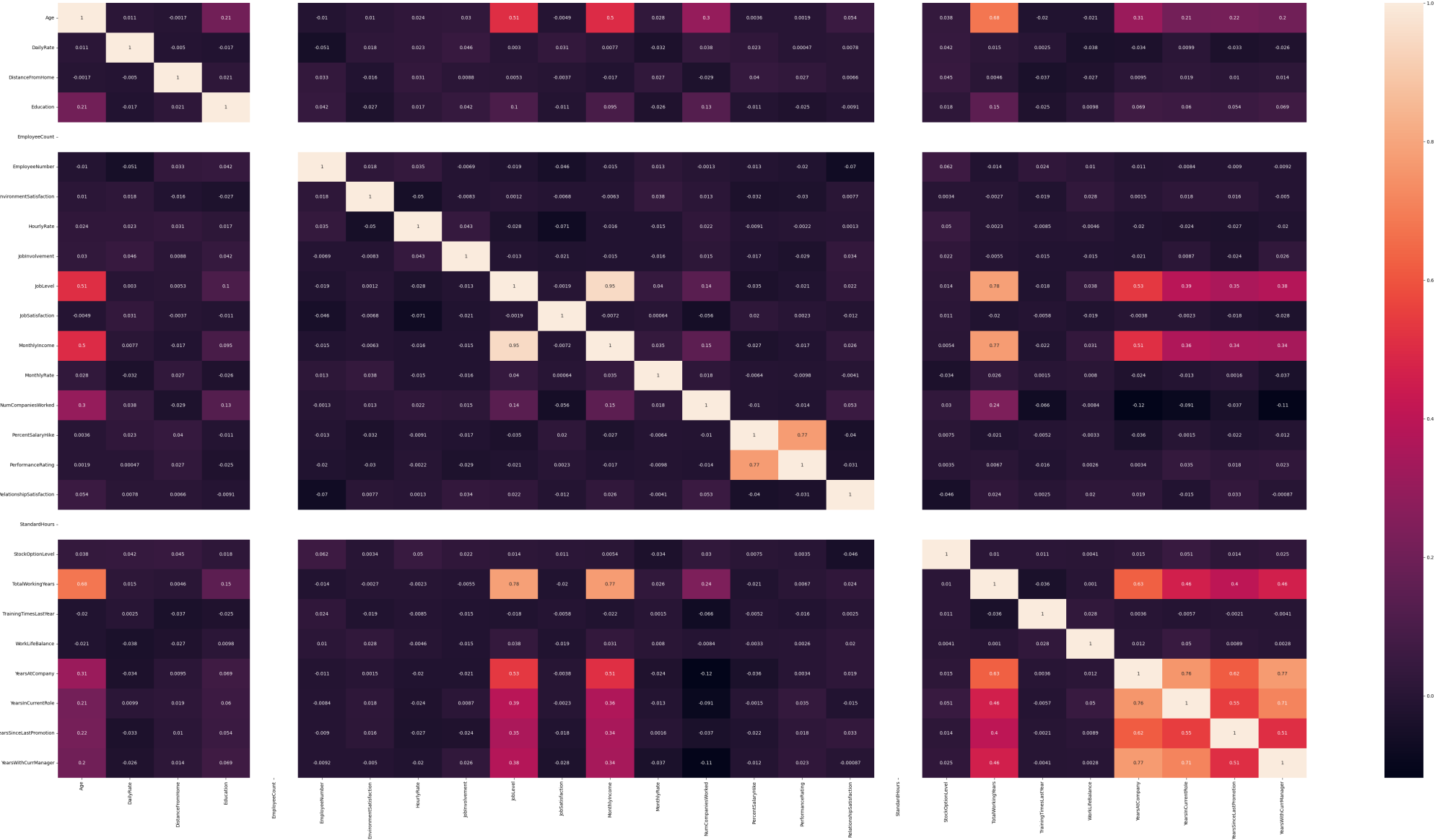
C:\Users\hansi\AppData\Local\Temp\ipykernel_17304\2244050776.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
corr=dataset.corr() #As Data visualization step in the question is given before encoding we are getting warning message.
```

In [139...

```
plt.subplots(figsize=(60,30))
sns.heatmap(corr,annot=True)
```

Out[139]: <Axes: >



Inference : This heatmap is a graphical representation of the correlation between variables in a dataset. The color intensity and annotations provide a quick way to identify the strength and direction of relationships between variables. It's used in exploratory data analysis to identify potential correlations among features.

5.Outlier detection

Outliers are those data points that are significantly different from the rest of the dataset. They are often abnormal observations that skew the data distribution, and arise due to inconsistent data entry, or erroneous observations.

Outlier Removal 3 methods:

1.IQR - Inter quartile range => $Q3 - Q1$

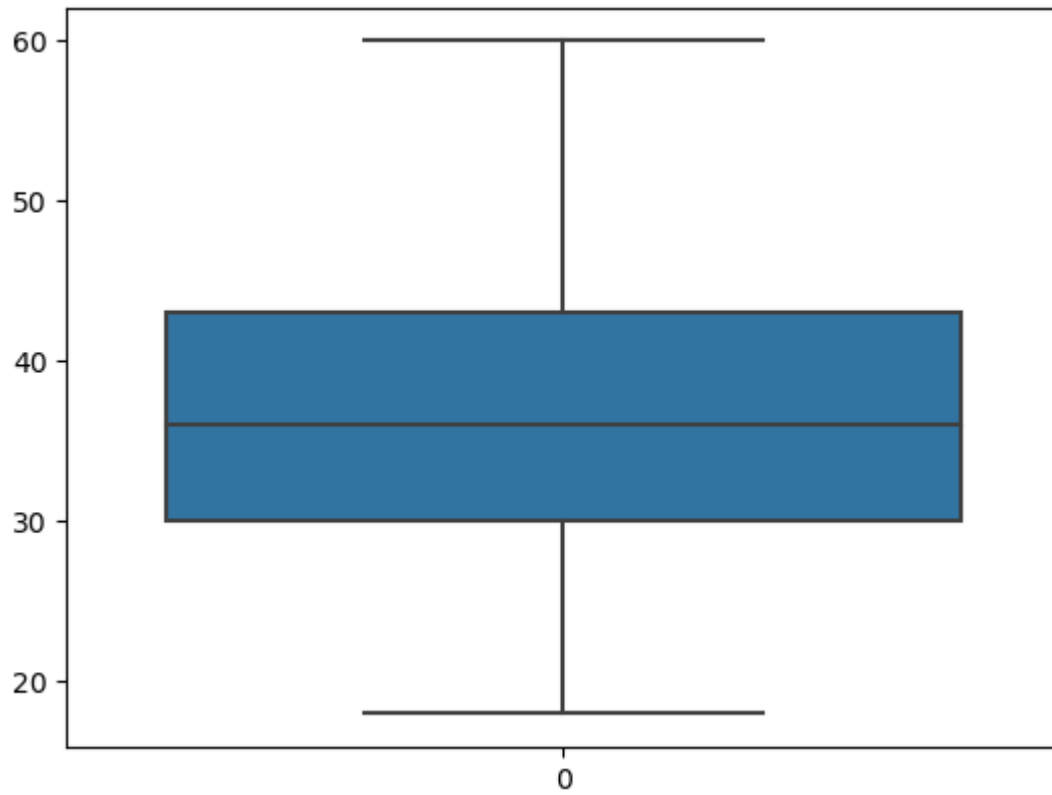
2.z Score

3.Percentile

Outlier Detection for Age

```
In [15]: sns.boxplot(dataset.Age)
```

```
Out[15]: <Axes: >
```

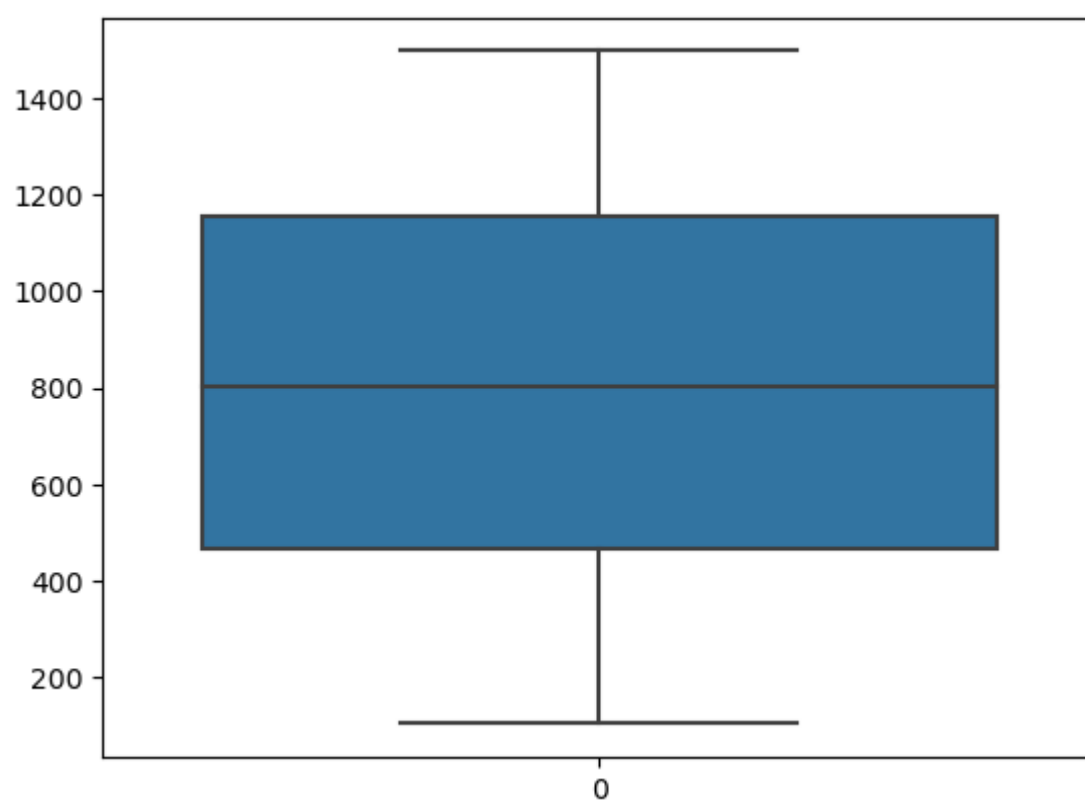


No outliers for Age

Outlier Detection for DailyRate

```
In [16]: sns.boxplot(dataset.DailyRate)
```

```
Out[16]: <Axes: >
```

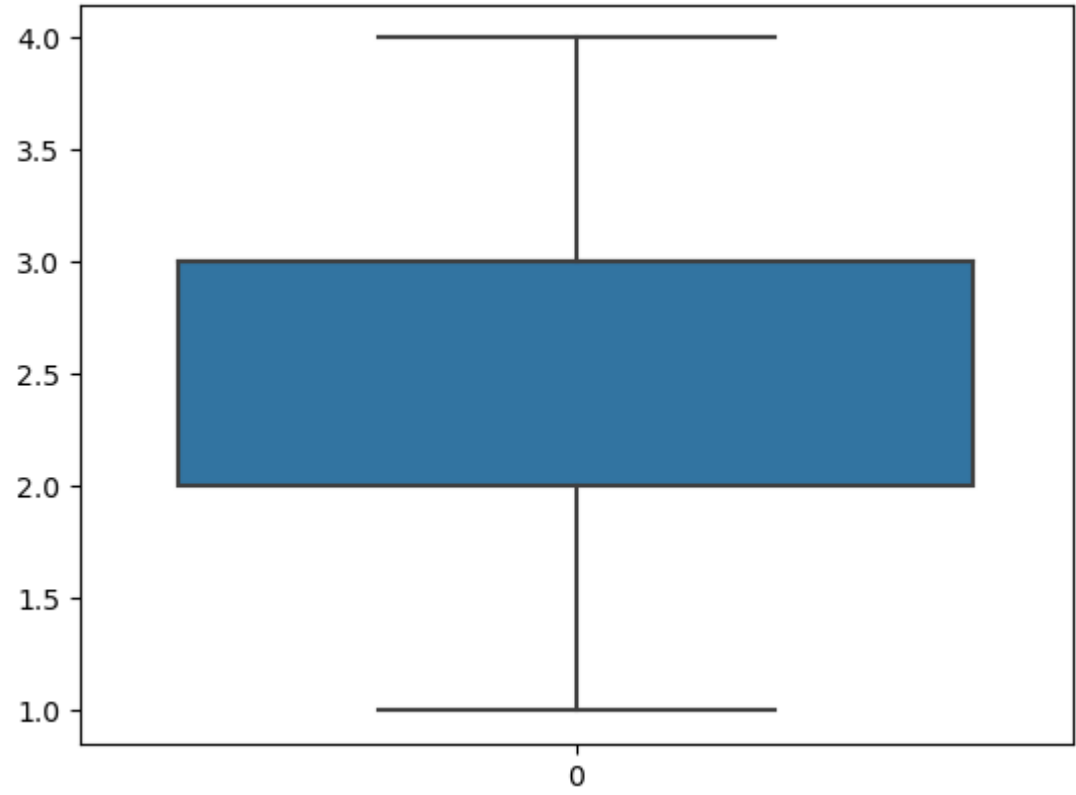


No outliers for Daily Rate

Outlier Detection for WorkLifeBalance

```
In [17]: sns.boxplot(dataset.WorkLifeBalance)
```

```
Out[17]: <Axes: >
```



No outliers for WorkLifeBalance

6.Seperation of dependent and independent variables

```
In [18]: #Independent Variables
x=dataset.drop(columns=['Attrition'], inplace=False)
x
```

Out[18]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...
0	41	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	...
1	49	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3	...
2	37	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4	...
3	33	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4	...
4	27	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1	...
...
1465	36	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	3	...
1466	39	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	4	...
1467	27	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	2	...
1468	49	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	4	...
1469	34	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	2	...

1470 rows × 34 columns



```
In [19]: x.head()
```

Out[19]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	Rel
0	41	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	...	
1	49	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	3	...	
2	37	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4	...	
3	33	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	4	...	
4	27	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	1	...	

5 rows × 34 columns



```
In [20]: #Dependent variable
y=dataset.iloc[:,1:2]
```

```
In [21]: y
```

Out[21]:

	Attrition
0	Yes
1	No
2	Yes
3	No
4	No
...	...
1465	No
1466	No
1467	No
1468	No
1469	No

1470 rows × 1 columns

```
In [22]: dataset.shape
```

Out[22]: (1470, 35)

```
In [23]: x.shape
```

Out[23]: (1470, 34)

```
In [24]: y.shape
```

Out[24]: (1470, 1)

7.Encoding

Convert the strings into numerical or binary format

- 1.Nominal Encoding :If varaible is not worried the about arrangement of data i)one hot encoding ii)one hot encoding with many categorical variables iii)mean encoding
- 2.Ordinal Encoding : Take care about the rank of the data i)label encoding ii)Target guided ordinal encoding

```
In [25]: dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                    1470 non-null   int64
1   Attrition                            1470 non-null   object
2   BusinessTravel                        1470 non-null   object
3   DailyRate                            1470 non-null   int64
4   Department                           1470 non-null   object
5   DistanceFromHome                     1470 non-null   int64
6   Education                             1470 non-null   int64
7   EducationField                        1470 non-null   object
8   EmployeeCount                        1470 non-null   int64
9   EmployeeNumber                       1470 non-null   int64
10  EnvironmentSatisfaction               1470 non-null   int64
11  Gender                               1470 non-null   object
12  HourlyRate                           1470 non-null   int64
13  JobInvolvement                       1470 non-null   int64
14  JobLevel                             1470 non-null   int64
15  JobRole                              1470 non-null   object
16  JobSatisfaction                      1470 non-null   int64
17  MaritalStatus                       1470 non-null   object
18  MonthlyIncome                       1470 non-null   int64
19  MonthlyRate                          1470 non-null   int64
20  NumCompaniesWorked                   1470 non-null   int64
21  Over18                              1470 non-null   object
22  OverTime                             1470 non-null   object
23  PercentSalaryHike                    1470 non-null   int64
24  PerformanceRating                    1470 non-null   int64
25  RelationshipSatisfaction              1470 non-null   int64
26  StandardHours                       1470 non-null   int64
27  StockOptionLevel                     1470 non-null   int64
28  TotalWorkingYears                    1470 non-null   int64
29  TrainingTimesLastYear                1470 non-null   int64
30  WorkLifeBalance                      1470 non-null   int64
31  YearsAtCompany                       1470 non-null   int64
32  YearsInCurrentRole                   1470 non-null   int64
33  YearsSinceLastPromotion              1470 non-null   int64
34  YearsWithCurrManager                 1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

Label Encoding on BusinessTravel

```
In [26]: from sklearn.preprocessing import LabelEncoder
```

```
In [27]: l=LabelEncoder()
```

```
In [28]: x.BusinessTravel=l.fit_transform(x.BusinessTravel)
x["BusinessTravel"]
```

```
Out[28]: 0      2
1      1
2      2
3      1
4      2
..
1465   1
1466   2
1467   2
1468   1
1469   2
Name: BusinessTravel, Length: 1470, dtype: int32
```

```
In [29]: x
```

Out[29]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	F
0	41	2	1102	Sales	1	2	Life Sciences	1	1	2	...	
1	49	1	279	Research & Development	8	1	Life Sciences	1	2	3	...	
2	37	2	1373	Research & Development	2	2	Other	1	4	4	...	
3	33	1	1392	Research & Development	3	4	Life Sciences	1	5	4	...	
4	27	2	591	Research & Development	2	1	Medical	1	7	1	...	
...	
1465	36	1	884	Research & Development	23	2	Medical	1	2061	3	...	
1466	39	2	613	Research & Development	6	1	Medical	1	2062	4	...	
1467	27	2	155	Research & Development	4	3	Life Sciences	1	2064	2	...	
1468	49	1	1023	Sales	2	3	Medical	1	2065	4	...	
1469	34	2	628	Research & Development	8	3	Medical	1	2068	2	...	

1470 rows × 34 columns



In [30]:

```
x["BusinessTravel"].value_counts()
```

Out[30]:

```
2    1043
1     277
0     150
Name: BusinessTravel, dtype: int64
```

Label Encoding on Department

In [31]:

```
x.Department=l.fit_transform(x.Department)
x["Department"]
```

Out[31]:

```
0     2
1     1
2     1
3     1
4     1
..
1465   1
1466   1
1467   1
1468   2
1469   1
Name: Department, Length: 1470, dtype: int32
```

In [32]:

```
x["Department"].value_counts()
```

Out[32]:

```
1     961
2     446
0      63
Name: Department, dtype: int64
```

Label Encoding on EducationField

In [33]:

```
x.EducationField=l.fit_transform(x.EducationField)
x["EducationField"]
```

Out[33]:

```
0     1
1     1
2     4
3     1
4     3
..
1465   3
1466   3
1467   1
1468   3
1469   3
Name: EducationField, Length: 1470, dtype: int32
```

In [34]:

```
x["EducationField"].value_counts()
```

Out[34]:

```
1     606
3     464
2     159
5     132
4      82
0      27
Name: EducationField, dtype: int64
```

In [35]:

```
x
```

Out[35]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	R
0	41	2	1102	2	1	2	1	1	1	2	...	
1	49	1	279	1	8	1	1	1	2	3	...	
2	37	2	1373	1	2	2	4	1	4	4	...	
3	33	1	1392	1	3	4	1	1	5	4	...	
4	27	2	591	1	2	1	3	1	7	1	...	
...	
1465	36	1	884	1	23	2	3	1	2061	3	...	
1466	39	2	613	1	6	1	3	1	2062	4	...	
1467	27	2	155	1	4	3	1	1	2064	2	...	
1468	49	1	1023	2	2	3	3	1	2065	4	...	
1469	34	2	628	1	8	3	3	1	2068	2	...	

1470 rows × 34 columns



Label Encoding on Gender,JobRole,MaritalStatus,Over18 & OverTime

In [36]:

```
# Label Encoding on Gender
x.Gender=l.fit_transform(x.Gender)
x["Gender"]
```

Out[36]:

0	0
1	1
2	1
3	0
4	1
	..
1465	1
1466	1
1467	1
1468	1
1469	1

Name: Gender, Length: 1470, dtype: int32

In [37]:

```
#Label Encoding on JobRole
x.JobRole=l.fit_transform(x.JobRole)
x["JobRole"]
```

Out[37]:

0	7
1	6
2	2
3	6
4	2
	..
1465	2
1466	0
1467	4
1468	7
1469	2

Name: JobRole, Length: 1470, dtype: int32

In [38]:

```
#Label Encoding on MaritalStatus
x.MaritalStatus=l.fit_transform(x.MaritalStatus)
x["MaritalStatus"]
```

Out[38]:

0	2
1	1
2	2
3	1
4	1
	..
1465	1
1466	1
1467	1
1468	1
1469	1

Name: MaritalStatus, Length: 1470, dtype: int32

In [39]:

```
#Label Encoding on Over18
x.Over18=l.fit_transform(x.Over18)
x["Over18"]
```

Out[39]:

0	0
1	0
2	0
3	0
4	0
	..
1465	0
1466	0
1467	0
1468	0
1469	0

Name: Over18, Length: 1470, dtype: int32

```
In [40]: #Label Encoding on OverTime
x.OverTime=l.fit_transform(x.OverTime)
x["OverTime"]
```

Out[40]:

0	1
1	0
2	1
3	1
4	0
..	
1465	0
1466	0
1467	1
1468	0
1469	0

Name: OverTime, Length: 1470, dtype: int32

```
In [41]: x
```

Out[41]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	R
0	41	2	1102	2	1	2	1	1	1	2	...	
1	49	1	279	1	8	1	1	1	2	3	...	
2	37	2	1373	1	2	2	4	1	4	4	...	
3	33	1	1392	1	3	4	1	1	5	4	...	
4	27	2	591	1	2	1	3	1	7	1	...	
...	
1465	36	1	884	1	23	2	3	1	2061	3	...	
1466	39	2	613	1	6	1	3	1	2062	4	...	
1467	27	2	155	1	4	3	1	1	2064	2	...	
1468	49	1	1023	2	2	3	3	1	2065	4	...	
1469	34	2	628	1	8	3	3	1	2068	2	...	

1470 rows × 34 columns



Label encoding on Dependent Variable Attrition

```
In [42]: #Label Encoding on Attrition
y["Attrition"]=l.fit_transform(y.Attrition)
y["Attrition"]
```

Out[42]:

0	1
1	0
2	1
3	0
4	0
..	
1465	0
1466	0
1467	0
1468	0
1469	0

Name: Attrition, Length: 1470, dtype: int32

We have training and testting data

example: 1000 rows

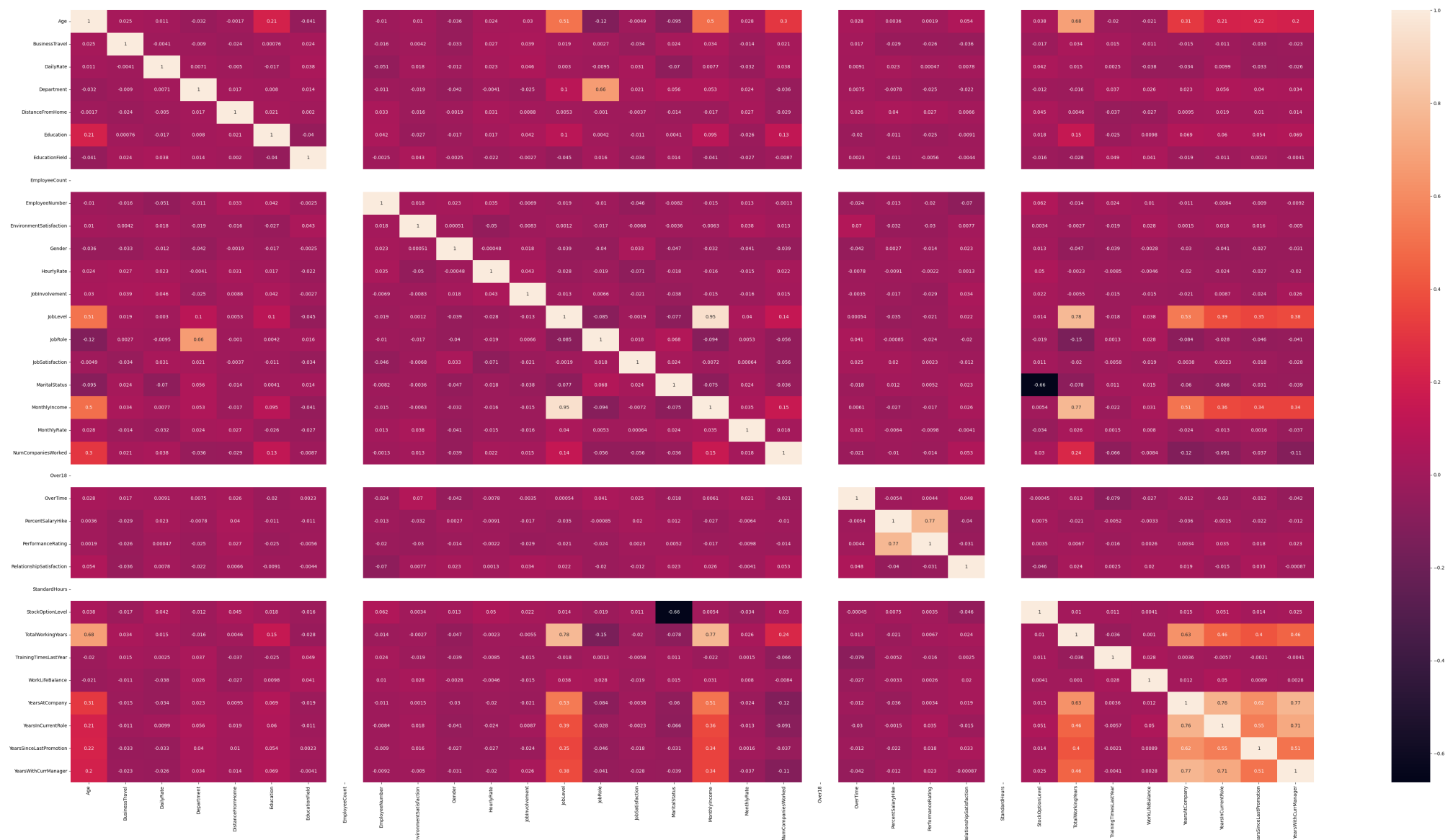
Training data : 70%-80%

Testing data : 20%-30% (checking performance)

Correlation after encoding

```
In [43]: corr = x.corr()
plt.subplots(figsize=(60,30))
sns.heatmap(corr,annot=True)
```

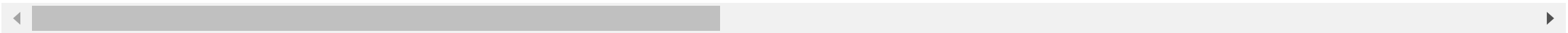
Out[43]: <Axes: >



```
In [44]: from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled=pd.DataFrame(ms.fit_transform(x),columns=x.columns)
```

```
Out[45]:
```

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction
0	0.547619	1.0	0.715820	1.0	0.000000	0.25	0.2	0.0	0.000000	0.333333
1	0.738095	0.5	0.126700	0.5	0.250000	0.00	0.2	0.0	0.000484	0.666667
2	0.452381	1.0	0.909807	0.5	0.035714	0.25	0.8	0.0	0.001451	1.000000
3	0.357143	0.5	0.923407	0.5	0.071429	0.75	0.2	0.0	0.001935	1.000000
4	0.214286	1.0	0.350036	0.5	0.035714	0.00	0.6	0.0	0.002903	0.000000
...
1465	0.428571	0.5	0.559771	0.5	0.785714	0.25	0.6	0.0	0.996613	0.666667
1466	0.500000	1.0	0.365784	0.5	0.178571	0.00	0.6	0.0	0.997097	1.000000
1467	0.214286	1.0	0.037938	0.5	0.107143	0.50	0.2	0.0	0.998065	0.333333
1468	0.738095	0.5	0.659270	1.0	0.035714	0.50	0.6	0.0	0.998549	1.000000
1469	0.380952	1.0	0.376521	0.5	0.250000	0.50	0.6	0.0	1.000000	0.333333



```
In [59]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_scaled,y,test_size=0.2,random_state=0)
```

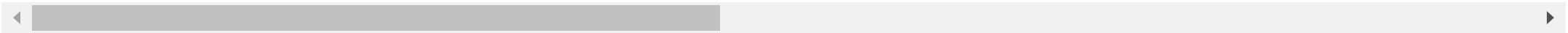
```
Out[60]: ((1176, 34), (294, 34), (1176, 1), (294, 1))
```

```
In [61]: x_train.head()
```

Out[61]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	
1374	0.952381	1.0	0.360057	1.0	0.714286	0.50	0.2	0.0	0.937107	1.000000	
1092	0.642857	1.0	0.607015	0.5	0.964286	0.50	1.0	0.0	0.747460	1.000000	
768	0.523810	1.0	0.141732	1.0	0.892857	0.50	0.4	0.0	0.515239	0.666667	
569	0.428571	0.0	0.953472	1.0	0.250000	0.75	0.2	0.0	0.381229	0.000000	
911	0.166667	0.5	0.355762	1.0	0.821429	0.00	0.2	0.0	0.615385	0.666667	

5 rows × 34 columns



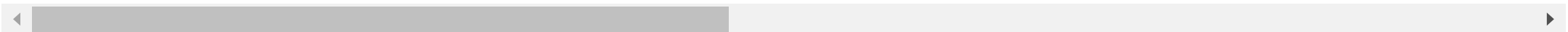
In [49]:

x_test

Out[49]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	...	R
442	36	0	635	2	10	4	3	1	592	2	...	
1091	33	2	575	1	25	3	1	1	1545	4	...	
981	35	1	662	2	18	4	2	1	1380	4	...	
785	40	2	1492	1	20	4	5	1	1092	1	...	
1332	29	1	459	1	24	2	1	1	1868	4	...	
...	
1439	36	2	557	2	3	3	3	1	2024	1	...	
481	34	2	254	1	1	2	1	1	649	2	...	
124	31	2	249	2	6	4	1	1	163	2	...	
198	38	2	1261	1	2	4	1	1	271	4	...	
1229	40	2	369	1	8	2	1	1	1724	2	...	

294 rows × 34 columns



In [50]:

y_train

Out[50]:

	Attrition
1374	0
1092	0
768	0
569	0
911	1
...	...
763	0
835	0
1216	0
559	0
684	0

1176 rows × 1 columns

In [51]:

y_test

Out[51]:

Attrition	
442	0
1091	0
981	1
785	0
1332	1
...	...
1439	0
481	0
124	1
198	0
1229	0

294 rows × 1 columns

In []:

MODEL BUILDING

- 1.Train the model using the training set
- 2.Test the model on testing set
- 3.Evaluation of the model

In [62]:

```
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

Initializing the model

In [63]:

```
model.fit(x_train,y_train)
```

C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\utils\validation.py:1184: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
 y = column_or_1d(y, warn=True)

Out[63]:

▼ LogisticRegression

LogisticRegression()

In [64]:

```
pred=model.predict(x_test)
```

In [65]:

```
pred
```

Out[65]:

```
array([0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0])
```

In [66]:

```
y_test
```

Out[66]:

	Attrition
442	0
1091	0
981	1
785	0
1332	1
...	...
1439	0
481	0
124	1
198	0
1229	0

294 rows × 1 columns

Evaluation of classification model

In [67]:

```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

In [68]:

```
accuracy_score(y_test,pred)
```

Out[68]:

```
0.8843537414965986
```

In [69]:

```
confusion_matrix(y_test,pred)
```

Out[69]:

```
array([[242,  3],
       [ 31, 18]], dtype=int64)
```

In [71]:

```
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.89	0.99	0.93	245
1	0.86	0.37	0.51	49
accuracy			0.88	294
macro avg	0.87	0.68	0.72	294
weighted avg	0.88	0.88	0.86	294

ROC CURVE

In [73]:

```
prob=model.predict_proba(x_test)[:,1]
```

In [74]:

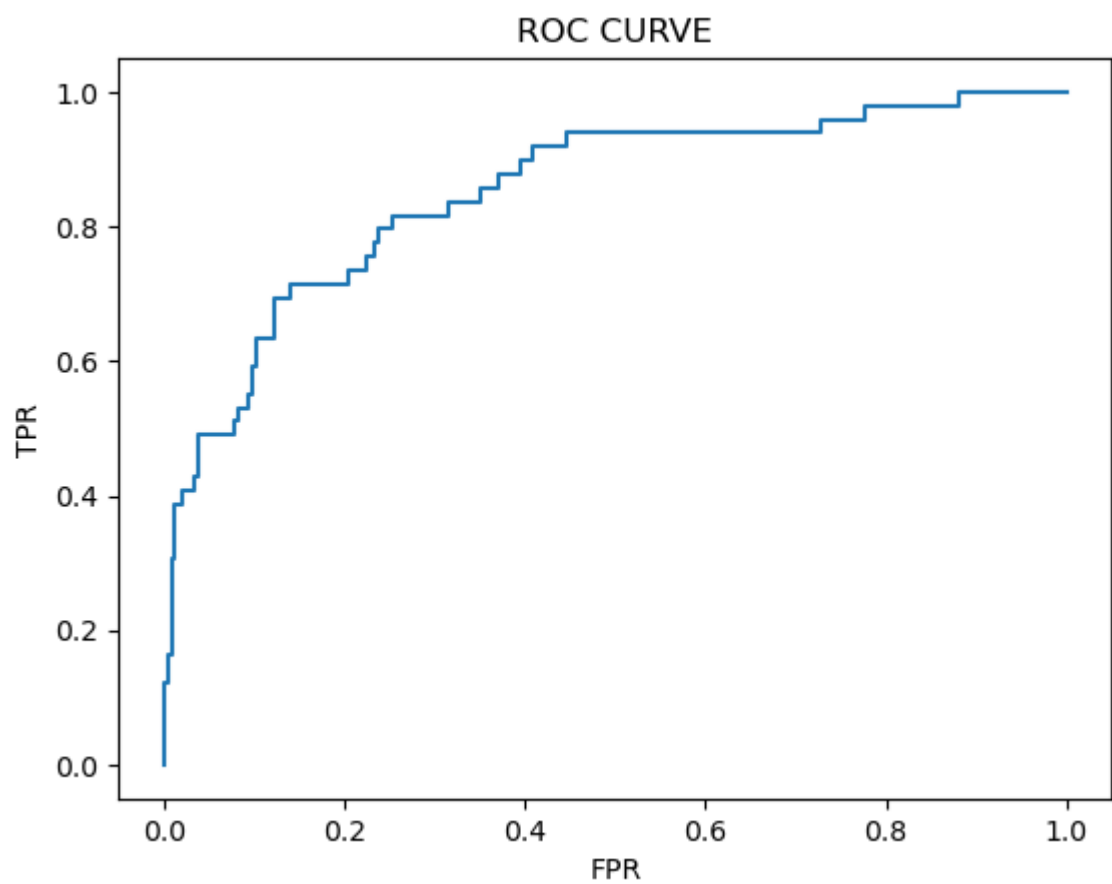
```
prob
```

```
Out[74]: array([0.16000127, 0.20600667, 0.31532384, 0.09242886, 0.63667551,
0.06153061, 0.61819432, 0.0757087 , 0.00841372, 0.3912069 ,
0.05398439, 0.33293123, 0.02020698, 0.67215483, 0.19786547,
0.03454902, 0.11043981, 0.17101703, 0.04477777, 0.22783614,
0.2335018 , 0.01553905, 0.06464492, 0.05029956, 0.58792413,
0.44849464, 0.07412714, 0.04460935, 0.67666632, 0.0584383 ,
0.01599026, 0.03521098, 0.06963085, 0.17397462, 0.07830857,
0.04288032, 0.08150424, 0.07106342, 0.03622137, 0.05223965,
0.04862098, 0.02091497, 0.01819361, 0.01362467, 0.02873997,
0.50236969, 0.41553218, 0.00306874, 0.73976412, 0.51382382,
0.09637213, 0.48845516, 0.08036228, 0.25757243, 0.66516772,
0.26308027, 0.01964858, 0.30198497, 0.02919946, 0.16038964,
0.02102747, 0.21670232, 0.13981568, 0.0358316 , 0.37208403,
0.03002317, 0.29091186, 0.16041142, 0.10437497, 0.08695177,
0.08217589, 0.30984518, 0.08531362, 0.07420689, 0.12268651,
0.06192552, 0.04640904, 0.07624712, 0.19738483, 0.03236316,
0.00884439, 0.0244108 , 0.13635803, 0.0260104 , 0.03341008,
0.08186888, 0.00499397, 0.03474852, 0.03858027, 0.14602694,
0.26167665, 0.16667357, 0.27400109, 0.24159565, 0.02160421,
0.17748606, 0.34076078, 0.28022482, 0.06914126, 0.05003806,
0.24437761, 0.74698271, 0.35438567, 0.01920627, 0.08778845,
0.03255847, 0.05461351, 0.15123251, 0.06843702, 0.13752637,
0.09584388, 0.04669882, 0.02493091, 0.15383171, 0.07081259,
0.03089296, 0.0537667 , 0.11554316, 0.00881616, 0.01263271,
0.17552253, 0.05045234, 0.08823238, 0.82995757, 0.03017756,
0.0236819 , 0.0087012 , 0.1349589 , 0.16474801, 0.05202613,
0.01524549, 0.29278083, 0.54767448, 0.34275448, 0.04629541,
0.38966344, 0.61333366, 0.14552367, 0.07402366, 0.24143471,
0.09418418, 0.0689069 , 0.10061956, 0.19346327, 0.20026293,
0.03004939, 0.14900424, 0.00348846, 0.11225149, 0.15843155,
0.06047573, 0.18601882, 0.06085869, 0.12221317, 0.03280184,
0.02738799, 0.06356425, 0.08302382, 0.01541716, 0.014665 ,
0.38517822, 0.01264231, 0.14961974, 0.80508787, 0.11598661,
0.2842811 , 0.17020143, 0.1530583 , 0.02764153, 0.00613226,
0.04191632, 0.09782393, 0.11551417, 0.10377982, 0.01779313,
0.14371315, 0.10615435, 0.10298963, 0.05132621, 0.09061081,
0.02897383, 0.09924087, 0.00512032, 0.75108423, 0.04296968,
0.04062134, 0.37518972, 0.04563128, 0.7251816 , 0.10671665,
0.36949086, 0.38146941, 0.32095493, 0.05266802, 0.08172004,
0.13947833, 0.04334317, 0.01469593, 0.26413988, 0.06330966,
0.1614747 , 0.15380517, 0.67152357, 0.05840793, 0.27891823,
0.04512564, 0.46033865, 0.00348431, 0.14068967, 0.02747401,
0.12714133, 0.17284246, 0.07341066, 0.10099827, 0.16870885,
0.02560842, 0.01824031, 0.08670796, 0.02834237, 0.13710215,
0.08778935, 0.2200061 , 0.73401148, 0.15938978, 0.4095449 ,
0.01513845, 0.11306309, 0.21497506, 0.32337575, 0.03409266,
0.04256318, 0.32157531, 0.05454465, 0.02348479, 0.16423352,
0.32696147, 0.22892063, 0.00877159, 0.08198819, 0.01156361,
0.1408691 , 0.29235147, 0.01270305, 0.17329916, 0.04081391,
0.04094165, 0.42771425, 0.34958286, 0.03766772, 0.12025286,
0.37698923, 0.3192629 , 0.79559338, 0.05385659, 0.21597037,
0.06383728, 0.00570991, 0.66018187, 0.35855286, 0.37783606,
0.36781398, 0.03554512, 0.21718203, 0.05943622, 0.06554485,
0.10081475, 0.00818713, 0.26591316, 0.42809675, 0.06542835,
0.09296803, 0.01259826, 0.14226651, 0.05072662, 0.02372258,
0.02586923, 0.06760427, 0.24315648, 0.26961432, 0.19831733,
0.2652296 , 0.0165923 , 0.15784236, 0.08398982, 0.02711775,
0.18750547, 0.00783535, 0.2844239 , 0.00270742, 0.02484969,
0.22585745, 0.72775605, 0.07691547, 0.26304359])
```

ROC_curve

```
In [75]: fpr,tpr,threshholds = roc_curve(y_test,probability)
```

```
In [76]: plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show()
```



DECISION TREE

```
In [77]: from sklearn.tree import DecisionTreeClassifier
dc=DecisionTreeClassifier()
```

```
In [78]: dc.fit(x_train,y_train)
```

```
Out[78]: ▾ DecisionTreeClassifier
DecisionTreeClassifier()
```

```
In [79]: pred=dc.predict(x_test)
```

```
In [80]: pred
```

```
Out[80]: array([0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0,
0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1,
0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 1, 0, 0, 0, 0, 0, 0])
```

```
In [81]: y_test
```

```
Out[81]:
```

Attrition	
442	0
1091	0
981	1
785	0
1332	1
...	...
1439	0
481	0
124	1
198	0
1229	0

294 rows × 1 columns

```
In [82]: dataset
```

Out[82]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSat
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	
...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	

1470 rows × 35 columns



In [85]:

```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

In [86]:

```
accuracy_score(y_test,pred)
```

Out[86]: 0.7517006802721088

In [87]:

```
confusion_matrix(y_test,pred)
```

Out[87]: array([[204, 41],
[32, 17]], dtype=int64)

In [89]:

```
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.86	0.83	0.85	245
1	0.29	0.35	0.32	49
accuracy			0.75	294
macro avg	0.58	0.59	0.58	294
weighted avg	0.77	0.75	0.76	294

In [91]:

```
probability=dc.predict_proba(x_test)[:,1]
```

In [92]:

```
probability
```

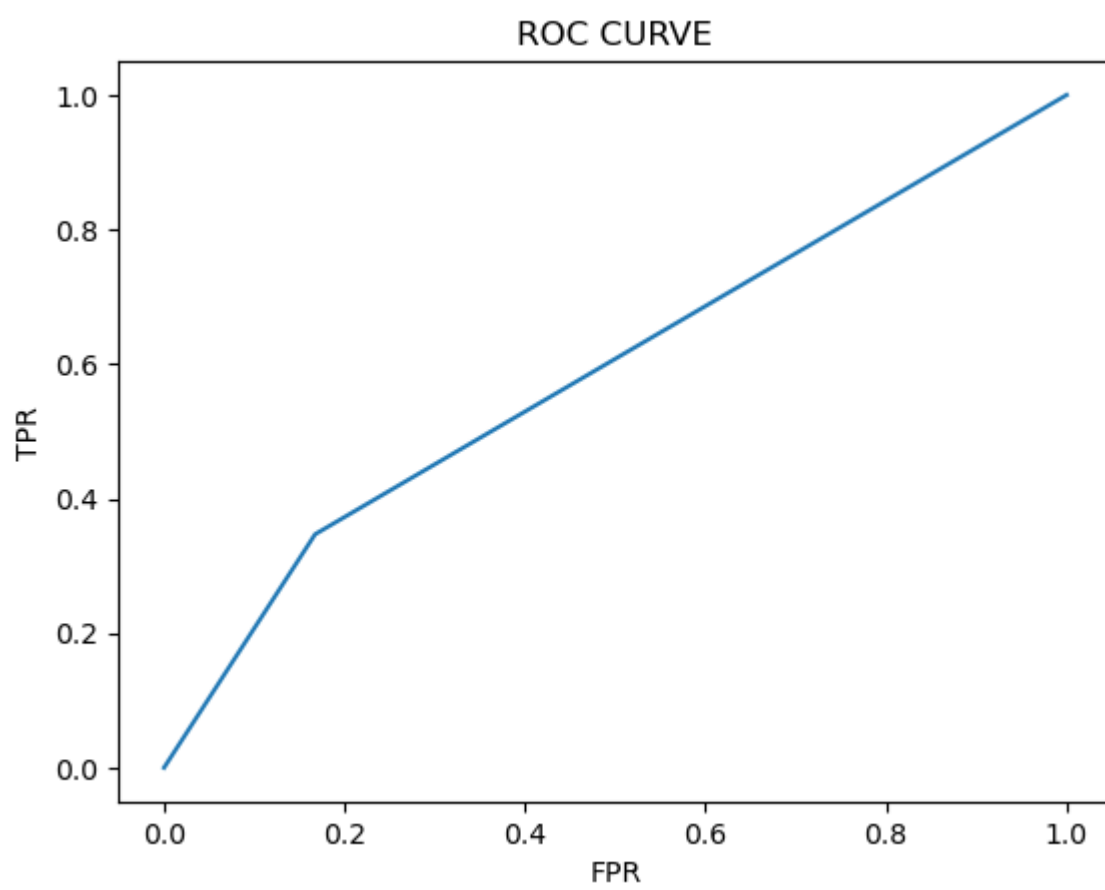
Out[92]: array([0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1.,
1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1., 0., 0., 0.,
0., 0., 0., 1., 0., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
0., 1., 1., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
1., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0., 1., 1., 1., 0.,
0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 1., 0., 1., 1., 0., 0., 0., 0., 1., 0., 0., 1.,
0., 0., 1., 0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0.,
1., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1., 1., 0.,
0., 1., 0., 0., 1., 0., 1., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0.,
0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0.,
0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 1., 0., 1., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 1., 0., 1., 1.,
0., 0., 0., 0., 0.])

In [93]:

```
fpr,tpr,threshsholds = roc_curve(y_test,probability)
```

In [94]:

```
plt.plot(fpr,tpr)  
plt.xlabel('FPR')  
plt.ylabel('TPR')  
plt.title('ROC CURVE')  
plt.show()
```



Hyper Parameter Tuning

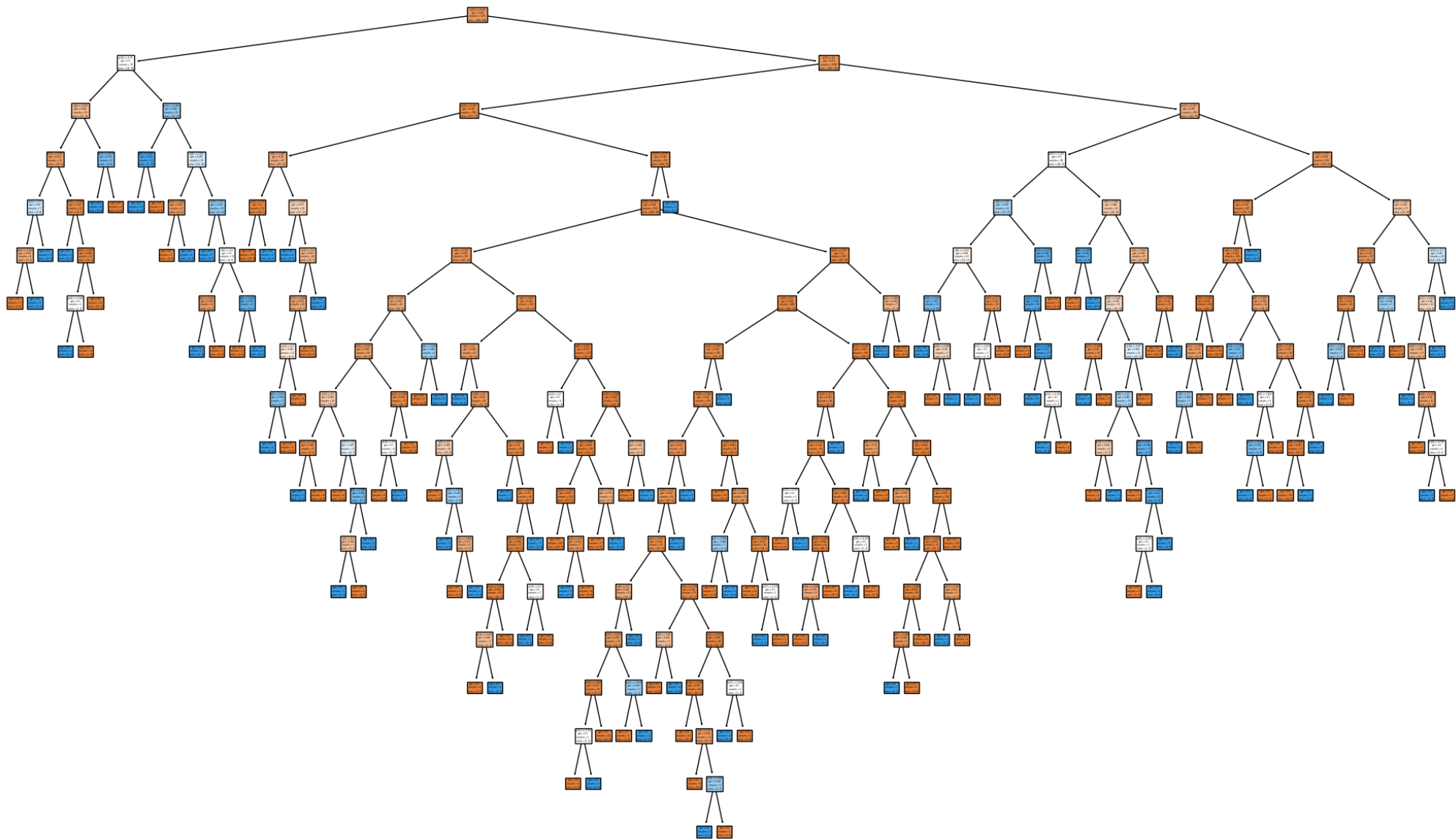
```
In [96]: from sklearn import tree
plt.figure(figsize=(25,15))
tree.plot_tree(dc,filled=True)
```


Out[96]: [Text(0.3255570652173913, 0.9722222222222222, 'x[27] <= 0.038\ngini = 0.269\nsamples = 1176\nvalue = [988, 188]'),
Text(0.08347826086956522, 0.9166666666666666, 'x[16] <= 0.75\ngini = 0.5\nsamples = 78\nvalue = [39, 39]'),
Text(0.05217391304347826, 0.8611111111111112, 'x[4] <= 0.554\ngini = 0.426\nsamples = 39\nvalue = [27, 12]'),
Text(0.034782608695652174, 0.8055555555555556, 'x[15] <= 0.167\ngini = 0.312\nsamples = 31\nvalue = [25, 6]'),
Text(0.020869565217391306, 0.75, 'x[21] <= 0.5\ngini = 0.49\nsamples = 7\nvalue = [3, 4]'),
Text(0.01391304347826087, 0.6944444444444444, 'x[22] <= 0.321\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.006956521739130435, 0.6388888888888888, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.020869565217391306, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.02782608695652174, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.04869565217391304, 0.75, 'x[19] <= 0.056\ngini = 0.153\nsamples = 24\nvalue = [22, 2]'),
Text(0.04173913043478261, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.05565217391304348, 0.6944444444444444, 'x[9] <= 0.167\ngini = 0.083\nsamples = 23\nvalue = [22, 1]'),
Text(0.04869565217391304, 0.6388888888888888, 'x[28] <= 0.583\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.04173913043478261, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.05565217391304348, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.06260869565217392, 0.6388888888888888, 'gini = 0.0\nsamples = 21\nvalue = [21, 0]'),
Text(0.06956521739130435, 0.8055555555555556, 'x[22] <= 0.679\ngini = 0.375\nsamples = 8\nvalue = [2, 6]'),
Text(0.06260869565217392, 0.75, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.07652173913043478, 0.75, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.11478260869565217, 0.8611111111111112, 'x[11] <= 0.364\ngini = 0.426\nsamples = 39\nvalue = [12, 27]'),
Text(0.09739130434782609, 0.8055555555555556, 'x[0] <= 0.369\ngini = 0.133\nsamples = 14\nvalue = [1, 13]'),
Text(0.09043478260869565, 0.75, 'gini = 0.0\nsamples = 13\nvalue = [0, 13]'),
Text(0.10434782608695652, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.13217391304347825, 0.8055555555555556, 'x[8] <= 0.105\ngini = 0.493\nsamples = 25\nvalue = [11, 14]'),
Text(0.11826086956521739, 0.75, 'x[22] <= 0.464\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.11130434782608696, 0.6944444444444444, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.12521739130434784, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.14608695652173914, 0.75, 'x[15] <= 0.5\ngini = 0.432\nsamples = 19\nvalue = [6, 13]'),
Text(0.1391304347826087, 0.6944444444444444, 'gini = 0.0\nsamples = 7\nvalue = [0, 7]'),
Text(0.15304347826086956, 0.6944444444444444, 'x[6] <= 0.4\ngini = 0.5\nsamples = 12\nvalue = [6, 6]'),
Text(0.1391304347826087, 0.6388888888888888, 'x[5] <= 0.125\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.13217391304347825, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.14608695652173914, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.16695652173913045, 0.6388888888888888, 'x[8] <= 0.249\ngini = 0.278\nsamples = 6\nvalue = [1, 5]'),
Text(0.16, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.17391304347826086, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue = [0, 5]'),
Text(0.5676358695652174, 0.9166666666666666, 'x[21] <= 0.5\ngini = 0.235\nsamples = 1098\nvalue = [949, 149]'),
Text(0.3196195652173913, 0.8611111111111112, 'x[29] <= 0.167\ngini = 0.162\nsamples = 798\nvalue = [727, 71]'),
Text(0.18782608695652173, 0.8055555555555556, 'x[8] <= 0.445\ngini = 0.38\nsamples = 47\nvalue = [35, 12]'),
Text(0.17391304347826086, 0.75, 'x[16] <= 0.75\ngini = 0.1\nsamples = 19\nvalue = [18, 1]'),
Text(0.16695652173913045, 0.6944444444444444, 'gini = 0.0\nsamples = 18\nvalue = [18, 0]'),
Text(0.1808695652173913, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.20173913043478262, 0.75, 'x[17] <= 0.094\ngini = 0.477\nsamples = 28\nvalue = [17, 11]'),
Text(0.19478260869565217, 0.6944444444444444, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.20869565217391303, 0.6944444444444444, 'x[32] <= 0.6\ngini = 0.413\nsamples = 24\nvalue = [17, 7]'),
Text(0.20173913043478262, 0.6388888888888888, 'x[11] <= 0.486\ngini = 0.351\nsamples = 22\nvalue = [17, 5]'),
Text(0.19478260869565217, 0.5833333333333334, 'x[24] <= 0.5\ngini = 0.496\nsamples = 11\nvalue = [6, 5]'),
Text(0.18782608695652173, 0.5277777777777778, 'x[0] <= 0.56\ngini = 0.408\nsamples = 7\nvalue = [2, 5]'),
Text(0.1808695652173913, 0.4722222222222222, 'gini = 0.0\nsamples = 5\nvalue = [0, 5]'),
Text(0.19478260869565217, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.20173913043478262, 0.5277777777777778, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.20869565217391303, 0.5833333333333334, 'gini = 0.0\nsamples = 11\nvalue = [11, 0]'),
Text(0.21565217391304348, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.4514130434782609, 0.8055555555555556, 'x[27] <= 0.975\ngini = 0.145\nsamples = 751\nvalue = [692, 59]'),
Text(0.4444565217391304, 0.75, 'x[30] <= 0.113\ngini = 0.143\nsamples = 750\nvalue = [692, 58]'),
Text(0.3143478260869565, 0.6944444444444444, 'x[9] <= 0.167\ngini = 0.218\nsamples = 257\nvalue = [225, 32]'),
Text(0.26956521739130435, 0.6388888888888888, 'x[33] <= 0.147\ngini = 0.355\nsamples = 65\nvalue = [50, 15]'),
Text(0.24695652173913044, 0.5833333333333334, 'x[33] <= 0.029\ngini = 0.303\nsamples = 59\nvalue = [48, 11]'),
Text(0.22260869565217392, 0.5277777777777778, 'x[12] <= 0.5\ngini = 0.463\nsamples = 22\nvalue = [14, 8]'),
Text(0.20869565217391303, 0.4722222222222222, 'x[11] <= 0.179\ngini = 0.198\nsamples = 9\nvalue = [8, 1]'),
Text(0.20173913043478262, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.21565217391304348, 0.4166666666666667, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.23652173913043478, 0.4722222222222222, 'x[11] <= 0.4\ngini = 0.497\nsamples = 13\nvalue = [6, 7]'),
Text(0.22956521739130434, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.24347826086956523, 0.4166666666666667, 'x[4] <= 0.286\ngini = 0.346\nsamples = 9\nvalue = [2, 7]'),
Text(0.23652173913043478, 0.3611111111111111, 'x[5] <= 0.5\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.22956521739130434, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.24347826086956523, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.25043478260869567, 0.3611111111111111, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.271304347826087, 0.5277777777777778, 'x[15] <= 0.167\ngini = 0.149\nsamples = 37\nvalue = [34, 3]'),
Text(0.2643478260869565, 0.4722222222222222, 'x[29] <= 0.5\ngini = 0.5\nsamples = 6\nvalue = [3, 3]'),
Text(0.2573913043478261, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.271304347826087, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.2782608695652174, 0.4722222222222222, 'gini = 0.0\nsamples = 31\nvalue = [31, 0]'),
Text(0.2921739130434783, 0.5833333333333334, 'x[8] <= 0.065\ngini = 0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.2852173913043478, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.2991304347826087, 0.5277777777777778, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.3591304347826087, 0.6388888888888888, 'x[0] <= 0.321\ngini = 0.161\nsamples = 192\nvalue = [175, 17]'),
Text(0.32, 0.5833333333333334, 'x[6] <= 0.1\ngini = 0.294\nsamples = 67\nvalue = [55, 12]'),
Text(0.3130434782608696, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.3269565217391304, 0.5277777777777778, 'x[29] <= 0.5\ngini = 0.26\nsamples = 65\nvalue = [55, 10]'),
Text(0.3026086956521739, 0.4722222222222222, 'x[6] <= 0.5\ngini = 0.469\nsamples = 16\nvalue = [10, 6]'),
Text(0.2956521739130435, 0.4166666666666667, 'gini = 0.0\nsamples = 7\nvalue = [7, 0]'),
Text(0.3095652173913043, 0.4166666666666667, 'x[9] <= 0.833\ngini = 0.444\nsamples = 9\nvalue = [3, 6]'),
Text(0.3026086956521739, 0.3611111111111111, 'gini = 0.0\nsamples = 5\nvalue = [0, 5]'),
Text(0.3165217391304348, 0.3611111111111111, 'x[2] <= 0.566\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.3095652173913043, 0.3055555555555556, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.3234782608695652, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.35130434782608694, 0.4722222222222222, 'x[2] <= 0.037\ngini = 0.15\nsamples = 49\nvalue = [45, 4]'),
Text(0.3443478260869565, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.3582608695652174, 0.4166666666666667, 'x[2] <= 0.938\ngini = 0.117\nsamples = 48\nvalue = [45, 3]'),
Text(0.35130434782608694, 0.3611111111111111, 'x[5] <= 0.875\ngini = 0.081\nsamples = 47\nvalue = [45, 2]'),
Text(0.3373913043478261, 0.3055555555555556, 'x[12] <= 0.167\ngini = 0.043\nsamples = 45\nvalue = [44, 1]'),
Text(0.33043478260869563, 0.25, 'x[14] <= 0.625\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.3234782608695652, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),

Text(0.3373913043478261, 0.1944444444444445, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.3443478260869565, 0.25, 'gini = 0.0\nsamples = 42\nvalue = [42, 0]'),
Text(0.3652173913043478, 0.3055555555555556, 'x[10] <= 0.5\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.3582608695652174, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.37217391304347824, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.3652173913043478, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.3982608695652174, 0.5833333333333334, 'x[8] <= 0.022\ngini = 0.077\nsamples = 125\nvalue = [120, 5]'),
Text(0.3791304347826087, 0.5277777777777778, 'x[14] <= 0.5\ngini = 0.5\nsamples = 4\nvalue = [2, 2]'),
Text(0.37217391304347824, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.38608695652173913, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.41739130434782606, 0.5277777777777778, 'x[18] <= 0.968\ngini = 0.048\nsamples = 121\nvalue = [118, 3]'),
Text(0.4, 0.4722222222222222, 'x[2] <= 0.98\ngini = 0.033\nsamples = 118\nvalue = [116, 2]'),
Text(0.38608695652173913, 0.4166666666666667, 'x[14] <= 0.938\ngini = 0.017\nsamples = 114\nvalue = [113, 1]'),
Text(0.3791304347826087, 0.3611111111111111, 'gini = 0.0\nsamples = 107\nvalue = [107, 0]'),
Text(0.39304347826086955, 0.3611111111111111, 'x[16] <= 0.25\ngini = 0.245\nsamples = 7\nvalue = [6, 1]'),
Text(0.38608695652173913, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4, 0.3055555555555556, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),
Text(0.41391304347826086, 0.4166666666666667, 'x[3] <= 0.75\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.40695652173913044, 0.3611111111111111, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.42086956521739133, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.43478260869565216, 0.4722222222222222, 'x[1] <= 0.75\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.42782608695652175, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.44173913043478263, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5745652173913044, 0.6944444444444444, 'x[30] <= 0.787\ngini = 0.1\nsamples = 493\nvalue = [467, 26]'),
Text(0.538695652173913, 0.6388888888888888, 'x[15] <= 0.5\ngini = 0.094\nsamples = 486\nvalue = [462, 24]'),
Text(0.48782608695652174, 0.5833333333333334, 'x[14] <= 0.938\ngini = 0.154\nsamples = 191\nvalue = [175, 16]'),
Text(0.4808695652173913, 0.5277777777777778, 'x[18] <= 0.481\ngini = 0.145\nsamples = 190\nvalue = [175, 15]'),
Text(0.46260869565217394, 0.4722222222222222, 'x[33] <= 0.794\ngini = 0.221\nsamples = 95\nvalue = [83, 12]'),
Text(0.45565217391304347, 0.4166666666666667, 'x[18] <= 0.47\ngini = 0.207\nsamples = 94\nvalue = [83, 11]'),
Text(0.44869565217391305, 0.3611111111111111, 'x[5] <= 0.375\ngini = 0.192\nsamples = 93\nvalue = [83, 10]'),
Text(0.4260869565217391, 0.3055555555555556, 'x[6] <= 0.9\ngini = 0.363\nsamples = 21\nvalue = [16, 5]'),
Text(0.4191304347826087, 0.25, 'x[17] <= 0.413\ngini = 0.266\nsamples = 19\nvalue = [16, 3]'),
Text(0.4052173913043478, 0.1944444444444445, 'x[19] <= 0.056\ngini = 0.117\nsamples = 16\nvalue = [15, 1]'),
Text(0.3982608695652174, 0.1388888888888889, 'x[15] <= 0.167\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.391304347826087, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.4052173913043478, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4121739130434783, 0.1388888888888889, 'gini = 0.0\nsamples = 14\nvalue = [14, 0]'),
Text(0.4330434782608696, 0.1944444444444445, 'x[24] <= 0.833\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4260869565217391, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.44, 0.1388888888888889, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.4330434782608696, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.47130434782608693, 0.3055555555555556, 'x[31] <= 0.139\ngini = 0.129\nsamples = 72\nvalue = [67, 5]'),
Text(0.4539130434782609, 0.25, 'x[8] <= 0.68\ngini = 0.444\nsamples = 6\nvalue = [4, 2]'),
Text(0.4469565217391304, 0.1944444444444445, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.4608695652173913, 0.1944444444444445, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.48869565217391303, 0.25, 'x[2] <= 0.958\ngini = 0.087\nsamples = 66\nvalue = [63, 3]'),
Text(0.4747826086956522, 0.1944444444444445, 'x[28] <= 0.583\ngini = 0.061\nsamples = 64\nvalue = [62, 2]'),
Text(0.4678260869565217, 0.1388888888888889, 'gini = 0.0\nsamples = 52\nvalue = [52, 0]'),
Text(0.4817391304347826, 0.1388888888888889, 'x[3] <= 0.75\ngini = 0.278\nsamples = 12\nvalue = [10, 2]'),
Text(0.4747826086956522, 0.08333333333333333, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),
Text(0.48869565217391303, 0.08333333333333333, 'x[18] <= 0.353\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4817391304347826, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.4956521739130435, 0.027777777777777776, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5026086956521739, 0.1944444444444445, 'x[22] <= 0.286\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.4956521739130435, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5095652173913043, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.46260869565217394, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.46956521739130436, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4991304347826087, 0.4722222222222222, 'x[19] <= 0.5\ngini = 0.061\nsamples = 95\nvalue = [92, 3]'),
Text(0.49217391304347824, 0.4166666666666667, 'gini = 0.0\nsamples = 76\nvalue = [76, 0]'),
Text(0.5060869565217392, 0.4166666666666667, 'x[8] <= 0.161\ngini = 0.266\nsamples = 19\nvalue = [16, 3]'),
Text(0.49217391304347824, 0.3611111111111111, 'x[2] <= 0.096\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4852173913043478, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.4991304347826087, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.52, 0.3611111111111111, 'x[31] <= 0.639\ngini = 0.117\nsamples = 16\nvalue = [15, 1]'),
Text(0.5130434782608696, 0.3055555555555556, 'gini = 0.0\nsamples = 14\nvalue = [14, 0]'),
Text(0.5269565217391304, 0.3055555555555556, 'x[27] <= 0.537\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.52, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5339130434782609, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.49478260869565216, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5895652173913043, 0.5833333333333334, 'x[22] <= 0.036\ngini = 0.053\nsamples = 295\nvalue = [287, 8]'),
Text(0.5652173913043478, 0.5277777777777778, 'x[32] <= 0.7\ngini = 0.159\nsamples = 46\nvalue = [42, 4]'),
Text(0.5582608695652174, 0.4722222222222222, 'x[12] <= 0.167\ngini = 0.124\nsamples = 45\nvalue = [42, 3]'),
Text(0.5408695652173913, 0.4166666666666667, 'x[13] <= 0.375\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.5339130434782609, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5478260869565217, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5756521739130435, 0.4166666666666667, 'x[27] <= 0.688\ngini = 0.089\nsamples = 43\nvalue = [41, 2]'),
Text(0.5617391304347826, 0.3611111111111111, 'x[14] <= 0.062\ngini = 0.048\nsamples = 41\nvalue = [40, 1]'),
Text(0.5547826086956522, 0.3055555555555556, 'x[18] <= 0.346\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.5478260869565217, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.5617391304347826, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.568695652173913, 0.3055555555555556, 'gini = 0.0\nsamples = 37\nvalue = [37, 0]'),
Text(0.5895652173913043, 0.3611111111111111, 'x[30] <= 0.212\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.5826086956521739, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5965217391304348, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5721739130434783, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6139130434782609, 0.5277777777777778, 'x[17] <= 0.056\ngini = 0.032\nsamples = 249\nvalue = [245, 4]'),
Text(0.5965217391304348, 0.4722222222222222, 'x[8] <= 0.33\ngini = 0.32\nsamples = 5\nvalue = [4, 1]'),
Text(0.5895652173913043, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6034782608695652, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.6313043478260869, 0.4722222222222222, 'x[2] <= 0.015\ngini = 0.024\nsamples = 244\nvalue = [241, 3]'),
Text(0.6173913043478261, 0.4166666666666667, 'x[22] <= 0.714\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.6104347826086957, 0.3611111111111111, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.6243478260869565, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6452173913043479, 0.4166666666666667, 'x[24] <= 0.167\ngini = 0.017\nsamples = 238\nvalue = [236, 2]'),

Text(0.6382608695652174, 0.3611111111111111, 'x[29] <= 0.833\ngini = 0.073\nsamples = 53\nvalue = [51, 2]'),
Text(0.6243478260869565, 0.3055555555555556, 'x[33] <= 0.088\ngini = 0.041\nsamples = 48\nvalue = [47, 1]'),
Text(0.6173913043478261, 0.25, 'x[0] <= 0.345\ngini = 0.245\nsamples = 7\nvalue = [6, 1]'),
Text(0.6104347826086957, 0.1944444444444445, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6243478260869565, 0.1944444444444445, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),
Text(0.6313043478260869, 0.25, 'gini = 0.0\nsamples = 41\nvalue = [41, 0]'),
Text(0.6521739130434783, 0.3055555555555556, 'x[17] <= 0.38\ngini = 0.32\nsamples = 5\nvalue = [4, 1]'),
Text(0.6452173913043479, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6591304347826087, 0.25, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.6521739130434783, 0.3611111111111111, 'gini = 0.0\nsamples = 185\nvalue = [185, 0]'),
Text(0.6104347826086957, 0.6388888888888888, 'x[2] <= 0.366\ngini = 0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.6034782608695652, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.6173913043478261, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.4583695652173913, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.8156521739130435, 0.8611111111111112, 'x[17] <= 0.157\ngini = 0.385\nsamples = 300\nvalue = [222, 78]'),
Text(0.7243478260869565, 0.8055555555555556, 'x[26] <= 0.167\ngini = 0.5\nsamples = 96\nvalue = [49, 47]'),
Text(0.6869565217391305, 0.75, 'x[4] <= 0.161\ngini = 0.459\nsamples = 42\nvalue = [15, 27]'),
Text(0.6591304347826087, 0.6944444444444444, 'x[8] <= 0.415\ngini = 0.499\nsamples = 23\nvalue = [12, 11]'),
Text(0.6382608695652174, 0.6388888888888888, 'x[18] <= 0.561\ngini = 0.355\nsamples = 13\nvalue = [3, 10]'),
Text(0.6313043478260869, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]'),
Text(0.6452173913043479, 0.5833333333333334, 'x[28] <= 0.583\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.6382608695652174, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.6521739130434783, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.68, 0.6388888888888888, 'x[27] <= 0.1\ngini = 0.18\nsamples = 10\nvalue = [9, 1]'),
Text(0.6730434782608695, 0.5833333333333334, 'x[11] <= 0.457\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.6660869565217391, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.68, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.6869565217391305, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.7147826086956521, 0.6944444444444444, 'x[13] <= 0.125\ngini = 0.266\nsamples = 19\nvalue = [3, 16]'),
Text(0.7078260869565217, 0.6388888888888888, 'x[11] <= 0.2\ngini = 0.198\nsamples = 18\nvalue = [2, 16]'),
Text(0.7008695652173913, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7147826086956521, 0.5833333333333334, 'x[31] <= 0.306\ngini = 0.111\nsamples = 17\nvalue = [1, 16]'),
Text(0.7078260869565217, 0.5277777777777778, 'gini = 0.0\nsamples = 15\nvalue = [0, 15]'),
Text(0.7217391304347827, 0.5277777777777778, 'x[12] <= 0.333\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.7147826086956521, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7286956521739131, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7217391304347827, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7617391304347826, 0.75, 'x[0] <= 0.202\ngini = 0.466\nsamples = 54\nvalue = [34, 20]'),
Text(0.7426086956521739, 0.6944444444444444, 'x[8] <= 0.164\ngini = 0.245\nsamples = 7\nvalue = [1, 6]'),
Text(0.7356521739130435, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7495652173913043, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.7808695652173913, 0.6944444444444444, 'x[2] <= 0.622\ngini = 0.418\nsamples = 47\nvalue = [33, 14]'),
Text(0.7634782608695653, 0.6388888888888888, 'x[2] <= 0.145\ngini = 0.482\nsamples = 32\nvalue = [19, 13]'),
Text(0.7495652173913043, 0.5833333333333334, 'x[17] <= 0.068\ngini = 0.18\nsamples = 10\nvalue = [9, 1]'),
Text(0.7426086956521739, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7565217391304347, 0.5277777777777778, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),
Text(0.7773913043478261, 0.5833333333333334, 'x[18] <= 0.87\ngini = 0.496\nsamples = 22\nvalue = [10, 12]'),
Text(0.7704347826086957, 0.5277777777777778, 'x[8] <= 0.41\ngini = 0.465\nsamples = 19\nvalue = [7, 12]'),
Text(0.7565217391304347, 0.4722222222222222, 'x[18] <= 0.715\ngini = 0.469\nsamples = 8\nvalue = [5, 3]'),
Text(0.7495652173913043, 0.4166666666666667, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.7634782608695653, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.7843478260869565, 0.4722222222222222, 'x[0] <= 0.25\ngini = 0.298\nsamples = 11\nvalue = [2, 9]'),
Text(0.7773913043478261, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7913043478260869, 0.4166666666666667, 'x[18] <= 0.202\ngini = 0.18\nsamples = 10\nvalue = [1, 9]'),
Text(0.7843478260869565, 0.3611111111111111, 'x[28] <= 0.417\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.7773913043478261, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7913043478260869, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7982608695652174, 0.3611111111111111, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]'),
Text(0.7843478260869565, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.7982608695652174, 0.6388888888888888, 'x[19] <= 0.944\ngini = 0.124\nsamples = 15\nvalue = [14, 1]'),
Text(0.7913043478260869, 0.5833333333333334, 'gini = 0.0\nsamples = 14\nvalue = [14, 0]'),
Text(0.8052173913043478, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.9069565217391304, 0.8055555555555556, 'x[16] <= 0.75\ngini = 0.258\nsamples = 204\nvalue = [173, 31]'),
Text(0.8521739130434782, 0.75, 'x[17] <= 0.992\ngini = 0.138\nsamples = 147\nvalue = [136, 11]'),
Text(0.8452173913043478, 0.6944444444444444, 'x[4] <= 0.482\ngini = 0.128\nsamples = 146\nvalue = [136, 10]'),
Text(0.8260869565217391, 0.6388888888888888, 'x[30] <= 0.063\ngini = 0.038\nsamples = 104\nvalue = [102, 2]'),
Text(0.8191304347826087, 0.5833333333333334, 'x[11] <= 0.193\ngini = 0.32\nsamples = 10\nvalue = [8, 2]'),
Text(0.8121739130434783, 0.5277777777777778, 'x[6] <= 0.7\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.8052173913043478, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.8191304347826087, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.8260869565217391, 0.5277777777777778, 'gini = 0.0\nsamples = 7\nvalue = [7, 0]'),
Text(0.8330434782608696, 0.5833333333333334, 'gini = 0.0\nsamples = 94\nvalue = [94, 0]'),
Text(0.8643478260869565, 0.6388888888888888, 'x[9] <= 0.167\ngini = 0.308\nsamples = 42\nvalue = [34, 8]'),
Text(0.8469565217391304, 0.5833333333333334, 'x[18] <= 0.194\ngini = 0.375\nsamples = 4\nvalue = [1, 3]'),
Text(0.84, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.8539130434782609, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.8817391304347826, 0.5833333333333334, 'x[0] <= 0.393\ngini = 0.229\nsamples = 38\nvalue = [33, 5]'),
Text(0.8678260869565217, 0.5277777777777778, 'x[4] <= 0.821\ngini = 0.5\nsamples = 6\nvalue = [3, 3]'),
Text(0.8608695652173913, 0.4722222222222222, 'x[22] <= 0.643\ngini = 0.375\nsamples = 4\nvalue = [1, 3]'),
Text(0.8539130434782609, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.8678260869565217, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.8747826086956522, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.8956521739130435, 0.5277777777777778, 'x[28] <= 0.917\ngini = 0.117\nsamples = 32\nvalue = [30, 2]'),
Text(0.888695652173913, 0.4722222222222222, 'x[8] <= 0.992\ngini = 0.062\nsamples = 31\nvalue = [30, 1]'),
Text(0.8817391304347826, 0.4166666666666667, 'gini = 0.0\nsamples = 30\nvalue = [30, 0]'),
Text(0.8956521739130435, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.9026086956521739, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.8591304347826086, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.9617391304347827, 0.75, 'x[14] <= 0.812\ngini = 0.456\nsamples = 57\nvalue = [37, 20]'),
Text(0.9373913043478261, 0.6944444444444444, 'x[32] <= 0.4\ngini = 0.238\nsamples = 29\nvalue = [25, 4]'),
Text(0.9234782608695652, 0.6388888888888888, 'x[8] <= 0.071\ngini = 0.142\nsamples = 26\nvalue = [24, 2]'),
Text(0.9165217391304348, 0.5833333333333334, 'x[33] <= 0.412\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.9095652173913044, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.9234782608695652, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.9304347826086956, 0.5833333333333334, 'gini = 0.0\nsamples = 23\nvalue = [23, 0]'),

Text(0.951304347826087, 0.6388888888888888, 'x[29] <= 0.833\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.9443478260869566, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.9582608695652174, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.9860869565217392, 0.6944444444444444, 'x[32] <= 0.1\ngini = 0.49\nsamples = 28\nvalue = [12, 16]'),
Text(0.9791304347826087, 0.6388888888888888, 'x[12] <= 0.833\ngini = 0.48\nsamples = 20\nvalue = [12, 8]'),
Text(0.9721739130434782, 0.5833333333333334, 'x[30] <= 0.013\ngini = 0.415\nsamples = 17\nvalue = [12, 5]'),
Text(0.9652173913043478, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.9791304347826087, 0.5277777777777778, 'x[18] <= 0.505\ngini = 0.32\nsamples = 15\nvalue = [12, 3]'),
Text(0.9721739130434782, 0.4722222222222222, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),
Text(0.9860869565217392, 0.4722222222222222, 'x[18] <= 0.706\ngini = 0.5\nsamples = 6\nvalue = [3, 3]'),
Text(0.9791304347826087, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.9930434782608696, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.9860869565217392, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.9930434782608696, 0.6388888888888888, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]')]



```
In [97]: from sklearn.model_selection import GridSearchCV
parameter={
'criterion':['gini','entropy'],
'splitter':['best','random'],
'max_depth':[1,2,3,4,5],
'max_features':['auto', 'sqrt', 'log2']
}
```

```
In [99]: grid_search=GridSearchCV(estimator=dc,param_grid=parameter,cv=5,scoring="accuracy")
```

```
In [100]: grid_search.fit(x_train,y_train)
```

```
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_validation.py:425: FitFailedWarning:
100 fits failed out of a total of 300.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.

Below are more details about the failures:
-----
100 fits failed with the following error:
Traceback (most recent call last):
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py", line 1144, in wrapper
    estimator._validate_params()
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py", line 637, in _validate_params
    validate_parameter_constraints(
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\utils\_param_validation.py", line 95, in validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features' parameter of DecisionTreeClassifier must be an int in the range [1, inf), a float in the range (0.0, 1.0], a str among {'sqrt', 'log2'} or None. Got 'auto' instead.

    warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_search.py:976: UserWarning: One or more of the test scores are non-finite: [
      nan      nan  0.84013704  0.84013704  0.84013704  0.84013704
      nan      nan  0.83759106  0.84013704  0.84183195  0.84013704
      nan      nan  0.8349982   0.84182834  0.83927876  0.83247746
      nan      nan  0.83588893  0.84609448  0.84097367  0.83929679
      nan      nan  0.83335377  0.85288136  0.83843851  0.83503065
      nan      nan  0.84013704  0.84013704  0.84013704  0.84013704
      nan      nan  0.8316264   0.84013704  0.84269023  0.84013704
      nan      nan  0.8469203   0.84013704  0.84351965  0.83758384
      nan      nan  0.84268662  0.83844573  0.8435485   0.83673278
      nan      nan  0.82228994  0.847793    0.82569059  0.84183556]
    warnings.warn(
```

Out[100]:

GridSearchCV

estimator: DecisionTreeClassifier

DecisionTreeClassifier

In [101...

grid_search.best_params_

Out[101]:

{'criterion': 'gini',
'max_depth': 5,
'max_features': 'sqrt',
'splitter': 'random'}

In [102...

dc_cv=DecisionTreeClassifier(criterion= 'gini',
max_depth= 5,
max_features= 'log2',
splitter= 'random')

In [104...

dc_cv.fit(x_train,y_train)

Out[104]:

DecisionTreeClassifier

DecisionTreeClassifier(max_depth=5, max_features='log2', splitter='random')

In [105...

print(classification_report(y_test,pred))

	precision	recall	f1-score	support
0	0.86	0.83	0.85	245
1	0.29	0.35	0.32	49
accuracy			0.75	294
macro avg	0.58	0.59	0.58	294
weighted avg	0.77	0.75	0.76	294

RANDOM FOREST

In [107...

from sklearn.ensemble import RandomForestClassifier
rc=RandomForestClassifier()

In [108...

forest_params = [{'max_depth': list(range(10, 15)), 'max_features': list(range(0,14))}]

In [109...

rc_cv= GridSearchCV(rfc,param_grid=forest_params,cv=10,scoring="accuracy")

In [111...

rc_cv.fit(x_train,y_train)

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

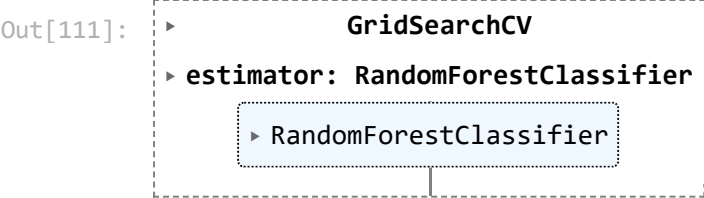
[illegible]


```
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_validation.py:425: FitFailedWarning:
50 fits failed out of a total of 700.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.
```

Below are more details about the failures:

```
-----
50 fits failed with the following error:
Traceback (most recent call last):
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py", line 1144, in wrapper
    estimator._validate_params()
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py", line 637, in _validate_params
    validate_parameter_constraints(
  File "C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\utils\_param_validation.py", line 95, in validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features' parameter of RandomForestClassifier must be an int in the range
[1, inf), a float in the range (0.0, 1.0], a str among {'sqrt', 'log2'} or None. Got 0 instead.

    warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\model_selection\_search.py:976: UserWarning: One or more of the test scores are non-fin
ite: [
      nan 0.84779806 0.85373751 0.85290453 0.85627264 0.8596842
0.85967695 0.86222657 0.86050992 0.85796755 0.86138635 0.85882225
0.85966247 0.85883674          nan 0.85119513 0.85628712 0.85545415
0.85969144 0.85627264 0.86734029 0.85714182 0.86222657 0.85713458
0.86221932 0.8596842  0.86307403 0.86307403          nan 0.84949297
0.85798928 0.85714907 0.85969144 0.85798928 0.85969144 0.85967695
0.8605389  0.86050992 0.85882225 0.85966247 0.85627988 0.86052441
      nan 0.85120238 0.85629436 0.85714182 0.85798928 0.86053165
0.86306678 0.85965522 0.85543242 0.85797479 0.85627264 0.85627988
0.85966971 0.85796031          nan 0.85034767 0.85290453 0.85544691
0.85798204 0.85799652 0.86051717 0.85543242 0.85966247 0.85797479
0.85204259 0.85966971 0.85710561 0.86054614]
    warnings.warn(
C:\Users\hansi\anaconda_3\Lib\site-packages\sklearn\base.py:1151: DataConversionWarning: A column-vector y was passed when a 1d array was e
xpected. Please change the shape of y to (n_samples,), for example using ravel().
    return fit_method(estimator, *args, **kwargs)
```



In [112...

pred=rc_cv.predict(x_test)

In [113...

print(classification_report(y_test,pred))

	precision	recall	f1-score	support
0	0.86	1.00	0.92	245
1	0.89	0.16	0.28	49
accuracy			0.86	294
macro avg	0.87	0.58	0.60	294
weighted avg	0.86	0.86	0.81	294

In [114...

rc_cv.best_params_

Out[114]: {'max_depth': 11, 'max_features': 6}

In []: