

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

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

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

Out[3]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sci
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sci
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Med
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sci
4	27	No	Travel_Rarely	591	Research & Development	2	1	Med

5 rows × 35 columns

```
In [4]: df.shape
```

```
Out[4]: (1470, 35)
```

```
In [5]: df.columns
```

```
Out[5]: Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

```
In [6]: 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
```

In [7]: df.describe()

Out[7]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000

8 rows × 26 columns

In [8]: `df.isnull().sum()`

```
Out[8]: 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
```

In [9]: `cor=df.corr()`
`cor`

C:\Temp\ipykernel_2552\3865792743.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
`cor=df.corr()`

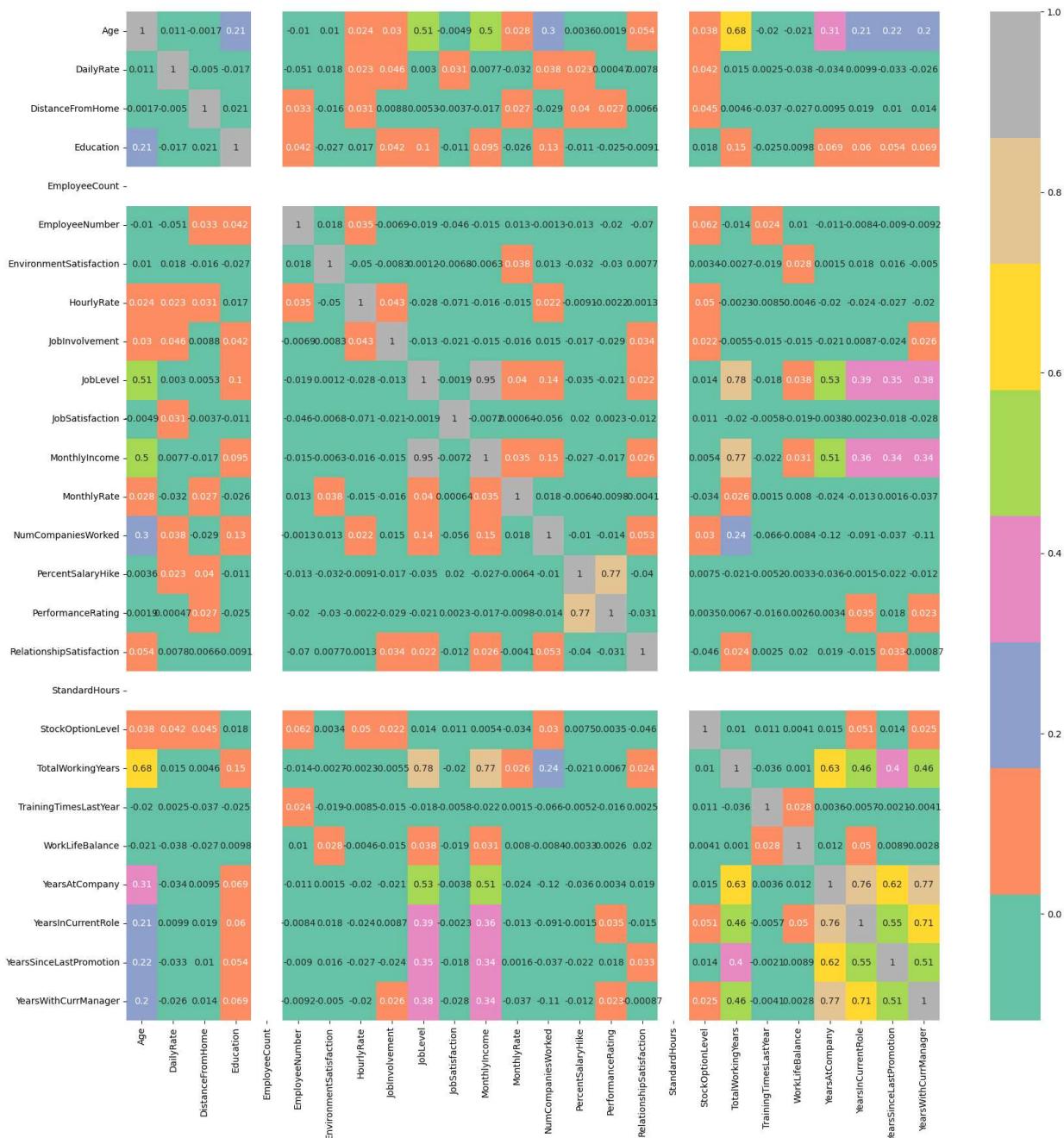
Out[9]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	Empl
Age	1.000000	0.010661	-0.001686	0.208034	NaN	NaN
DailyRate	0.010661	1.000000	-0.004985	-0.016806	NaN	NaN
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	NaN	NaN
Education	0.208034	-0.016806	0.021042	1.000000	NaN	NaN
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN
EmployeeNumber	-0.010145	-0.050990	0.032916	0.042070	NaN	NaN
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	NaN	NaN
HourlyRate	0.024287	0.023381	0.031131	0.016775	NaN	NaN
JobInvolvement	0.029820	0.046135	0.008783	0.042438	NaN	NaN
JobLevel	0.509604	0.002966	0.005303	0.101589	NaN	NaN
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	NaN	NaN
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	NaN	NaN
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	NaN	NaN
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	NaN	NaN
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	NaN	NaN
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	NaN	NaN
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	NaN	NaN
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	NaN	NaN
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	NaN	NaN
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	NaN	NaN
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	NaN	NaN
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	NaN	NaN
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	NaN	NaN
YearsSinceLastPromotion	0.216513	-0.032229	0.010029	0.054254	NaN	NaN
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	NaN	NaN

26 rows × 26 columns

```
In [10]: plt.figure(figsize=(20,20))
sns.heatmap(cor, annot=True, cmap='Set2')
```

Out[10]: <Axes: >



```
In [11]: df.unique()
```

```
Out[11]:
```

Age	43
Attrition	2
BusinessTravel	3
DailyRate	886
Department	3
DistanceFromHome	29
Education	5
EducationField	6
EmployeeCount	1
EmployeeNumber	1470
EnvironmentSatisfaction	4
Gender	2
HourlyRate	71
JobInvolvement	4
JobLevel	5
JobRole	9
JobSatisfaction	4
MaritalStatus	3
MonthlyIncome	1349
MonthlyRate	1427
NumCompaniesWorked	10
Over18	1
Overtime	2
PercentSalaryHike	15
PerformanceRating	2
RelationshipSatisfaction	4
StandardHours	1
StockOptionLevel	4
TotalWorkingYears	40
TrainingTimesLastYear	7
WorkLifeBalance	4
YearsAtCompany	37
YearsInCurrentRole	19
YearsSinceLastPromotion	16
YearsWithCurrManager	18

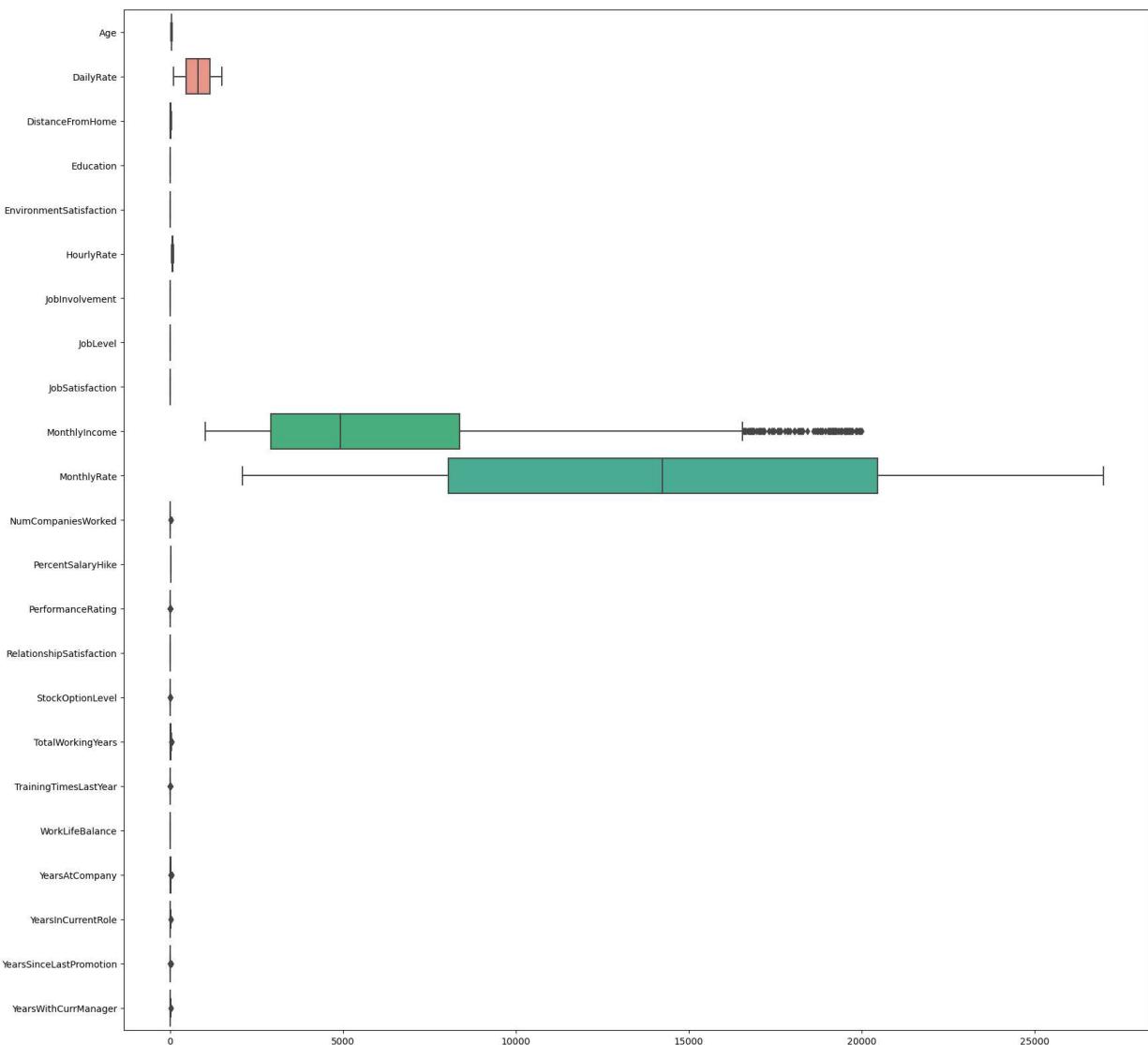
dtype: int64

```
In [72]: df.drop(columns=['EmployeeCount', 'Over18', 'StandardHours', 'EmployeeNumber'], inplace=True)
```

```
In [73]: numeric=['Age', 'DailyRate', 'DistanceFromHome', 'Education', 'EnvironmentSatisfaction', 'HourlyRate', 'JobLevel', 'JobSatisfaction', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked', 'PerformanceRating', 'RelationshipSatisfaction', 'StockOptionLevel', 'TotalWorkingYears', 'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion']
```

```
In [74]: plt.figure(figsize=[20,20])
sns.boxplot(df[numeric], orient='h')
```

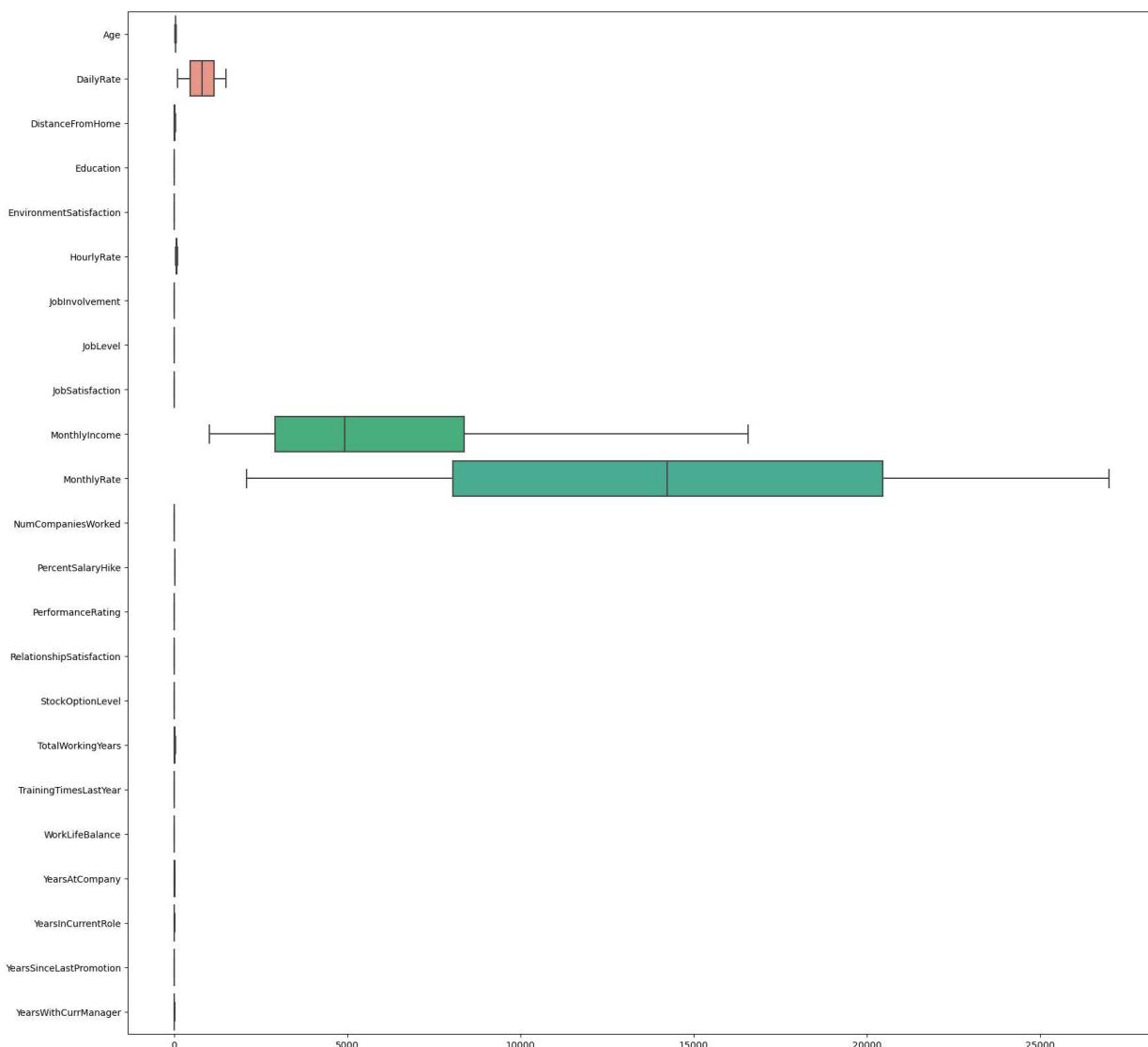
```
Out[74]: <Axes: >
```



```
In [75]: for i in numeric:
    q1=df[i].quantile(0.25)
    q3=df[i].quantile(0.75)
    IQR=q3-q1
    upper_limit=q3+1.5*IQR
    lower_limit=q1-1.5*IQR
    df[i]=np.where(df[i]>upper_limit,upper_limit,np.where(df[i]<lower_limit,lower_limit,df[i]))
```

```
In [76]: plt.figure(figsize=[20,20])
sns.boxplot(df[numeric],orient='h')
```

Out[76]: <Axes: >



In [77]: `df.head()`

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
0	41.0	Yes	Travel_Rarely	1102.0	Sales		1.0	2.0
1	49.0	No	Travel_Frequently	279.0	Research & Development		8.0	1.0
2	37.0	Yes	Travel_Rarely	1373.0	Research & Development		2.0	2.0
3	33.0	No	Travel_Frequently	1392.0	Research & Development		3.0	4.0
4	27.0	No	Travel_Rarely	591.0	Research & Development		2.0	1.0

5 rows × 31 columns

In [78]: `df.columns`

```
Out[78]: Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
       'DistanceFromHome', 'Education', 'EducationField',
       'EnvironmentSatisfaction', 'Gender', 'HourlyRate', 'JobInvolvement',
       'JobLevel', 'JobRole', 'JobSatisfaction', 'MaritalStatus',
       'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked', 'OverTime',
       'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager'],
      dtype='object')
```

```
In [79]: x=df.drop(columns=['Attrition'],axis=1)
```

```
In [80]: x.head()
```

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Env
0	41.0	Travel_Rarely	1102.0	Sales		1.0	2.0	Life Sciences
1	49.0	Travel_Frequently	279.0	Research & Development		8.0	1.0	Life Sciences
2	37.0	Travel_Rarely	1373.0	Research & Development		2.0	2.0	Other
3	33.0	Travel_Frequently	1392.0	Research & Development		3.0	4.0	Life Sciences
4	27.0	Travel_Rarely	591.0	Research & Development		2.0	1.0	Medical

5 rows × 30 columns

```
In [81]: x.shape
```

```
Out[81]: (1470, 30)
```

```
In [82]: type(x)
```

```
Out[82]: pandas.core.frame.DataFrame
```

```
In [83]: y=df.Attrition
```

```
In [84]: y.head()
```

```
Out[84]: 0    Yes
1    No
2    Yes
3    No
4    No
Name: Attrition, dtype: object
```

```
In [85]: y.shape
```

```
Out[85]: (1470,)
```

```
In [86]: type(y)
```

```
Out[86]: pandas.core.series.Series
```

```
In [87]: x.columns
```

```
Out[87]: Index(['Age', 'BusinessTravel', 'DailyRate', 'Department', 'DistanceFromHome',
   'Education', 'EducationField', 'EnvironmentSatisfaction', 'Gender',
   'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole',
   'JobSatisfaction', 'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
   'NumCompaniesWorked', 'OverTime', 'PercentSalaryHike',
   'PerformanceRating', 'RelationshipSatisfaction', 'StockOptionLevel',
   'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
   'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
   'YearsWithCurrManager'],
  dtype='object')
```

```
In [88]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
In [89]: x.Gender=le.fit_transform(x.Gender)
m1=dict(zip(le.classes_,range(len(le.classes_))))
```

```
In [90]: m1
```

```
Out[90]: {'Female': 0, 'Male': 1}
```

```
In [91]: x.OverTime=le.fit_transform(x.OverTime)
m2=dict(zip(le.classes_,range(len(le.classes_))))
m2
```

```
Out[91]: {'No': 0, 'Yes': 1}
```

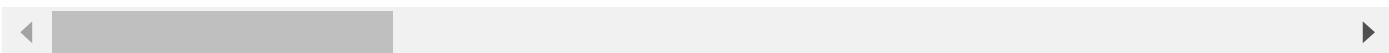
```
In [92]: col=['BusinessTravel','Department','EducationField','JobRole','MaritalStatus']
x_encode=pd.get_dummies(x[col],drop_first=True)
```

```
In [93]: x_encode
```

Out[93]:

	BusinessTravel_Travel_Frequently	BusinessTravel_Travel_Rarely	Department_Research & Development	Department1
0	0	1	0	
1	1	0	1	
2	0	1	1	
3	1	0	1	
4	0	1	1	
...
1465	1	0	1	
1466	0	1	1	
1467	0	1	1	
1468	1	0	0	
1469	0	1	1	

1470 rows × 19 columns



In [94]: `x=pd.concat([x,x_encode],axis=1)`
`x`

Out[94]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField
0	41.0	Travel_Rarely	1102.0	Sales	1.0	2.0	Life Sciences
1	49.0	Travel_Frequently	279.0	Research & Development	8.0	1.0	Life Sciences
2	37.0	Travel_Rarely	1373.0	Research & Development	2.0	2.0	Other
3	33.0	Travel_Frequently	1392.0	Research & Development	3.0	4.0	Life Sciences
4	27.0	Travel_Rarely	591.0	Research & Development	2.0	1.0	Medical
...
1465	36.0	Travel_Frequently	884.0	Research & Development	23.0	2.0	Medical
1466	39.0	Travel_Rarely	613.0	Research & Development	6.0	1.0	Medical
1467	27.0	Travel_Rarely	155.0	Research & Development	4.0	3.0	Life Sciences
1468	49.0	Travel_Frequently	1023.0	Sales	2.0	3.0	Medical
1469	34.0	Travel_Rarely	628.0	Research & Development	8.0	3.0	Medical

1470 rows × 49 columns

--	--

In [95]: `x.drop(x[col],axis=1,inplace=True)`In [96]: `x.shape`

Out[96]: (1470, 44)

In [97]: `x.head()`

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Gender	HourlyRate	JobLevel
0	41.0	1102.0	1.0	2.0	2.0	0	94.0	
1	49.0	279.0	8.0	1.0	3.0	1	61.0	
2	37.0	1373.0	2.0	2.0	4.0	1	92.0	
3	33.0	1392.0	3.0	4.0	4.0	0	56.0	
4	27.0	591.0	2.0	1.0	1.0	1	40.0	

5 rows × 44 columns

--	--

```
In [98]: from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled=pd.DataFrame(ms.fit_transform(x),columns=x.columns)
```

```
In [99]: x_scaled.head()
```

Out[99]:

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	Gender	HourlyRate
0	0.547619	0.715820	0.000000	0.25	0.333333	0.0	0.914286
1	0.738095	0.126700	0.250000	0.00	0.666667	1.0	0.442857
2	0.452381	0.909807	0.035714	0.25	1.000000	1.0	0.885714
3	0.357143	0.923407	0.071429	0.75	1.000000	0.0	0.371429
4	0.214286	0.350036	0.035714	0.00	0.000000	1.0	0.142857

5 rows × 44 columns

```
In [100... from sklearn.model_selection import train_test_split
```

```
In [101... x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

Out[101]: ((1176, 44), (294, 44), (1176,), (294,))

Decision Tree classifier

```
In [102... from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()
```

```
In [103... dtc.fit(x_train,y_train)
```

Out[103]:

DecisionTreeClassifier()

```
In [104... y_pred=dtc.predict(x_test)
```

```
In [105... y_pred
```

```
Out[105]: array(['No', 'No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'Yes', 'Yes',
       'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
       'Yes', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No'],
      dtype=object)
```

```
In [106... diff=pd.DataFrame({ "Actual":y_test,"Predicted":y_pred})
diff
```

Out[106]:

	Actual	Predicted
442	No	No
1091	No	No
981	Yes	No
785	No	Yes
1332	Yes	No
...
1439	No	No
481	No	No
124	Yes	No
198	No	No
1229	No	No

294 rows × 2 columns

```
In [107... from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_
```

```
In [108... accuracy_score(y_test,y_pred)
```

```
Out[108]: 0.7585034013605442
```

```
In [109... confusion_matrix(y_test,y_pred)
```

```
Out[109]: array([[205,  40],  
   [ 31,  18]], dtype=int64)
```

```
In [110... pd.crosstab(y_test,y_pred)
```

```
Out[110]: col_0  No  Yes
```

Attrition

No	205	40
Yes	31	18

```
In [111... print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
No	0.87	0.84	0.85	245
Yes	0.31	0.37	0.34	49
accuracy			0.76	294
macro avg	0.59	0.60	0.59	294
weighted avg	0.78	0.76	0.77	294

```
In [112... from sklearn import tree  
plt.figure(figsize=(25,15))  
tree.plot_tree(dtc,filled=True)
```

```

Out[112]: [Text(0.3251313025210084, 0.9705882352941176, 'x[18] <= 1.5\ngini = 0.269\nsamples = 1176\nvalue = [988, 188']),
Text(0.08907563025210084, 0.9117647058823529, 'x[43] <= 0.5\ngini = 0.5\nsamples = 78\nvalue = [39, 39']),
Text(0.05042016806722689, 0.8529411764705882, 'x[2] <= 16.5\ngini = 0.426\nsamples = 39\nvalue = [27, 12']),
Text(0.03361344537815126, 0.7941176470588235, 'x[25] <= 0.5\ngini = 0.312\nsamples = 31\nvalue = [25, 6']),
Text(0.020168067226890758, 0.7352941176470589, 'x[9] <= 1.5\ngini = 0.198\nsamples = 27\nvalue = [24, 3']),
Text(0.013445378151260505, 0.6764705882352942, 'x[13] <= 0.5\ngini = 0.5\nsamples = 6\nvalue = [3, 3']),
Text(0.0067226890756302525, 0.6176470588235294, 'gini = 0.0\nsamples = 3\nvalue = [3, 0']),
Text(0.020168067226890758, 0.6176470588235294, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]),
Text(0.02689075630252101, 0.6764705882352942, 'gini = 0.0\nsamples = 21\nvalue = [21, 0']),
Text(0.047058823529411764, 0.7352941176470589, 'x[3] <= 2.5\ngini = 0.375\nsamples = 4\nvalue = [1, 3']),
Text(0.040336134453781515, 0.6764705882352942, 'gini = 0.0\nsamples = 1\nvalue = [1, 0']),
Text(0.05378151260504202, 0.6764705882352942, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]),
Text(0.06722689075630252, 0.7941176470588235, 'x[14] <= 20.5\ngini = 0.375\nsamples = 8\nvalue = [2, 6']),
Text(0.06050420168067227, 0.7352941176470589, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]),
Text(0.07394957983193277, 0.7352941176470589, 'gini = 0.0\nsamples = 2\nvalue = [2, 0']),
Text(0.12773109243697478, 0.8529411764705882, 'x[39] <= 0.5\ngini = 0.426\nsamples = 39\nvalue = [12, 27']),
Text(0.10756302521008404, 0.7941176470588235, 'x[31] <= 0.5\ngini = 0.26\nsamples = 26\nvalue = [4, 22']),
Text(0.08739495798319327, 0.7352941176470589, 'x[6] <= 96.0\ngini = 0.095\nsamples = 20\nvalue = [1, 19']),
Text(0.08067226890756303, 0.6764705882352942, 'gini = 0.0\nsamples = 18\nvalue = [0, 18]),
Text(0.09411764705882353, 0.6764705882352942, 'x[4] <= 3.5\ngini = 0.5\nsamples = 2\nvalue = [1, 1']),
Text(0.08739495798319327, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]),
Text(0.10084033613445378, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]),
Text(0.12773109243697478, 0.7352941176470589, 'x[2] <= 9.5\ngini = 0.5\nsamples = 6\nvalue = [3, 3]),
Text(0.12100840336134454, 0.6764705882352942, 'x[1] <= 1063.0\ngini = 0.375\nsamples = 4\nvalue = [1, 3]),
Text(0.11428571428571428, 0.6176470588235294, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]),
Text(0.12773109243697478, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]),
Text(0.13445378151260504, 0.6764705882352942, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]),
Text(0.14789915966386555, 0.7941176470588235, 'x[16] <= 1.5\ngini = 0.473\nsamples = 13\nvalue = [8, 5']),
Text(0.1411764705882353, 0.7352941176470589, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]),
Text(0.1546218487394958, 0.7352941176470589, 'x[6] <= 34.5\ngini = 0.32\nsamples = 1\nvalue = [8, 2]),

```

```

Text(0.14789915966386555, 0.6764705882352942, 'gini = 0.0\nsamples = 2\nvalue = [0,
2']),
Text(0.16134453781512606, 0.6764705882352942, 'gini = 0.0\nsamples = 8\nvalue = [8,
0']),
Text(0.561186974789916, 0.9117647058823529, 'x[13] <= 0.5\ngini = 0.235\nsamples = 1
098\nvalue = [949, 149']'),
Text(0.314390756302521, 0.8529411764705882, 'x[20] <= 1.5\ngini = 0.162\nsamples = 7
98\nvalue = [727, 71']'),
Text(0.20504201680672268, 0.7941176470588235, 'x[1] <= 1146.0\ngini = 0.38\nsamples
= 47\nvalue = [35, 12']'),
Text(0.19831932773109243, 0.7352941176470589, 'x[7] <= 2.5\ngini = 0.463\nsamples =
33\nvalue = [21, 12']'),
Text(0.17478991596638654, 0.6764705882352942, 'x[2] <= 13.5\ngini = 0.42\nsamples =
10\nvalue = [3, 7']'),
Text(0.16806722689075632, 0.6176470588235294, 'gini = 0.0\nsamples = 6\nvalue = [0,
6']'),
Text(0.1815126050420168, 0.6176470588235294, 'x[9] <= 1.5\ngini = 0.375\nsamples = 4
\nvalue = [3, 1']'),
Text(0.17478991596638654, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']'),
Text(0.18823529411764706, 0.5588235294117647, 'gini = 0.0\nsamples = 3\nvalue = [3,
0']'),
Text(0.2218487394957983, 0.6764705882352942, 'x[19] <= 3.5\ngini = 0.34\nsamples = 2
3\nvalue = [18, 5']'),
Text(0.20840336134453782, 0.6176470588235294, 'x[23] <= 5.0\ngini = 0.117\nsamples =
16\nvalue = [15, 1']'),
Text(0.20168067226890757, 0.5588235294117647, 'gini = 0.0\nsamples = 15\nvalue = [1
5, 0']'),
Text(0.21512605042016808, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']'),
Text(0.23529411764705882, 0.6176470588235294, 'x[24] <= 9.0\ngini = 0.49\nsamples =
7\nvalue = [3, 4']'),
Text(0.22857142857142856, 0.5588235294117647, 'gini = 0.0\nsamples = 4\nvalue = [0,
4']'),
Text(0.24201680672268908, 0.5588235294117647, 'gini = 0.0\nsamples = 3\nvalue = [3,
0']'),
Text(0.21176470588235294, 0.7352941176470589, 'gini = 0.0\nsamples = 14\nvalue = [1
4, 0']'),
Text(0.4237394957983193, 0.7941176470588235, 'x[21] <= 4.5\ngini = 0.145\nsamples =
751\nvalue = [692, 59']'),
Text(0.33865546218487397, 0.7352941176470589, 'x[4] <= 1.5\ngini = 0.218\nsamples =
257\nvalue = [225, 32']'),
Text(0.30084033613445377, 0.6764705882352942, 'x[24] <= 2.5\ngini = 0.355\nsamples =
65\nvalue = [50, 15']'),
Text(0.27899159663865547, 0.6176470588235294, 'x[24] <= 0.5\ngini = 0.303\nsamples =
59\nvalue = [48, 11']'),
Text(0.25546218487394956, 0.5588235294117647, 'x[7] <= 2.5\ngini = 0.463\nsamples =
22\nvalue = [14, 8']'),
Text(0.24201680672268908, 0.5, 'x[6] <= 42.5\ngini = 0.198\nsamples = 9\nvalue = [8,
1']'),
Text(0.23529411764705882, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']'),
Text(0.24873949579831933, 0.4411764705882353, 'gini = 0.0\nsamples = 8\nvalue = [8,
0']'),
Text(0.2689075630252101, 0.5, 'x[6] <= 58.0\ngini = 0.497\nsamples = 13\nvalue = [6,
7']'),
Text(0.26218487394957984, 0.4411764705882353, 'gini = 0.0\nsamples = 4\nvalue = [4,
0']'),
Text(0.27563025210084036, 0.4411764705882353, 'x[2] <= 9.0\ngini = 0.346\nsamples =
9\nvalue = [2, 7']'),

```

```

Text(0.2689075630252101, 0.38235294117647056, 'x[40] <= 0.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
Text(0.26218487394957984, 0.3235294117647059, 'gini = 0.0\nsamples = 2\nvalue = [2,
0']),
Text(0.27563025210084036, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.2823529411764706, 0.38235294117647056, 'gini = 0.0\nsamples = 6\nvalue = [0,
6']),
Text(0.3025210084033613, 0.5588235294117647, 'x[9] <= 1.5\ngini = 0.149\nsamples = 3
7\nvalue = [34, 3']),
Text(0.2957983193277311, 0.5, 'x[21] <= 3.5\ngini = 0.5\nsamples = 6\nvalue = [3,
3']),
Text(0.28907563025210087, 0.4411764705882353, 'gini = 0.0\nsamples = 3\nvalue = [0,
3']),
Text(0.3025210084033613, 0.4411764705882353, 'gini = 0.0\nsamples = 3\nvalue = [3,
0']),
Text(0.3092436974789916, 0.5, 'gini = 0.0\nsamples = 31\nvalue = [31, 0']),
Text(0.3226890756302521, 0.6176470588235294, 'x[7] <= 1.5\ngini = 0.444\nsamples = 6
\nvalue = [2, 4']),
Text(0.31596638655462184, 0.5588235294117647, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]),
Text(0.32941176470588235, 0.5588235294117647, 'x[5] <= 0.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
Text(0.3226890756302521, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [2, 0']),
Text(0.33613445378151263, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
Text(0.3764705882352941, 0.6764705882352942, 'x[29] <= 0.5\ngini = 0.161\nsamples =
192\nvalue = [175, 17']),
Text(0.3697478991596639, 0.6176470588235294, 'x[20] <= 2.5\ngini = 0.24\nsamples = 1
22\nvalue = [105, 17']),
Text(0.3563025210084034, 0.5588235294117647, 'x[6] <= 74.0\ngini = 0.463\nsamples =
22\nvalue = [14, 8']),
Text(0.3495798319327731, 0.5, 'x[0] <= 35.5\ngini = 0.444\nsamples = 12\nvalue = [4,
8']),
Text(0.34285714285714286, 0.4411764705882353, 'x[1] <= 277.5\ngini = 0.198\nsamples =
9\nvalue = [1, 8']),
Text(0.33613445378151263, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.3495798319327731, 0.38235294117647056, 'gini = 0.0\nsamples = 8\nvalue = [0,
8]),
Text(0.3563025210084034, 0.4411764705882353, 'gini = 0.0\nsamples = 3\nvalue = [3,
0']),
Text(0.3630252100840336, 0.5, 'gini = 0.0\nsamples = 10\nvalue = [10, 0']),
Text(0.3831932773109244, 0.5588235294117647, 'x[1] <= 112.5\ngini = 0.164\nsamples =
100\nvalue = [91, 9']),
Text(0.3764705882352941, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
Text(0.3899159663865546, 0.5, 'x[10] <= 3079.0\ngini = 0.149\nsamples = 99\nvalue =
[91, 8']),
Text(0.3764705882352941, 0.4411764705882353, 'x[3] <= 4.5\ngini = 0.283\nsamples = 4
1\nvalue = [34, 7']),
Text(0.3697478991596639, 0.38235294117647056, 'x[0] <= 52.5\ngini = 0.224\nsamples =
39\nvalue = [34, 5']),
Text(0.3630252100840336, 0.3235294117647059, 'x[10] <= 3012.0\ngini = 0.188\nsamples =
38\nvalue = [34, 4']),
Text(0.3563025210084034, 0.2647058823529412, 'x[18] <= 18.0\ngini = 0.149\nsamples =
37\nvalue = [34, 3']),
Text(0.3495798319327731, 0.20588235294117646, 'x[10] <= 2751.5\ngini = 0.105\nsamples =
36\nvalue = [34, 2']),
Text(0.34285714285714286, 0.14705882352941177, 'gini = 0.0\nsamples = 29\nvalue = [2
9, 0']),
Text(0.3563025210084034, 0.14705882352941177, 'x[14] <= 13.5\ngini = 0.408\nsamples

```

```

= 7\nvalue = [5, 2']),
Text(0.3495798319327731, 0.08823529411764706, 'gini = 0.0\nsamples = 2\nvalue = [0,
2']),
Text(0.3630252100840336, 0.08823529411764706, 'gini = 0.0\nsamples = 5\nvalue = [5,
0']),
Text(0.3630252100840336, 0.20588235294117646, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.3697478991596639, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.3764705882352941, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.3831932773109244, 0.38235294117647056, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]),
Text(0.40336134453781514, 0.4411764705882353, 'x[1] <= 1458.5\ngini = 0.034\nsamples
= 58\nvalue = [57, 1']),
Text(0.39663865546218485, 0.38235294117647056, 'gini = 0.0\nsamples = 57\nvalue = [5
7, 0']),
Text(0.41008403361344536, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.3831932773109244, 0.6176470588235294, 'gini = 0.0\nsamples = 70\nvalue = [70,
0]),
Text(0.5088235294117647, 0.7352941176470589, 'x[10] <= 6685.0\ngini = 0.103\nsamples
= 494\nvalue = [467, 27']),
Text(0.46218487394957986, 0.6764705882352942, 'x[24] <= 13.5\ngini = 0.06\nsamples =
291\nvalue = [282, 9']),
Text(0.44201680672268906, 0.6176470588235294, 'x[6] <= 70.5\ngini = 0.054\nsamples =
289\nvalue = [281, 8']),
Text(0.4218487394957983, 0.5588235294117647, 'x[12] <= 6.5\ngini = 0.012\nsamples =
169\nvalue = [168, 1']),
Text(0.4151260504201681, 0.5, 'gini = 0.0\nsamples = 151\nvalue = [151, 0']),
Text(0.42857142857142855, 0.5, 'x[10] <= 2840.5\ngini = 0.105\nsamples = 18\nvalue =
[17, 1]),
Text(0.4218487394957983, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.43529411764705883, 0.4411764705882353, 'gini = 0.0\nsamples = 17\nvalue = [1
7, 0]),
Text(0.46218487394957986, 0.5588235294117647, 'x[41] <= 0.5\ngini = 0.11\nsamples =
120\nvalue = [113, 7']),
Text(0.45546218487394957, 0.5, 'x[22] <= 4.5\ngini = 0.096\nsamples = 119\nvalue =
[113, 6']),
Text(0.44873949579831934, 0.4411764705882353, 'x[25] <= 0.5\ngini = 0.183\nsamples =
59\nvalue = [53, 6']),
Text(0.4235294117647059, 0.38235294117647056, 'x[11] <= 22835.0\ngini = 0.115\nsampl
es = 49\nvalue = [46, 3']),
Text(0.40672268907563025, 0.3235294117647059, 'x[10] <= 2064.0\ngini = 0.081\nsample
s = 47\nvalue = [45, 2']),
Text(0.39327731092436974, 0.2647058823529412, 'x[19] <= 1.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
Text(0.3865546218487395, 0.20588235294117646, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.4, 0.20588235294117646, 'gini = 0.0\nsamples = 2\nvalue = [2, 0']),
Text(0.42016806722689076, 0.2647058823529412, 'x[37] <= 0.5\ngini = 0.044\nsamples =
44\nvalue = [43, 1']),
Text(0.4134453781512605, 0.20588235294117646, 'gini = 0.0\nsamples = 41\nvalue = [4
1, 0']),
Text(0.426890756302521, 0.20588235294117646, 'x[14] <= 13.0\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
Text(0.42016806722689076, 0.14705882352941177, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.4336134453781513, 0.14705882352941177, 'gini = 0.0\nsamples = 2\nvalue = [2,

```

```

0]'),
Text(0.4403361344537815, 0.3235294117647059, 'x[18] <= 7.5\ngini = 0.5\nsamples = 2
\nvalue = [1, 1']),
Text(0.4336134453781513, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.4470588235294118, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.47394957983193275, 0.38235294117647056, 'x[9] <= 2.5\ngini = 0.42\nsamples =
10\nvalue = [7, 3']),
Text(0.4672268907563025, 0.3235294117647059, 'x[5] <= 0.5\ngini = 0.375\nsamples = 4
\nvalue = [1, 3']),
Text(0.46050420168067224, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.47394957983193275, 0.2647058823529412, 'gini = 0.0\nsamples = 3\nvalue = [0,
3']),
Text(0.48067226890756304, 0.3235294117647059, 'gini = 0.0\nsamples = 6\nvalue = [6,
0']),
Text(0.46218487394957986, 0.4411764705882353, 'gini = 0.0\nsamples = 60\nvalue = [6
0, 0]),
Text(0.4689075630252101, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
Text(0.4823529411764706, 0.6176470588235294, 'x[6] <= 69.0\ngini = 0.5\nsamples = 2
\nvalue = [1, 1']),
Text(0.4756302521008403, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.4890756302521008, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.5554621848739496, 0.6764705882352942, 'x[10] <= 10497.0\ngini = 0.162\nsample
s = 203\nvalue = [185, 18']),
Text(0.5210084033613446, 0.6176470588235294, 'x[1] <= 148.5\ngini = 0.275\nsamples =
97\nvalue = [81, 16']),
Text(0.5025210084033613, 0.5588235294117647, 'x[19] <= 2.5\ngini = 0.48\nsamples = 5
\nvalue = [2, 3']),
Text(0.4957983193277311, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [2, 0']),
Text(0.5092436974789916, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [0, 3']),
Text(0.5394957983193277, 0.5588235294117647, 'x[0] <= 43.5\ngini = 0.243\nsamples =
92\nvalue = [79, 13']),
Text(0.5226890756302521, 0.5, 'x[11] <= 25806.0\ngini = 0.157\nsamples = 70\nvalue =
[64, 6']),
Text(0.5159663865546219, 0.4411764705882353, 'x[6] <= 98.0\ngini = 0.134\nsamples =
69\nvalue = [64, 5']),
Text(0.5092436974789916, 0.38235294117647056, 'x[10] <= 6704.0\ngini = 0.111\nsample
s = 68\nvalue = [64, 4']),
Text(0.5025210084033613, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5159663865546219, 0.3235294117647059, 'x[24] <= 0.5\ngini = 0.086\nsamples =
67\nvalue = [64, 3']),
Text(0.5025210084033613, 0.2647058823529412, 'x[19] <= 2.5\ngini = 0.5\nsamples = 2
\nvalue = [1, 1']),
Text(0.4957983193277311, 0.20588235294117646, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]),
Text(0.5092436974789916, 0.20588235294117646, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5294117647058824, 0.2647058823529412, 'x[9] <= 1.5\ngini = 0.06\nsamples = 65
\nvalue = [63, 2']),
Text(0.5226890756302521, 0.20588235294117646, 'x[33] <= 0.5\ngini = 0.208\nsamples =
17\nvalue = [15, 2']),
Text(0.5159663865546219, 0.14705882352941177, 'x[2] <= 1.5\ngini = 0.117\nsamples =
16\nvalue = [15, 1']),
Text(0.5092436974789916, 0.08823529411764706, 'x[7] <= 2.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),

```

```

Text(0.5025210084033613, 0.029411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.5159663865546219, 0.029411764705882353, 'gini = 0.0\nsamples = 2\nvalue = [2,
0']),
Text(0.5226890756302521, 0.08823529411764706, 'gini = 0.0\nsamples = 13\nvalue = [1
3, 0']),
Text(0.5294117647058824, 0.14705882352941177, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5361344537815126, 0.20588235294117646, 'gini = 0.0\nsamples = 48\nvalue = [4
8, 0']),
Text(0.5226890756302521, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5294117647058824, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5563025210084034, 0.5, 'x[2] <= 8.5\ngini = 0.434\nsamples = 22\nvalue = [15,
7']),
Text(0.5428571428571428, 0.4411764705882353, 'x[10] <= 10474.0\ngini = 0.231\nsample
s = 15\nvalue = [13, 2']),
Text(0.5361344537815126, 0.38235294117647056, 'x[12] <= 0.5\ngini = 0.133\nsamples =
14\nvalue = [13, 1']),
Text(0.5294117647058824, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5428571428571428, 0.3235294117647059, 'gini = 0.0\nsamples = 13\nvalue = [13,
0]),
Text(0.5495798319327732, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.5697478991596638, 0.4411764705882353, 'x[22] <= 10.5\ngini = 0.408\nsamples =
7\nvalue = [2, 5']),
Text(0.5630252100840336, 0.38235294117647056, 'gini = 0.0\nsamples = 5\nvalue = [0,
5]),
Text(0.5764705882352941, 0.38235294117647056, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]),
Text(0.5899159663865546, 0.6176470588235294, 'x[7] <= 1.5\ngini = 0.037\nsamples = 1
06\nvalue = [104, 2']),
Text(0.5764705882352941, 0.5588235294117647, 'x[24] <= 9.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
Text(0.5697478991596638, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
Text(0.5831932773109244, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [2, 0']),
Text(0.6033613445378151, 0.5588235294117647, 'x[17] <= 2.25\ngini = 0.019\nsamples =
103\nvalue = [102, 1']),
Text(0.5966386554621849, 0.5, 'gini = 0.0\nsamples = 99\nvalue = [99, 0']),
Text(0.6100840336134454, 0.5, 'x[19] <= 3.5\ngini = 0.375\nsamples = 4\nvalue = [3,
1]),
Text(0.6033613445378151, 0.4411764705882353, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]),
Text(0.6168067226890757, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.8079831932773109, 0.8529411764705882, 'x[10] <= 3995.0\ngini = 0.385\nsamples
= 300\nvalue = [222, 78']),
Text(0.7084033613445379, 0.7941176470588235, 'x[17] <= 0.5\ngini = 0.5\nsamples = 96
\nvalue = [49, 47']),
Text(0.6739495798319328, 0.7352941176470589, 'x[2] <= 5.5\ngini = 0.459\nsamples = 4
2\nvalue = [15, 27']),
Text(0.6504201680672269, 0.6764705882352942, 'x[11] <= 12300.5\ngini = 0.499\nsample
s = 23\nvalue = [12, 11']),
Text(0.6369747899159663, 0.6176470588235294, 'x[31] <= 0.5\ngini = 0.426\nsamples =
13\nvalue = [4, 9']),
Text(0.6302521008403361, 0.5588235294117647, 'x[19] <= 1.5\ngini = 0.298\nsamples =
11\nvalue = [2, 9']),
Text(0.6235294117647059, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1, 0']),

```

```

Text(0.6369747899159663, 0.5, 'x[11] <= 9772.0\ngini = 0.18\nsamples = 10\nvalue =
[1, 9']),
Text(0.6302521008403361, 0.4411764705882353, 'gini = 0.0\nsamples = 8\nvalue = [0,
8']),
Text(0.6436974789915967, 0.4411764705882353, 'x[1] <= 1008.5\ngini = 0.5\nsamples =
2\nvalue = [1, 1']),
Text(0.6369747899159663, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.6504201680672269, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
Text(0.6436974789915967, 0.5588235294117647, 'gini = 0.0\nsamples = 2\nvalue = [2,
0']),
Text(0.6638655462184874, 0.6176470588235294, 'x[10] <= 3055.5\ngini = 0.32\nsamples
= 10\nvalue = [8, 2']),
Text(0.6571428571428571, 0.5588235294117647, 'gini = 0.0\nsamples = 7\nvalue = [7,
0']),
Text(0.6705882352941176, 0.5588235294117647, 'x[6] <= 47.0\ngini = 0.444\nsamples =
3\nvalue = [1, 2']),
Text(0.6638655462184874, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1, 0']),
Text(0.6773109243697479, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0, 2']),
Text(0.6974789915966386, 0.6764705882352942, 'x[8] <= 1.5\ngini = 0.266\nsamples = 1
9\nvalue = [3, 16']),
Text(0.6907563025210084, 0.6176470588235294, 'x[6] <= 44.0\ngini = 0.198\nsamples =
18\nvalue = [2, 16']),
Text(0.6840336134453782, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [1,
0']),
Text(0.6974789915966386, 0.5588235294117647, 'x[23] <= 6.5\ngini = 0.111\nsamples =
17\nvalue = [1, 16']),
Text(0.6907563025210084, 0.5, 'gini = 0.0\nsamples = 15\nvalue = [0, 15']),
Text(0.704201680672269, 0.5, 'x[19] <= 1.25\ngini = 0.5\nsamples = 2\nvalue = [1,
1']),
Text(0.6974789915966386, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]),
Text(0.7109243697478992, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
Text(0.704201680672269, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]),
Text(0.7428571428571429, 0.7352941176470589, 'x[0] <= 26.5\ngini = 0.466\nsamples =
54\nvalue = [34, 20']),
Text(0.7243697478991596, 0.6764705882352942, 'x[0] <= 22.5\ngini = 0.245\nsamples =
7\nvalue = [1, 6']),
Text(0.7176470588235294, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]),
Text(0.7310924369747899, 0.6176470588235294, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]),
Text(0.761344537815126, 0.6764705882352942, 'x[1] <= 971.5\ngini = 0.418\nsamples =
47\nvalue = [33, 14']),
Text(0.7445378151260504, 0.6176470588235294, 'x[1] <= 304.0\ngini = 0.482\nsamples =
32\nvalue = [19, 13']),
Text(0.7310924369747899, 0.5588235294117647, 'x[21] <= 9.5\ngini = 0.18\nsamples = 1
0\nvalue = [9, 1]),
Text(0.7243697478991596, 0.5, 'gini = 0.0\nsamples = 9\nvalue = [9, 0']),
Text(0.7378151260504202, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
Text(0.7579831932773109, 0.5588235294117647, 'x[11] <= 23769.5\ngini = 0.496\nsample
s = 22\nvalue = [10, 12']),
Text(0.7512605042016807, 0.5, 'x[29] <= 0.5\ngini = 0.465\nsamples = 19\nvalue = [7,
12']),
Text(0.7378151260504202, 0.4411764705882353, 'x[1] <= 851.0\ngini = 0.298\nsamples =
11\nvalue = [2, 9']),
Text(0.7310924369747899, 0.38235294117647056, 'gini = 0.0\nsamples = 8\nvalue = [0,

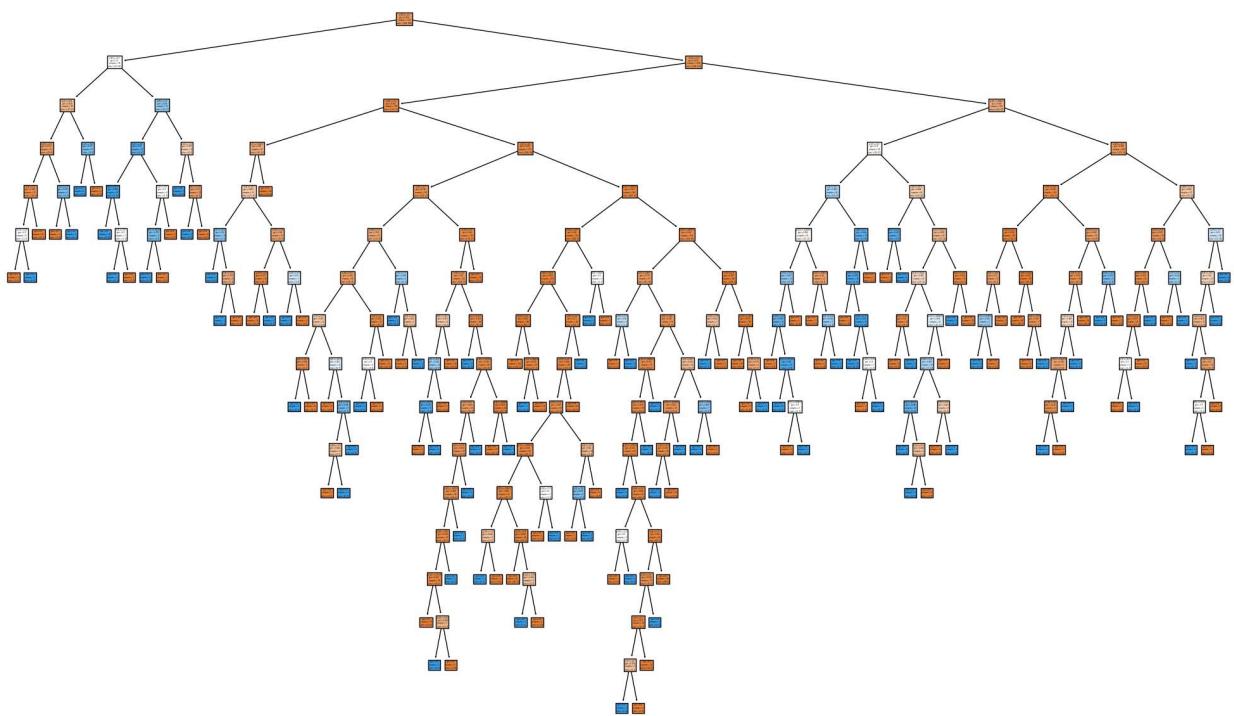
```

```

8'],
  Text(0.7445378151260504, 0.38235294117647056, 'x[31] <= 0.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1']),
  Text(0.7378151260504202, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
  Text(0.7512605042016807, 0.3235294117647059, 'gini = 0.0\nsamples = 2\nvalue = [2,
0']),
  Text(0.7647058823529411, 0.4411764705882353, 'x[35] <= 0.5\ngini = 0.469\nsamples =
8\nvalue = [5, 3']),
  Text(0.7579831932773109, 0.38235294117647056, 'gini = 0.0\nsamples = 5\nvalue = [5,
0']),
  Text(0.7714285714285715, 0.38235294117647056, 'gini = 0.0\nsamples = 3\nvalue = [0,
3']),
  Text(0.7647058823529411, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [3, 0']),
  Text(0.7781512605042017, 0.6176470588235294, 'x[6] <= 34.5\ngini = 0.124\nsamples =
15\nvalue = [14, 1']),
  Text(0.7714285714285715, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
  Text(0.7848739495798319, 0.5588235294117647, 'gini = 0.0\nsamples = 14\nvalue = [14,
0']),
  Text(0.907563025210084, 0.7941176470588235, 'x[43] <= 0.5\ngini = 0.258\nsamples = 2
04\nvalue = [173, 31']),
  Text(0.8521008403361344, 0.7352941176470589, 'x[2] <= 14.5\ngini = 0.138\nsamples =
147\nvalue = [136, 11']),
  Text(0.8184873949579832, 0.6764705882352942, 'x[21] <= 2.5\ngini = 0.056\nsamples =
105\nvalue = [102, 3']),
  Text(0.8050420168067227, 0.6176470588235294, 'x[6] <= 43.5\ngini = 0.32\nsamples = 1
0\nvalue = [8, 2']),
  Text(0.7983193277310925, 0.5588235294117647, 'x[1] <= 1238.5\ngini = 0.444\nsamples =
3\nvalue = [1, 2']),
  Text(0.7915966386554621, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0, 2']),
  Text(0.8050420168067227, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1, 0']),
  Text(0.8117647058823529, 0.5588235294117647, 'gini = 0.0\nsamples = 7\nvalue = [7,
0']),
  Text(0.8319327731092437, 0.6176470588235294, 'x[0] <= 54.0\ngini = 0.021\nsamples =
95\nvalue = [94, 1']),
  Text(0.8252100840336134, 0.5588235294117647, 'gini = 0.0\nsamples = 88\nvalue = [88,
0']),
  Text(0.838655462184874, 0.5588235294117647, 'x[6] <= 77.0\ngini = 0.245\nsamples = 7
\nvalue = [6, 1']),
  Text(0.8319327731092437, 0.5, 'gini = 0.0\nsamples = 6\nvalue = [6, 0']),
  Text(0.8453781512605042, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1']),
  Text(0.8857142857142857, 0.6764705882352942, 'x[30] <= 0.5\ngini = 0.308\nsamples =
42\nvalue = [34, 8']),
  Text(0.8722689075630252, 0.6176470588235294, 'x[4] <= 2.5\ngini = 0.229\nsamples = 3
8\nvalue = [33, 5']),
  Text(0.865546218487395, 0.5588235294117647, 'x[12] <= 5.0\ngini = 0.486\nsamples = 1
2\nvalue = [7, 5']),
  Text(0.8588235294117647, 0.5, 'x[10] <= 11547.0\ngini = 0.346\nsamples = 9\nvalue =
[7, 2']),
  Text(0.8521008403361344, 0.4411764705882353, 'x[18] <= 6.0\ngini = 0.219\nsamples =
8\nvalue = [7, 1']),
  Text(0.8453781512605042, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0,
1']),
  Text(0.8588235294117647, 0.38235294117647056, 'gini = 0.0\nsamples = 7\nvalue = [7,
0']),
  Text(0.865546218487395, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]),
  Text(0.8722689075630252, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [0, 3']),
  Text(0.8789915966386554, 0.5588235294117647, 'gini = 0.0\nsamples = 26\nvalue = [26,

```

```
0]'),
Text(0.8991596638655462, 0.6176470588235294, 'x[14] <= 17.5\n gini = 0.375\n samples =
4\n value = [1, 3]'),
Text(0.892436974789916, 0.5588235294117647, 'gini = 0.0\n samples = 3\n value = [0,
3]'),
Text(0.9058823529411765, 0.5588235294117647, 'gini = 0.0\n samples = 1\n value = [1,
0]'),
Text(0.9630252100840336, 0.7352941176470589, 'x[40] <= 0.5\n gini = 0.456\n samples =
57\n value = [37, 20]'),
Text(0.9394957983193277, 0.6764705882352942, 'x[23] <= 6.0\n gini = 0.238\n samples =
29\n value = [25, 4]'),
Text(0.9260504201680673, 0.6176470588235294, 'x[6] <= 97.5\n gini = 0.142\n samples =
26\n value = [24, 2]'),
Text(0.9193277310924369, 0.5588235294117647, 'x[7] <= 1.5\n gini = 0.077\n samples = 2
5\n value = [24, 1]'),
Text(0.9126050420168067, 0.5, 'x[2] <= 6.0\n gini = 0.5\n samples = 2\n value = [1,
1]'),
Text(0.9058823529411765, 0.4411764705882353, 'gini = 0.0\n samples = 1\n value = [1,
0]'),
Text(0.9193277310924369, 0.4411764705882353, 'gini = 0.0\n samples = 1\n value = [0,
1]'),
Text(0.9260504201680673, 0.5, 'gini = 0.0\n samples = 23\n value = [23, 0]'),
Text(0.9327731092436975, 0.5588235294117647, 'gini = 0.0\n samples = 1\n value = [0,
1]'),
Text(0.9529411764705882, 0.6176470588235294, 'x[7] <= 2.5\n gini = 0.444\n samples = 3
\n value = [1, 2]'),
Text(0.946218487394958, 0.5588235294117647, 'gini = 0.0\n samples = 1\n value = [1,
0]'),
Text(0.9596638655462185, 0.5588235294117647, 'gini = 0.0\n samples = 2\n value = [0,
2]'),
Text(0.9865546218487395, 0.6764705882352942, 'x[23] <= 1.5\n gini = 0.49\n samples = 2
8\n value = [12, 16]'),
Text(0.9798319327731092, 0.6176470588235294, 'x[2] <= 23.5\n gini = 0.48\n samples = 2
0\n value = [12, 8]'),
Text(0.973109243697479, 0.5588235294117647, 'x[2] <= 1.5\n gini = 0.415\n samples = 17
\n value = [12, 5]'),
Text(0.9663865546218487, 0.5, 'gini = 0.0\n samples = 2\n value = [0, 2]'),
Text(0.9798319327731092, 0.5, 'x[4] <= 1.5\n gini = 0.32\n samples = 15\n value = [12,
3]'),
Text(0.973109243697479, 0.4411764705882353, 'x[0] <= 35.0\n gini = 0.5\n samples = 6\n
value = [3, 3]'),
Text(0.9663865546218487, 0.38235294117647056, 'gini = 0.0\n samples = 3\n value = [0,
3]'),
Text(0.9798319327731092, 0.38235294117647056, 'gini = 0.0\n samples = 3\n value = [3,
0]'),
Text(0.9865546218487395, 0.4411764705882353, 'gini = 0.0\n samples = 9\n value = [9,
0]'),
Text(0.9865546218487395, 0.5588235294117647, 'gini = 0.0\n samples = 3\n value = [0,
3]'),
Text(0.9932773109243698, 0.6176470588235294, 'gini = 0.0\n samples = 8\n value = [0,
8]')
```



```
In [113... from sklearn.model_selection import GridSearchCV
parameter={
    'criterion':['gini','entropy'],
    'splitter':['best','random'],
    'max_depth':[1,2,3,4,5],
    'max_features':['auto', 'sqrt', 'log2']

}

In [114... grid_search=GridSearchCV(estimator=dtc,param_grid=parameter,cv=5,scoring="accuracy")

In [115... grid_search.fit(x_train,y_train)
```

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

Below are more details about the failures:

```
100 fits failed with the following error:
Traceback (most recent call last):
  File "D:\anac\Lib\site-packages\sklearn\model_selection\_validation.py", line 729,
in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "D:\anac\Lib\site-packages\sklearn\base.py", line 1145, in wrapper
    estimator._validate_params()
  File "D:\anac\Lib\site-packages\sklearn\base.py", line 638, in _validate_params
    validate_parameter_constraints()
  File "D:\anac\Lib\site-packages\sklearn\utils\_param_validation.py", line 95, in va
lidate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features' parameter o
f DecisionTreeClassifier must be an int in the range [1, inf), a float in the range
(0.0, 1.0], a str among {'sqrt', 'log2'} or None. Got 'auto' instead.

    warnings.warn(some_fits_failed_message, FitFailedWarning)
D:\anac\Lib\site-packages\sklearn\model_selection\_search.py:979: UserWarning: One or
more of the test scores are non-finite: [      nan      nan  0.84013704  0.84013704
 0.84013704  0.84013704
      nan      nan  0.84269023  0.83588172  0.83928597  0.84013704
      nan      nan  0.83842409  0.84012982  0.83248828  0.84013704
      nan      nan  0.84099171  0.83927876  0.8401154   0.84183916
      nan      nan  0.84438154  0.83163361  0.82821132  0.8273855
      nan      nan  0.84013704  0.84013704  0.84013704  0.84013704
      nan      nan  0.8409881   0.83843491  0.83928597  0.83758384
      nan      nan  0.84097728  0.83928597  0.84183916  0.84183916
      nan      nan  0.8384313   0.83504508  0.83928237  0.83417959
      nan      nan  0.82143887  0.83842048  0.83164803  0.83247386]
```

Out[115]:

```
▶      GridSearchCV
▶ estimator: DecisionTreeClassifier
    ▶ DecisionTreeClassifier
```

In [116...]

```
best_params=grid_search.best_params_
best_params
```

Out[116]:

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

In [117...]

```
dtc_cv=DecisionTreeClassifier(**best_params)
dtc_cv.fit(x_train,y_train)
```

Out[117]:

```
▼ DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=5, max_features='sqrt')
```

In [118...]

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

In [119...]

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

	precision	recall	f1-score	support
No	0.85	0.98	0.91	245
Yes	0.60	0.12	0.20	49
accuracy			0.84	294
macro avg	0.72	0.55	0.56	294
weighted avg	0.81	0.84	0.79	294

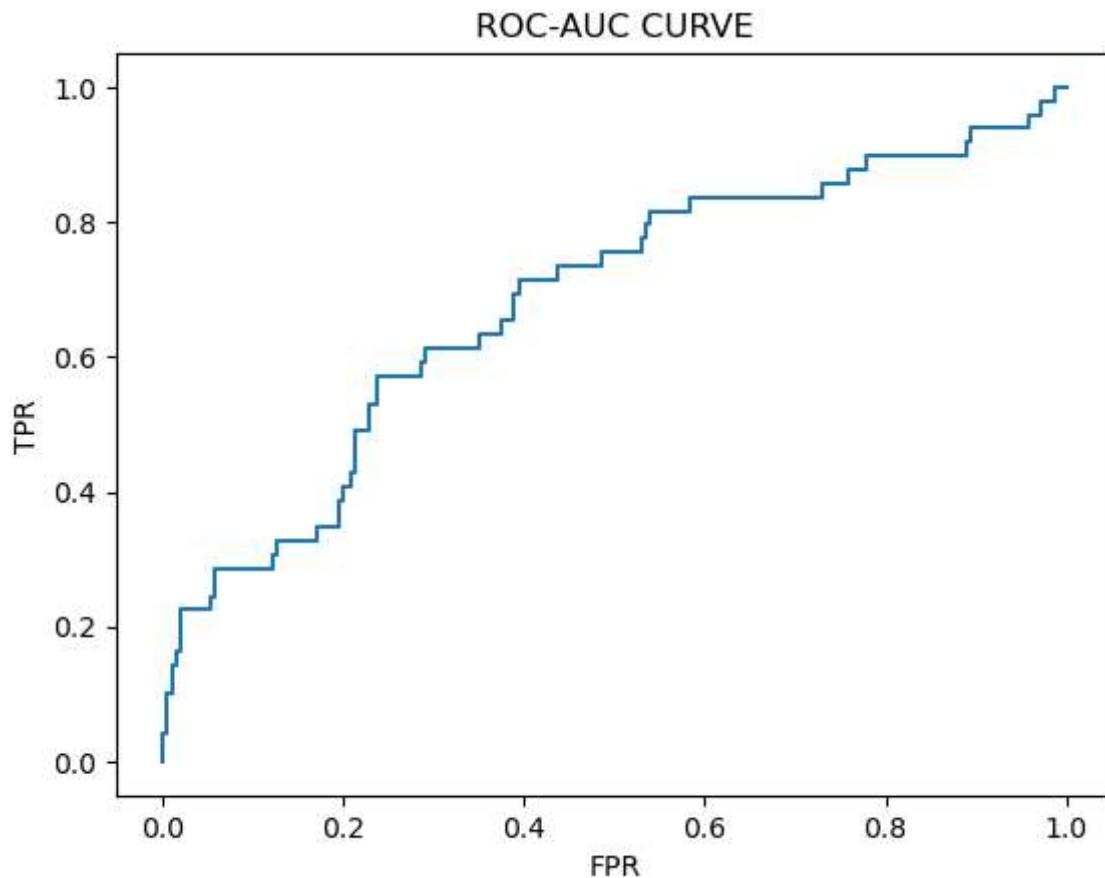
In [120...]

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

```
Out[120]: array([0.08999511, 0.29555124, 0.26244307, 0.08991852, 0.4326493 ,  
    0.2955407 , 0.06055374, 0.07123524, 0.02269479, 0.17942702,  
    0.02435243, 0.317242 , 0.12218491, 0.13694187, 0.07265891,  
    0.08278287, 0.10997225, 0.15562138, 0.10216015, 0.19202755,  
    0.28049572, 0.0244915 , 0.12998219, 0.18983644, 0.37957492,  
    0.07152787, 0.1589406 , 0.04706495, 0.33367402, 0.11777139,  
    0.03926955, 0.06616821, 0.13607836, 0.03089669, 0.10297573,  
    0.08054468, 0.05055832, 0.24779596, 0.05785782, 0.33410343,  
    0.03023164, 0.0911186 , 0.02417157, 0.02539921, 0.04975705,  
    0.21825976, 0.2865197 , 0.07226293, 0.37572481, 0.22185117,  
    0.08865693, 0.20849442, 0.13292131, 0.09523314, 0.17680313,  
    0.09450062, 0.16202248, 0.10567899, 0.17256962, 0.2387301 ,  
    0.06541198, 0.30739986, 0.04094804, 0.10813906, 0.25899322,  
    0.08718281, 0.10922285, 0.09371157, 0.15791503, 0.19795304,  
    0.02569542, 0.30494357, 0.20065601, 0.24623769, 0.20292391,  
    0.28493382, 0.16858508, 0.04449592, 0.35074745, 0.05324709,  
    0.02336322, 0.07119952, 0.24189106, 0.12382623, 0.08620189,  
    0.12957883, 0.01221489, 0.10576623, 0.11598709, 0.18617529,  
    0.32807017, 0.19083055, 0.30403466, 0.18300404, 0.06586486,  
    0.29707869, 0.33102258, 0.25985136, 0.16977209, 0.08714701,  
    0.19552502, 0.34120333, 0.04587907, 0.07210308, 0.16508533,  
    0.10038997, 0.05412877, 0.3073702 , 0.12967562, 0.11195285,  
    0.10182736, 0.0183611 , 0.08083995, 0.16622153, 0.15418735,  
    0.0678074 , 0.05449195, 0.12411172, 0.03230875, 0.0565903 ,  
    0.2235917 , 0.15311919, 0.087823 , 0.41019233, 0.09716099,  
    0.01038681, 0.04806032, 0.22558672, 0.18193589, 0.05417683,  
    0.01716002, 0.19557016, 0.23567198, 0.14258709, 0.04729506,  
    0.03498608, 0.20340339, 0.08304733, 0.15843408, 0.082499 ,  
    0.0903528 , 0.0857862 , 0.08991551, 0.08436046, 0.04662925,  
    0.12330939, 0.02177829, 0.00919274, 0.03236404, 0.13329946,  
    0.11674901, 0.14156237, 0.19567851, 0.01025776, 0.02915897,  
    0.05723929, 0.02835169, 0.21375908, 0.06364801, 0.04449479,  
    0.01994729, 0.02209586, 0.27110199, 0.45938449, 0.07121079,  
    0.23606694, 0.05985707, 0.23050211, 0.28152716, 0.03288207,  
    0.07511987, 0.06666555, 0.10946387, 0.37616496, 0.01509448,  
    0.30156161, 0.18832063, 0.03026292, 0.17839845, 0.10110998,  
    0.190615 , 0.08979994, 0.02572184, 0.20813859, 0.1471404 ,  
    0.17261704, 0.22709376, 0.05715968, 0.36197623, 0.32867 ,  
    0.32063466, 0.31712776, 0.19584048, 0.15737967, 0.25533286,  
    0.060488 , 0.09861639, 0.04502715, 0.32284015, 0.05449148,  
    0.19866029, 0.24762068, 0.33571949, 0.16216912, 0.29150677,  
    0.13743857, 0.32691385, 0.03630853, 0.16001811, 0.04891474,  
    0.15775034, 0.28860612, 0.09081728, 0.19839663, 0.10794551,  
    0.13360027, 0.01977318, 0.14876907, 0.13201554, 0.23826818,  
    0.16208067, 0.29357259, 0.43413371, 0.02442261, 0.22163881,  
    0.14226426, 0.18468954, 0.18879816, 0.3457911 , 0.08840592,  
    0.02887249, 0.17454853, 0.17237776, 0.11015596, 0.0794893 ,  
    0.23157808, 0.32383649, 0.06169046, 0.14867717, 0.04954239,  
    0.14995523, 0.24007068, 0.01219199, 0.16835347, 0.06045922,  
    0.14148966, 0.17004406, 0.1418339 , 0.07587647, 0.05012165,  
    0.34678611, 0.05809208, 0.58511068, 0.11970087, 0.11706498,  
    0.16958423, 0.01063199, 0.39063657, 0.3525142 , 0.14772956,  
    0.1394325 , 0.07900901, 0.08205538, 0.21418867, 0.09703341,  
    0.20973871, 0.05918156, 0.16010086, 0.21264622, 0.13349089,  
    0.14176302, 0.06908087, 0.33540582, 0.13756998, 0.06586227,  
    0.04492486, 0.08112604, 0.11063329, 0.27267983, 0.2244248 ,  
    0.2720098 , 0.13017097, 0.20078604, 0.1029593 , 0.00869502,  
    0.16956319, 0.06629013, 0.22306265, 0.00870342, 0.07448481,  
    0.15434505, 0.15020701, 0.18824581, 0.18443288])
```

```
In [121]: y_test = y_test.map({'Yes': 1, 'No': 0}).astype(int)
fpr,tpr,thresholds=roc_curve(y_test,probability, pos_label=1)

In [122]: plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC-AUC CURVE')
plt.show()
```



Logistic Regression

```
In [43]: from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

```
In [44]: model.fit(x_train,y_train)
```

D:\anac\Lib\site-packages\sklearn\linear_model_logistic.py:460: ConvergenceWarning:
lbfgs failed to converge (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT.

Increase the number of iterations (`max_iter`) or scale the data as shown in:
<https://scikit-learn.org/stable/modules/preprocessing.html>
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
`n_iter_i = _check_optimize_result(`

Out[44]: `LogisticRegression`
`LogisticRegression()`

```
In [45]: y_pred=model.predict(x_test)
```

```
In [46]: y_pred
```

```
In [47]: diff1=pd.DataFrame({"Actual":y_test,"Predicted":y_pred})  
diff1
```

Out[47]:

	Actual	Predicted
442	No	No
1091	No	No
981	Yes	No
785	No	No
1332	Yes	No
...
1439	No	No
481	No	No
124	Yes	No
198	No	No
1229	No	No

294 rows × 2 columns

In [48]: `from sklearn.metrics import confusion_matrix, accuracy_score, classification_report, roc_`In [49]: `accuracy_score(y_pred, y_test)`

Out[49]: 0.8367346938775511

In [50]: `confusion_matrix(y_test, y_pred)`Out[50]: `array([[245, 0], [48, 1]], dtype=int64)`In [51]: `pd.crosstab(y_test, y_pred)`

Out[51]:

	col_0	No	Yes
Attrition			
No	245	0	
Yes	48	1	

In [52]: `print(classification_report(y_test, y_pred))`

	precision	recall	f1-score	support
No	0.84	1.00	0.91	245
Yes	1.00	0.02	0.04	49
accuracy			0.84	294
macro avg	0.92	0.51	0.48	294
weighted avg	0.86	0.84	0.77	294

```
In [55]: probability=model.predict_proba(x_test)[:,1]
probability
```

```
Out[55]: array([ 0.08999511,  0.29555124,  0.26244307,  0.08991852,  0.4326493 ,  
   0.2955407 ,  0.06055374,  0.07123524,  0.02269479,  0.17942702,  
   0.02435243,  0.317242 ,  0.12218491,  0.13694187,  0.07265891,  
   0.08278287,  0.10997225,  0.15562138,  0.10216015,  0.19202755,  
   0.28049572,  0.0244915 ,  0.12998219,  0.18983644,  0.37957492,  
   0.07152787,  0.1589406 ,  0.04706495,  0.33367402,  0.11777139,  
   0.03926955,  0.06616821,  0.13607836,  0.03089669,  0.10297573,  
   0.08054468,  0.05055832,  0.24779596,  0.05785782,  0.33410343,  
   0.03023164,  0.0911186 ,  0.02417157,  0.02539921,  0.04975705,  
   0.21825976,  0.2865197 ,  0.07226293,  0.37572481,  0.22185117,  
   0.08865693,  0.20849442,  0.13292131,  0.09523314,  0.17680313,  
   0.09450062,  0.16202248,  0.10567899,  0.17256962,  0.2387301 ,  
   0.06541198,  0.30739986,  0.04094804,  0.10813906,  0.25899322,  
   0.08718281,  0.10922285,  0.09371157,  0.15791503,  0.19795304,  
   0.02569542,  0.30494357,  0.20065601,  0.24623769,  0.20292391,  
   0.28493382,  0.16858508,  0.04449592,  0.35074745,  0.05324709,  
   0.02336322,  0.07119952,  0.24189106,  0.12382623,  0.08620189,  
   0.12957883,  0.01221489,  0.10576623,  0.11598709,  0.18617529,  
   0.32807017,  0.19083055,  0.30403466,  0.18300404,  0.06586486,  
   0.29707869,  0.33102258,  0.25985136,  0.16977209,  0.08714701,  
   0.19552502,  0.34120333,  0.04587907,  0.07210308,  0.16508533,  
   0.10038997,  0.05412877,  0.3073702 ,  0.12967562,  0.11195285,  
   0.10182736,  0.0183611 ,  0.08083995,  0.16622153,  0.15418735,  
   0.0678074 ,  0.05449195,  0.12411172,  0.03230875,  0.0565903 ,  
   0.2235917 ,  0.15311919,  0.087823 ,  0.41019233,  0.09716099,  
   0.01038681,  0.04806032,  0.22558672,  0.18193589,  0.05417683,  
   0.01716002,  0.19557016,  0.23567198,  0.14258709,  0.04729506,  
   0.03498608,  0.20340339,  0.08304733,  0.15843408,  0.082499 ,  
   0.0903528 ,  0.0857862 ,  0.08991551,  0.08436046,  0.04662925,  
   0.12330939,  0.02177829,  0.00919274,  0.03236404,  0.13329946,  
   0.11674901,  0.14156237,  0.19567851,  0.01025776,  0.02915897,  
   0.05723929,  0.02835169,  0.21375908,  0.06364801,  0.04449479,  
   0.01994729,  0.02209586,  0.27110199,  0.45938449,  0.07121079,  
   0.23606694,  0.05985707,  0.23050211,  0.28152716,  0.03288207,  
   0.07511987,  0.06666555,  0.10946387,  0.37616496,  0.01509448,  
   0.30156161,  0.18832063,  0.03026292,  0.17839845,  0.10110998,  
   0.190615 ,  0.08979994,  0.02572184,  0.20813859,  0.1471404 ,  
   0.17261704,  0.22709376,  0.05715968,  0.36197623,  0.32867 ,  
   0.32063466,  0.31712776,  0.19584048,  0.15737967,  0.25533286,  
   0.060488 ,  0.09861639,  0.04502715,  0.32284015,  0.05449148,  
   0.19866029,  0.24762068,  0.33571949,  0.16216912,  0.29150677,  
   0.13743857,  0.32691385,  0.03630853,  0.16001811,  0.04891474,  
   0.15775034,  0.28860612,  0.09081728,  0.19839663,  0.10794551,  
   0.13360027,  0.01977318,  0.14876907,  0.13201554,  0.23826818,  
   0.16208067,  0.29357259,  0.43413371,  0.02442261,  0.22163881,  
   0.14226426,  0.18468954,  0.18879816,  0.3457911 ,  0.08840592,  
   0.02887249,  0.17454853,  0.17237776,  0.11015596,  0.0794893 ,  
   0.23157808,  0.32383649,  0.06169046,  0.14867717,  0.04954239,  
   0.14995523,  0.24007068,  0.01219199,  0.16835347,  0.06045922,  
   0.14148966,  0.17004406,  0.1418339 ,  0.07587647,  0.05012165,  
   0.34678611,  0.05809208,  0.58511068,  0.11970087,  0.11706498,  
   0.16958423,  0.01063199,  0.39063657,  0.3525142 ,  0.14772956,  
   0.1394325 ,  0.07900901,  0.08205538,  0.21418867,  0.09703341,  
   0.20973871,  0.05918156,  0.16010086,  0.21264622,  0.13349089,  
   0.14176302,  0.06908087,  0.33540582,  0.13756998,  0.06586227,  
   0.04492486,  0.08112604,  0.11063329,  0.27267983,  0.2244248 ,  
   0.2720098 ,  0.13017097,  0.20078604,  0.1029593 ,  0.00869502,  
   0.16956319,  0.06629013,  0.22306265,  0.00870342,  0.07448481,  
   0.15434505,  0.15020701,  0.18824581,  0.18443288])
```

```
In [58]: y_test = y_test.map({'Yes': 1, 'No': 0}).astype(int)
fpr,tpr,thresholds=roc_curve(y_test,probability, pos_label=1)
```

```
In [59]: plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC-AUC CURVE')
plt.show()
```

