S AVINASH GUPTA

21BCE7754

SLOT MORNING AI ML EXTERNSHIP

Assignment : Perform Data-preprocessing for HR-Employee-Attrition.

 # Connecting the drive Throgh the following Syntax

```
from google.colab import drive
drive.mount('/content/drive/')
```

```
Drive already mounted at /content/drive/; to attempt to forcibly
remount, call drive.mount("/content/drive/", force_remount=True).
```

 # Importing Nesscary Libraies

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
```

Specifying the OS path to save the files in the current locations

```
import os
os.chdir('/content/drive/MyDrive/Smart_bridge_AI_ML')
```

 # Reading the Dataset throught the following Syntax

```
dataset=pd.read_csv("/content/drive/MyDrive/Smart_bridge_AI_ML/
Datasets/HR-Employee-Attrition.csv")

dataset
```

```
      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
...   ...       ...                ...         ...
...
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
4                    2          1        Medical              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
4                  7  ...                         4            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
4                    1                  6                      3
...                ...                ...                    ...
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
4                  3               2                  2
...              ...             ...                ...
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
4                           2                     2
...                       ...                   ...
1465                        0                     3
1466                        1                     7
1467                        0                     3
1468                        0                     8
1469                        1                     2

[1470 rows x 35 columns]
```

```
dataset.shape # Specify the shape to find the Number of rows and Cols
```

```
(1470, 35)
```

```
dataset.info() # Here using this we can find that whether it was belong to caterogical or numeric
```

```
<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

dataset.describe() # here we can find deep info regarding dataset mean
median and correlation values.

                Age    DailyRate  DistanceFromHome    Education
EmployeeCount  \
count  1470.000000  1470.000000       1470.000000  1470.000000
1470.0
mean     36.923810   802.485714          9.192517     2.912925
1.0
```

```
std       9.135373   403.509100          8.106864     1.024165
0.0
min      18.000000   102.000000          1.000000     1.000000
1.0
25%      30.000000   465.000000          2.000000     2.000000
1.0
50%      36.000000   802.000000          7.000000     3.000000
1.0
75%      43.000000  1157.000000         14.000000     4.000000
1.0
max      60.000000  1499.000000         29.000000     5.000000
1.0

       EmployeeNumber  EnvironmentSatisfaction   HourlyRate
JobInvolvement  \
count     1470.000000              1470.000000  1470.000000
1470.000000
mean      1024.865306                 2.721769    65.891156
2.729932
std        602.024335                 1.093082    20.329428
0.711561
min          1.000000                 1.000000    30.000000
1.000000
25%        491.250000                 2.000000    48.000000
2.000000
50%       1020.500000                 3.000000    66.000000
3.000000
75%       1555.750000                 4.000000    83.750000
3.000000
max       2068.000000                 4.000000   100.000000
4.000000

          JobLevel  ...  RelationshipSatisfaction  StandardHours  \
count  1470.000000  ...               1470.000000         1470.0
mean      2.063946  ...                  2.712245           80.0
std       1.106940  ...                  1.081209            0.0
min       1.000000  ...                  1.000000           80.0
25%       1.000000  ...                  2.000000           80.0
50%       2.000000  ...                  3.000000           80.0
75%       3.000000  ...                  4.000000           80.0
max       5.000000  ...                  4.000000           80.0

       StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
count       1470.000000        1470.000000            1470.000000
mean           0.793878          11.279592               2.799320
std            0.852077           7.780782               1.289271
min            0.000000           0.000000               0.000000
25%            0.000000           6.000000               2.000000
50%            1.000000          10.000000               3.000000
75%            1.000000          15.000000               3.000000
```

```
max             3.000000          40.000000                6.000000

       WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
count      1470.000000     1470.000000         1470.000000
mean          2.761224        7.008163            4.229252
std           0.706476        6.126525            3.623137
min           1.000000        0.000000            0.000000
25%           2.000000        3.000000            2.000000
50%           3.000000        5.000000            3.000000
75%           3.000000        9.000000            7.000000
max           4.000000       40.000000           18.000000

       YearsSinceLastPromotion  YearsWithCurrManager
count              1470.000000           1470.000000
mean                  2.187755              4.123129
std                   3.222430              3.568136
min                   0.000000              0.000000
25%                   0.000000              2.000000
50%                   1.000000              3.000000
75%                   3.000000              7.000000
max                  15.000000             17.000000

[8 rows x 26 columns]
```

 # Checking NULL VALUES HERE

```
dataset.isnull().any() # using this we can findout whehther we have
null values are not
```

```
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
```

```
dataset.isnull().sum() # if we have null values here we use this
synatx to find out how many are there.
```

```
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
```

No need to Handling Null values Because in the given dataset dont have any null values

```
dataset.corr() # here we find the relation between the variable using this values if it was postive and near to 1 it means highly related to each other and vic versa range from -1 to 1
```

```
<ipython-input-98-d7e5f659ce3d>: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.
  dataset.corr() # here we find the relation between the variable using this values if it was postive and near to 1 it means highly related to each other and vic versa range from -1 to 1
```

	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]
```

```
dataset.corr().Age.sort_values(ascending=False) # making them in
ascending other to understand easy
```

```
<ipython-input-99-214b84a6c8f0>: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.
  dataset.corr().Age.sort_values(ascending=False) # making them in
ascending other to understand easy

Age                         1.000000
TotalWorkingYears           0.680381
JobLevel                    0.509604
MonthlyIncome               0.497855
YearsAtCompany              0.311309
NumCompaniesWorked          0.299635
YearsSinceLastPromotion     0.216513
YearsInCurrentRole          0.212901
Education                   0.208034
YearsWithCurrManager        0.202089
RelationshipSatisfaction    0.053535
StockOptionLevel            0.037510
JobInvolvement              0.029820
MonthlyRate                 0.028051
HourlyRate                  0.024287
DailyRate                   0.010661
EnvironmentSatisfaction     0.010146
PercentSalaryHike           0.003634
PerformanceRating           0.001904
DistanceFromHome           -0.001686
JobSatisfaction            -0.004892
EmployeeNumber             -0.010145
TrainingTimesLastYear      -0.019621
WorkLifeBalance            -0.021490
EmployeeCount                    NaN
StandardHours                    NaN
Name: Age, dtype: float64
```

```
dataset.corr().TotalWorkingYears.sort_values(ascending=False)  # It seems like less Important and and its values also very small.
```

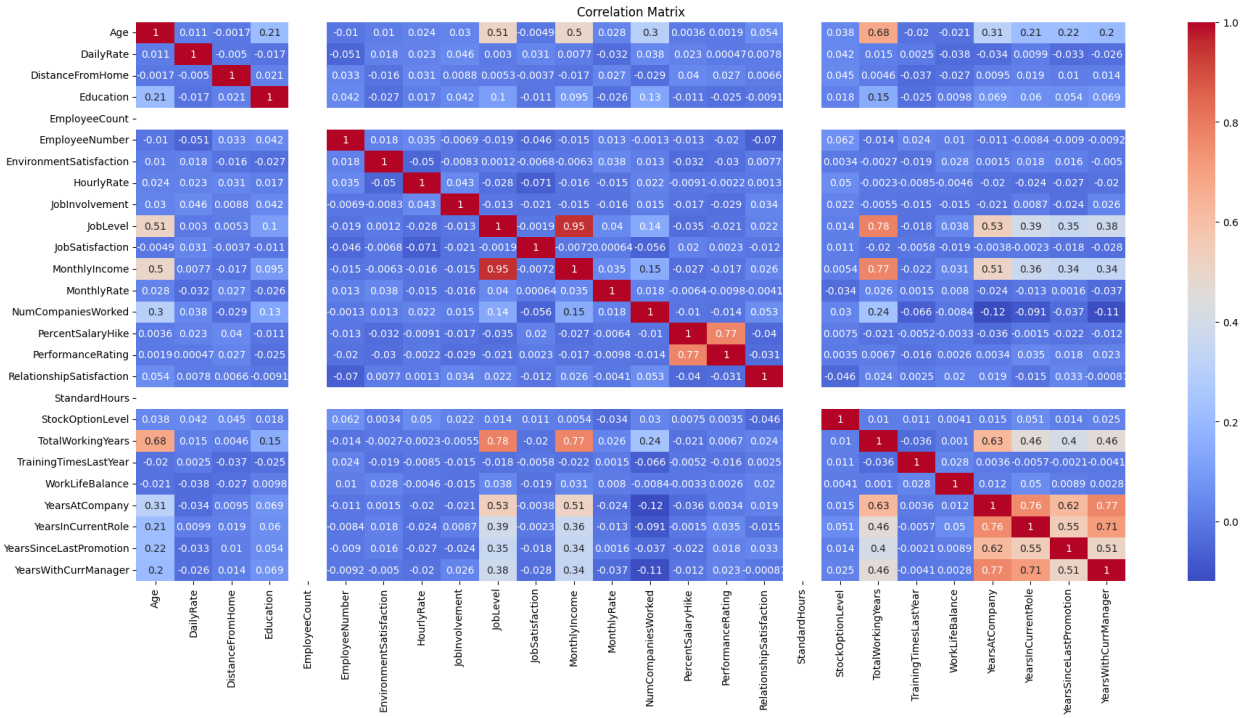
```
<ipython-input-100-9341c25c3ff3>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  dataset.corr().TotalWorkingYears.sort_values(ascending=False)  # It seems like less Important and and its values also very small.
```

```
TotalWorkingYears            1.000000
JobLevel                     0.782208
MonthlyIncome                0.772893
Age                          0.680381
YearsAtCompany               0.628133
YearsInCurrentRole           0.460365
YearsWithCurrManager         0.459188
YearsSinceLastPromotion      0.404858
NumCompaniesWorked           0.237639
Education                    0.148280
MonthlyRate                  0.026442
RelationshipSatisfaction     0.024054
DailyRate                    0.014515
StockOptionLevel             0.010136
PerformanceRating            0.006744
DistanceFromHome             0.004628
WorkLifeBalance              0.001008
HourlyRate                  -0.002334
EnvironmentSatisfaction     -0.002693
JobInvolvement              -0.005533
EmployeeNumber              -0.014365
JobSatisfaction             -0.020185
PercentSalaryHike           -0.020608
TrainingTimesLastYear       -0.035662
EmployeeCount                     NaN
StandardHours                     NaN
Name: TotalWorkingYears, dtype: float64
```

```
corr_matrix = dataset.corr()

# Create a larger figure
plt.figure(figsize=(22, 10))  # Adjust the width and height as needed

# Create the heatmap with annotations
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')

# Add a title
plt.title('Correlation Matrix')
```

```
# Show the plot
plt.show()
```

```
<ipython-input-101-f358d6eaa217>:1: FutureWarning: The default value
of numeric_only in DataFrame.corr is deprecated. In a future version,
it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  corr_matrix = dataset.corr()
```



Dropping the cloumns which have less threshold values.

```
import pandas as pd

correlation_matrix = dataset.corr()

# Set the correlation threshold value
threshold = 0.01 # using abs function we can find which are like
greater than 0.01 value

# Find columns with correlations greater than or less than the
threshold
columns_to_keep = [col for col in correlation_matrix.columns if
                   any(abs(correlation_matrix[col]) > threshold)]
```

```
# Create a new DataFrame with the selected columns
dataset_updated = dataset[columns_to_keep]

dataset.head()
```

```
   Age  DailyRate  DistanceFromHome  Education  EmployeeNumber  \
0   41       1102                 1          2               1
1   49        279                 8          1               2
2   37       1373                 2          2               4
3   33       1392                 3          4               5
4   27        591                 2          1               7

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  PerformanceRating
RelationshipSatisfaction  \
0                4  ...                  3                          1

1                2  ...                  4                          4

2                3  ...                  3                          2

3                3  ...                  3                          3

4                2  ...                  3                          4


   StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                 0                  8                      0
1                 1                 10                      3
2                 0                  7                      3
3                 0                  8                      3
4                 1                  6                      3

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

   YearsSinceLastPromotion  YearsWithCurrManager
0                        0                     5
1                        1                     7
2                        0                     0
3                        3                     0
```

```
4                         2                       2

[5 rows x 24 columns]
```

```
import pandas as pd
# HERE WE ARE ADDING IT BECAUSE OF IN ABOVE WE CHECK THE COLS OF INT64
and we update the dataset which having only int64 values so we have to
add object col here.
# Add the 'Attrition' column back to 'dataset_updated' which is key
interest to the dataset.
dataset_updated['Attrition'] = dataset['Attrition']

# Now, 'dataset_updated' contains the 'Attrition' column

dataset_updated.head()
```

```
   Age  DailyRate  DistanceFromHome  Education  EmployeeNumber  \
0   41       1102                 1          2               1
1   49        279                 8          1               2
2   37       1373                 2          2               4
3   33       1392                 3          4               5
4   27        591                 2          1               7

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  RelationshipSatisfaction  StockOptionLevel  \
0                4  ...                         1                 0
1                2  ...                         4                 1
2                3  ...                         2                 0
3                3  ...                         3                 0
4                2  ...                         4                 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


   Attrition
0        Yes
1         No
2        Yes
3         No
4         No

[5 rows x 25 columns]

dataset_updated.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 25 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   DailyRate                 1470 non-null   int64
 2   DistanceFromHome          1470 non-null   int64
 3   Education                 1470 non-null   int64
 4   EmployeeNumber            1470 non-null   int64
 5   EnvironmentSatisfaction   1470 non-null   int64
 6   HourlyRate                1470 non-null   int64
 7   JobInvolvement            1470 non-null   int64
 8   JobLevel                  1470 non-null   int64
 9   JobSatisfaction           1470 non-null   int64
 10  MonthlyIncome             1470 non-null   int64
 11  MonthlyRate               1470 non-null   int64
 12  NumCompaniesWorked        1470 non-null   int64
 13  PercentSalaryHike         1470 non-null   int64
 14  PerformanceRating         1470 non-null   int64
 15  RelationshipSatisfaction  1470 non-null   int64
 16  StockOptionLevel          1470 non-null   int64
 17  TotalWorkingYears         1470 non-null   int64
 18  TrainingTimesLastYear     1470 non-null   int64
 19  WorkLifeBalance           1470 non-null   int64
 20  YearsAtCompany            1470 non-null   int64
 21  YearsInCurrentRole        1470 non-null   int64
```

```
 22  YearsSinceLastPromotion   1470 non-null   int64
 23  YearsWithCurrManager      1470 non-null   int64
 24  Attrition                 1470 non-null   object
dtypes: int64(24), object(1)
memory usage: 287.2+ KB
```

```
dataset_updated.shape
```

```
(1470, 25)
```

```
dataset_updated.corr()
```

```
<ipython-input-120-b5164f5c9c8d>: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.
  dataset_updated.corr()
```

	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
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
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

                          EmployeeNumber  EnvironmentSatisfaction
HourlyRate  \
Age                            -0.010145                 0.010146
0.024287
DailyRate                      -0.050990                 0.018355
0.023381
DistanceFromHome                0.032916                -0.016075
0.031131
Education                       0.042070                -0.027128
0.016775
EmployeeNumber                  1.000000                 0.017621
0.035179
EnvironmentSatisfaction         0.017621                 1.000000   -
0.049857
HourlyRate                      0.035179                -0.049857
1.000000
JobInvolvement                 -0.006888                -0.008278
0.042861
JobLevel                       -0.018519                 0.001212   -
0.027853
JobSatisfaction                -0.046247                -0.006784   -
0.071335
MonthlyIncome                  -0.014829                -0.006259   -
0.015794
MonthlyRate                     0.012648                 0.037600   -
0.015297
NumCompaniesWorked             -0.001251                 0.012594
0.022157
PercentSalaryHike              -0.012944                -0.031701   -
0.009062
```

```
PerformanceRating               -0.020359                -0.029548   -
0.002172
RelationshipSatisfaction        -0.069861                 0.007665
0.001330
StockOptionLevel                 0.062227                 0.003432
0.050263
TotalWorkingYears               -0.014365                -0.002693   -
0.002334
TrainingTimesLastYear            0.023603                -0.019359   -
0.008548
WorkLifeBalance                  0.010309                 0.027627   -
0.004607
YearsAtCompany                  -0.011240                 0.001458   -
0.019582
YearsInCurrentRole              -0.008416                 0.018007   -
0.024106
YearsSinceLastPromotion         -0.009019                 0.016194   -
0.026716
YearsWithCurrManager            -0.009197                -0.004999   -
0.020123

                           JobInvolvement  JobLevel
JobSatisfaction  ...  \
Age                              0.029820  0.509604         -
0.004892  ...
DailyRate                        0.046135  0.002966
0.030571  ...
DistanceFromHome                 0.008783  0.005303         -
0.003669  ...
Education                        0.042438  0.101589         -
0.011296  ...
EmployeeNumber                  -0.006888 -0.018519         -
0.046247  ...
EnvironmentSatisfaction         -0.008278  0.001212         -
0.006784  ...
HourlyRate                       0.042861 -0.027853         -
0.071335  ...
JobInvolvement                   1.000000 -0.012630         -
0.021476  ...
JobLevel                        -0.012630  1.000000         -
0.001944  ...
JobSatisfaction                 -0.021476 -0.001944
1.000000  ...
MonthlyIncome                   -0.015271  0.950300         -
0.007157  ...
MonthlyRate                     -0.016322  0.039563
0.000644  ...
NumCompaniesWorked               0.015012  0.142501         -
0.055699  ...
```

```
PercentSalaryHike               -0.017205 -0.034730
0.020002  ...
PerformanceRating               -0.029071 -0.021222
0.002297  ...
RelationshipSatisfaction         0.034297  0.021642        -
0.012454  ...
StockOptionLevel                 0.021523  0.013984
0.010690  ...
TotalWorkingYears               -0.005533  0.782208        -
0.020185  ...
TrainingTimesLastYear           -0.015338 -0.018191        -
0.005779  ...
WorkLifeBalance                 -0.014617  0.037818        -
0.019459  ...
YearsAtCompany                  -0.021355  0.534739        -
0.003803  ...
YearsInCurrentRole               0.008717  0.389447        -
0.002305  ...
YearsSinceLastPromotion         -0.024184  0.353885        -
0.018214  ...
YearsWithCurrManager             0.025976  0.375281        -
0.027656  ...

                          PerformanceRating  RelationshipSatisfaction
\
Age                                0.001904                  0.053535

DailyRate                          0.000473                  0.007846

DistanceFromHome                   0.027110                  0.006557

Education                         -0.024539                 -0.009118

EmployeeNumber                    -0.020359                 -0.069861

EnvironmentSatisfaction           -0.029548                  0.007665

HourlyRate                        -0.002172                  0.001330

JobInvolvement                    -0.029071                  0.034297

JobLevel                          -0.021222                  0.021642

JobSatisfaction                    0.002297                 -0.012454

MonthlyIncome                     -0.017120                  0.025873

MonthlyRate                       -0.009811                 -0.004085

NumCompaniesWorked                -0.014095                  0.052733
```

PercentSalaryHike	0.773550	-0.040490
PerformanceRating	1.000000	-0.031351
RelationshipSatisfaction	-0.031351	1.000000
StockOptionLevel	0.003506	-0.045952
TotalWorkingYears	0.006744	0.024054
TrainingTimesLastYear	-0.015579	0.002497
WorkLifeBalance	0.002572	0.019604
YearsAtCompany	0.003435	0.019367
YearsInCurrentRole	0.034986	-0.015123
YearsSinceLastPromotion	0.017896	0.033493
YearsWithCurrManager	0.022827	-0.000867

	StockOptionLevel	TotalWorkingYears	\
Age	0.037510	0.680381	
DailyRate	0.042143	0.014515	
DistanceFromHome	0.044872	0.004628	
Education	0.018422	0.148280	
EmployeeNumber	0.062227	-0.014365	
EnvironmentSatisfaction	0.003432	-0.002693	
HourlyRate	0.050263	-0.002334	
JobInvolvement	0.021523	-0.005533	
JobLevel	0.013984	0.782208	
JobSatisfaction	0.010690	-0.020185	
MonthlyIncome	0.005408	0.772893	
MonthlyRate	-0.034323	0.026442	
NumCompaniesWorked	0.030075	0.237639	
PercentSalaryHike	0.007528	-0.020608	
PerformanceRating	0.003506	0.006744	
RelationshipSatisfaction	-0.045952	0.024054	
StockOptionLevel	1.000000	0.010136	
TotalWorkingYears	0.010136	1.000000	
TrainingTimesLastYear	0.011274	-0.035662	
WorkLifeBalance	0.004129	0.001008	
YearsAtCompany	0.015058	0.628133	
YearsInCurrentRole	0.050818	0.460365	
YearsSinceLastPromotion	0.014352	0.404858	
YearsWithCurrManager	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
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
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	
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	
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	

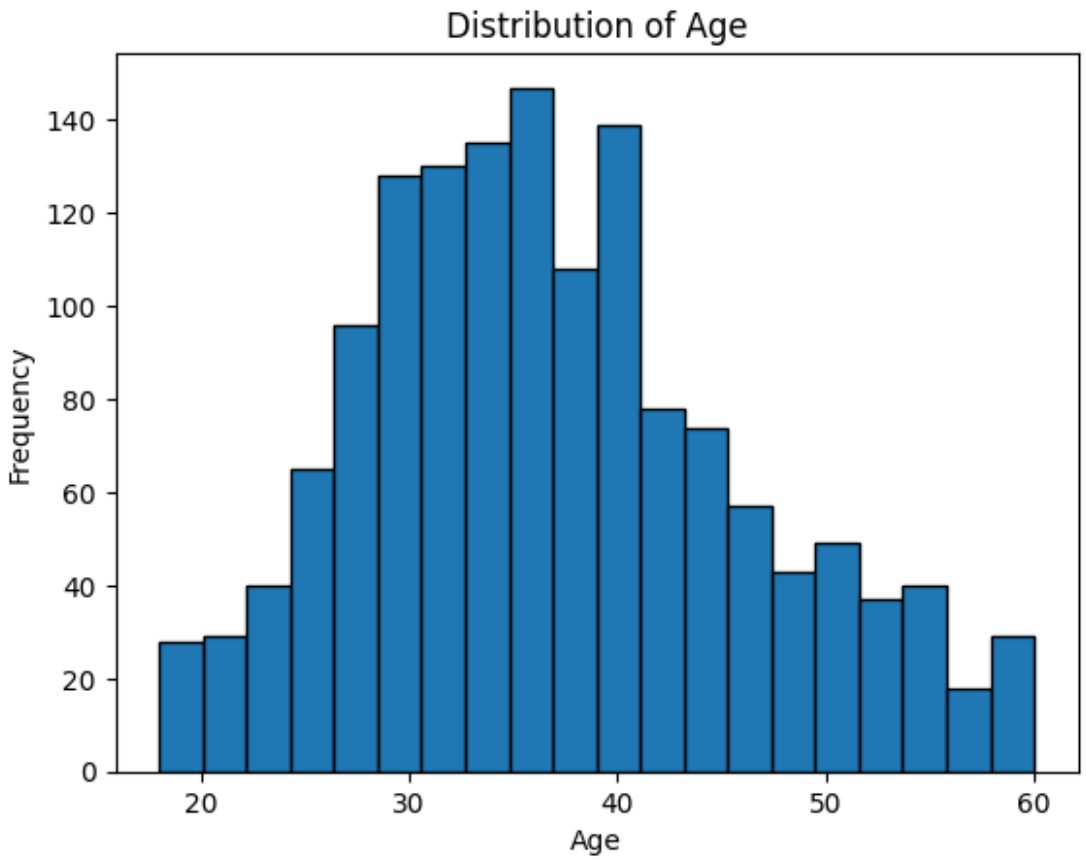
```
YearsWithCurrManager             0.769212             0.714365

                          YearsSinceLastPromotion
YearsWithCurrManager
Age                                   0.216513
0.202089
DailyRate                            -0.033229             -
0.026363
DistanceFromHome                      0.010029
0.014406
Education                             0.054254
0.069065
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
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

[24 rows x 24 columns]

dataset_updated.head()

   Age  DailyRate  DistanceFromHome  Education  EmployeeNumber  \
0   41       1102                 1          2               1
1   49        279                 8          1               2
2   37       1373                 2          2               4
3   33       1392                 3          4               5
4   27        591                 2          1               7

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                        2          94               3         2
1                        3          61               2         2
2                        4          92               2         1
3                        4          56               3         1
4                        1          40               3         1

   JobSatisfaction  ...  RelationshipSatisfaction  StockOptionLevel  \
0                4  ...                         1                 0
1                2  ...                         4                 1
2                3  ...                         2                 0
3                3  ...                         3                 0
4                2  ...                         4                 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
```

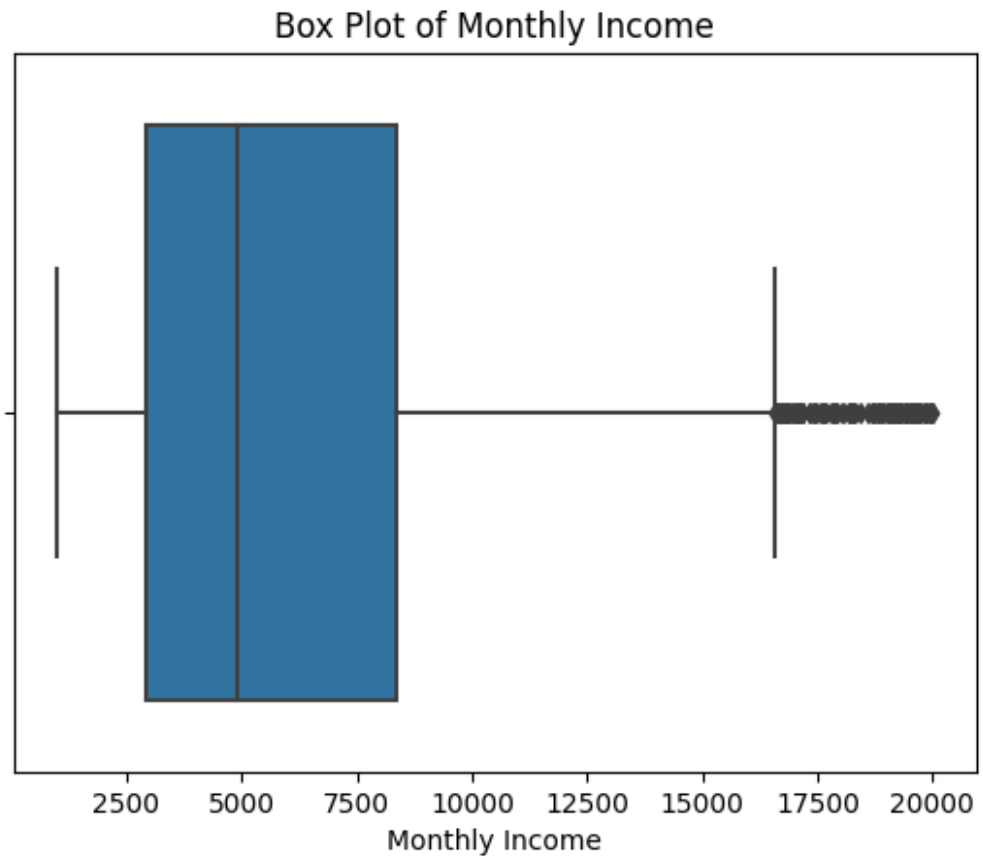
```
4                    2                              2                        2


   Attrition
0        Yes
1         No
2        Yes
3         No
4         No

[5 rows x 25 columns]
```

Data Visualization

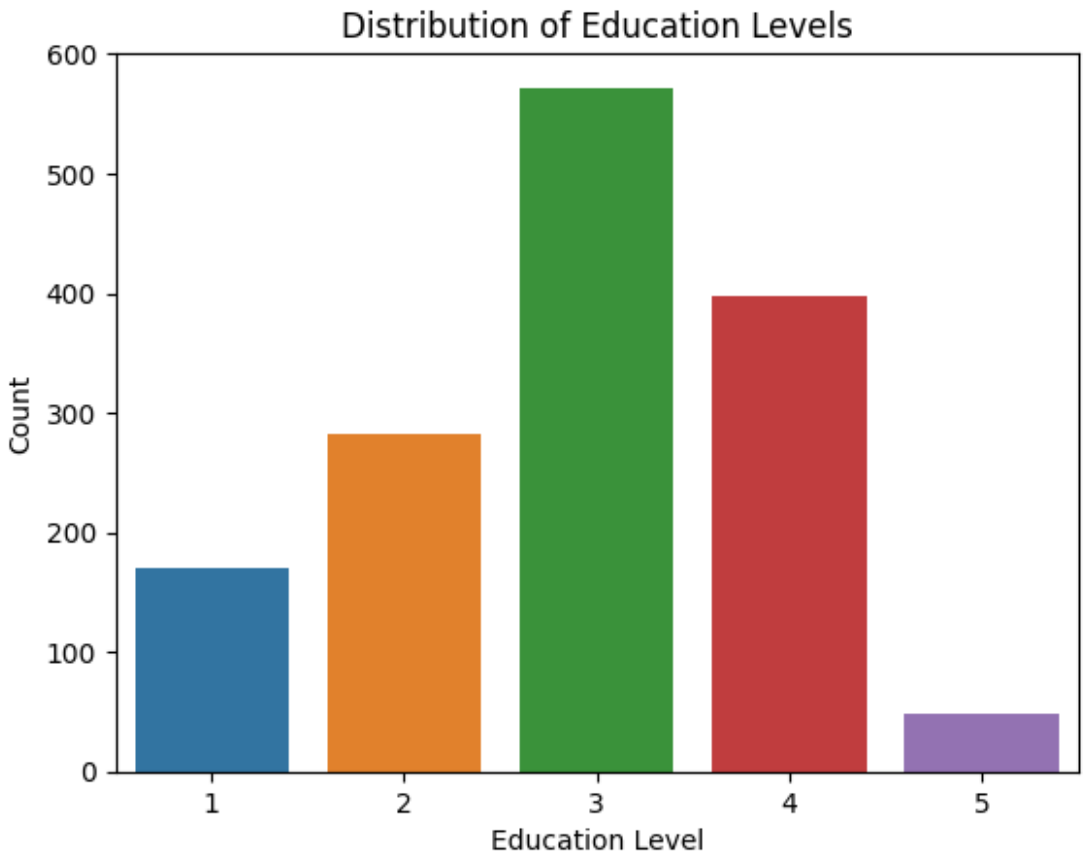
```
import matplotlib.pyplot as plt

# Example: Histogram of Age
plt.hist(dataset_updated['Age'], bins=20, edgecolor='k')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Distribution of Age')
plt.show()
```



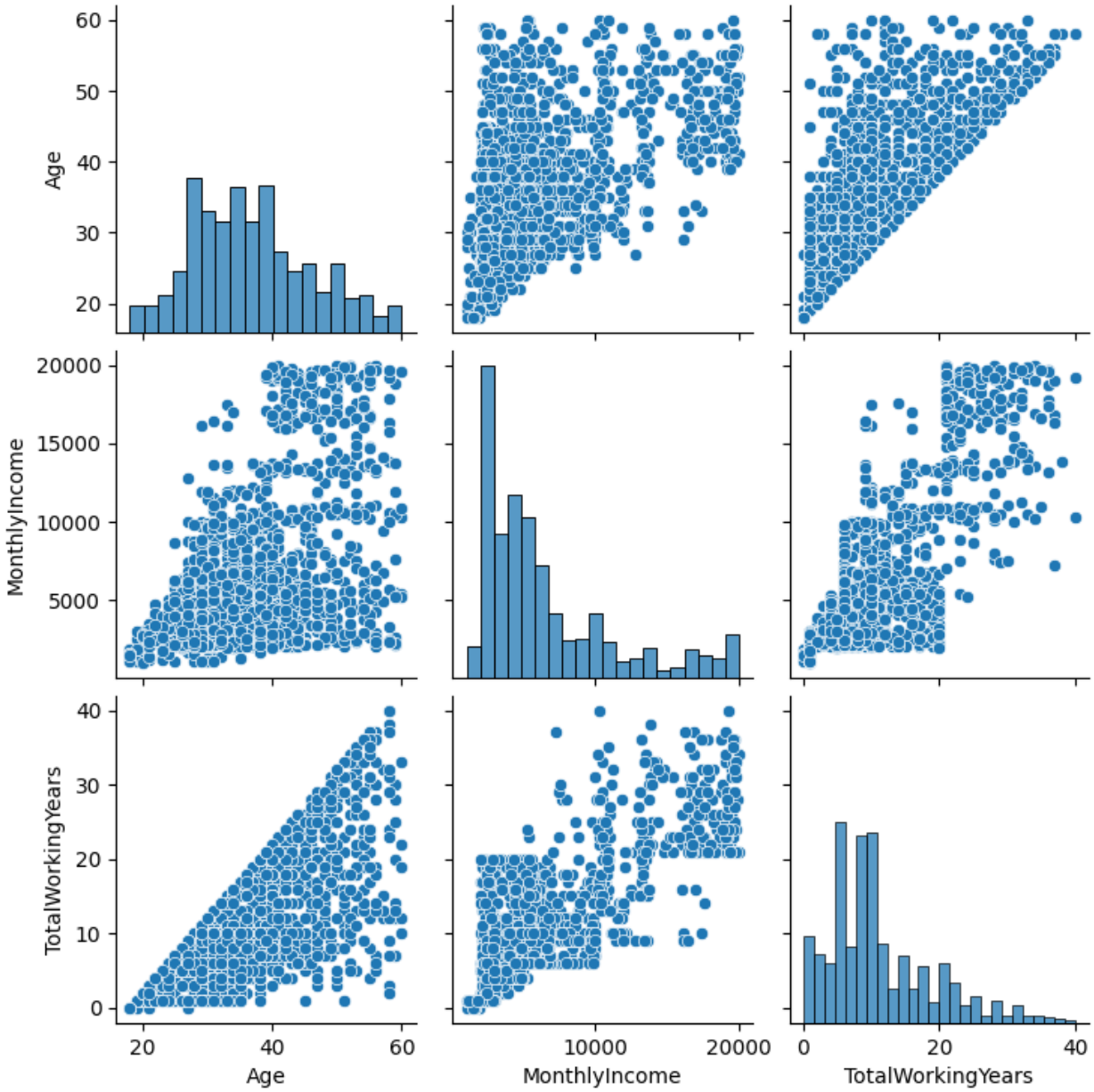

```
import seaborn as sns
# here we detected the oulier so we have to handle the outliers.
# Example: Box plot of MonthlyIncome
sns.boxplot(x='MonthlyIncome', data=dataset_updated)
plt.xlabel('Monthly Income')
plt.title('Box Plot of Monthly Income')
plt.show()
```



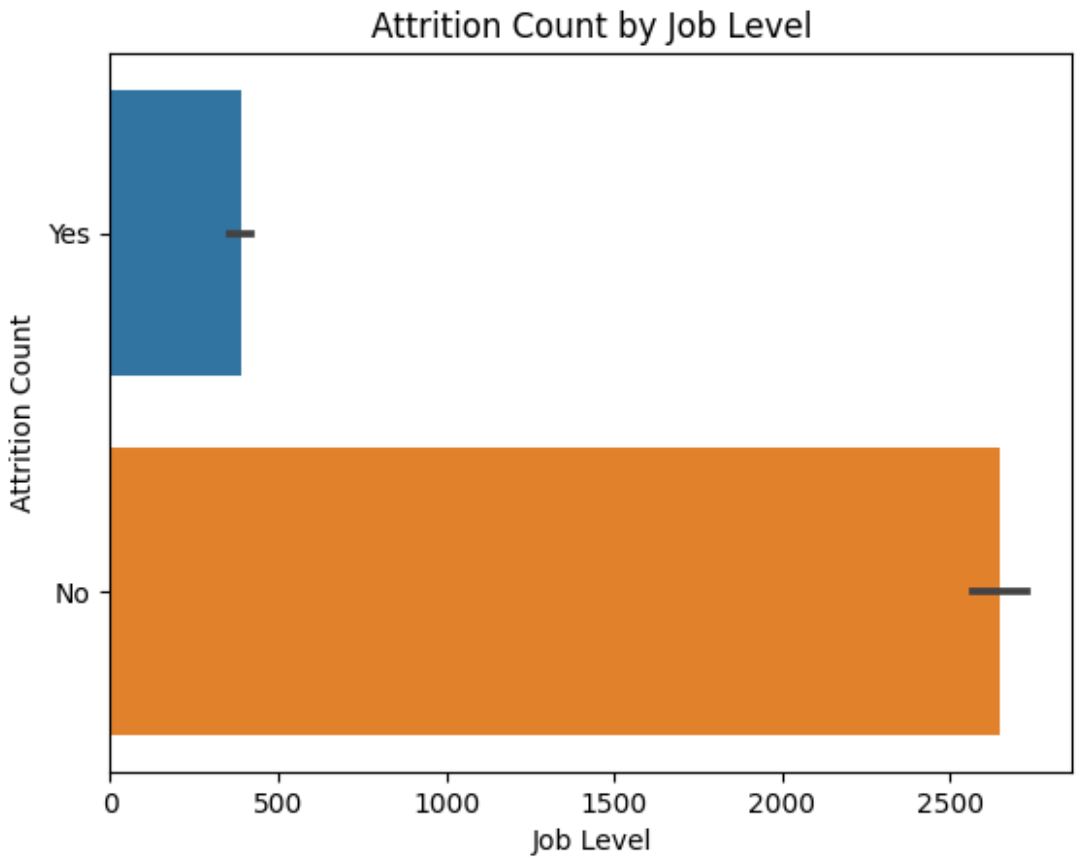

```
# Example: Count plot of Education
sns.countplot(x='Education', data=dataset_updated)
plt.xlabel('Education Level')
plt.ylabel('Count')
plt.title('Distribution of Education Levels')
plt.show()
```




```
# Example: Pair plot of a subset of numerical variables
sns.pairplot(dataset_updated[['Age', 'MonthlyIncome',
'TotalWorkingYears']])
plt.show()
```

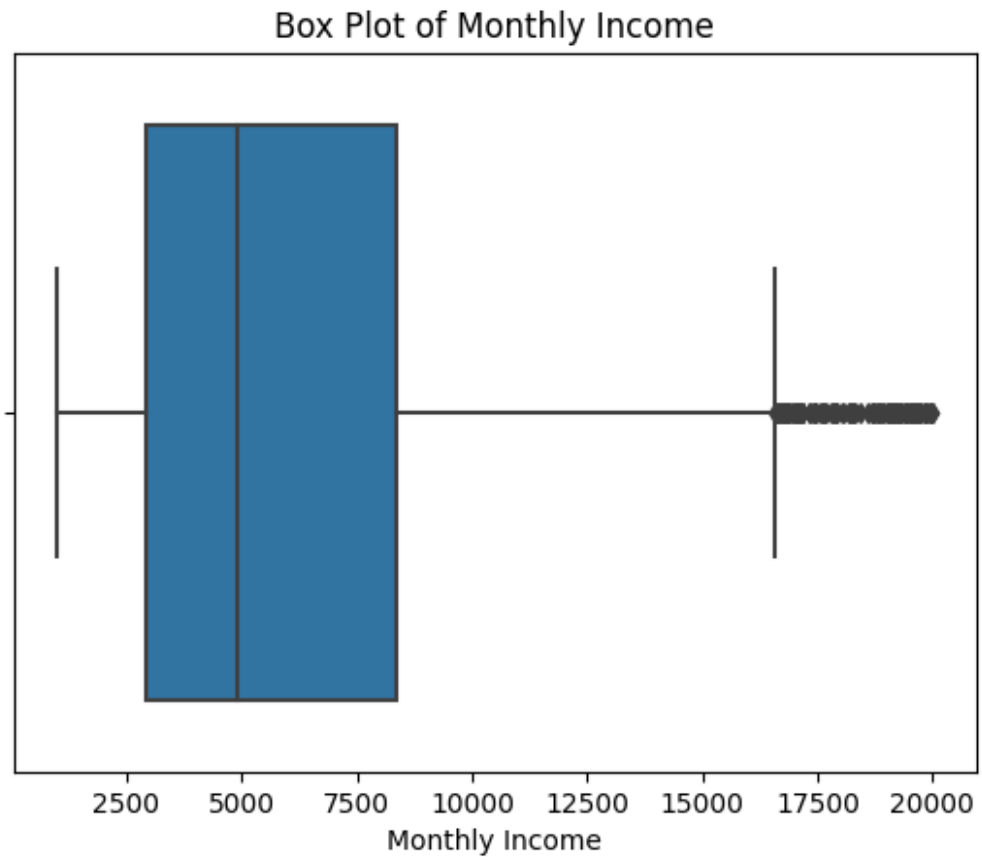



```
import seaborn as sns
import matplotlib.pyplot as plt

# Example: Bar plot of Attrition by Job Level
sns.barplot(x='JobLevel', y='Attrition', data=dataset_updated,
estimator=sum)
plt.xlabel('Job Level')
plt.ylabel('Attrition Count')
plt.title('Attrition Count by Job Level')
plt.show()
```



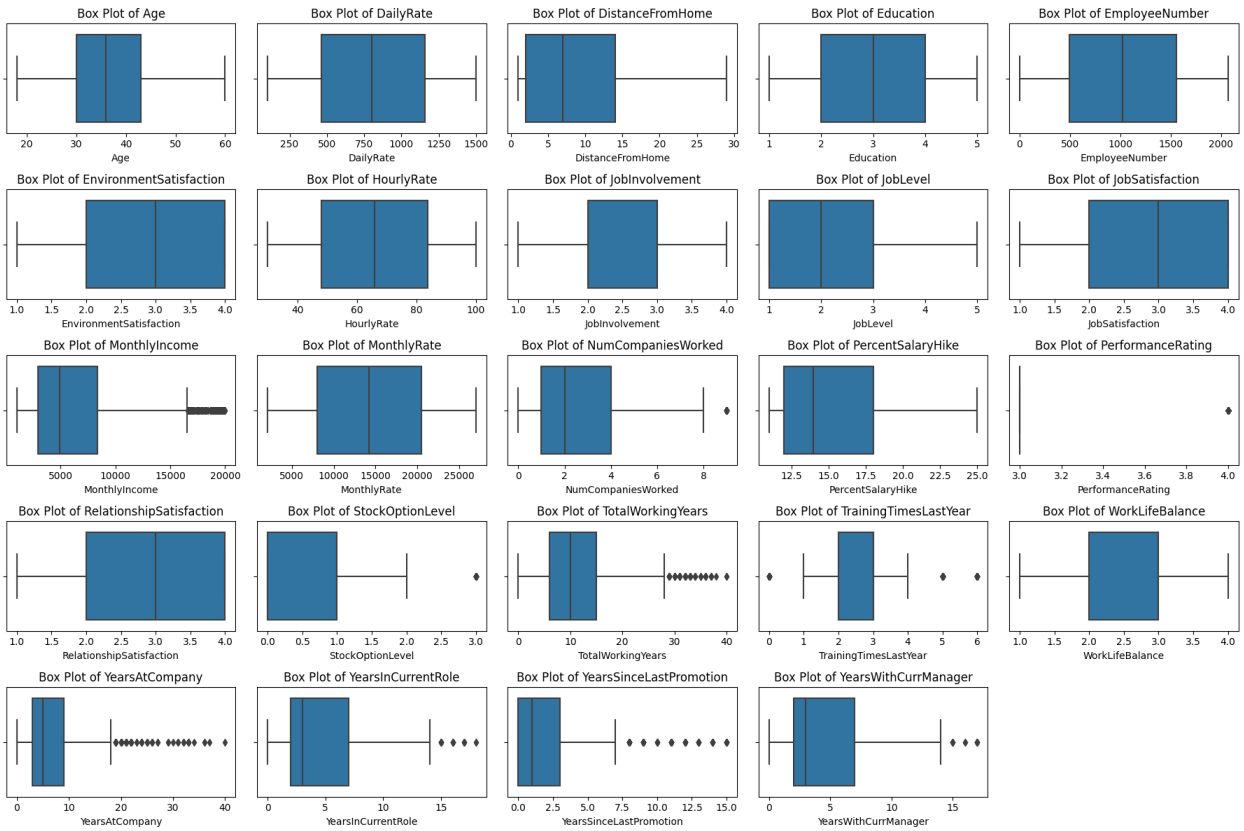
Outlier Detections

```
import seaborn as sns
# here we detected the oulier so we hav eto handle the outliers.
# Example: Box plot of MonthlyIncome
sns.boxplot(x='MonthlyIncome', data=dataset_updated)
plt.xlabel('Monthly Income')
plt.title('Box Plot of Monthly Income')
plt.show()
```




```
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming 'dataset_updated' is your DataFrame

# List of all numerical columns
numerical_columns = ['Age', 'DailyRate', 'DistanceFromHome',
'Education', 'EmployeeNumber',
                     'EnvironmentSatisfaction', 'HourlyRate',
'JobInvolvement', 'JobLevel',
                     'JobSatisfaction', 'MonthlyIncome',
'MonthlyRate', 'NumCompaniesWorked',
                     'PercentSalaryHike', 'PerformanceRating',
'RelationshipSatisfaction',
                     'StockOptionLevel', 'TotalWorkingYears',
'TrainingTimesLastYear',
                     'WorkLifeBalance', 'YearsAtCompany',
'YearsInCurrentRole',
                     'YearsSinceLastPromotion',
'YearsWithCurrManager']

# Create a subplot grid for box plots
plt.figure(figsize=(18, 12))
```

```
for i, column in enumerate(numerical_columns, 1):
    plt.subplot(5, 5, i)  # 5 rows, 5 columns grid (adjust as needed)
    sns.boxplot(x=dataset_updated[column])
    plt.xlabel(column)
    plt.title(f'Box Plot of {column}')

# Adjust subplot layout
plt.tight_layout()

# Show the box plots
plt.show()
```




```
# we observe the outliers in YearAtCompany,YearsInCurrentRole,YearsSinceLasrPromotion,YearsWithCurrManager,TrainingTimesLastYear,TotalWorkingYears,StockOptionLevel,MonthlyIncome,NumcompaniesWorked...etc...
```

HANDLING OUTLIERS

```
import pandas as pd
from scipy import stats

# Define a threshold for identifying outliers (e.g., Z-score threshold)
```

```
z_score_threshold = 3

# Create a copy of the dataset to preserve the original data
dataset_no_outliers = dataset_updated.copy()

# Iterate through numerical columns and remove outliers
for column in dataset_updated.select_dtypes(include=['int64']):
    # Calculate Z-scores for the column
    z_scores = stats.zscore(dataset_updated[column])

    # Find data points with Z-scores greater than the threshold
    outliers = dataset_updated[column][abs(z_scores) >
z_score_threshold]

    # Remove outliers from the dataset
    dataset_no_outliers =
dataset_no_outliers[~dataset_no_outliers[column].isin(outliers)]

# The 'dataset_no_outliers' DataFrame now contains the dataset with
outliers removed for all numerical columns.

dataset_updated.Age.shape
```

```
(1470,)
```

```
dataset_no_outliers.Age.shape  # we reduced some rows which are
detects as outliers
```

```
(1387,)
```

Splitting Dataset Like Dependent and Independent variables

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder

# Create a copy of the original dataset
data = dataset_updated.copy()

# Separate the dependent variable (target) from the independent
variables (features)
X = data.drop("Attrition", axis=1)  # Independent variables (features)
y = data["Attrition"]  # Dependent variable (target)
```

#Perform Encoding to change the catergorical values to Numerical vaules

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

```
# Use LabelEncoder to convert "Attrition" to numerical values (0 and 1)
le = LabelEncoder()
y = le.fit_transform(y)

y
```

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

```
X
```

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

      EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                           2          94               3         2
1                           3          61               2         2
2                           4          92               2         1
3                           4          56               3         1
4                           1          40               3         1
...                       ...         ...             ...       ...
1465                        3          41               4         2
1466                        4          42               2         3
1467                        2          87               4         2
1468                        4          63               2         2
1469                        2          82               4         2

      JobSatisfaction  ...  PerformanceRating
RelationshipSatisfaction  \
0                   4  ...                  3
1
1                   2  ...                  4
4
2                   3  ...                  3
2
3                   3  ...                  3
3
4                   2  ...                  3
4
...               ...  ...                ...                        ..
```

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

```
      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                    0                  8                      0
1                    1                 10                      3
2                    0                  7                      3
3                    0                  8                      3
4                    1                  6                      3
...                ...                ...                    ...
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
4                   3               2                   2
...               ...             ...                 ...
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
4                           2                     2
...                       ...                   ...
1465                        0                     3
1466                        1                     7
1467                        0                     3
1468                        0                     8
1469                        1                     2
```

```
[1470 rows x 24 columns]

Warning: Total number of columns (24) exceeds max_columns (20)
limiting to first (20) columns.
```

Performing the Feature Scaling here where to make them equal measure while calcuting

```
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()

X_Scaled=pd.DataFrame(ms.fit_transform(X),columns=X.columns)

X_Scaled
```

```
           Age  DailyRate  DistanceFromHome  Education  EmployeeNumber
\
0     0.547619   0.715820          0.000000       0.25        0.000000

1     0.738095   0.126700          0.250000       0.00        0.000484

2     0.452381   0.909807          0.035714       0.25        0.001451

3     0.357143   0.923407          0.071429       0.75        0.001935

4     0.214286   0.350036          0.035714       0.00        0.002903

...        ...        ...               ...        ...             ...

1465  0.428571   0.559771          0.785714       0.25        0.996613

1466  0.500000   0.365784          0.178571       0.00        0.997097

1467  0.214286   0.037938          0.107143       0.50        0.998065

1468  0.738095   0.659270          0.035714       0.50        0.998549

1469  0.380952   0.376521          0.250000       0.50        1.000000


      EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                    0.333333    0.914286        0.666667      0.25
1                    0.666667    0.442857        0.333333      0.25
2                    1.000000    0.885714        0.333333      0.00
3                    1.000000    0.371429        0.666667      0.00
4                    0.000000    0.142857        0.666667      0.00
...                       ...         ...             ...       ...
1465                 0.666667    0.157143        1.000000      0.25
```

```
1466                 1.000000    0.171429        0.333333      0.50
1467                 0.333333    0.814286        1.000000      0.25
1468                 1.000000    0.471429        0.333333      0.25
1469                 0.333333    0.742857        1.000000      0.25

      JobSatisfaction  ...  PerformanceRating
RelationshipSatisfaction  \
0            1.000000  ...                0.0
0.000000
1            0.333333  ...                1.0
1.000000
2            0.666667  ...                0.0
0.333333
3            0.666667  ...                0.0
0.666667
4            0.333333  ...                0.0
1.000000
...               ...  ...                ...                          ..
.
1465         1.000000  ...                0.0
0.666667
1466         0.000000  ...                0.0
0.000000
1467         0.333333  ...                1.0
0.333333
1468         0.333333  ...                0.0
1.000000
1469         0.666667  ...                0.0
0.000000

      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             0.333333              0.150               0.500000
...                ...                ...                    ...
1465          0.333333              0.425               0.500000
1466          0.333333              0.225               0.833333
1467          0.333333              0.150               0.000000
1468          0.000000              0.425               0.500000
1469          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             0.666667           0.050            0.111111
...                ...             ...                 ...
```

```
1465         0.666667           0.125            0.111111
1466         0.666667           0.175            0.388889
1467         0.666667           0.150            0.111111
1468         0.333333           0.225            0.333333
1469         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                    0.133333              0.117647
...                       ...                   ...
1465                 0.000000              0.176471
1466                 0.066667              0.411765
1467                 0.000000              0.176471
1468                 0.000000              0.470588
1469                 0.066667              0.117647

[1470 rows x 24 columns]
```

Spliting Dataset into Train and Test for futher evalution

```
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)

print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)

(1176, 24) (294, 24) (1176,) (294,)

x_train.head()

           Age  DailyRate  DistanceFromHome  Education  EmployeeNumber
\
1374  0.952381   0.360057          0.714286       0.50        0.937107

1092  0.642857   0.607015          0.964286       0.50        0.747460

768   0.523810   0.141732          0.892857       0.50        0.515239

569   0.428571   0.953472          0.250000       0.75        0.381229

911   0.166667   0.355762          0.821429       0.00        0.615385


      EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
```

```
1374                 1.000000    0.600000        0.666667      0.75
1092                 1.000000    0.957143        0.666667      0.00
768                  0.666667    0.628571        0.666667      0.25
569                  0.000000    0.657143        0.333333      0.50
911                  0.666667    0.614286        0.000000      0.00

      JobSatisfaction  ...  PerformanceRating
RelationshipSatisfaction  \
1374              1.0  ...                0.0
0.666667
1092              1.0  ...                1.0
1.000000
768               0.0  ...                0.0
0.333333
569               0.0  ...                0.0
0.333333
911               1.0  ...                0.0
1.000000

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
1374          0.333333              0.725               0.333333
1092          0.333333              0.200               0.500000
768           0.333333              0.200               0.500000
569           0.000000              0.250               0.166667
911           0.000000              0.025               0.666667

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
1374         0.333333           0.025            0.000000
1092         0.666667           0.125            0.222222
768          0.333333           0.175            0.388889
569          0.666667           0.250            0.388889
911          0.666667           0.025            0.000000

      YearsSinceLastPromotion  YearsWithCurrManager
1374                 0.000000              0.000000
1092                 0.000000              0.176471
768                  0.466667              0.294118
569                  0.000000              0.529412
911                  0.066667              0.000000

[5 rows x 24 columns]
```

```
# Decision Tree

from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier()

model.fit(x_train,y_train)

DecisionTreeClassifier()
```

```
pred=model.predict(x_test)

pred

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

y_test

array([0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0,
       0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0,
       1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
1,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 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, 1, 0, 0, 1,
0,
```

```
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
0,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 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, 1,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
0,
       0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0,
       0, 1, 0, 0, 0, 1, 0, 0])

data

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

      EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  \
0                           2          94               3         2
1                           3          61               2         2
2                           4          92               2         1
3                           4          56               3         1
4                           1          40               3         1
...                       ...         ...             ...       ...
1465                        3          41               4         2
1466                        4          42               2         3
1467                        2          87               4         2
1468                        4          63               2         2
1469                        2          82               4         2

      JobSatisfaction  ...  RelationshipSatisfaction  StockOptionLevel
\
0                   4  ...                         1                 0

1                   2  ...                         4                 1

2                   3  ...                         2                 0

3                   3  ...                         3                 0
```

```
4                    2  ...                         4                  1

...                ...  ...                       ...                ...

1465                 4  ...                         3                  1

1466                 1  ...                         1                  1

1467                 2  ...                         2                  1

1468                 2  ...                         4                  0

1469                 3  ...                         1                  0


      TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  \
0                     8                      0                1
1                    10                      3                3
2                     7                      3                3
3                     8                      3                3
4                     6                      3                3
...                 ...                    ...              ...
1465                 17                      3                3
1466                  9                      5                3
1467                  6                      0                3
1468                 17                      3                2
1469                  6                      3                4

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

      YearsWithCurrManager  Attrition
0                        5        Yes
1                        7         No
2                        0        Yes
3                        0         No
4                        2         No
...                    ...        ...
1465                     3         No
1466                     7         No
```

```
1467                      3         No
1468                      8         No
1469                      2         No

[1470 rows x 25 columns]

Warning: Total number of columns (25) exceeds max_columns (20)
limiting to first (20) columns.
```

```
model.predict(ms.transform([[41,1102, 1,   2,   1,   2,   94,  3,
     2,   4,   5993,19479,     8,2,3,4,5,11,   3,   1,   0,   8
     ,0,  1]]))  # we are using ms to transform the input to scaled
values.
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but MinMaxScaler was
fitted with feature names
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but
DecisionTreeClassifier was fitted with feature names
  warnings.warn(

array([1])
```

```
from sklearn.metrics import
accuracy_score,confusion_matrix,classification_report,roc_auc_score,ro
c_curve

accuracy_score(y_test,pred)
```

```
0.7346938775510204
```

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

```
array([[203,  42],
       [ 36,  13]])
```

```
pd.crosstab(y_test,pred)
```

```
col_0    0   1
row_0
0      203  42
1       36  13
```

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

```
              precision    recall  f1-score   support

           0       0.85      0.83      0.84       245
           1       0.24      0.27      0.25        49
```

```
    accuracy                           0.73       294
   macro avg       0.54      0.55      0.54       294
weighted avg       0.75      0.73      0.74       294
```

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

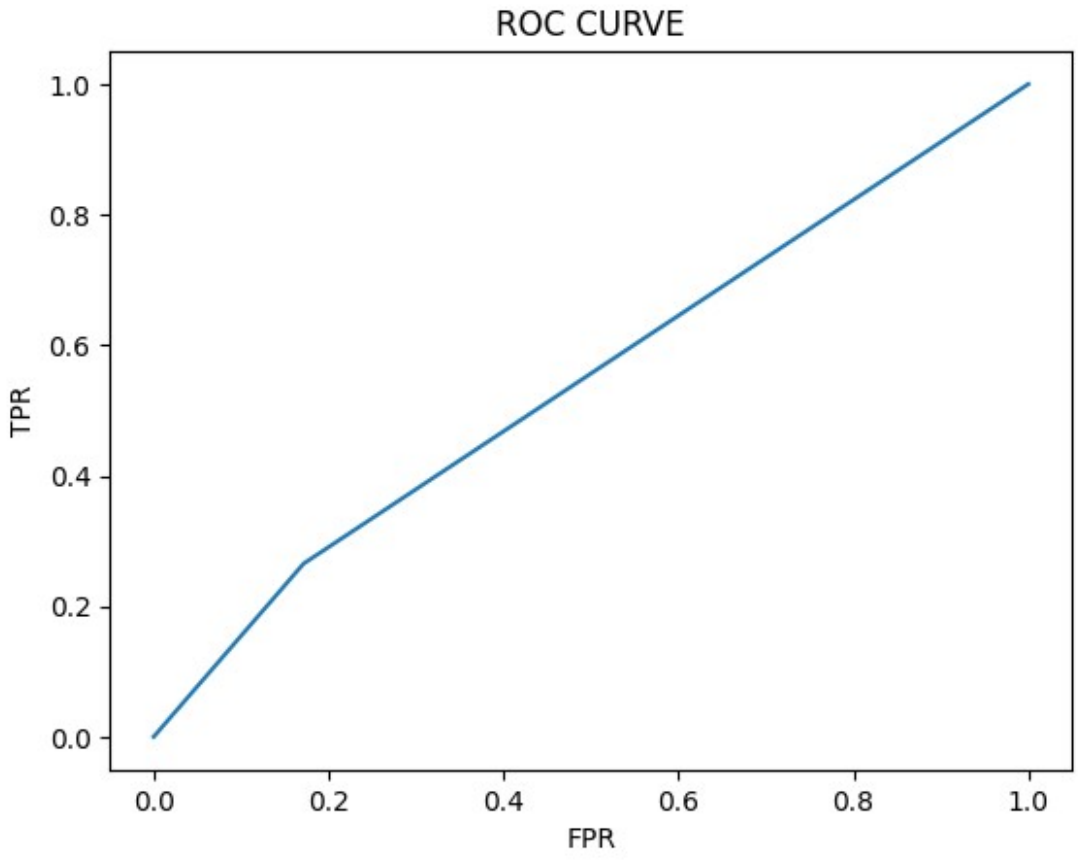
```
Probability
```

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

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

```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
```

```
plt.title('ROC CURVE')
plt.show()
```




```
from sklearn import tree
plt.figure(figsize=(25,15))
tree.plot_tree(model,filled=True)
```

```
[Text(0.28559802827380953, 0.9722222222222222, 'x[17] <= 0.038\ngini =
0.269\nsamples = 1176\nvalue = [988, 188]'),
 Text(0.06101190476190476, 0.9166666666666666, 'x[0] <= 0.202\ngini =
0.5\nsamples = 78\nvalue = [39, 39]'),
 Text(0.023809523809523808, 0.8611111111111112, 'x[6] <= 0.364\ngini =
0.439\nsamples = 40\nvalue = [13, 27]'),
 Text(0.011904761904761904, 0.8055555555555556, 'x[2] <= 0.018\ngini =
0.142\nsamples = 13\nvalue = [1, 12]'),
 Text(0.005952380952380952, 0.75, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.017857142857142856, 0.75, 'gini = 0.0\nsamples = 12\nvalue =
[0, 12]'),
 Text(0.03571428571428571, 0.8055555555555556, 'x[9] <= 0.167\ngini =
0.494\nsamples = 27\nvalue = [12, 15]'),
 Text(0.02976190476190476, 0.75, 'gini = 0.0\nsamples = 4\nvalue = [0,
```

```
4]'),
 Text(0.041666666666666664, 0.75, 'x[3] <= 0.125\ngini = 0.499\
nsamples = 23\nvalue = [12, 11]'),
 Text(0.02976190476190476, 0.6944444444444444, 'x[4] <= 0.271\ngini =
0.278\nsamples = 6\nvalue = [1, 5]'),
 Text(0.023809523809523808, 0.6388888888888888, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.03571428571428571, 0.6388888888888888, 'gini = 0.0\nsamples =
5\nvalue = [0, 5]'),
 Text(0.05357142857142857, 0.6944444444444444, 'x[19] <= 0.167\ngini =
0.457\nsamples = 17\nvalue = [11, 6]'),
 Text(0.047619047619047616, 0.6388888888888888, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.05952380952380952, 0.6388888888888888, 'x[2] <= 0.411\ngini =
0.391\nsamples = 15\nvalue = [11, 4]'),
 Text(0.05357142857142857, 0.5833333333333334, 'x[10] <= 0.089\ngini =
0.494\nsamples = 9\nvalue = [5, 4]'),
 Text(0.047619047619047616, 0.5277777777777778, 'x[1] <= 0.797\ngini =
0.408\nsamples = 7\nvalue = [5, 2]'),
 Text(0.041666666666666664, 0.4722222222222222, 'x[4] <= 0.775\ngini =
0.278\nsamples = 6\nvalue = [5, 1]'),
 Text(0.03571428571428571, 0.4166666666666667, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.047619047619047616, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.05357142857142857, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.05952380952380952, 0.5277777777777778, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.06547619047619048, 0.5833333333333334, 'gini = 0.0\nsamples =
6\nvalue = [6, 0]'),
 Text(0.09821428571428571, 0.8611111111111112, 'x[5] <= 0.5\ngini =
0.432\nsamples = 38\nvalue = [26, 12]'),
 Text(0.08333333333333333, 0.8055555555555556, 'x[6] <= 0.643\ngini =
0.49\nsamples = 14\nvalue = [6, 8]'),
 Text(0.07142857142857142, 0.75, 'x[19] <= 0.5\ngini = 0.346\nsamples
= 9\nvalue = [2, 7]'),
 Text(0.06547619047619048, 0.6944444444444444, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.07738095238095238, 0.6944444444444444, 'gini = 0.0\nsamples =
7\nvalue = [0, 7]'),
 Text(0.09523809523809523, 0.75, 'x[19] <= 0.5\ngini = 0.32\nsamples =
5\nvalue = [4, 1]'),
 Text(0.08928571428571429, 0.6944444444444444, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.10119047619047619, 0.6944444444444444, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.1130952380952381, 0.8055555555555556, 'x[1] <= 0.178\ngini =
0.278\nsamples = 24\nvalue = [20, 4]'),
```

```
 Text(0.10714285714285714, 0.75, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.11904761904761904, 0.75, 'x[4] <= 0.796\ngini = 0.165\nsamples
= 22\nvalue = [20, 2]'),
 Text(0.1130952380952381, 0.6944444444444444, 'gini = 0.0\nsamples =
17\nvalue = [17, 0]'),
 Text(0.125, 0.6944444444444444, 'x[7] <= 0.5\ngini = 0.48\nsamples =
5\nvalue = [3, 2]'),
 Text(0.11904761904761904, 0.6388888888888888, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.13095238095238096, 0.6388888888888888, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.5101841517857143, 0.9166666666666666, 'x[8] <= 0.125\ngini =
0.235\nsamples = 1098\nvalue = [949, 149]'),
 Text(0.2564174107142857, 0.8611111111111112, 'x[19] <= 0.167\ngini =
0.337\nsamples = 364\nvalue = [286, 78]'),
 Text(0.1488095238095238, 0.8055555555555556, 'x[1] <= 0.885\ngini =
0.499\nsamples = 25\nvalue = [12, 13]'),
 Text(0.14285714285714285, 0.75, 'x[10] <= 0.087\ngini = 0.455\
nsamples = 20\nvalue = [7, 13]'),
 Text(0.13690476190476192, 0.6944444444444444, 'gini = 0.0\nsamples =
7\nvalue = [0, 7]'),
 Text(0.1488095238095238, 0.6944444444444444, 'x[1] <= 0.269\ngini =
0.497\nsamples = 13\nvalue = [7, 6]'),
 Text(0.14285714285714285, 0.6388888888888888, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.15476190476190477, 0.6388888888888888, 'x[11] <= 0.251\ngini =
0.375\nsamples = 8\nvalue = [2, 6]'),
 Text(0.1488095238095238, 0.5833333333333334, 'x[23] <= 0.324\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.14285714285714285, 0.5277777777777778, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.15476190476190477, 0.5277777777777778, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.16071428571428573, 0.5833333333333334, 'gini = 0.0\nsamples =
5\nvalue = [0, 5]'),
 Text(0.15476190476190477, 0.75, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
 Text(0.3640252976190476, 0.8055555555555556, 'x[16] <= 0.167\ngini =
0.31\nsamples = 339\nvalue = [274, 65]'),
 Text(0.30115327380952384, 0.75, 'x[13] <= 0.75\ngini = 0.394\nsamples
= 152\nvalue = [111, 41]'),
 Text(0.24516369047619047, 0.6944444444444444, 'x[13] <= 0.036\ngini =
0.366\nsamples = 141\nvalue = [107, 34]'),
 Text(0.18452380952380953, 0.6388888888888888, 'x[11] <= 0.752\ngini =
0.499\nsamples = 27\nvalue = [14, 13]'),
 Text(0.17857142857142858, 0.5833333333333334, 'x[11] <= 0.247\ngini =
0.483\nsamples = 22\nvalue = [9, 13]'),
 Text(0.16666666666666666, 0.5277777777777778, 'x[6] <= 0.679\ngini =
```

```
0.463\nsamples = 11\nvalue = [7, 4]'),
 Text(0.16071428571428573, 0.4722222222222222, 'x[9] <= 0.5\ngini =
0.444\nsamples = 6\nvalue = [2, 4]'),
 Text(0.15476190476190477, 0.4166666666666667, 'gini = 0.0\nsamples =
4\nvalue = [0, 4]'),
 Text(0.16666666666666666, 0.4166666666666667, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.17261904761904762, 0.4722222222222222, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.19047619047619047, 0.5277777777777778, 'x[0] <= 0.393\ngini =
0.298\nsamples = 11\nvalue = [2, 9]'),
 Text(0.18452380952380953, 0.4722222222222222, 'gini = 0.0\nsamples =
8\nvalue = [0, 8]'),
 Text(0.19642857142857142, 0.4722222222222222, 'x[23] <= 0.088\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.19047619047619047, 0.4166666666666667, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.20238095238095238, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.19047619047619047, 0.5833333333333334, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.30580357142857145, 0.6388888888888888, 'x[4] <= 0.68\ngini =
0.301\nsamples = 114\nvalue = [93, 21]'),
 Text(0.27827380952380953, 0.5833333333333334, 'x[4] <= 0.64\ngini =
0.365\nsamples = 75\nvalue = [57, 18]'),
 Text(0.27232142857142855, 0.5277777777777778, 'x[0] <= 0.369\ngini =
0.342\nsamples = 73\nvalue = [57, 16]'),
 Text(0.24107142857142858, 0.4722222222222222, 'x[2] <= 0.054\ngini =
0.427\nsamples = 42\nvalue = [29, 13]'),
 Text(0.21428571428571427, 0.4166666666666667, 'x[1] <= 0.189\ngini =
0.486\nsamples = 12\nvalue = [5, 7]'),
 Text(0.20833333333333334, 0.3611111111111111, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.22023809523809523, 0.3611111111111111, 'x[1] <= 0.765\ngini =
0.346\nsamples = 9\nvalue = [2, 7]'),
 Text(0.21428571428571427, 0.3055555555555556, 'gini = 0.0\nsamples =
6\nvalue = [0, 6]'),
 Text(0.2261904761904762, 0.3055555555555556, 'x[10] <= 0.103\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.22023809523809523, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.23214285714285715, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.26785714285714285, 0.4166666666666667, 'x[10] <= 0.083\ngini =
0.32\nsamples = 30\nvalue = [24, 6]'),
 Text(0.25595238095238093, 0.3611111111111111, 'x[4] <= 0.513\ngini =
0.473\nsamples = 13\nvalue = [8, 5]'),
 Text(0.25, 0.3055555555555556, 'x[10] <= 0.072\ngini = 0.397\nsamples
= 11\nvalue = [8, 3]'),
```

```
 Text(0.24404761904761904, 0.25, 'gini = 0.0\nsamples = 7\nvalue = [7,
0]'),
 Text(0.25595238095238093, 0.25, 'x[1] <= 0.684\ngini = 0.375\nsamples
= 4\nvalue = [1, 3]'),
 Text(0.25, 0.19444444444444445, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.2619047619047619, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.2619047619047619, 0.3055555555555556, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.27976190476190477, 0.3611111111111111, 'x[4] <= 0.062\ngini =
0.111\nsamples = 17\nvalue = [16, 1]'),
 Text(0.27380952380952384, 0.3055555555555556, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.2857142857142857, 0.3055555555555556, 'gini = 0.0\nsamples =
16\nvalue = [16, 0]'),
 Text(0.30357142857142855, 0.4722222222222222, 'x[4] <= 0.023\ngini =
0.175\nsamples = 31\nvalue = [28, 3]'),
 Text(0.2976190476190476, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.30952380952380953, 0.4166666666666667, 'x[22] <= 0.733\ngini =
0.124\nsamples = 30\nvalue = [28, 2]'),
 Text(0.30357142857142855, 0.3611111111111111, 'x[12] <= 0.556\ngini =
0.067\nsamples = 29\nvalue = [28, 1]'),
 Text(0.2976190476190476, 0.3055555555555556, 'gini = 0.0\nsamples =
25\nvalue = [25, 0]'),
 Text(0.30952380952380953, 0.3055555555555556, 'x[0] <= 0.619\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.30357142857142855, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.31547619047619047, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.31547619047619047, 0.3611111111111111, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.28422619047619047, 0.5277777777777778, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.3333333333333333, 0.5833333333333334, 'x[17] <= 0.063\ngini =
0.142\nsamples = 39\nvalue = [36, 3]'),
 Text(0.32142857142857145, 0.5277777777777778, 'x[13] <= 0.5\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.31547619047619047, 0.4722222222222222, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.3273809523809524, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.34523809523809523, 0.5277777777777778, 'x[0] <= 0.798\ngini =
0.054\nsamples = 36\nvalue = [35, 1]'),
 Text(0.3392857142857143, 0.4722222222222222, 'gini = 0.0\nsamples =
35\nvalue = [35, 0]'),
 Text(0.35119047619047616, 0.4722222222222222, 'gini = 0.0\nsamples =
```

```
1\nvalue = [0, 1]'),
 Text(0.35714285714285715, 0.6944444444444444, 'x[4] <= 0.523\ngini =
0.463\nsamples = 11\nvalue = [4, 7]'),
 Text(0.35119047619047616, 0.6388888888888888, 'x[6] <= 0.136\ngini =
0.219\nsamples = 8\nvalue = [1, 7]'),
 Text(0.34523809523809523, 0.5833333333333334, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.35714285714285715, 0.5833333333333334, 'gini = 0.0\nsamples =
7\nvalue = [0, 7]'),
 Text(0.3630952380952381, 0.6388888888888888, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.42689732142857145, 0.75, 'x[6] <= 0.136\ngini = 0.224\nsamples
= 187\nvalue = [163, 24]'),
 Text(0.3869047619047619, 0.6944444444444444, 'x[12] <= 0.944\ngini =
0.444\nsamples = 24\nvalue = [16, 8]'),
 Text(0.38095238095238093, 0.6388888888888888, 'x[15] <= 0.167\ngini =
0.363\nsamples = 21\nvalue = [16, 5]'),
 Text(0.36904761904761907, 0.5833333333333334, 'x[23] <= 0.088\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
 Text(0.3630952380952381, 0.5277777777777778, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.375, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.39285714285714285, 0.5833333333333334, 'x[21] <= 0.361\ngini =
0.208\nsamples = 17\nvalue = [15, 2]'),
 Text(0.3869047619047619, 0.5277777777777778, 'x[12] <= 0.056\ngini =
0.117\nsamples = 16\nvalue = [15, 1]'),
 Text(0.38095238095238093, 0.4722222222222222, 'x[15] <= 0.833\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.375, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.3869047619047619, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.39285714285714285, 0.4722222222222222, 'gini = 0.0\nsamples =
14\nvalue = [14, 0]'),
 Text(0.39880952380952384, 0.5277777777777778, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.39285714285714285, 0.6388888888888888, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
 Text(0.46688988095238093, 0.6944444444444444, 'x[0] <= 0.202\ngini =
0.177\nsamples = 163\nvalue = [147, 16]'),
 Text(0.42857142857142855, 0.6388888888888888, 'x[12] <= 0.5\ngini =
0.365\nsamples = 25\nvalue = [19, 6]'),
 Text(0.4226190476190476, 0.5833333333333334, 'x[11] <= 0.723\ngini =
0.287\nsamples = 23\nvalue = [19, 4]'),
 Text(0.4107142857142857, 0.5277777777777778, 'x[1] <= 0.941\ngini =
0.105\nsamples = 18\nvalue = [17, 1]'),
 Text(0.40476190476190477, 0.4722222222222222, 'gini = 0.0\nsamples =
17\nvalue = [17, 0]'),
```

```
 Text(0.4166666666666667, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.43452380952380953, 0.5277777777777778, 'x[5] <= 0.167\ngini =
0.48\nsamples = 5\nvalue = [2, 3]'),
 Text(0.42857142857142855, 0.4722222222222222, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.44047619047619047, 0.4722222222222222, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
 Text(0.43452380952380953, 0.5833333333333334, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.5052083333333334, 0.6388888888888888, 'x[17] <= 0.138\ngini =
0.134\nsamples = 138\nvalue = [128, 10]'),
 Text(0.47470238095238093, 0.5833333333333334, 'x[23] <= 0.029\ngini =
0.258\nsamples = 46\nvalue = [39, 7]'),
 Text(0.4583333333333333, 0.5277777777777778, 'x[5] <= 0.833\ngini =
0.5\nsamples = 8\nvalue = [4, 4]'),
 Text(0.4523809523809524, 0.4722222222222222, 'x[11] <= 0.298\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(0.44642857142857145, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.4583333333333333, 0.4166666666666667, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.4642857142857143, 0.4722222222222222, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
 Text(0.49107142857142855, 0.5277777777777778, 'x[3] <= 0.125\ngini =
0.145\nsamples = 38\nvalue = [35, 3]'),
 Text(0.47619047619047616, 0.4722222222222222, 'x[10] <= 0.111\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.47023809523809523, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.48214285714285715, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.5059523809523809, 0.4722222222222222, 'x[12] <= 0.556\ngini =
0.105\nsamples = 36\nvalue = [34, 2]'),
 Text(0.49404761904761907, 0.4166666666666667, 'x[18] <= 0.75\ngini =
0.059\nsamples = 33\nvalue = [32, 1]'),
 Text(0.4880952380952381, 0.3611111111111111, 'gini = 0.0\nsamples =
27\nvalue = [27, 0]'),
 Text(0.5, 0.3611111111111111, 'x[17] <= 0.063\ngini = 0.278\nsamples
= 6\nvalue = [5, 1]'),
 Text(0.49404761904761907, 0.3055555555555556, 'x[4] <= 0.377\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.4880952380952381, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.5, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.5059523809523809, 0.3055555555555556, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.5178571428571429, 0.4166666666666667, 'x[20] <= 0.063\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
```

```
 Text(0.5119047619047619, 0.3611111111111111, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.5238095238095238, 0.3611111111111111, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.5357142857142857, 0.5833333333333334, 'x[1] <= 0.024\ngini =
0.063\nsamples = 92\nvalue = [89, 3]'),
 Text(0.5238095238095238, 0.5277777777777778, 'x[0] <= 0.381\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.5178571428571429, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.5297619047619048, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.5476190476190477, 0.5277777777777778, 'x[13] <= 0.036\ngini =
0.043\nsamples = 90\nvalue = [88, 2]'),
 Text(0.5416666666666666, 0.4722222222222222, 'x[4] <= 0.689\ngini =
0.32\nsamples = 10\nvalue = [8, 2]'),
 Text(0.5357142857142857, 0.4166666666666667, 'gini = 0.0\nsamples =
7\nvalue = [7, 0]'),
 Text(0.5476190476190477, 0.4166666666666667, 'x[11] <= 0.233\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.5416666666666666, 0.3611111111111111, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.5535714285714286, 0.3611111111111111, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.5535714285714286, 0.4722222222222222, 'gini = 0.0\nsamples =
80\nvalue = [80, 0]'),
 Text(0.7639508928571429, 0.8611111111111112, 'x[5] <= 0.167\ngini =
0.175\nsamples = 734\nvalue = [663, 71]'),
 Text(0.6026785714285714, 0.8055555555555556, 'x[7] <= 0.167\ngini =
0.327\nsamples = 136\nvalue = [108, 28]'),
 Text(0.5773809523809523, 0.75, 'x[18] <= 0.25\ngini = 0.219\nsamples
= 8\nvalue = [1, 7]'),
 Text(0.5714285714285714, 0.6944444444444444, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.5833333333333334, 0.6944444444444444, 'gini = 0.0\nsamples =
7\nvalue = [0, 7]'),
 Text(0.6279761904761905, 0.75, 'x[23] <= 0.029\ngini = 0.274\nsamples
= 128\nvalue = [107, 21]'),
 Text(0.5952380952380952, 0.6944444444444444, 'x[10] <= 0.541\ngini =
0.495\nsamples = 20\nvalue = [11, 9]'),
 Text(0.5892857142857143, 0.6388888888888888, 'x[1] <= 0.433\ngini =
0.48\nsamples = 15\nvalue = [6, 9]'),
 Text(0.5773809523809523, 0.5833333333333334, 'x[13] <= 0.679\ngini =
0.32\nsamples = 10\nvalue = [2, 8]'),
 Text(0.5714285714285714, 0.5277777777777778, 'x[3] <= 0.625\ngini =
0.198\nsamples = 9\nvalue = [1, 8]'),
 Text(0.5654761904761905, 0.4722222222222222, 'gini = 0.0\nsamples =
7\nvalue = [0, 7]'),
 Text(0.5773809523809523, 0.4722222222222222, 'x[21] <= 0.028\ngini =
```

```
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.5714285714285714, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.5833333333333334, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.5833333333333334, 0.5277777777777778, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.6011904761904762, 0.5833333333333334, 'x[8] <= 0.375\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(0.5952380952380952, 0.5277777777777778, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
 Text(0.6071428571428571, 0.5277777777777778, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.6011904761904762, 0.6388888888888888, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.6607142857142857, 0.6944444444444444, 'x[2] <= 0.268\ngini =
0.198\nsamples = 108\nvalue = [96, 12]'),
 Text(0.6369047619047619, 0.6388888888888888, 'x[1] <= 0.086\ngini =
0.082\nsamples = 70\nvalue = [67, 3]'),
 Text(0.625, 0.5833333333333334, 'x[0] <= 0.464\ngini = 0.5\nsamples =
4\nvalue = [2, 2]'),
 Text(0.6190476190476191, 0.5277777777777778, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.6309523809523809, 0.5277777777777778, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.6488095238095238, 0.5833333333333334, 'x[19] <= 0.167\ngini =
0.03\nsamples = 66\nvalue = [65, 1]'),
 Text(0.6428571428571429, 0.5277777777777778, 'x[21] <= 0.083\ngini =
0.278\nsamples = 6\nvalue = [5, 1]'),
 Text(0.6369047619047619, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.6488095238095238, 0.4722222222222222, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
 Text(0.6547619047619048, 0.5277777777777778, 'gini = 0.0\nsamples =
60\nvalue = [60, 0]'),
 Text(0.6845238095238095, 0.6388888888888888, 'x[0] <= 0.595\ngini =
0.361\nsamples = 38\nvalue = [29, 9]'),
 Text(0.6726190476190477, 0.5833333333333334, 'x[0] <= 0.405\ngini =
0.238\nsamples = 29\nvalue = [25, 4]'),
 Text(0.6666666666666666, 0.5277777777777778, 'x[22] <= 0.033\ngini =
0.444\nsamples = 12\nvalue = [8, 4]'),
 Text(0.6607142857142857, 0.4722222222222222, 'x[10] <= 0.418\ngini =
0.32\nsamples = 5\nvalue = [1, 4]'),
 Text(0.6547619047619048, 0.4166666666666667, 'gini = 0.0\nsamples =
4\nvalue = [0, 4]'),
 Text(0.6666666666666666, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.6726190476190477, 0.4722222222222222, 'gini = 0.0\nsamples =
7\nvalue = [7, 0]'),
```

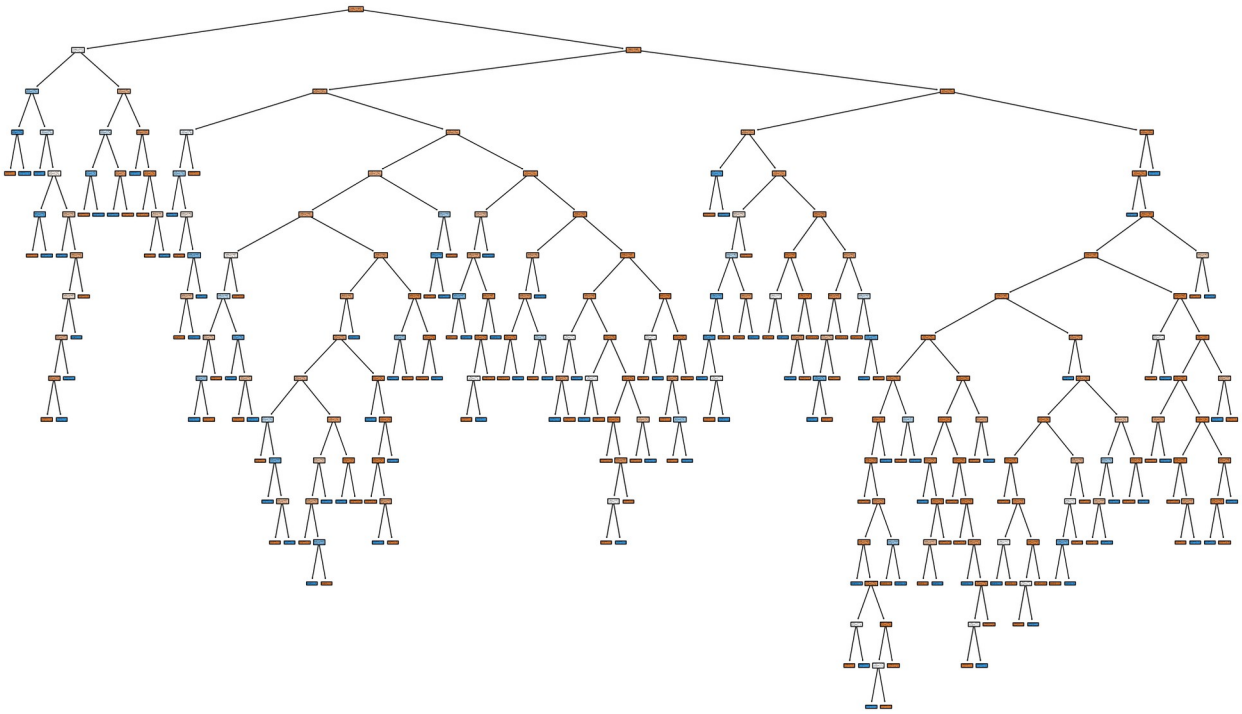
```
 Text(0.6785714285714286, 0.5277777777777778, 'gini = 0.0\nsamples =
17\nvalue = [17, 0]'),
 Text(0.6964285714285714, 0.5833333333333334, 'x[11] <= 0.214\ngini =
0.494\nsamples = 9\nvalue = [4, 5]'),
 Text(0.6904761904761905, 0.5277777777777778, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.7023809523809523, 0.5277777777777778, 'x[11] <= 0.857\ngini =
0.278\nsamples = 6\nvalue = [1, 5]'),
 Text(0.6964285714285714, 0.4722222222222222, 'gini = 0.0\nsamples =
5\nvalue = [0, 5]'),
 Text(0.7083333333333334, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.9252232142857143, 0.8055555555555556, 'x[17] <= 0.975\ngini =
0.133\nsamples = 598\nvalue = [555, 43]'),
 Text(0.9192708333333334, 0.75, 'x[4] <= 0.003\ngini = 0.128\nsamples
= 596\nvalue = [555, 41]'),
 Text(0.9133184523809523, 0.6944444444444444, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.9252232142857143, 0.6944444444444444, 'x[6] <= 0.993\ngini =
0.125\nsamples = 595\nvalue = [555, 40]'),
 Text(0.8802083333333334, 0.6388888888888888, 'x[17] <= 0.388\ngini =
0.121\nsamples = 590\nvalue = [552, 38]'),
 Text(0.8080357142857143, 0.5833333333333334, 'x[10] <= 0.332\ngini =
0.153\nsamples = 383\nvalue = [351, 32]'),
 Text(0.7485119047619048, 0.5277777777777778, 'x[16] <= 0.167\ngini =
0.105\nsamples = 271\nvalue = [256, 15]'),
 Text(0.7202380952380952, 0.4722222222222222, 'x[2] <= 0.857\ngini =
0.169\nsamples = 107\nvalue = [97, 10]'),
 Text(0.7083333333333334, 0.4166666666666667, 'x[1] <= 0.982\ngini =
0.128\nsamples = 102\nvalue = [95, 7]'),
 Text(0.7023809523809523, 0.3611111111111111, 'x[5] <= 0.833\ngini =
0.112\nsamples = 101\nvalue = [95, 6]'),
 Text(0.6964285714285714, 0.3055555555555556, 'gini = 0.0\nsamples =
62\nvalue = [62, 0]'),
 Text(0.7083333333333334, 0.3055555555555556, 'x[22] <= 0.367\ngini =
0.26\nsamples = 39\nvalue = [33, 6]'),
 Text(0.6964285714285714, 0.25, 'x[19] <= 0.167\ngini = 0.198\nsamples
= 36\nvalue = [32, 4]'),
 Text(0.6904761904761905, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7023809523809523, 0.19444444444444445, 'x[20] <= 0.038\ngini =
0.157\nsamples = 35\nvalue = [32, 3]'),
 Text(0.6904761904761905, 0.1388888888888889, 'x[0] <= 0.464\ngini =
0.5\nsamples = 4\nvalue = [2, 2]'),
 Text(0.6845238095238095, 0.08333333333333333, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.6964285714285714, 0.08333333333333333, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.7142857142857143, 0.1388888888888889, 'x[0] <= 0.202\ngini =
```

```
0.062\nsamples = 31\nvalue = [30, 1]'),
 Text(0.7083333333333334, 0.08333333333333333, 'x[20] <= 0.15\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.7023809523809523, 0.027777777777777776, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7142857142857143, 0.027777777777777776, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.7202380952380952, 0.08333333333333333, 'gini = 0.0\nsamples =
29\nvalue = [29, 0]'),
 Text(0.7202380952380952, 0.25, 'x[7] <= 0.167\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.7142857142857143, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.7261904761904762, 0.19444444444444445, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.7142857142857143, 0.3611111111111111, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7321428571428571, 0.4166666666666667, 'x[6] <= 0.3\ngini =
0.48\nsamples = 5\nvalue = [2, 3]'),
 Text(0.7261904761904762, 0.3611111111111111, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.7380952380952381, 0.3611111111111111, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
 Text(0.7767857142857143, 0.4722222222222222, 'x[18] <= 0.917\ngini =
0.059\nsamples = 164\nvalue = [159, 5]'),
 Text(0.7619047619047619, 0.4166666666666667, 'x[13] <= 0.036\ngini =
0.049\nsamples = 160\nvalue = [156, 4]'),
 Text(0.75, 0.3611111111111111, 'x[7] <= 0.167\ngini = 0.208\nsamples
= 17\nvalue = [15, 2]'),
 Text(0.7440476190476191, 0.3055555555555556, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7559523809523809, 0.3055555555555556, 'x[2] <= 0.071\ngini =
0.117\nsamples = 16\nvalue = [15, 1]'),
 Text(0.75, 0.25, 'x[22] <= 0.033\ngini = 0.444\nsamples = 3\nvalue =
[2, 1]'),
 Text(0.7440476190476191, 0.19444444444444445, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.7559523809523809, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7619047619047619, 0.25, 'gini = 0.0\nsamples = 13\nvalue =
[13, 0]'),
 Text(0.7738095238095238, 0.3611111111111111, 'x[6] <= 0.693\ngini =
0.028\nsamples = 143\nvalue = [141, 2]'),
 Text(0.7678571428571429, 0.3055555555555556, 'gini = 0.0\nsamples =
99\nvalue = [99, 0]'),
 Text(0.7797619047619048, 0.3055555555555556, 'x[13] <= 0.393\ngini =
0.087\nsamples = 44\nvalue = [42, 2]'),
 Text(0.7738095238095238, 0.25, 'gini = 0.0\nsamples = 31\nvalue =
[31, 0]'),
 Text(0.7857142857142857, 0.25, 'x[6] <= 0.736\ngini = 0.26\nsamples =
```

```
13\nvalue = [11, 2]'),
 Text(0.7797619047619048, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7916666666666666, 0.19444444444444445, 'x[9] <= 0.167\ngini =
0.153\nsamples = 12\nvalue = [11, 1]'),
 Text(0.7857142857142857, 0.1388888888888889, 'x[3] <= 0.625\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.7797619047619048, 0.08333333333333333, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.7916666666666666, 0.08333333333333333, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.7976190476190477, 0.1388888888888889, 'gini = 0.0\nsamples =
10\nvalue = [10, 0]'),
 Text(0.7916666666666666, 0.4166666666666667, 'x[2] <= 0.661\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.7857142857142857, 0.3611111111111111, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.7976190476190477, 0.3611111111111111, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.8675595238095238, 0.5277777777777778, 'x[10] <= 0.334\ngini =
0.257\nsamples = 112\nvalue = [95, 17]'),
 Text(0.8616071428571429, 0.4722222222222222, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.8735119047619048, 0.4722222222222222, 'x[11] <= 0.779\ngini =
0.236\nsamples = 110\nvalue = [95, 15]'),
 Text(0.8422619047619048, 0.4166666666666667, 'x[2] <= 0.696\ngini =
0.148\nsamples = 87\nvalue = [80, 7]'),
 Text(0.8154761904761905, 0.3611111111111111, 'x[22] <= 0.233\ngini =
0.078\nsamples = 74\nvalue = [71, 3]'),
 Text(0.8095238095238095, 0.3055555555555556, 'gini = 0.0\nsamples =
50\nvalue = [50, 0]'),
 Text(0.8214285714285714, 0.3055555555555556, 'x[17] <= 0.237\ngini =
0.219\nsamples = 24\nvalue = [21, 3]'),
 Text(0.8095238095238095, 0.25, 'x[11] <= 0.421\ngini = 0.5\nsamples =
4\nvalue = [2, 2]'),
 Text(0.8035714285714286, 0.19444444444444445, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.8154761904761905, 0.19444444444444445, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.8333333333333334, 0.25, 'x[0] <= 0.298\ngini = 0.095\nsamples
= 20\nvalue = [19, 1]'),
 Text(0.8273809523809523, 0.19444444444444445, 'x[4] <= 0.433\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.8214285714285714, 0.1388888888888889, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.8333333333333334, 0.1388888888888889, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.8392857142857143, 0.19444444444444445, 'gini = 0.0\nsamples =
18\nvalue = [18, 0]'),
```

```
 Text(0.8690476190476191, 0.3611111111111111, 'x[18] <= 0.417\ngini =
0.426\nsamples = 13\nvalue = [9, 4]'),
 Text(0.8630952380952381, 0.3055555555555556, 'x[9] <= 0.5\ngini =
0.5\nsamples = 8\nvalue = [4, 4]'),
 Text(0.8571428571428571, 0.25, 'x[11] <= 0.338\ngini = 0.32\nsamples
= 5\nvalue = [1, 4]'),
 Text(0.8511904761904762, 0.19444444444444445, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.8630952380952381, 0.19444444444444445, 'gini = 0.0\nsamples =
4\nvalue = [0, 4]'),
 Text(0.8690476190476191, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.875, 0.3055555555555556, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
 Text(0.9047619047619048, 0.4166666666666667, 'x[16] <= 0.167\ngini =
0.454\nsamples = 23\nvalue = [15, 8]'),
 Text(0.8928571428571429, 0.3611111111111111, 'x[22] <= 0.3\ngini =
0.463\nsamples = 11\nvalue = [4, 7]'),
 Text(0.8869047619047619, 0.3055555555555556, 'x[4] <= 0.886\ngini =
0.444\nsamples = 6\nvalue = [4, 2]'),
 Text(0.8809523809523809, 0.25, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(0.8928571428571429, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.8988095238095238, 0.3055555555555556, 'gini = 0.0\nsamples =
5\nvalue = [0, 5]'),
 Text(0.9166666666666666, 0.3611111111111111, 'x[23] <= 0.5\ngini =
0.153\nsamples = 12\nvalue = [11, 1]'),
 Text(0.9107142857142857, 0.3055555555555556, 'gini = 0.0\nsamples =
11\nvalue = [11, 0]'),
 Text(0.9226190476190477, 0.3055555555555556, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.9523809523809523, 0.5833333333333334, 'x[4] <= 0.029\ngini =
0.056\nsamples = 207\nvalue = [201, 6]'),
 Text(0.9345238095238095, 0.5277777777777778, 'x[5] <= 0.833\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.9285714285714286, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
 Text(0.9404761904761905, 0.4722222222222222, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.9702380952380952, 0.5277777777777778, 'x[4] <= 0.986\ngini =
0.048\nsamples = 205\nvalue = [200, 5]'),
 Text(0.9523809523809523, 0.4722222222222222, 'x[11] <= 0.012\ngini =
0.039\nsamples = 202\nvalue = [198, 4]'),
 Text(0.9345238095238095, 0.4166666666666667, 'x[11] <= 0.009\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.9285714285714286, 0.3611111111111111, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.9404761904761905, 0.3611111111111111, 'gini = 0.0\nsamples =
```

```
1\nvalue = [0, 1]'),
 Text(0.9702380952380952, 0.4166666666666667, 'x[20] <= 0.562\ngini =
0.03\nsamples = 198\nvalue = [195, 3]'),
 Text(0.9523809523809523, 0.3611111111111111, 'x[10] <= 0.992\ngini =
0.011\nsamples = 180\nvalue = [179, 1]'),
 Text(0.9464285714285714, 0.3055555555555556, 'gini = 0.0\nsamples =
176\nvalue = [176, 0]'),
 Text(0.9583333333333334, 0.3055555555555556, 'x[11] <= 0.749\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.9523809523809523, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.9642857142857143, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.9880952380952381, 0.3611111111111111, 'x[4] <= 0.829\ngini =
0.198\nsamples = 18\nvalue = [16, 2]'),
 Text(0.9821428571428571, 0.3055555555555556, 'x[10] <= 0.498\ngini =
0.111\nsamples = 17\nvalue = [16, 1]'),
 Text(0.9761904761904762, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.9880952380952381, 0.25, 'gini = 0.0\nsamples = 16\nvalue =
[16, 0]'),
 Text(0.9940476190476191, 0.3055555555555556, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.9880952380952381, 0.4722222222222222, 'x[13] <= 0.179\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.9821428571428571, 0.4166666666666667, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
 Text(0.9940476190476191, 0.4166666666666667, 'gini = 0.0\nsamples =
2\nvalue = [2, 0]'),
 Text(0.9702380952380952, 0.6388888888888888, 'x[1] <= 0.845\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
 Text(0.9642857142857143, 0.5833333333333334, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
 Text(0.9761904761904762, 0.5833333333333334, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
 Text(0.9311755952380952, 0.75, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]')]
```



Increasing the performance using hyper parameters

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

grid_search=GridSearchCV(estimator=model,param_grid=parameters,cv=5,sc
oring="accuracy")

grid_search
```

```
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')
```

```
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:
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(
/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(
```

```
/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(
/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(
/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(
/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(
```

```
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')
```

```
grid_search.best_params_
```

```
{'criterion': 'gini',
 'max_depth': 4,
 'max_features': 'sqrt',
 'splitter': 'random'}
```

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

```
DecisionTreeClassifier(criterion='entropy', max_depth=3,
max_features='sqrt')
```

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

```
pred
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

```
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 1, 0, 0])
```

```
print(classification_report(y_test,pred))# now accuracy increas to 84
percentage.
```

```
              precision    recall  f1-score   support

           0       0.84      0.99      0.91       245
           1       0.67      0.08      0.15        49

    accuracy                           0.84       294
   macro avg       0.76      0.54      0.53       294
weighted avg       0.81      0.84      0.78       294
```

Logistic Regression:

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

model.fit(x_train,y_train)

LogisticRegression()

pred=model.predict(x_test)

pred

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
```

```
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0])

y_test

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

model.predict(ms.transform([[41,1102, 1,   2,   1,   2,   94,  3,
     2,   4,   5993,19479,     8,2,3,4,5,11,   3,   1,   0,   8
     ,0,  1]]))  # we are using ms to transform the input to scaled
values.

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but MinMaxScaler was
fitted with feature names
```

```
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439:
UserWarning: X does not have valid feature names, but
LogisticRegression was fitted with feature names
  warnings.warn(
```

```
array([0])
```

```
from sklearn.metrics import
accuracy_score,confusion_matrix,classification_report,roc_auc_score,ro
c_curve
```

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

```
0.8401360544217688
```

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

```
array([[241,   4],
       [ 43,   6]])
```

```
pd.crosstab(y_test,pred)
```

```
col_0    0  1
row_0
0      241  4
1       43  6
```

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

```
              precision    recall  f1-score   support

           0       0.85      0.98      0.91       245
           1       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
```