

assignment-22-sep-1

September 28, 2023

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

```
[2]: df=pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
```

```
[3]: df.head()
```

```
[3]:   Age Attrition   BusinessTravel   DailyRate   Department \
0   41      Yes   Travel_Rarely      1102      Sales
1   49      No   Travel_Frequently      279  Research & Development
2   37      Yes   Travel_Rarely      1373  Research & Development
3   33      No   Travel_Frequently      1392  Research & Development
4   27      No   Travel_Rarely      591   Research & Development

      DistanceFromHome   Education   EducationField   EmployeeCount   EmployeeNumber \
0              1          2   Life Sciences              1              1
1              8          1   Life Sciences              1              2
2              2          2          Other              1              4
3              3          4   Life Sciences              1              5
4              2          1          Medical              1              7

      ...   RelationshipSatisfaction   StandardHours   StockOptionLevel \
0   ...              1              80              0
1   ...              4              80              1
2   ...              2              80              0
3   ...              3              80              0
4   ...              4              80              1

      TotalWorkingYears   TrainingTimesLastYear   WorkLifeBalance   YearsAtCompany \
0              8              0              1              6
1             10              3              3             10
2              7              3              3              0
3              8              3              3              8
4              6              3              3              2
```

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
4	2	2	2

[5 rows x 35 columns]

```
[4]: df.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

```

```
[5]: df.shape
```

```
[5]: (1470, 35)
```

```
[6]: df. Attrition.value_counts()
```

```

[6]: No      1233
     Yes      237
     Name: Attrition, dtype: int64

```

```
[7]: df.corr()
```

<ipython-input-7-2f6f6606aa2c>: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.

```
df.corr()
```

```

[7]:
      Age  DailyRate  DistanceFromHome  Education \
Age      1.000000   0.010661         -0.001686   0.208034
DailyRate 0.010661   1.000000         -0.004985  -0.016806
DistanceFromHome -0.001686 -0.004985         1.000000   0.021042
Education  0.208034 -0.016806         0.021042   1.000000
EmployeeCount      NaN         NaN           NaN         NaN
EmployeeNumber -0.010145 -0.050990         0.032916   0.042070
EnvironmentSatisfaction 0.010146  0.018355        -0.016075  -0.027128
HourlyRate      0.024287  0.023381         0.031131   0.016775
JobInvolvement  0.029820  0.046135         0.008783   0.042438
JobLevel       0.509604  0.002966         0.005303   0.101589
JobSatisfaction -0.004892  0.030571        -0.003669  -0.011296
MonthlyIncome   0.497855  0.007707        -0.017014   0.094961
MonthlyRate     0.028051 -0.032182         0.027473  -0.026084
NumCompaniesWorked  0.299635  0.038153        -0.029251   0.126317
PercentSalaryHike  0.003634  0.022704         0.040235  -0.011111
PerformanceRating  0.001904  0.000473         0.027110  -0.024539
RelationshipSatisfaction 0.053535  0.007846         0.006557  -0.009118
StandardHours      NaN         NaN           NaN         NaN
StockOptionLevel  0.037510  0.042143         0.044872   0.018422
TotalWorkingYears  0.680381  0.014515         0.004628   0.148280
TrainingTimesLastYear -0.019621  0.002453        -0.036942  -0.025100
WorkLifeBalance  -0.021490 -0.037848        -0.026556   0.009819

```

YearsAtCompany	0.311309	-0.034055	0.009508	0.069114
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065

	EmployeeCount	EmployeeNumber \
Age	NaN	-0.010145
DailyRate	NaN	-0.050990
DistanceFromHome	NaN	0.032916
Education	NaN	0.042070
EmployeeCount	NaN	NaN
EmployeeNumber	NaN	1.000000
EnvironmentSatisfaction	NaN	0.017621
HourlyRate	NaN	0.035179
JobInvolvement	NaN	-0.006888
JobLevel	NaN	-0.018519
JobSatisfaction	NaN	-0.046247
MonthlyIncome	NaN	-0.014829
MonthlyRate	NaN	0.012648
NumCompaniesWorked	NaN	-0.001251
PercentSalaryHike	NaN	-0.012944
PerformanceRating	NaN	-0.020359
RelationshipSatisfaction	NaN	-0.069861
StandardHours	NaN	NaN
StockOptionLevel	NaN	0.062227
TotalWorkingYears	NaN	-0.014365
TrainingTimesLastYear	NaN	0.023603
WorkLifeBalance	NaN	0.010309
YearsAtCompany	NaN	-0.011240
YearsInCurrentRole	NaN	-0.008416
YearsSinceLastPromotion	NaN	-0.009019
YearsWithCurrManager	NaN	-0.009197

	EnvironmentSatisfaction	HourlyRate	JobInvolvement \
Age	0.010146	0.024287	0.029820
DailyRate	0.018355	0.023381	0.046135
DistanceFromHome	-0.016075	0.031131	0.008783
Education	-0.027128	0.016775	0.042438
EmployeeCount	NaN	NaN	NaN
EmployeeNumber	0.017621	0.035179	-0.006888
EnvironmentSatisfaction	1.000000	-0.049857	-0.008278
HourlyRate	-0.049857	1.000000	0.042861
JobInvolvement	-0.008278	0.042861	1.000000
JobLevel	0.001212	-0.027853	-0.012630
JobSatisfaction	-0.006784	-0.071335	-0.021476
MonthlyIncome	-0.006259	-0.015794	-0.015271
MonthlyRate	0.037600	-0.015297	-0.016322

NumCompaniesWorked	0.012594	0.022157	0.015012
PercentSalaryHike	-0.031701	-0.009062	-0.017205
PerformanceRating	-0.029548	-0.002172	-0.029071
RelationshipSatisfaction	0.007665	0.001330	0.034297
StandardHours	NaN	NaN	NaN
StockOptionLevel	0.003432	0.050263	0.021523
TotalWorkingYears	-0.002693	-0.002334	-0.005533
TrainingTimesLastYear	-0.019359	-0.008548	-0.015338
WorkLifeBalance	0.027627	-0.004607	-0.014617
YearsAtCompany	0.001458	-0.019582	-0.021355
YearsInCurrentRole	0.018007	-0.024106	0.008717
YearsSinceLastPromotion	0.016194	-0.026716	-0.024184
YearsWithCurrManager	-0.004999	-0.020123	0.025976

	JobLevel	...	RelationshipSatisfaction	\
Age	0.509604	...	0.053535	
DailyRate	0.002966	...	0.007846	
DistanceFromHome	0.005303	...	0.006557	
Education	0.101589	...	-0.009118	
EmployeeCount	NaN	...	NaN	
EmployeeNumber	-0.018519	...	-0.069861	
EnvironmentSatisfaction	0.001212	...	0.007665	
HourlyRate	-0.027853	...	0.001330	
JobInvolvement	-0.012630	...	0.034297	
JobLevel	1.000000	...	0.021642	
JobSatisfaction	-0.001944	...	-0.012454	
MonthlyIncome	0.950300	...	0.025873	
MonthlyRate	0.039563	...	-0.004085	
NumCompaniesWorked	0.142501	...	0.052733	
PercentSalaryHike	-0.034730	...	-0.040490	
PerformanceRating	-0.021222	...	-0.031351	
RelationshipSatisfaction	0.021642	...	1.000000	
StandardHours	NaN	...	NaN	
StockOptionLevel	0.013984	...	-0.045952	
TotalWorkingYears	0.782208	...	0.024054	
TrainingTimesLastYear	-0.018191	...	0.002497	
WorkLifeBalance	0.037818	...	0.019604	
YearsAtCompany	0.534739	...	0.019367	
YearsInCurrentRole	0.389447	...	-0.015123	
YearsSinceLastPromotion	0.353885	...	0.033493	
YearsWithCurrManager	0.375281	...	-0.000867	

	StandardHours	StockOptionLevel	TotalWorkingYears	\
Age	NaN	0.037510	0.680381	
DailyRate	NaN	0.042143	0.014515	
DistanceFromHome	NaN	0.044872	0.004628	
Education	NaN	0.018422	0.148280	

EmployeeCount	NaN	NaN	NaN
EmployeeNumber	NaN	0.062227	-0.014365
EnvironmentSatisfaction	NaN	0.003432	-0.002693
HourlyRate	NaN	0.050263	-0.002334
JobInvolvement	NaN	0.021523	-0.005533
JobLevel	NaN	0.013984	0.782208
JobSatisfaction	NaN	0.010690	-0.020185
MonthlyIncome	NaN	0.005408	0.772893
MonthlyRate	NaN	-0.034323	0.026442
NumCompaniesWorked	NaN	0.030075	0.237639
PercentSalaryHike	NaN	0.007528	-0.020608
PerformanceRating	NaN	0.003506	0.006744
RelationshipSatisfaction	NaN	-0.045952	0.024054
StandardHours	NaN	NaN	NaN
StockOptionLevel	NaN	1.000000	0.010136
TotalWorkingYears	NaN	0.010136	1.000000
TrainingTimesLastYear	NaN	0.011274	-0.035662
WorkLifeBalance	NaN	0.004129	0.001008
YearsAtCompany	NaN	0.015058	0.628133
YearsInCurrentRole	NaN	0.050818	0.460365
YearsSinceLastPromotion	NaN	0.014352	0.404858
YearsWithCurrManager	NaN	0.024698	0.459188

	TrainingTimesLastYear	WorkLifeBalance \
Age	-0.019621	-0.021490
DailyRate	0.002453	-0.037848
DistanceFromHome	-0.036942	-0.026556
Education	-0.025100	0.009819
EmployeeCount	NaN	NaN
EmployeeNumber	0.023603	0.010309
EnvironmentSatisfaction	-0.019359	0.027627
HourlyRate	-0.008548	-0.004607
JobInvolvement	-0.015338	-0.014617
JobLevel	-0.018191	0.037818
JobSatisfaction	-0.005779	-0.019459
MonthlyIncome	-0.021736	0.030683
MonthlyRate	0.001467	0.007963
NumCompaniesWorked	-0.066054	-0.008366
PercentSalaryHike	-0.005221	-0.003280
PerformanceRating	-0.015579	0.002572
RelationshipSatisfaction	0.002497	0.019604
StandardHours	NaN	NaN
StockOptionLevel	0.011274	0.004129
TotalWorkingYears	-0.035662	0.001008
TrainingTimesLastYear	1.000000	0.028072
WorkLifeBalance	0.028072	1.000000
YearsAtCompany	0.003569	0.012089

YearsInCurrentRole	-0.005738	0.049856
YearsSinceLastPromotion	-0.002067	0.008941
YearsWithCurrManager	-0.004096	0.002759

	YearsAtCompany	YearsInCurrentRole \
Age	0.311309	0.212901
DailyRate	-0.034055	0.009932
DistanceFromHome	0.009508	0.018845
Education	0.069114	0.060236
EmployeeCount	NaN	NaN
EmployeeNumber	-0.011240	-0.008416
EnvironmentSatisfaction	0.001458	0.018007
HourlyRate	-0.019582	-0.024106
JobInvolvement	-0.021355	0.008717
JobLevel	0.534739	0.389447
JobSatisfaction	-0.003803	-0.002305
MonthlyIncome	0.514285	0.363818
MonthlyRate	-0.023655	-0.012815
NumCompaniesWorked	-0.118421	-0.090754
PercentSalaryHike	-0.035991	-0.001520
PerformanceRating	0.003435	0.034986
RelationshipSatisfaction	0.019367	-0.015123
StandardHours	NaN	NaN
StockOptionLevel	0.015058	0.050818
TotalWorkingYears	0.628133	0.460365
TrainingTimesLastYear	0.003569	-0.005738
WorkLifeBalance	0.012089	0.049856
YearsAtCompany	1.000000	0.758754
YearsInCurrentRole	0.758754	1.000000
YearsSinceLastPromotion	0.618409	0.548056
YearsWithCurrManager	0.769212	0.714365

	YearsSinceLastPromotion	YearsWithCurrManager
Age	0.216513	0.202089
DailyRate	-0.033229	-0.026363
DistanceFromHome	0.010029	0.014406
Education	0.054254	0.069065
EmployeeCount	NaN	NaN
EmployeeNumber	-0.009019	-0.009197
EnvironmentSatisfaction	0.016194	-0.004999
HourlyRate	-0.026716	-0.020123
JobInvolvement	-0.024184	0.025976
JobLevel	0.353885	0.375281
JobSatisfaction	-0.018214	-0.027656
MonthlyIncome	0.344978	0.344079
MonthlyRate	0.001567	-0.036746
NumCompaniesWorked	-0.036814	-0.110319

PercentSalaryHike	-0.022154	-0.011985
PerformanceRating	0.017896	0.022827
RelationshipSatisfaction	0.033493	-0.000867
StandardHours	NaN	NaN
StockOptionLevel	0.014352	0.024698
TotalWorkingYears	0.404858	0.459188
TrainingTimesLastYear	-0.002067	-0.004096
WorkLifeBalance	0.008941	0.002759
YearsAtCompany	0.618409	0.769212
YearsInCurrentRole	0.548056	0.714365
YearsSinceLastPromotion	1.000000	0.510224
YearsWithCurrManager	0.510224	1.000000

[26 rows x 26 columns]

```
[8]: df.isnull().any()
```

```
[8]: 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

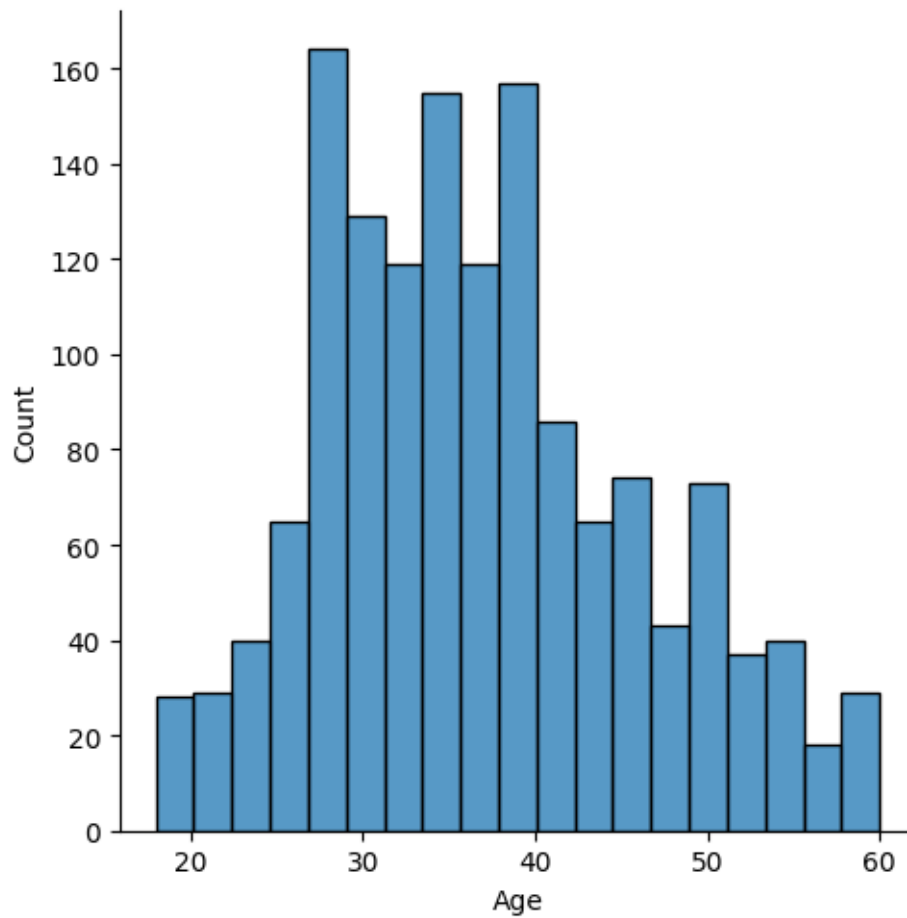
dtype: bool

```
[9]: df.isnull().sum()
```

```
[9]: 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
```

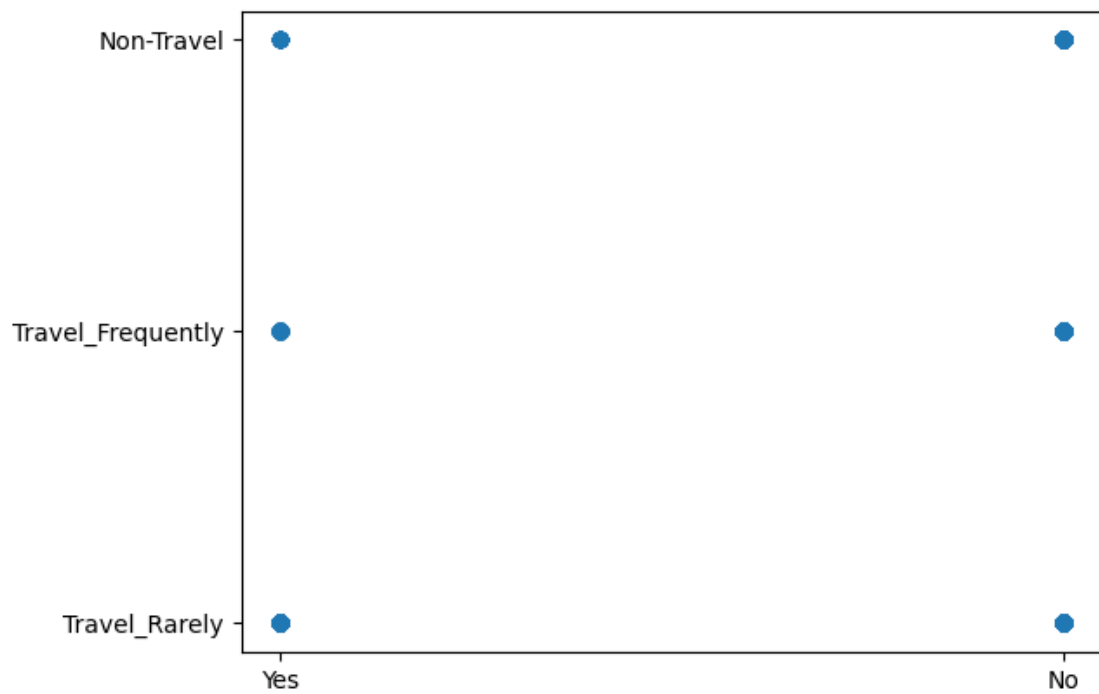
```
[10]: sns.displot(df["Age"])
```

```
[10]: <seaborn.axisgrid.FacetGrid at 0x7ab6c484c7c0>
```



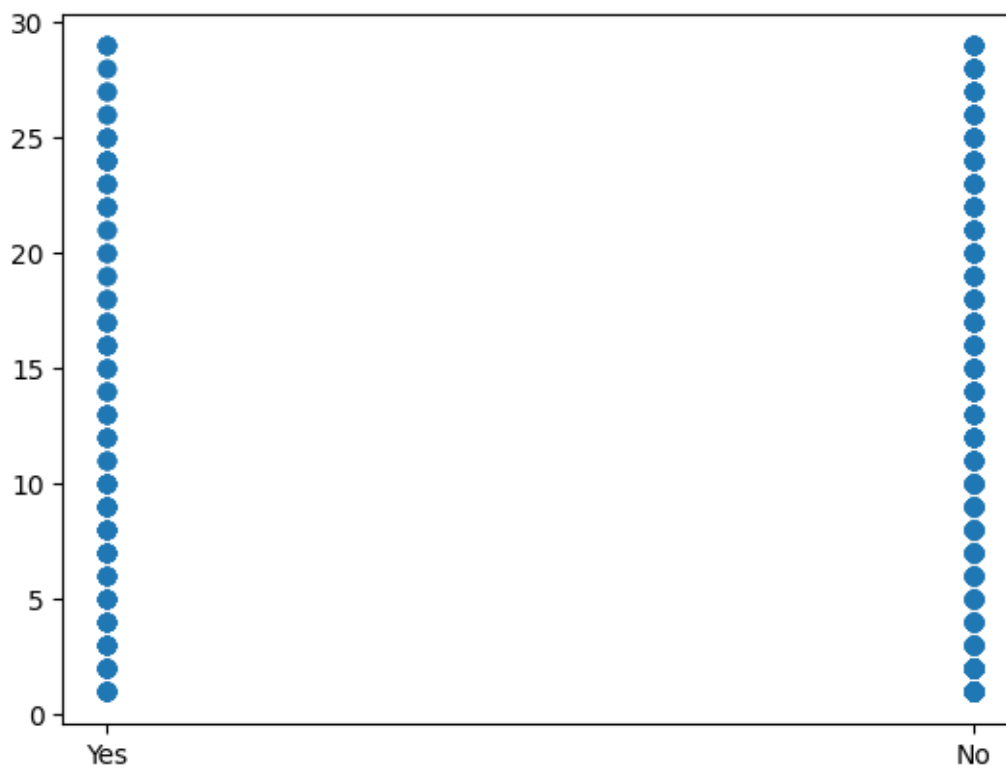
```
[11]: plt.scatter(df['Attrition'],df['BusinessTravel'])
```

```
[11]: <matplotlib.collections.PathCollection at 0x7ab6c25dfc40>
```



```
[12]: plt.scatter(df['Attrition'],df['DistanceFromHome'])
```

```
[12]: <matplotlib.collections.PathCollection at 0x7ab6c265ae60>
```



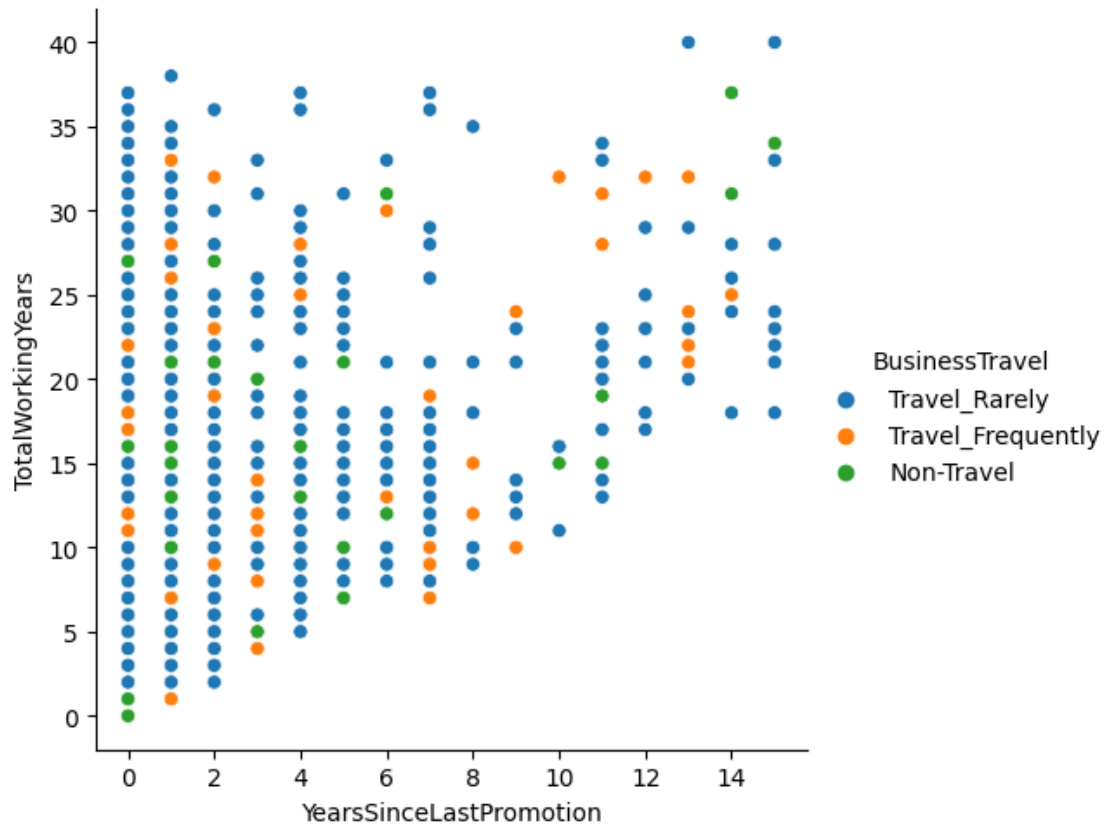
```
[13]: plt.scatter(df['Attrition'],df['StandardHours'])
```

```
[13]: <matplotlib.collections.PathCollection at 0x7ab6c24d5990>
```



```
[14]: sns.  
      ↪relplot(x="YearsSinceLastPromotion",y="TotalWorkingYears",data=df,hue="BusinessTravel")
```

```
[14]: <seaborn.axisgrid.FacetGrid at 0x7ab6c2506d10>
```

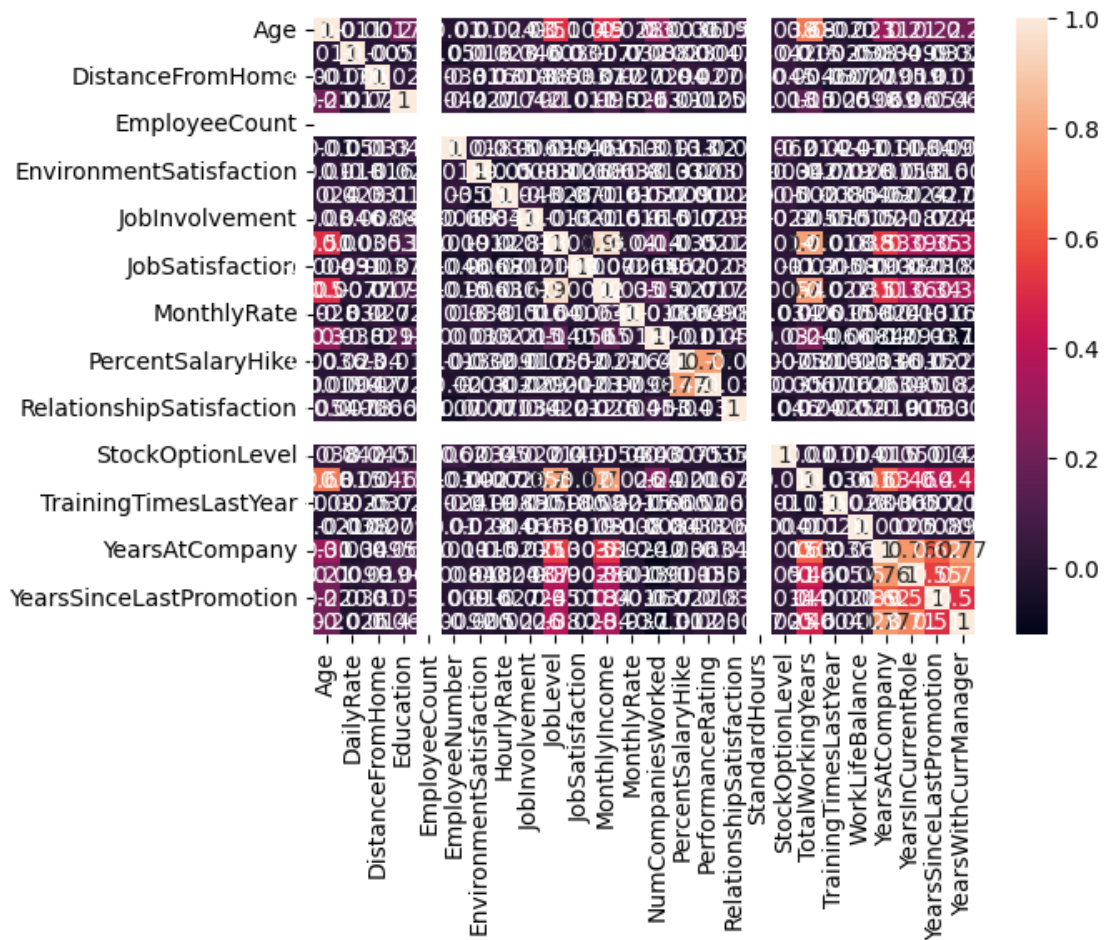


```
[15]: sns.heatmap(df.corr(),annot=True)
```

<ipython-input-15-8df7bcac526d>: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.

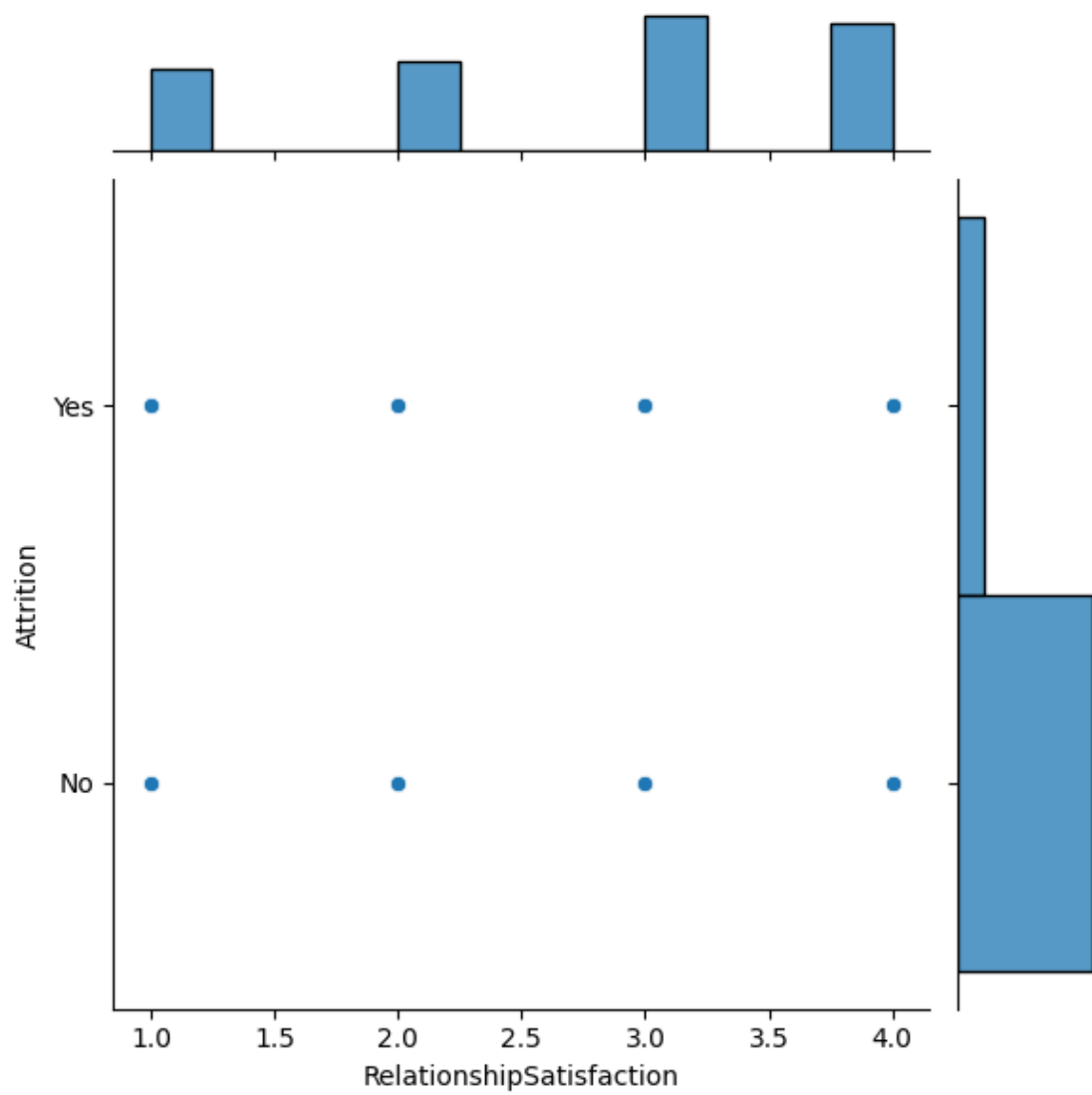
```
sns.heatmap(df.corr(),annot=True)
```

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



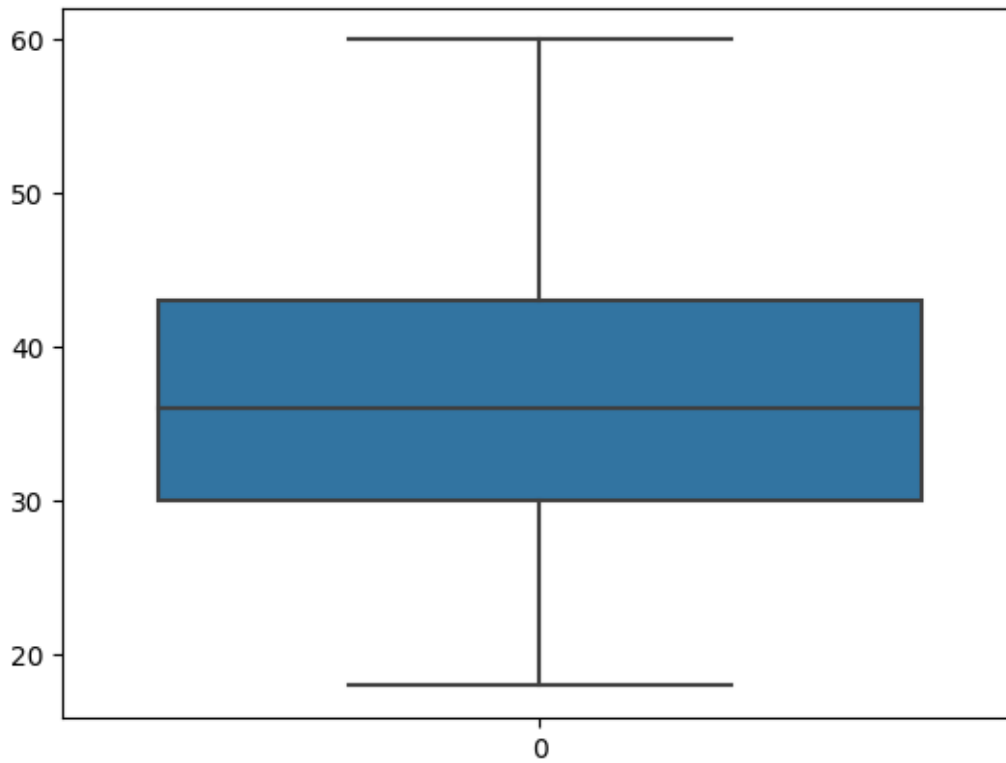
```
[16]: sns.jointplot(x="RelationshipSatisfaction",y="Attrition",data=df)
```

```
[16]: <seaborn.axisgrid.JointGrid at 0x7ab6c0aebb20>
```



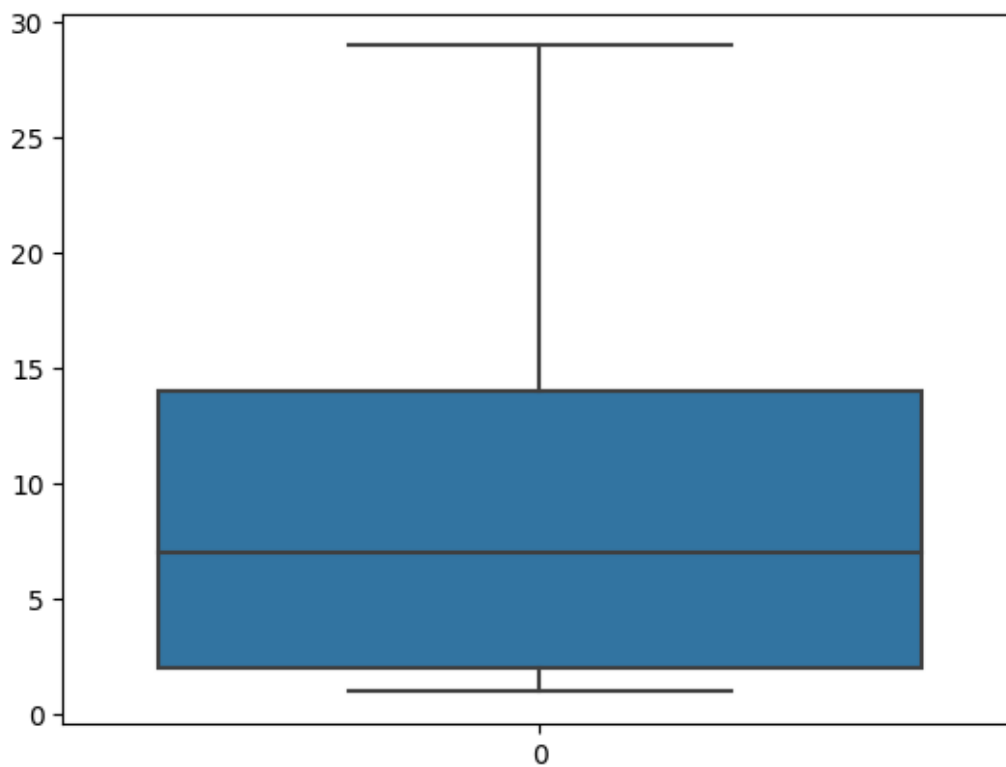
```
[17]: sns.boxplot(df.Age)
```

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

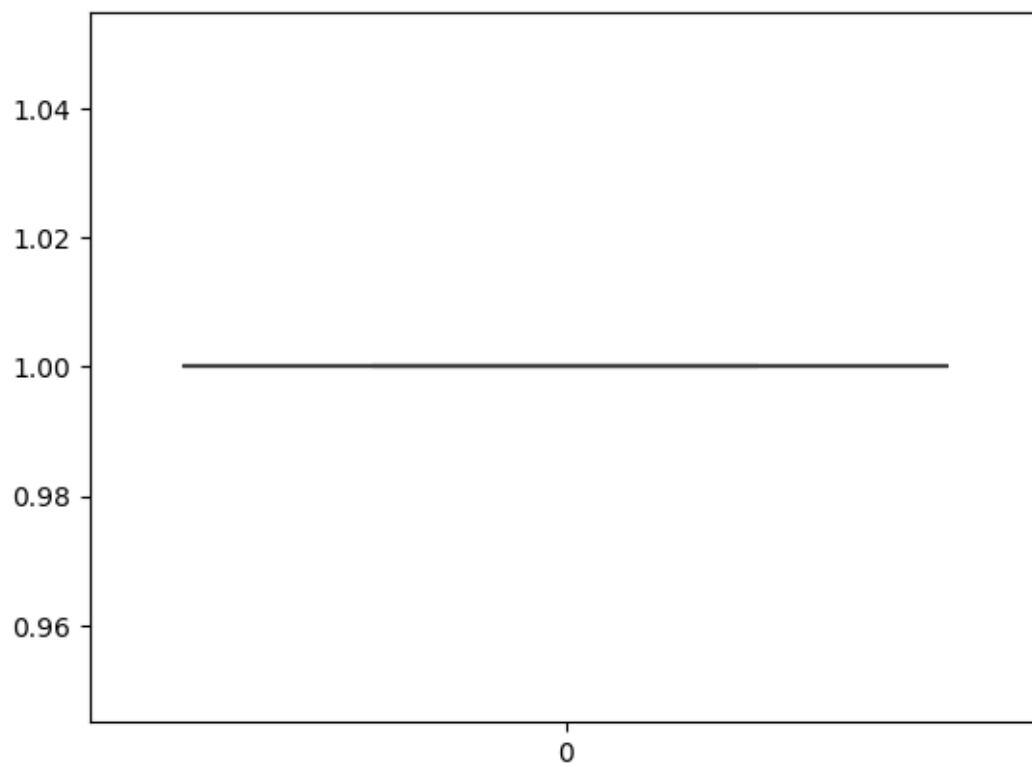
```
[18]: sns.boxplot(df.DistanceFromHome)
```

```
[18]: <Axes: >
```



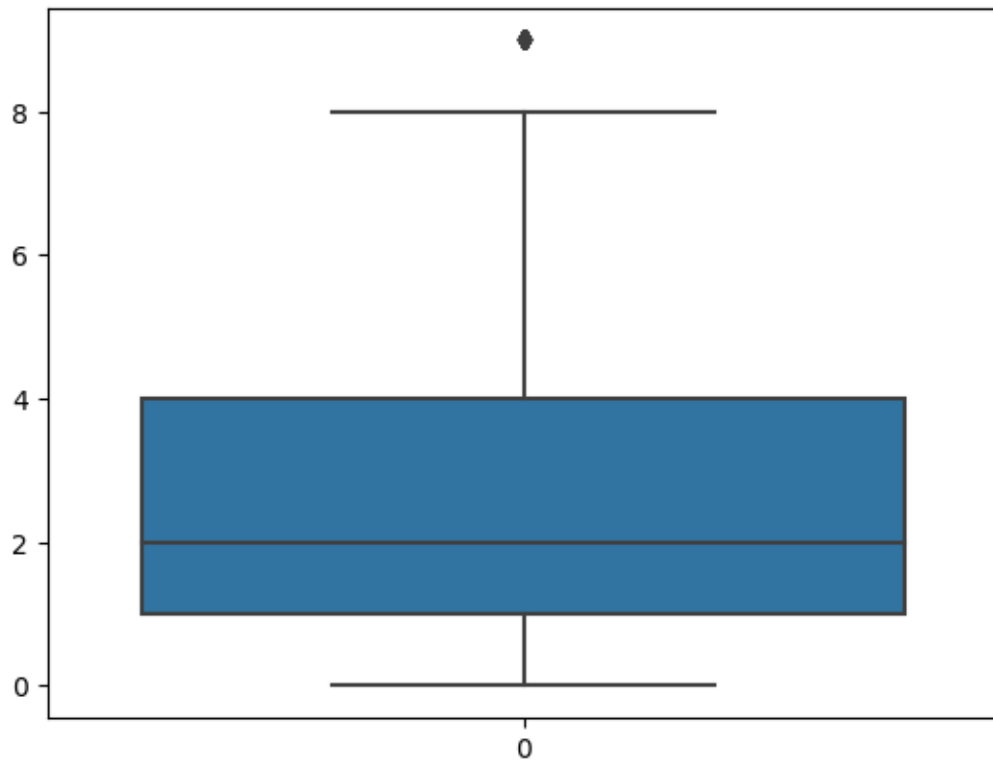
```
[19]: sns.boxplot(df.EmployeeCount )
```

```
[19]: <Axes: >
```



```
[20]: sns.boxplot(df.NumCompaniesWorked)
```

```
[20]: <Axes: >
```



```
[21]: q1=df.NumCompaniesWorked.quantile(0.25)
      q3=df.NumCompaniesWorked.quantile(0.75)
```

```
[22]: print(q1)
      print(q3)
```

```
1.0
4.0
```

```
[23]: IQR=q3-q1
```

```
[24]: IQR
```

```
[24]: 3.0
```

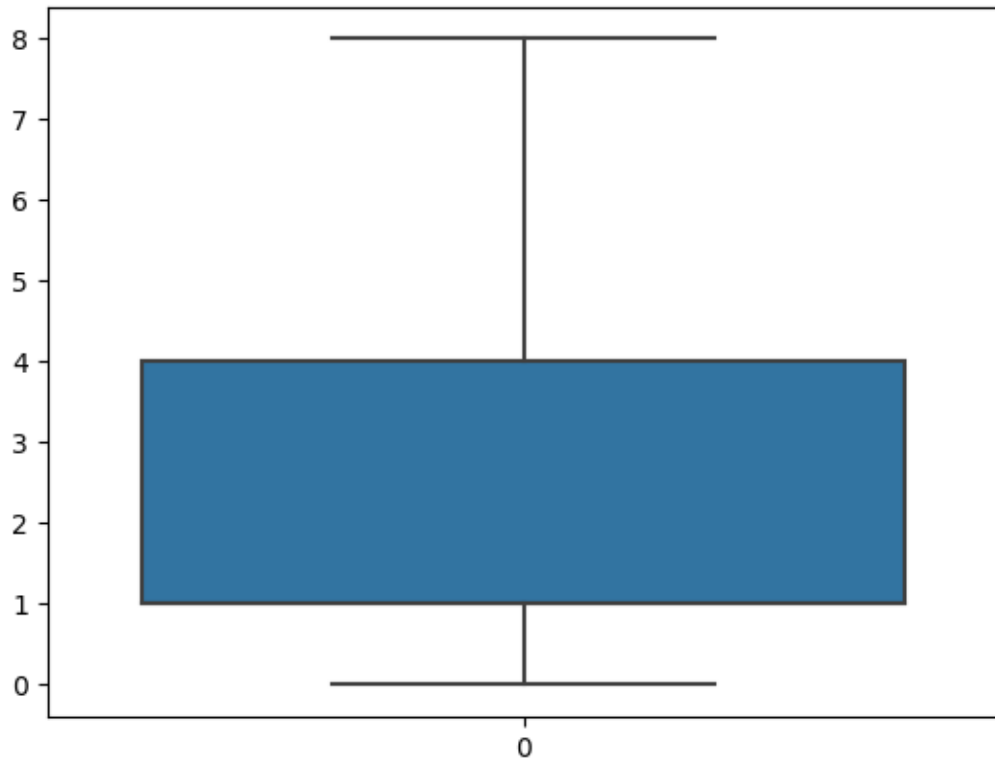
```
[25]: upper_limit=q3+1.5*IQR
      upper_limit
```

```
[25]: 8.5
```

```
[26]: df=df[df.NumCompaniesWorked<upper_limit]
```

```
[27]: sns.boxplot(df.NumCompaniesWorked)
```

```
[27]: <Axes: >
```



```
[28]: #dependent variable  
y=df.Attrition
```

```
[29]: y.head()
```

```
[29]: 0    Yes  
1    No  
2    Yes  
3    No  
5    No  
Name: Attrition, dtype: object
```

```
[30]: #independent variable  
x=df.drop(["Attrition"],axis=1)
```

```
[31]: x.head()
```

```

[31]: Age      BusinessTravel  DailyRate      Department \
0    41      Travel_Rarely      1102      Sales
1    49  Travel_Frequently      279  Research & Development
2    37      Travel_Rarely      1373  Research & Development
3    33  Travel_Frequently      1392  Research & Development
5    32  Travel_Frequently      1005  Research & Development

      DistanceFromHome  Education  EducationField  EmployeeCount  EmployeeNumber \
0                1        2  Life Sciences            1            1
1                8        1  Life Sciences            1            2
2                2        2      Other              1            4
3                3        4  Life Sciences            1            5
5                2        2  Life Sciences            1            8

      EnvironmentSatisfaction  ... RelationshipSatisfaction  StandardHours \
0                2  ...              1            80
1                3  ...              4            80
2                4  ...              2            80
3                4  ...              3            80
5                4  ...              3            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance \
0                0            8            0            1
1                1           10            3            3
2                0            7            3            3
3                0            8            3            3
5                0            8            2            2

      YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion \
0                6            4            0
1               10            7            1
2                0            0            0
3                8            7            3
5                7            7            3

      YearsWithCurrManager
0                5
1                7
2                0
3                0
5                6

[5 rows x 34 columns]

```

```
[32]: x.shape
```

```
[32]: (1418, 34)
```

```
[33]: y.shape
```

```
[33]: (1418,)
```

```
[34]: df.head()
```

```
[34]:
```

	Age	Attrition	BusinessTravel	DailyRate	Department	\
0	41	Yes	Travel_Rarely	1102		Sales
1	49	No	Travel_Frequently	279	Research & Development	
2	37	Yes	Travel_Rarely	1373	Research & Development	
3	33	No	Travel_Frequently	1392	Research & Development	
5	32	No	Travel_Frequently	1005	Research & Development	

	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	\
0	1	2	Life Sciences	1	1	
1	8	1	Life Sciences	1	2	
2	2	2	Other	1	4	
3	3	4	Life Sciences	1	5	
5	2	2	Life Sciences	1	8	

	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
0	...	1	80	0
1	...	4	80	1
2	...	2	80	0
3	...	3	80	0
5	...	3	80	0

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
0	8	0	1	6	
1	10	3	3	10	
2	7	3	3	0	
3	8	3	3	8	
5	8	2	2	7	

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
5	7	3	6

```
[5 rows x 35 columns]
```

```
[35]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x["BusinessTravel"]=le.fit_transform(x["BusinessTravel"])
```

```
[36]: x["BusinessTravel"]
```

```
[36]: 0      2
      1      1
      2      2
      3      1
      5      1
      ..
     1465    1
     1466    2
     1467    2
     1468    1
     1469    2
      Name: BusinessTravel, Length: 1418, dtype: int64
```

```
[37]: x.head()
```

```
[37]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  \
0   41                2      1102      Sales                1
1   49                1       279  Research & Development      8
2   37                2     1373  Research & Development      2
3   33                1     1392  Research & Development      3
5   32                1     1005  Research & Development      2

      Education EducationField  EmployeeCount  EmployeeNumber  \
0           2  Life Sciences                1                1
1           1  Life Sciences                1                2
2           2      Other                1                4
3           4  Life Sciences                1                5
5           2  Life Sciences                1                8

      EnvironmentSatisfaction  ... RelationshipSatisfaction  StandardHours  \
0                2  ...                1                80
1                3  ...                4                80
2                4  ...                2                80
3                4  ...                3                80
5                4  ...                3                80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  \
0                0                8                0                1
1                1               10                3                3
2                0                7                3                3
3                0                8                3                3
5                0                8                2                2

      YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
0                6                4                0
```


1	10	7	1
2	0	0	0
3	8	7	3
5	7	7	3

	YearsWithCurrManager
0	5
1	7
2	0
3	0
5	6

[5 rows x 34 columns]

```
[38]: x["Department"]=le.fit_transform(x["Department"])
```

```
[39]: x["Department"]
```

```
[39]: 0      2
      1      1
      2      1
      3      1
      5      1
      ..
1465    1
1466    1
1467    1
1468    2
1469    1
Name: Department, Length: 1418, dtype: int64
```

```
[40]: x.head()
```

```
[40]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41                2      1102           2                1         2
1   49                1       279           1                8         1
2   37                2      1373           1                2         2
3   33                1      1392           1                3         4
5   32                1      1005           1                2         2

   EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction  ...  \
0   Life Sciences              1                1                      2  ...
1   Life Sciences              1                2                      3  ...
2           Other              1                4                      4  ...
3   Life Sciences              1                5                      4  ...
5   Life Sciences              1                8                      4  ...
```

	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
0	1	80	0	
1	4	80	1	
2	2	80	0	
3	3	80	0	
5	3	80	0	

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
0	8	0	1	6	
1	10	3	3	10	
2	7	3	3	0	
3	8	3	3	8	
5	8	2	2	7	

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
5	7	3	6

[5 rows x 34 columns]

```
[41]: x["EducationField"]=le.fit_transform(x["EducationField"])
```

```
[42]: x["EducationField"]
```

```
[42]: 0      1
      1      1
      2      4
      3      1
      5      1
      ..
1465    3
1466    3
1467    1
1468    3
1469    3
Name: EducationField, Length: 1418, dtype: int64
```

```
[43]: x.head()
```

```
[43]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41             2      1102           2             1           2
1   49             1       279           1             8           1
2   37             2      1373           1             2           2
3   33             1      1392           1             3           4
```

5	32	1	1005	1	2	2
---	----	---	------	---	---	---

	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	\
0	1	1	1		2
1	1	1	2		3
2	4	1	4		4
3	1	1	5		4
5	1	1	8		4

	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
0	...	1	80	0
1	...	4	80	1
2	...	2	80	0
3	...	3	80	0
5	...	3	80	0

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
0	8	0	1		6
1	10	3	3		10
2	7	3	3		0
3	8	3	3		8
5	8	2	2		7

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
5	7	3	6

[5 rows x 34 columns]

```
[44]: non_numeric_columns = x.select_dtypes(exclude=['number']).columns
```

```
[45]: print(non_numeric_columns)
```

```
Index(['Gender', 'JobRole', 'MaritalStatus', 'Over18', 'OverTime'],
      dtype='object')
```

```
[46]: x["Gender"] = le.fit_transform(x["Gender"])
```

```
[47]: x["Gender"]
```

```
[47]: 0    0
      1    1
      2    1
      3    0
```

```

5      1
..
1465   1
1466   1
1467   1
1468   1
1469   1
Name: Gender, Length: 1418, dtype: int64

```

```
[48]: x.head()
```

```

[48]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education \
0    41                2      1102           2                1         2
1    49                1       279           1                8         1
2    37                2     1373           1                2         2
3    33                1     1392           1                3         4
5    32                1     1005           1                2         2

      EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction \
0                  1              1              1                      2
1                  1              1              2                      3
2                  4              1              4                      4
3                  1              1              5                      4
5                  1              1              8                      4

      ...  RelationshipSatisfaction  StandardHours  StockOptionLevel \
0  ...                          1              80                0
1  ...                          4              80                1
2  ...                          2              80                0
3  ...                          3              80                0
5  ...                          3              80                0

      TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany \
0                  8              0              1              6
1                 10              3              3             10
2                  7              3              3              0
3                  8              3              3              8
5                  8              2              2              7

      YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4              0              5
1                  7              1              7
2                  0              0              0
3                  7              3              0
5                  7              3              6

```

```
[5 rows x 34 columns]
```

```
[49]: x["JobRole"]=le.fit_transform(x["JobRole"])
```

```
[50]: x["JobRole"]
```

```
[50]: 0      7
      1      6
      2      2
      3      6
      5      2
      ..
     1465     2
     1466     0
     1467     4
     1468     7
     1469     2
      Name: JobRole, Length: 1418, dtype: int64
```

```
[51]: x.head()
```

```
[51]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41                2      1102           2                1         2
1   49                1       279           1                8         1
2   37                2     1373           1                2         2
3   33                1     1392           1                3         4
5   32                1     1005           1                2         2

      EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction  \
0                1                1                1                        2
1                1                1                2                        3
2                4                1                4                        4
3                1                1                5                        4
5                1                1                8                        4

      ...  RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0  ...                1                80                0
1  ...                4                80                1
2  ...                2                80                0
3  ...                3                80                0
5  ...                3                80                0

      TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  \
0                8                0                1                6
1               10                3                3               10
2                7                3                3                0
3                8                3                3                8
5                8                2                2                7
```

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
5	7	3	6

[5 rows x 34 columns]

```
[52]: x["MaritalStatus"]=le.fit_transform(x["MaritalStatus"])
```

```
[53]: x["MaritalStatus"]
```

```
[53]: 0      2
      1      1
      2      2
      3      1
      5      2
      ..
     1465     1
     1466     1
     1467     1
     1468     1
     1469     1
      Name: MaritalStatus, Length: 1418, dtype: int64
```

```
[54]: x.head()
```

```
[54]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41                2      1102           2                1         2
1   49                1       279           1                8         1
2   37                2     1373           1                2         2
3   33                1     1392           1                3         4
5   32                1     1005           1                2         2

      EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction  \
0                1                1                1                2
1                1                1                2                3
2                4                1                4                4
3                1                1                5                4
5                1                1                8                4

      ...  RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0   ...                1                80                0
1   ...                4                80                1
2   ...                2                80                0
3   ...                3                80                0
```

```

5 ...          3          80          0

      TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany \
0           8           0           1           6
1          10           3           3          10
2           7           3           3           0
3           8           3           3           8
5           8           2           2           7

      YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0           4           0           5
1           7           1           7
2           0           0           0
3           7           3           0
5           7           3           6

[5 rows x 34 columns]

```

```
[55]: x["Over18"]=le.fit_transform(x["Over18"])
```

```
[56]: x["Over18"]
```

```

[56]: 0      0
      1      0
      2      0
      3      0
      5      0
      ..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Over18, Length: 1418, dtype: int64

```

```
[57]: x.head()
```

```

[57]:   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education \
0   41             2      1102           2             1           2
1   49             1       279           1             8           1
2   37             2     1373           1             2           2
3   33             1     1392           1             3           4
5   32             1     1005           1             2           2

      EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction \
0           1           1           1           2
1           1           1           2           3

```

2	4	1	4	4
3	1	1	5	4
5	1	1	8	4

	RelationshipSatisfaction	StandardHours	StockOptionLevel	\
0	...	1	80	0
1	...	4	80	1
2	...	2	80	0
3	...	3	80	0
5	...	3	80	0

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
0	8	0	1	6	
1	10	3	3	10	
2	7	3	3	0	
3	8	3	3	8	
5	8	2	2	7	

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
5	7	3	6

[5 rows x 34 columns]

```
[58]: x["OverTime"]=le.fit_transform(x["OverTime"])
```

```
[59]: x["OverTime"]
```

```
[59]: 0      1
      1      0
      2      1
      3      1
      5      0
      ..
     1465     0
     1466     0
     1467     1
     1468     0
     1469     0
      Name: OverTime, Length: 1418, dtype: int64
```

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



```
[61]: x_scaled
```

```
[61]:      Age  BusinessTravel  DailyRate  Department  DistanceFromHome  \
0      0.547619          1.0    0.715820          1.0          0.000000
1      0.738095          0.5    0.126700          0.5          0.250000
2      0.452381          1.0    0.909807          0.5          0.035714
3      0.357143          0.5    0.923407          0.5          0.071429
4      0.333333          0.5    0.646385          0.5          0.035714
...      ...          ...          ...          ...          ...
1413    0.428571          0.5    0.559771          0.5          0.785714
1414    0.500000          1.0    0.365784          0.5          0.178571
1415    0.214286          1.0    0.037938          0.5          0.107143
1416    0.738095          0.5    0.659270          1.0          0.035714
1417    0.380952          1.0    0.376521          0.5          0.250000

      Education  EducationField  EmployeeCount  EmployeeNumber  \
0          0.25          0.2          0.0          0.000000
1          0.00          0.2          0.0          0.000484
2          0.25          0.8          0.0          0.001451
3          0.75          0.2          0.0          0.001935
4          0.25          0.2          0.0          0.003387
...      ...          ...          ...          ...
1413        0.25          0.6          0.0          0.996613
1414        0.00          0.6          0.0          0.997097
1415        0.50          0.2          0.0          0.998065
1416        0.50          0.6          0.0          0.998549
1417        0.50          0.6          0.0          1.000000

      EnvironmentSatisfaction  ...  RelationshipSatisfaction  StandardHours  \
0          0.333333  ...          0.000000          0.0
1          0.666667  ...          1.000000          0.0
2          1.000000  ...          0.333333          0.0
3          1.000000  ...          0.666667          0.0
4          1.000000  ...          0.666667          0.0
...      ...  ...          ...          ...
1413        0.666667  ...          0.666667          0.0
1414        1.000000  ...          0.000000          0.0
1415        0.333333  ...          0.333333          0.0
1416        1.000000  ...          1.000000          0.0
1417        0.333333  ...          0.000000          0.0

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0          0.000000          0.200          0.000000
1          0.333333          0.250          0.500000
2          0.000000          0.175          0.500000
3          0.000000          0.200          0.500000
4          0.000000          0.200          0.333333
```

```

...
1413      0.333333      0.425      0.500000
1414      0.333333      0.225      0.833333
1415      0.333333      0.150      0.000000
1416      0.000000      0.425      0.500000
1417      0.000000      0.150      0.500000

```

```

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0      0.000000      0.150      0.222222
1      0.666667      0.250      0.388889
2      0.666667      0.000      0.000000
3      0.666667      0.200      0.388889
4      0.333333      0.175      0.388889
...
1413      0.666667      0.125      0.111111
1414      0.666667      0.175      0.388889
1415      0.666667      0.150      0.111111
1416      0.333333      0.225      0.333333
1417      1.000000      0.100      0.166667

```

```

      YearsSinceLastPromotion  YearsWithCurrManager
0      0.000000      0.294118
1      0.066667      0.411765
2      0.000000      0.000000
3      0.200000      0.000000
4      0.200000      0.352941
...
1413      0.000000      0.176471
1414      0.066667      0.411765
1415      0.000000      0.176471
1416      0.000000      0.470588
1417      0.066667      0.117647

```

[1418 rows x 34 columns]

```

[62]: 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)

```

```

[63]: print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)

```

(1134, 34) (284, 34) (1134,) (284,)

1 Logistic Regression

```
[64]: from sklearn.linear_model import LogisticRegression
      model=LogisticRegression()
```

```
[65]: model.fit(x_train,y_train)
```

```
[65]: LogisticRegression()
```

```
[66]: pred=model.predict(x_test)
```

```
[67]: pred
```

[illegible]

```
[68]: y_test
```

```
[68]: 451      No
      639      No
      832      No
      1287     No
      1277     No
```

```

..
521    No
550    No
1113   No
335    No
917    No
Name: Attrition, Length: 284, dtype: object

```

```
[69]: df
```

```

[69]:   Age  Attrition  BusinessTravel  DailyRate  Department \
0    41         Yes   Travel_Rarely    1102         Sales
1    49          No  Travel_Frequently     279  Research & Development
2    37         Yes   Travel_Rarely    1373  Research & Development
3    33          No  Travel_Frequently    1392  Research & Development
5    32          No  Travel_Frequently    1005  Research & Development
...  ...
1465  36          No  Travel_Frequently     884  Research & Development
1466  39          No   Travel_Rarely     613  Research & Development
1467  27          No   Travel_Rarely     155  Research & Development
1468  49          No  Travel_Frequently    1023         Sales
1469  34          No   Travel_Rarely     628  Research & Development

      DistanceFromHome  Education  EducationField  EmployeeCount \
0                    1          2  Life Sciences             1
1                    8          1  Life Sciences             1
2                    2          2         Other             1
3                    3          4  Life Sciences             1
5                    2          2  Life Sciences             1
...
1465                23          2         Medical             1
1466                 6          1         Medical             1
1467                 4          3  Life Sciences             1
1468                 2          3         Medical             1
1469                 8          3         Medical             1

      EmployeeNumber  ...  RelationshipSatisfaction  StandardHours \
0                  1  ...                        1              80
1                  2  ...                        4              80
2                  4  ...                        2              80
3                  5  ...                        3              80
5                  8  ...                        3              80
...
1465              2061  ...                        3              80
1466              2062  ...                        1              80
1467              2064  ...                        2              80
1468              2065  ...                        4              80

```

1469	2068	...	1	80
	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
0	0	8	0	
1	1	10	3	
2	0	7	3	
3	0	8	3	
5	0	8	2	
...	
1465	1	17	3	
1466	1	9	5	
1467	1	6	0	
1468	0	17	3	
1469	0	6	3	

	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	\
0	1	6	4	
1	3	10	7	
2	3	0	0	
3	3	8	7	
5	2	7	7	
...	
1465	3	5	2	
1466	3	7	7	
1467	3	6	2	
1468	2	9	6	
1469	4	4	3	

	YearsSinceLastPromotion	YearsWithCurrManager
0	0	5
1	1	7
2	0	0
3	3	0
5	3	6
...
1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

[1418 rows x 35 columns]

```
[70]: from sklearn.metrics import
      ↪ accuracy_score, confusion_matrix, classification_report, roc_auc_score, roc_curve
```

```
[71]: accuracy_score(y_test, pred)
```

```
[71]: 0.8697183098591549
```

```
[72]: confusion_matrix(y_test,pred)
```

```
[72]: array([[237,  3],
          [ 34, 10]])
```

```
[73]: pd.crosstab(y_test,pred)
```

```
[73]: col_0      No  Yes
Attrition
No      237    3
Yes     34   10
```

```
[74]: print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
No	0.87	0.99	0.93	240
Yes	0.77	0.23	0.35	44
accuracy			0.87	284
macro avg	0.82	0.61	0.64	284
weighted avg	0.86	0.87	0.84	284

```
[75]: #ROC_AUC Curve
```

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

```
[75]: array([0.13657895, 0.03742004, 0.08053736, 0.08659374, 0.023358 ,
          0.10563069, 0.13815154, 0.00229225, 0.06771379, 0.12744425,
          0.08172802, 0.05965762, 0.0638561 , 0.12855128, 0.2275486 ,
          0.08936636, 0.06240484, 0.09603478, 0.21199145, 0.05717384,
          0.01180209, 0.00367791, 0.07898725, 0.02473968, 0.11962886,
          0.12904799, 0.0184306 , 0.0365714 , 0.02049336, 0.10008116,
          0.16143025, 0.03099261, 0.05571065, 0.04469354, 0.21600549,
          0.42230677, 0.2197372 , 0.5227653 , 0.18101958, 0.10182865,
          0.03088844, 0.18054679, 0.08248226, 0.01733578, 0.19733818,
          0.06725397, 0.01197982, 0.01366601, 0.02702768, 0.18659878,
          0.04323244, 0.00445696, 0.05192806, 0.1866853 , 0.1632088 ,
          0.27853238, 0.07437663, 0.09816652, 0.00573849, 0.00449716,
          0.0059488 , 0.03111943, 0.00839901, 0.00669404, 0.04253402,
          0.18695255, 0.19941885, 0.03278527, 0.00238087, 0.01663221,
          0.58136087, 0.1578733 , 0.21711936, 0.03898385, 0.04521495,
          0.03220001, 0.06616953, 0.19809653, 0.10991992, 0.22934288,
```

```

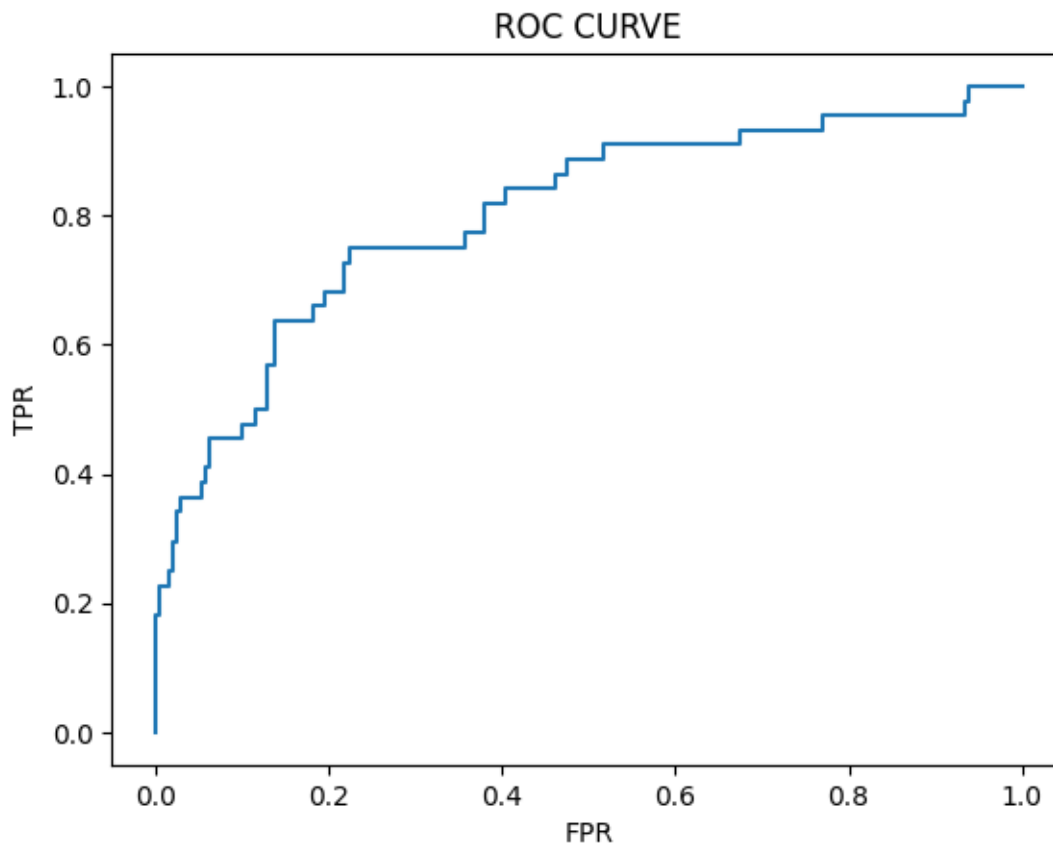
0.05904098, 0.02037218, 0.66970453, 0.26829173, 0.08216447,
0.04010601, 0.11590138, 0.27057603, 0.22694055, 0.20450222,
0.56793147, 0.22053355, 0.36393157, 0.01755166, 0.01233427,
0.01492107, 0.2081514 , 0.12205625, 0.40315397, 0.04856193,
0.07330096, 0.25379683, 0.14516211, 0.28647266, 0.02781388,
0.18391223, 0.26396952, 0.01946723, 0.28598072, 0.04347479,
0.15563751, 0.13357455, 0.00963796, 0.02116195, 0.07528362,
0.05922541, 0.11977388, 0.00903596, 0.36455439, 0.05168354,
0.20310448, 0.01231492, 0.05158269, 0.57453501, 0.07656055,
0.03508536, 0.30385493, 0.0309728 , 0.42983322, 0.02371366,
0.05130702, 0.02103465, 0.04602763, 0.01905589, 0.32734204,
0.19614051, 0.06294798, 0.0186783 , 0.00440507, 0.12521514,
0.35937712, 0.01824423, 0.03851794, 0.36623505, 0.0761209 ,
0.26592758, 0.03553327, 0.02772604, 0.0193432 , 0.28332535,
0.31642215, 0.02571374, 0.12136821, 0.32580669, 0.13472202,
0.06624905, 0.08617629, 0.03661786, 0.01839348, 0.15357873,
0.39926896, 0.71257736, 0.89315923, 0.00546009, 0.00246771,
0.02778452, 0.05857899, 0.36399558, 0.01646451, 0.14794275,
0.47711028, 0.03384135, 0.01739 , 0.04238425, 0.20976761,
0.54481958, 0.02510394, 0.01863455, 0.24136931, 0.06312414,
0.03643677, 0.00616726, 0.1100783 , 0.15064248, 0.07821613,
0.10409581, 0.20971698, 0.13795456, 0.28657845, 0.02226441,
0.23272876, 0.23596972, 0.16844684, 0.00414635, 0.03126561,
0.44815074, 0.01643598, 0.10900941, 0.01603778, 0.0333788 ,
0.27797218, 0.14158042, 0.05577601, 0.09399929, 0.24091949,
0.09998247, 0.01242131, 0.02205424, 0.1890573 , 0.06235382,
0.09115454, 0.00728886, 0.19906759, 0.1575069 , 0.20840636,
0.13738917, 0.05410298, 0.18636277, 0.08545779, 0.2373784 ,
0.04893286, 0.28718093, 0.07707427, 0.25024676, 0.11690009,
0.05663235, 0.06336832, 0.1402614 , 0.09635028, 0.5603858 ,
0.07966128, 0.18409077, 0.00949154, 0.04702311, 0.16756119,
0.03001824, 0.51420487, 0.00555785, 0.09370631, 0.01171392,
0.12695966, 0.03659918, 0.3821563 , 0.13188418, 0.17530265,
0.19609367, 0.10202889, 0.74603311, 0.05622724, 0.15448205,
0.17941515, 0.07061508, 0.07724554, 0.11220407, 0.19871038,
0.08215216, 0.00188234, 0.15323164, 0.06851284, 0.02069078,
0.71737346, 0.17804198, 0.15215912, 0.00469619, 0.23093543,
0.03742954, 0.06874542, 0.45373149, 0.6448183 , 0.09910567,
0.3574995 , 0.02215789, 0.00967421, 0.07067802, 0.35407627,
0.31550123, 0.01930184, 0.08248221, 0.07689043, 0.01921869,
0.13324521, 0.08754501, 0.22298726, 0.42007529])

```

```
[76]: y_test_encoded = le.fit_transform(y_test)
```

```
[77]: fpr, tpr, threshholds = roc_curve (y_test_encoded, probability)
```

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



2 Decision Tree

```
[81]: from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()
```

```
[82]: dtc.fit(x_train,y_train)
```

```
[82]: DecisionTreeClassifier()
```

```
[84]: pred=dtc.predict(x_test)
```

```
[85]: pred
```



```
[85]: array(['No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes',
        'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'Yes', 'No', 'No', 'No',
        'Yes', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes',
        'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'Yes', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes',
        'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes',
        'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'Yes',
        'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'No',
        'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'Yes',
        'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No',
        'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'Yes',
        'No', 'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
        'No', 'Yes', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No',
        'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No',
        'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
        'No', 'No', 'No', 'No', 'No', 'No', 'No'], dtype=object)
```

```
[86]: y_test
```

```
[86]: 451    No
      639    No
      832    No
     1287    No
     1277    No
      ..
     521    No
     550    No
    1113    No
     335    No
     917    No
      Name: Attrition, Length: 284, dtype: object
```

```
[87]: df
```

```
[87]:   Age Attrition BusinessTravel DailyRate Department \
0    41      Yes      Travel_Rarely    1102      Sales
```

1	49	No	Travel_Frequently	279	Research & Development
2	37	Yes	Travel_Rarely	1373	Research & Development
3	33	No	Travel_Frequently	1392	Research & Development
5	32	No	Travel_Frequently	1005	Research & Development
...
1465	36	No	Travel_Frequently	884	Research & Development
1466	39	No	Travel_Rarely	613	Research & Development
1467	27	No	Travel_Rarely	155	Research & Development
1468	49	No	Travel_Frequently	1023	Sales
1469	34	No	Travel_Rarely	628	Research & Development

	DistanceFromHome	Education	EducationField	EmployeeCount	\
0	1	2	Life Sciences	1	
1	8	1	Life Sciences	1	
2	2	2	Other	1	
3	3	4	Life Sciences	1	
5	2	2	Life Sciences	1	
...	
1465	23	2	Medical	1	
1466	6	1	Medical	1	
1467	4	3	Life Sciences	1	
1468	2	3	Medical	1	
1469	8	3	Medical	1	

	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	\
0	1	...	1	80	
1	2	...	4	80	
2	4	...	2	80	
3	5	...	3	80	
5	8	...	3	80	
...	
1465	2061	...	3	80	
1466	2062	...	1	80	
1467	2064	...	2	80	
1468	2065	...	4	80	
1469	2068	...	1	80	

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
0	0	8	0	
1	1	10	3	
2	0	7	3	
3	0	8	3	
5	0	8	2	
...	
1465	1	17	3	
1466	1	9	5	
1467	1	6	0	

1468	0	17	3
1469	0	6	3

	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	\
0	1	6	4	
1	3	10	7	
2	3	0	0	
3	3	8	7	
5	2	7	7	
...	
1465	3	5	2	
1466	3	7	7	
1467	3	6	2	
1468	2	9	6	
1469	4	4	3	

	YearsSinceLastPromotion	YearsWithCurrManager
0	0	5
1	1	7
2	0	0
3	3	0
5	3	6
...
1465	0	3
1466	1	7
1467	0	3
1468	0	8
1469	1	2

[1418 rows x 35 columns]

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

```
[105]: probability
```

```
[105]: array([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., 1., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1.,
0., 0., 0., 1., 0., 1., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
0., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1.,
1., 1., 0., 0., 1., 0., 0., 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.,
1., 0., 0., 0., 0., 0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0.,
1., 0., 0., 1., 0., 1., 0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 0.,
0., 0., 0., 1., 1., 0., 0., 0., 1., 0., 0., 0., 0., 1., 1., 0., 0.,
```

```

0., 0., 0., 1., 0., 0., 0., 1., 1., 0., 1., 0., 0., 1., 0., 0., 0.,
1., 1., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 1., 1., 0., 1., 0.,
0., 0., 0., 1., 0., 0., 0., 0., 0., 1., 1., 0., 0., 0., 1., 0., 0.,
0., 0., 1., 0., 0., 0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]

```

```

[106]: from sklearn import tree
plt.figure(figsize=(25,15))
tree.plot_tree(dtc,filled=True)

```

```

[106]: [Text(0.4609086715867159, 0.9722222222222222, 'x[21] <= 0.5\ngini =
0.268\nsamples = 1134\nvalue = [953, 181]'),
Text(0.15751845018450183, 0.9166666666666666, 'x[17] <= 0.048\ngini =
0.185\nsamples = 806\nvalue = [723, 83]'),
Text(0.014760147601476014, 0.8611111111111112, 'x[2] <= 0.169\ngini =
0.5\nsamples = 18\nvalue = [9, 9]'),
Text(0.007380073800738007, 0.8055555555555556, 'gini = 0.0\nsamples = 4\nvalue
= [0, 4]'),
Text(0.02214022140221402, 0.8055555555555556, 'x[15] <= 0.167\ngini =
0.459\nsamples = 14\nvalue = [9, 5]'),
Text(0.014760147601476014, 0.75, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.02952029520295203, 0.75, 'x[0] <= 0.262\ngini = 0.298\nsamples =
11\nvalue = [9, 2]'),
Text(0.02214022140221402, 0.6944444444444444, 'gini = 0.0\nsamples = 8\nvalue =
[8, 0]'),
Text(0.03690036900369004, 0.6944444444444444, 'x[14] <= 0.875\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.02952029520295203, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.04428044280442804, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.3002767527675277, 0.8611111111111112, 'x[26] <= 0.167\ngini =
0.17\nsamples = 788\nvalue = [714, 74]'),
Text(0.14990774907749077, 0.8055555555555556, 'x[29] <= 0.167\ngini =
0.249\nsamples = 337\nvalue = [288, 49]'),
Text(0.08118081180811808, 0.75, 'x[4] <= 0.25\ngini = 0.498\nsamples =
17\nvalue = [9, 8]'),
Text(0.06642066420664207, 0.6944444444444444, 'x[32] <= 0.333\ngini =
0.32\nsamples = 10\nvalue = [8, 2]'),
Text(0.05904059040590406, 0.6388888888888888, 'gini = 0.0\nsamples = 8\nvalue =
[8, 0]'),
Text(0.07380073800738007, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.0959409594095941, 0.6944444444444444, 'x[4] <= 0.696\ngini =
0.245\nsamples = 7\nvalue = [1, 6]'),
Text(0.08856088560885608, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue =
[0, 6]'),

```

```

Text(0.1033210332103321, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.21863468634686348, 0.75, 'x[6] <= 0.9\ngini = 0.223\nsamples =
320\nvalue = [279, 41]'),
Text(0.16051660516605165, 0.6944444444444444, 'x[15] <= 0.167\ngini =
0.185\nsamples = 290\nvalue = [260, 30]'),
Text(0.11808118081180811, 0.6388888888888888, 'x[30] <= 0.163\ngini =
0.346\nsamples = 54\nvalue = [42, 12]'),
Text(0.0996309963099631, 0.5833333333333334, 'x[19] <= 0.188\ngini =
0.493\nsamples = 25\nvalue = [14, 11]'),
Text(0.08487084870848709, 0.5277777777777778, 'x[5] <= 0.75\ngini =
0.165\nsamples = 11\nvalue = [10, 1]'),
Text(0.07749077490774908, 0.4722222222222222, 'gini = 0.0\nsamples = 10\nvalue
= [10, 0]'),
Text(0.09225092250922509, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.11439114391143912, 0.5277777777777778, 'x[30] <= 0.063\ngini =
0.408\nsamples = 14\nvalue = [4, 10]'),
Text(0.1070110701107011, 0.4722222222222222, 'x[0] <= 0.357\ngini =
0.444\nsamples = 6\nvalue = [4, 2]'),
Text(0.0996309963099631, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.11439114391143912, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
Text(0.12177121771217712, 0.4722222222222222, 'gini = 0.0\nsamples = 8\nvalue =
[0, 8]'),
Text(0.13653136531365315, 0.5833333333333334, 'x[17] <= 0.095\ngini =
0.067\nsamples = 29\nvalue = [28, 1]'),
Text(0.12915129151291513, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.14391143911439114, 0.5277777777777778, 'gini = 0.0\nsamples = 28\nvalue
= [28, 0]'),
Text(0.2029520295202952, 0.6388888888888888, 'x[12] <= 0.167\ngini =
0.141\nsamples = 236\nvalue = [218, 18]'),
Text(0.16605166051660517, 0.5833333333333334, 'x[2] <= 0.217\ngini =
0.444\nsamples = 12\nvalue = [8, 4]'),
Text(0.15867158671586715, 0.5277777777777778, 'gini = 0.0\nsamples = 6\nvalue =
[6, 0]'),
Text(0.17343173431734318, 0.5277777777777778, 'x[24] <= 0.167\ngini =
0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.16605166051660517, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.18081180811808117, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
Text(0.23985239852398524, 0.5833333333333334, 'x[4] <= 0.982\ngini =
0.117\nsamples = 224\nvalue = [210, 14]'),
Text(0.2177121771217712, 0.5277777777777778, 'x[2] <= 0.939\ngini =

```

```

0.104\nsamples = 219\nvalue = [207, 12]'),
  Text(0.19557195571955718, 0.4722222222222222, 'x[2] <= 0.044\ngini =
0.089\nsamples = 214\nvalue = [204, 10]'),
  Text(0.17343173431734318, 0.4166666666666667, 'x[18] <= 0.905\ngini =
0.305\nsamples = 16\nvalue = [13, 3]'),
  Text(0.16605166051660517, 0.3611111111111111, 'x[32] <= 0.4\ngini =
0.133\nsamples = 14\nvalue = [13, 1]'),
  Text(0.15867158671586715, 0.3055555555555556, 'gini = 0.0\nsamples = 13\nvalue
= [13, 0]'),
  Text(0.17343173431734318, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.18081180811808117, 0.3611111111111111, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
  Text(0.2177121771217712, 0.4166666666666667, 'x[6] <= 0.1\ngini =
0.068\nsamples = 198\nvalue = [191, 7]'),
  Text(0.19557195571955718, 0.3611111111111111, 'x[9] <= 0.333\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
  Text(0.1881918819188192, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.2029520295202952, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
  Text(0.23985239852398524, 0.3611111111111111, 'x[33] <= 0.088\ngini =
0.06\nsamples = 195\nvalue = [189, 6]'),
  Text(0.2177121771217712, 0.3055555555555556, 'x[4] <= 0.643\ngini =
0.145\nsamples = 51\nvalue = [47, 4]'),
  Text(0.2029520295202952, 0.25, 'x[4] <= 0.054\ngini = 0.083\nsamples =
46\nvalue = [44, 2]'),
  Text(0.19557195571955718, 0.19444444444444445, 'x[19] <= 0.812\ngini =
0.278\nsamples = 12\nvalue = [10, 2]'),
  Text(0.1881918819188192, 0.1388888888888889, 'x[4] <= 0.018\ngini =
0.165\nsamples = 11\nvalue = [10, 1]'),
  Text(0.18081180811808117, 0.08333333333333333, 'gini = 0.0\nsamples = 8\nvalue
= [8, 0]'),
  Text(0.19557195571955718, 0.08333333333333333, 'x[0] <= 0.405\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
  Text(0.1881918819188192, 0.027777777777777776, 'gini = 0.0\nsamples = 1\nvalue
= [0, 1]'),
  Text(0.2029520295202952, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue
= [2, 0]'),
  Text(0.2029520295202952, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.21033210332103322, 0.19444444444444445, 'gini = 0.0\nsamples = 34\nvalue
= [34, 0]'),
  Text(0.23247232472324722, 0.25, 'x[14] <= 0.375\ngini = 0.48\nsamples =
5\nvalue = [3, 2]'),
  Text(0.22509225092250923, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue
= [0, 2]'),

```

```

Text(0.23985239852398524, 0.19444444444444445, 'gini = 0.0\nsamples = 3\nvalue
= [3, 0]'),
Text(0.26199261992619927, 0.30555555555555556, 'x[0] <= 0.655\ngini =
0.027\nsamples = 144\nvalue = [142, 2]'),
Text(0.25461254612546125, 0.25, 'gini = 0.0\nsamples = 115\nvalue = [115, 0]'),
Text(0.2693726937269373, 0.25, 'x[24] <= 0.167\ngini = 0.128\nsamples =
29\nvalue = [27, 2]'),
Text(0.25461254612546125, 0.19444444444444445, 'x[12] <= 0.5\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.24723247232472326, 0.13888888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.26199261992619927, 0.13888888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.28413284132841327, 0.19444444444444445, 'x[0] <= 0.69\ngini =
0.071\nsamples = 27\nvalue = [26, 1]'),
Text(0.2767527675276753, 0.13888888888888889, 'x[1] <= 0.75\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.2693726937269373, 0.08333333333333333, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.28413284132841327, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue
= [0, 1]'),
Text(0.2915129151291513, 0.13888888888888889, 'gini = 0.0\nsamples = 24\nvalue =
[24, 0]'),
Text(0.23985239852398524, 0.4722222222222222, 'x[30] <= 0.113\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.23247232472324722, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.24723247232472326, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.26199261992619927, 0.5277777777777778, 'x[1] <= 0.75\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.25461254612546125, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.2693726937269373, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.2767527675276753, 0.6944444444444444, 'x[0] <= 0.274\ngini =
0.464\nsamples = 30\nvalue = [19, 11]'),
Text(0.26199261992619927, 0.6388888888888888, 'x[22] <= 0.036\ngini =
0.32\nsamples = 10\nvalue = [2, 8]'),
Text(0.25461254612546125, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.2693726937269373, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue =
[0, 8]'),
Text(0.2915129151291513, 0.6388888888888888, 'x[11] <= 0.321\ngini =
0.255\nsamples = 20\nvalue = [17, 3]'),
Text(0.28413284132841327, 0.5833333333333334, 'x[11] <= 0.25\ngini =
0.48\nsamples = 5\nvalue = [2, 3]'),

```

```

Text(0.2767527675276753, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.2915129151291513, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
Text(0.2988929889298893, 0.5833333333333334, 'gini = 0.0\nsamples = 15\nvalue =
[15, 0]'),
Text(0.45064575645756455, 0.8055555555555556, 'x[30] <= 0.787\ngini =
0.105\nsamples = 451\nvalue = [426, 25]'),
Text(0.40498154981549817, 0.75, 'x[17] <= 0.072\ngini = 0.094\nsamples =
447\nvalue = [425, 22]'),
Text(0.33579335793357934, 0.6944444444444444, 'x[29] <= 0.5\ngini =
0.249\nsamples = 55\nvalue = [47, 8]'),
Text(0.3210332103321033, 0.6388888888888888, 'x[11] <= 0.129\ngini =
0.475\nsamples = 18\nvalue = [11, 7]'),
Text(0.31365313653136534, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
Text(0.3284132841328413, 0.5833333333333334, 'x[11] <= 0.507\ngini =
0.391\nsamples = 15\nvalue = [11, 4]'),
Text(0.3210332103321033, 0.5277777777777778, 'gini = 0.0\nsamples = 9\nvalue =
[9, 0]'),
Text(0.33579335793357934, 0.5277777777777778, 'x[8] <= 0.383\ngini =
0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.3284132841328413, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.34317343173431736, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
Text(0.3505535055350554, 0.6388888888888888, 'x[8] <= 0.022\ngini =
0.053\nsamples = 37\nvalue = [36, 1]'),
Text(0.34317343173431736, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.35793357933579334, 0.5833333333333334, 'gini = 0.0\nsamples = 36\nvalue
= [36, 0]'),
Text(0.474169741697417, 0.6944444444444444, 'x[2] <= 0.997\ngini =
0.069\nsamples = 392\nvalue = [378, 14]'),
Text(0.4354243542435424, 0.6388888888888888, 'x[22] <= 0.036\ngini =
0.064\nsamples = 390\nvalue = [377, 13]'),
Text(0.3800738007380074, 0.5833333333333334, 'x[17] <= 0.352\ngini =
0.183\nsamples = 59\nvalue = [53, 6]'),
Text(0.36531365313653136, 0.5277777777777778, 'x[4] <= 0.071\ngini =
0.053\nsamples = 37\nvalue = [36, 1]'),
Text(0.35793357933579334, 0.4722222222222222, 'x[32] <= 0.033\ngini =
0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.3505535055350554, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
Text(0.36531365313653136, 0.4166666666666667, 'x[6] <= 0.8\ngini = 0.5\nsamples
= 2\nvalue = [1, 1]'),
Text(0.35793357933579334, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =

```



```

[0, 1]'),
Text(0.3726937269372694, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.3726937269372694, 0.4722222222222222, 'gini = 0.0\nsamples = 31\nvalue =
[31, 0]'),
Text(0.3948339483394834, 0.5277777777777778, 'x[2] <= 0.486\ngini =
0.351\nsamples = 22\nvalue = [17, 5]'),
Text(0.3874538745387454, 0.4722222222222222, 'x[33] <= 0.353\ngini =
0.494\nsamples = 9\nvalue = [4, 5]'),
Text(0.3800738007380074, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
Text(0.3948339483394834, 0.4166666666666667, 'x[27] <= 0.275\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
Text(0.3874538745387454, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.4022140221402214, 0.3611111111111111, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
Text(0.4022140221402214, 0.4722222222222222, 'gini = 0.0\nsamples = 13\nvalue =
[13, 0]'),
Text(0.4907749077490775, 0.5833333333333334, 'x[8] <= 0.986\ngini =
0.041\nsamples = 331\nvalue = [324, 7]'),
Text(0.46863468634686345, 0.5277777777777778, 'x[26] <= 0.833\ngini =
0.036\nsamples = 327\nvalue = [321, 6]'),
Text(0.44649446494464945, 0.4722222222222222, 'x[2] <= 0.117\ngini =
0.02\nsamples = 294\nvalue = [291, 3]'),
Text(0.42435424354243545, 0.4166666666666667, 'x[2] <= 0.092\ngini =
0.114\nsamples = 33\nvalue = [31, 2]'),
Text(0.41697416974169743, 0.3611111111111111, 'gini = 0.0\nsamples = 30\nvalue
= [30, 0]'),
Text(0.4317343173431734, 0.3611111111111111, 'x[1] <= 0.25\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.42435424354243545, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.43911439114391143, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.46863468634686345, 0.4166666666666667, 'x[0] <= 0.179\ngini =
0.008\nsamples = 261\nvalue = [260, 1]'),
Text(0.4612546125461255, 0.3611111111111111, 'x[19] <= 0.438\ngini =
0.153\nsamples = 12\nvalue = [11, 1]'),
Text(0.45387453874538747, 0.3055555555555556, 'gini = 0.0\nsamples = 10\nvalue
= [10, 0]'),
Text(0.46863468634686345, 0.3055555555555556, 'x[6] <= 0.4\ngini = 0.5\nsamples
= 2\nvalue = [1, 1]'),
Text(0.4612546125461255, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.47601476014760147, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.47601476014760147, 0.3611111111111111, 'gini = 0.0\nsamples = 249\nvalue
= [249, 0]'),

```

```

Text(0.4907749077490775, 0.4722222222222222, 'x[2] <= 0.117\ngini =
0.165\nsamples = 33\nvalue = [30, 3]'),
Text(0.4833948339483395, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.4981549815498155, 0.4166666666666667, 'x[24] <= 0.167\ngini =
0.117\nsamples = 32\nvalue = [30, 2]'),
Text(0.4907749077490775, 0.3611111111111111, 'x[4] <= 0.268\ngini =
0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.4833948339483395, 0.3055555555555556, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
Text(0.4981549815498155, 0.3055555555555556, 'x[11] <= 0.679\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4907749077490775, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.5055350553505535, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5055350553505535, 0.3611111111111111, 'gini = 0.0\nsamples = 25\nvalue =
[25, 0]'),
Text(0.5129151291512916, 0.5277777777777778, 'x[14] <= 0.812\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.5055350553505535, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.5202952029520295, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.5129151291512916, 0.6388888888888888, 'x[0] <= 0.25\ngini = 0.5\nsamples
= 2\nvalue = [1, 1]'),
Text(0.5055350553505535, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.5202952029520295, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.496309963099631, 0.75, 'x[2] <= 0.59\ngini = 0.375\nsamples = 4\nvalue =
[1, 3]'),
Text(0.488929889298893, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
Text(0.503690036900369, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.7642988929889298, 0.9166666666666666, 'x[17] <= 0.13\ngini =
0.419\nsamples = 328\nvalue = [230, 98]'),
Text(0.6411439114391144, 0.8611111111111112, 'x[26] <= 0.167\ngini =
0.498\nsamples = 107\nvalue = [50, 57]'),
Text(0.5904059040590406, 0.8055555555555556, 'x[4] <= 0.161\ngini =
0.427\nsamples = 55\nvalue = [17, 38]'),
Text(0.5571955719557196, 0.75, 'x[8] <= 0.4\ngini = 0.499\nsamples = 27\nvalue
= [14, 13]'),
Text(0.5350553505535055, 0.6944444444444444, 'x[18] <= 0.561\ngini =
0.337\nsamples = 14\nvalue = [3, 11]'),
Text(0.5276752767527675, 0.6388888888888888, 'gini = 0.0\nsamples = 9\nvalue =
[0, 9]'),
Text(0.5424354243542435, 0.6388888888888888, 'x[9] <= 0.333\ngini =

```

```

0.48\nsamples = 5\nvalue = [3, 2]'),
  Text(0.5350553505535055, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
  Text(0.5498154981549815, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
  Text(0.5793357933579336, 0.6944444444444444, 'x[2] <= 0.33\ngini =
0.26\nsamples = 13\nvalue = [11, 2]'),
  Text(0.5719557195571956, 0.6388888888888888, 'x[1] <= 0.75\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
  Text(0.5645756457564576, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
  Text(0.5793357933579336, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.5867158671586716, 0.6388888888888888, 'gini = 0.0\nsamples = 10\nvalue =
[10, 0]'),
  Text(0.6236162361623616, 0.75, 'x[4] <= 0.875\ngini = 0.191\nsamples =
28\nvalue = [3, 25]'),
  Text(0.6162361623616236, 0.6944444444444444, 'x[12] <= 0.833\ngini =
0.137\nsamples = 27\nvalue = [2, 25]'),
  Text(0.6014760147601476, 0.6388888888888888, 'x[19] <= 0.062\ngini =
0.077\nsamples = 25\nvalue = [1, 24]'),
  Text(0.5940959409594095, 0.5833333333333334, 'x[8] <= 0.358\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
  Text(0.5867158671586716, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.6014760147601476, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.6088560885608856, 0.5833333333333334, 'gini = 0.0\nsamples = 23\nvalue =
[0, 23]'),
  Text(0.6309963099630996, 0.6388888888888888, 'x[32] <= 0.067\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
  Text(0.6236162361623616, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.6383763837638377, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.6309963099630996, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.6918819188191881, 0.8055555555555556, 'x[33] <= 0.029\ngini =
0.464\nsamples = 52\nvalue = [33, 19]'),
  Text(0.6678966789667896, 0.75, 'x[2] <= 0.717\ngini = 0.457\nsamples =
17\nvalue = [6, 11]'),
  Text(0.6531365313653137, 0.6944444444444444, 'x[5] <= 0.125\ngini =
0.18\nsamples = 10\nvalue = [1, 9]'),
  Text(0.6457564575645757, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.6605166051660517, 0.6388888888888888, 'gini = 0.0\nsamples = 9\nvalue =
[0, 9]'),

```

```

Text(0.6826568265682657, 0.6944444444444444, 'x[30] <= 0.013\ngini =
0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.6752767527675276, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.6900369003690037, 0.6388888888888888, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
Text(0.7158671586715867, 0.75, 'x[17] <= 0.057\ngini = 0.353\nsamples =
35\nvalue = [27, 8]'),
Text(0.7084870848708487, 0.6944444444444444, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.7232472324723247, 0.6944444444444444, 'x[32] <= 0.433\ngini =
0.298\nsamples = 33\nvalue = [27, 6]'),
Text(0.7047970479704797, 0.6388888888888888, 'x[0] <= 0.202\ngini =
0.231\nsamples = 30\nvalue = [26, 4]'),
Text(0.6900369003690037, 0.5833333333333334, 'x[15] <= 0.833\ngini =
0.5\nsamples = 4\nvalue = [2, 2]'),
Text(0.6826568265682657, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.6974169741697417, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.7195571955719557, 0.5833333333333334, 'x[19] <= 0.562\ngini =
0.142\nsamples = 26\nvalue = [24, 2]'),
Text(0.7121771217712177, 0.5277777777777778, 'gini = 0.0\nsamples = 21\nvalue =
[21, 0]'),
Text(0.7269372693726938, 0.5277777777777778, 'x[18] <= 0.772\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.7195571955719557, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.7343173431734318, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.7416974169741697, 0.6388888888888888, 'x[8] <= 0.399\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.7343173431734318, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.7490774907749077, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
Text(0.8874538745387454, 0.8611111111111112, 'x[16] <= 0.75\ngini =
0.302\nsamples = 221\nvalue = [180, 41]'),
Text(0.8376383763837638, 0.8055555555555556, 'x[19] <= 0.562\ngini =
0.192\nsamples = 158\nvalue = [141, 17]'),
Text(0.8007380073800738, 0.75, 'x[0] <= 0.119\ngini = 0.136\nsamples =
136\nvalue = [126, 10]'),
Text(0.7933579335793358, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.8081180811808119, 0.6944444444444444, 'x[4] <= 0.589\ngini =
0.124\nsamples = 135\nvalue = [126, 9]'),
Text(0.7785977859778598, 0.6388888888888888, 'x[28] <= 0.917\ngini =

```

```

0.039\nsamples = 101\nvalue = [99, 2]'),
  Text(0.7638376383763837, 0.5833333333333334, 'x[12] <= 0.167\ngini =
0.02\nsamples = 97\nvalue = [96, 1]'),
  Text(0.7564575645756457, 0.5277777777777778, 'x[26] <= 0.167\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
  Text(0.7490774907749077, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.7638376383763837, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
  Text(0.7712177121771218, 0.5277777777777778, 'gini = 0.0\nsamples = 93\nvalue =
[93, 0]'),
  Text(0.7933579335793358, 0.5833333333333334, 'x[10] <= 0.5\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
  Text(0.7859778597785978, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
  Text(0.8007380073800738, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.8376383763837638, 0.6388888888888888, 'x[14] <= 0.812\ngini =
0.327\nsamples = 34\nvalue = [27, 7]'),
  Text(0.8228782287822878, 0.5833333333333334, 'x[0] <= 0.345\ngini =
0.137\nsamples = 27\nvalue = [25, 2]'),
  Text(0.8154981549815498, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.8302583025830258, 0.5277777777777778, 'x[4] <= 0.982\ngini =
0.074\nsamples = 26\nvalue = [25, 1]'),
  Text(0.8228782287822878, 0.4722222222222222, 'gini = 0.0\nsamples = 24\nvalue =
[24, 0]'),
  Text(0.8376383763837638, 0.4722222222222222, 'x[0] <= 0.667\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
  Text(0.8302583025830258, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
  Text(0.8450184501845018, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
  Text(0.8523985239852399, 0.5833333333333334, 'x[2] <= 0.715\ngini =
0.408\nsamples = 7\nvalue = [2, 5]'),
  Text(0.8450184501845018, 0.5277777777777778, 'gini = 0.0\nsamples = 5\nvalue =
[0, 5]'),
  Text(0.8597785977859779, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
  Text(0.8745387453874539, 0.75, 'x[26] <= 0.167\ngini = 0.434\nsamples =
22\nvalue = [15, 7]'),
  Text(0.8671586715867159, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
  Text(0.8819188191881919, 0.6944444444444444, 'x[10] <= 0.5\ngini =
0.332\nsamples = 19\nvalue = [15, 4]'),
  Text(0.8745387453874539, 0.6388888888888888, 'gini = 0.0\nsamples = 10\nvalue =
[10, 0]'),

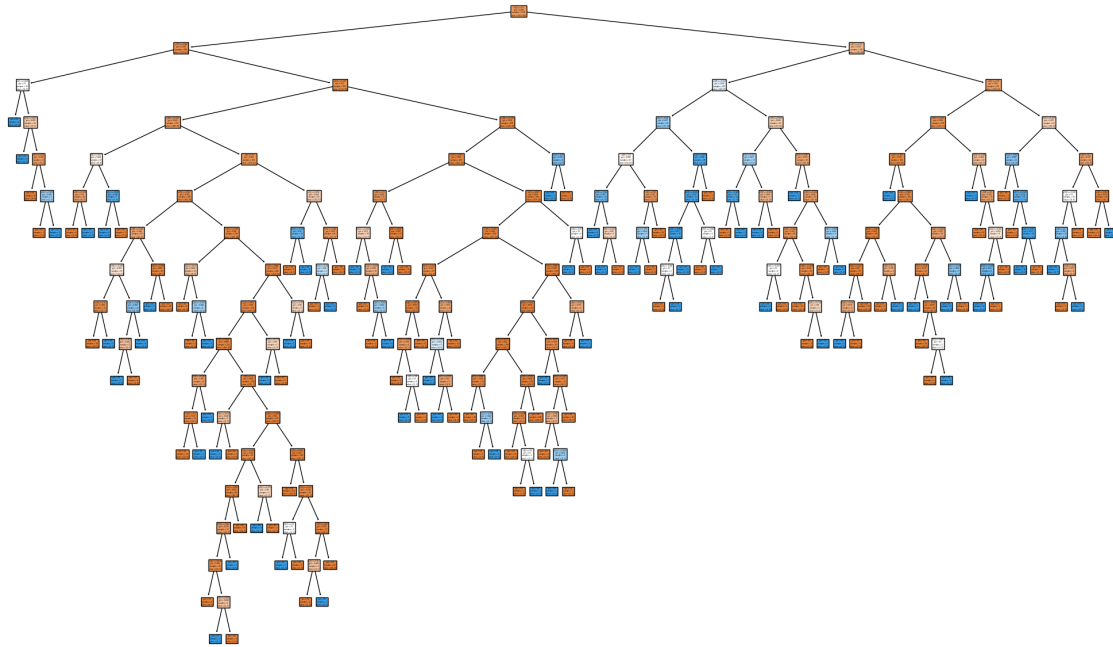
```

```

Text(0.8892988929889298, 0.6388888888888888, 'x[22] <= 0.321\ngini =
0.494\nsamples = 9\nvalue = [5, 4]'),
Text(0.8819188191881919, 0.5833333333333334, 'x[19] <= 0.875\ngini =
0.32\nsamples = 5\nvalue = [1, 4]'),
Text(0.8745387453874539, 0.5277777777777778, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
Text(0.8892988929889298, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.8966789667896679, 0.5833333333333334, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
Text(0.9372693726937269, 0.8055555555555556, 'x[27] <= 0.237\ngini =
0.472\nsamples = 63\nvalue = [39, 24]'),
Text(0.9040590405904059, 0.75, 'x[2] <= 0.069\ngini = 0.408\nsamples =
21\nvalue = [6, 15]'),
Text(0.8966789667896679, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.9114391143911439, 0.6944444444444444, 'x[11] <= 0.129\ngini =
0.278\nsamples = 18\nvalue = [3, 15]'),
Text(0.9040590405904059, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
Text(0.9188191881918819, 0.6388888888888888, 'x[11] <= 0.95\ngini =
0.117\nsamples = 16\nvalue = [1, 15]'),
Text(0.9114391143911439, 0.5833333333333334, 'gini = 0.0\nsamples = 15\nvalue =
[0, 15]'),
Text(0.9261992619926199, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
Text(0.9704797047970479, 0.75, 'x[15] <= 0.5\ngini = 0.337\nsamples = 42\nvalue
= [33, 9]'),
Text(0.955719557195572, 0.6944444444444444, 'x[18] <= 0.678\ngini =
0.5\nsamples = 16\nvalue = [8, 8]'),
Text(0.948339483394834, 0.6388888888888888, 'x[6] <= 0.5\ngini = 0.397\nsamples
= 11\nvalue = [3, 8]'),
Text(0.940959409594096, 0.5833333333333334, 'gini = 0.0\nsamples = 7\nvalue =
[0, 7]'),
Text(0.955719557195572, 0.5833333333333334, 'x[32] <= 0.4\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.948339483394834, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
Text(0.9630996309963099, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
Text(0.9630996309963099, 0.6388888888888888, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
Text(0.985239852398524, 0.6944444444444444, 'x[4] <= 0.911\ngini =
0.074\nsamples = 26\nvalue = [25, 1]'),
Text(0.977859778597786, 0.6388888888888888, 'gini = 0.0\nsamples = 25\nvalue =
[25, 0]'),
Text(0.992619926199262, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =

```

[0, 1]'))



```
[107]: 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']
}
```

```
[110]: grid_search=GridSearchCV(estimator=dtc,param_grid=parameter,cv=5,scoring="accuracy")
```

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

```
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
```



```

warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.

```

```

removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(

```

```

/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(

```

```
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:  
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be  
removed in 1.3. To keep the past behaviour, explicitly set  
`max_features='sqrt'`.  
    warnings.warn(  
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
```



```

warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.

```

[illegible]

```

/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
    warnings.warn(

```



```
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269:
FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be
removed in 1.3. To keep the past behaviour, explicitly set
`max_features='sqrt'`.
warnings.warn(
```

```
[111]: GridSearchCV(cv=5, estimator=DecisionTreeClassifier(),
                  param_grid={'criterion': ['gini', 'entropy'],
                              'max_depth': [1, 2, 3, 4, 5],
                              'max_features': ['auto', 'sqrt', 'log2'],
                              'splitter': ['best', 'random']},
                  scoring='accuracy')
```

```
[112]: grid_search.best_params_
```

```
[112]: {'criterion': 'gini',
        'max_depth': 5,
        'max_features': 'sqrt',
        'splitter': 'random'}
```

```
[113]: dtc_cv=DecisionTreeClassifier(criterion= 'entropy',
        max_depth=3,
        max_features='sqrt',
        splitter='best')
dtc_cv.fit(x_train,y_train)
```

```
[113]: DecisionTreeClassifier(criterion='entropy', max_depth=3, max_features='sqrt')
```

```
[114]: pred=dtc_cv.predict(x_test)
```

```
[115]: print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
No	0.85	1.00	0.92	240
Yes	0.00	0.00	0.00	44
accuracy			0.85	284
macro avg	0.42	0.50	0.46	284
weighted avg	0.71	0.85	0.77	284

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
```

0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

```
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344:
UndefinedMetricWarning: Precision and F-score are ill-defined and being set to
0.0 in labels with no predicted samples. Use `zero_division` parameter to
control this behavior.
_warn_prf(average, modifier, msg_start, len(result))
```

3 Random Forest

```
[116]: from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
```

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

```
[118]: rfc_cv= GridSearchCV(rfc,param_grid=forest_params,cv=10,scoring="accuracy")
```

```
[119]: rfc_cv.fit(x_train,y_train)
```

```
/usr/local/lib/python3.10/dist-
packages/sklearn/model_selection/_validation.py:378: 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 "/usr/local/lib/python3.10/dist-
packages/sklearn/model_selection/_validation.py", line 686, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py",
line 340, in fit
    self._validate_params()
  File "/usr/local/lib/python3.10/dist-packages/sklearn/base.py", line 600, in
_validate_params
    validate_parameter_constraints(
```

```

File "/usr/local/lib/python3.10/dist-
packages/sklearn/utils/_param_validation.py", line 97, 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', 'auto' (deprecated), 'log2'}
or None. Got 0 instead.

```

```

warnings.warn(some_fits_failed_message, FitFailedWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_search.py:952:
UserWarning: One or more of the test scores are non-finite: [      nan
0.84391399 0.85010092 0.85187083 0.85182425 0.8518553
0.85272473 0.85363298 0.85276355 0.85363298 0.85801894 0.85625679
0.8606971 0.85274026      nan 0.84479118 0.84565285 0.85272473
0.85360969 0.85978885 0.85448688 0.85363298 0.85451017 0.85627232
0.85538736 0.85366403 0.85715727 0.85716504      nan 0.84480671
0.85186306 0.85447912 0.85360193 0.85538736 0.85801894 0.85276355
0.8553796 0.85626456 0.85891942 0.85803447 0.86071262 0.85718056
      nan 0.84744605 0.8492082 0.85537184 0.86242043 0.85537184
0.8553796 0.85714951 0.85803447 0.85627232 0.85540289 0.85981214
0.85980438 0.85627232      nan 0.84391399 0.84830772 0.85714951
0.85274802 0.85625679 0.85450241 0.85539512 0.85626456 0.85360969
0.85981214 0.85097035 0.85803447 0.85628784]
warnings.warn(

```

```

[119]: GridSearchCV(cv=10, estimator=RandomForestClassifier(),
    param_grid=[{'max_depth': [10, 11, 12, 13, 14],
                  'max_features': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
                                   12, 13]}],
    scoring='accuracy')

```

```

[120]: pred=rfc_cv.predict(x_test)

```

```

[121]: print(classification_report(y_test,pred))

```

	precision	recall	f1-score	support
No	0.86	1.00	0.92	240
Yes	0.86	0.14	0.24	44
accuracy			0.86	284
macro avg	0.86	0.57	0.58	284
weighted avg	0.86	0.86	0.82	284

```

[122]: rfc_cv.best_params_

```

```
[122]: {'max_depth': 13, 'max_features': 4}
```

```
[ ]:
```