assignment4

September 26, 2023

- Data Preprocessing.

o Import the Libraries.
o Importing the dataset.
o Checking for Null Values.
o Data Visualization.
o Outlier Detection
o Splitting Dependent and Independent variables
o- Encoding
o Feature Scaling.
o Splitting Data into Train and Test.

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

Data Collection.

o Collect the dataset or Create the dataset

[4]:
```
df = pd.read_csv("/content/WA_Fn-UseC_-HR-Employee-Attrition.csv")
df
```

[4]:

	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]
```

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

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

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
```

```
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

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

[7]:

	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]
```

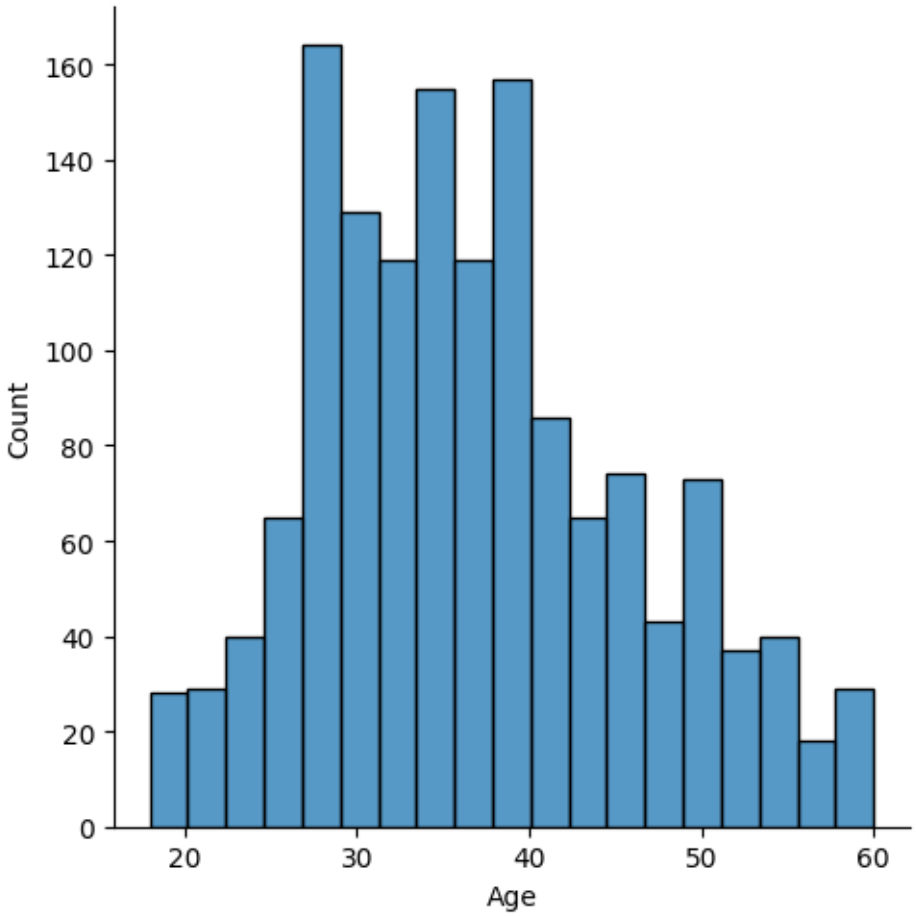
[8]:
```
df.isnull().any()
```

[8]:
```
Age                         False
Attrition                   False
BusinessTravel              False
DailyRate                   False
Department                  False
DistanceFromHome            False
Education                   False
EducationField              False
```

```
EmployeeCount               False
EmployeeNumber              False
EnvironmentSatisfaction     False
Gender                      False
HourlyRate                  False
JobInvolvement              False
JobLevel                    False
JobRole                     False
JobSatisfaction             False
MaritalStatus               False
MonthlyIncome               False
MonthlyRate                 False
NumCompaniesWorked          False
Over18                      False
OverTime                    False
PercentSalaryHike           False
PerformanceRating           False
RelationshipSatisfaction    False
StandardHours               False
StockOptionLevel            False
TotalWorkingYears           False
TrainingTimesLastYear       False
WorkLifeBalance             False
YearsAtCompany              False
YearsInCurrentRole          False
YearsSinceLastPromotion     False
YearsWithCurrManager        False
dtype: bool
```

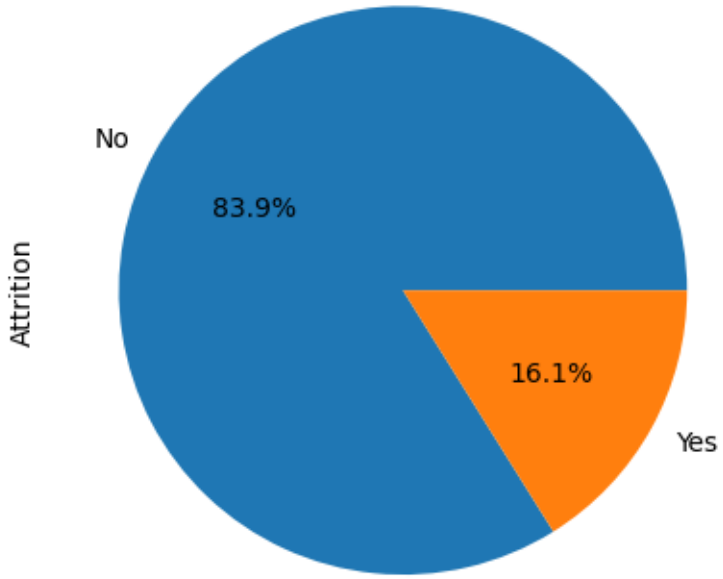
[9]:
```
sns.displot(df["Age"])
```

[9]:
```
<seaborn.axisgrid.FacetGrid at 0x7fbaf144eda0>
```



[10]:
```
df.Attrition.value_counts().plot(kind="pie",autopct="%1.1f%%")
```

[10]: <Axes: ylabel='Attrition'>



[11]:
```
df.corr()
```

```
<ipython-input-11-2f6f6606aa2c>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.corr()
```

[11]:

	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

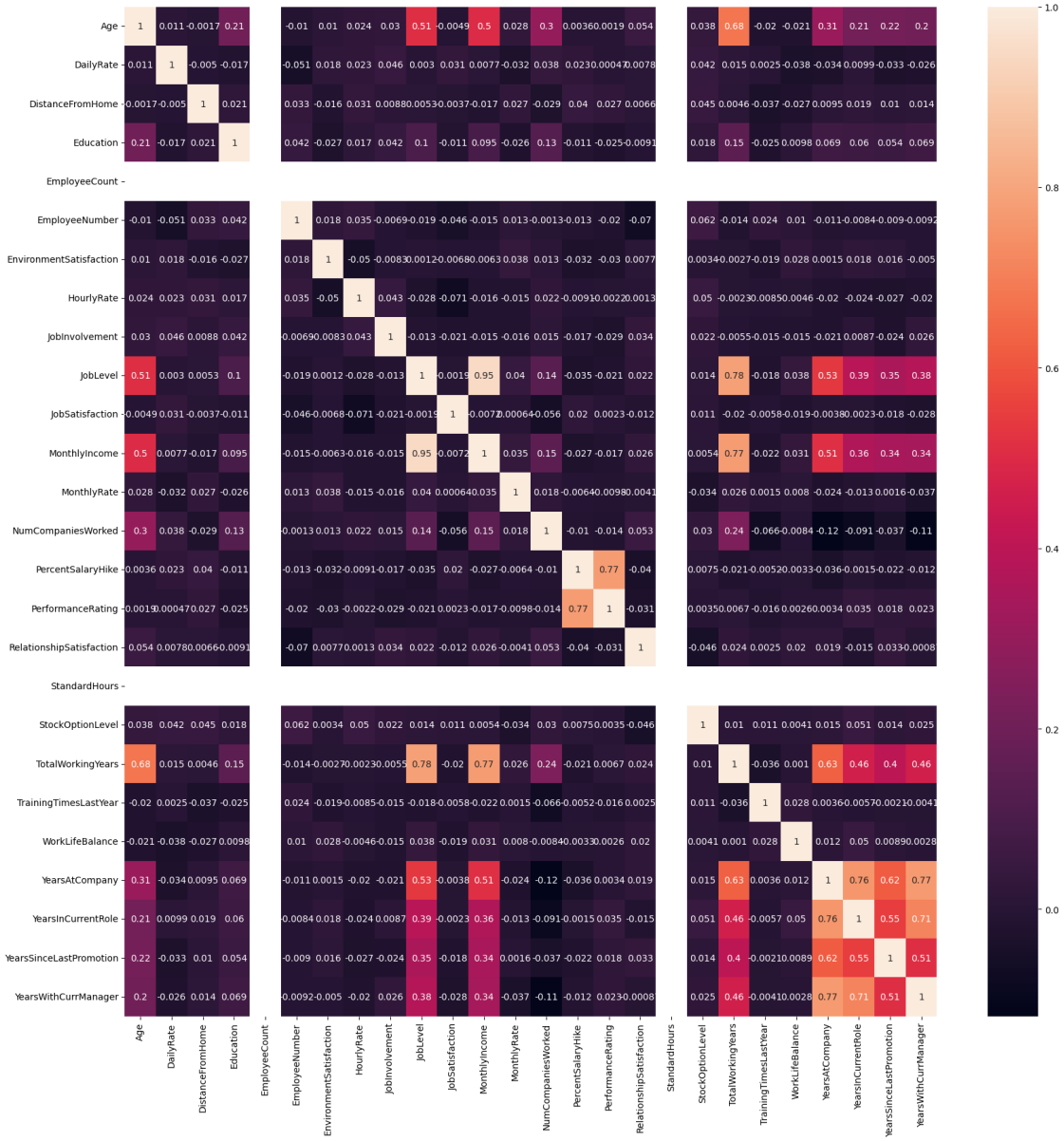
EnvironmentSatisfaction	0.016194	-0.004999
HourlyRate	-0.026716	-0.020123
JobInvolvement	-0.024184	0.025976
JobLevel	0.353885	0.375281
JobSatisfaction	-0.018214	-0.027656
MonthlyIncome	0.344978	0.344079
MonthlyRate	0.001567	-0.036746
NumCompaniesWorked	-0.036814	-0.110319
PercentSalaryHike	-0.022154	-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]

[12]:
```
plt.subplots(figsize=(20,20))
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-12-427abebf5887>:2: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr(),annot=True)
```

[12]: <Axes: >



[13]:
```
df.head()
```

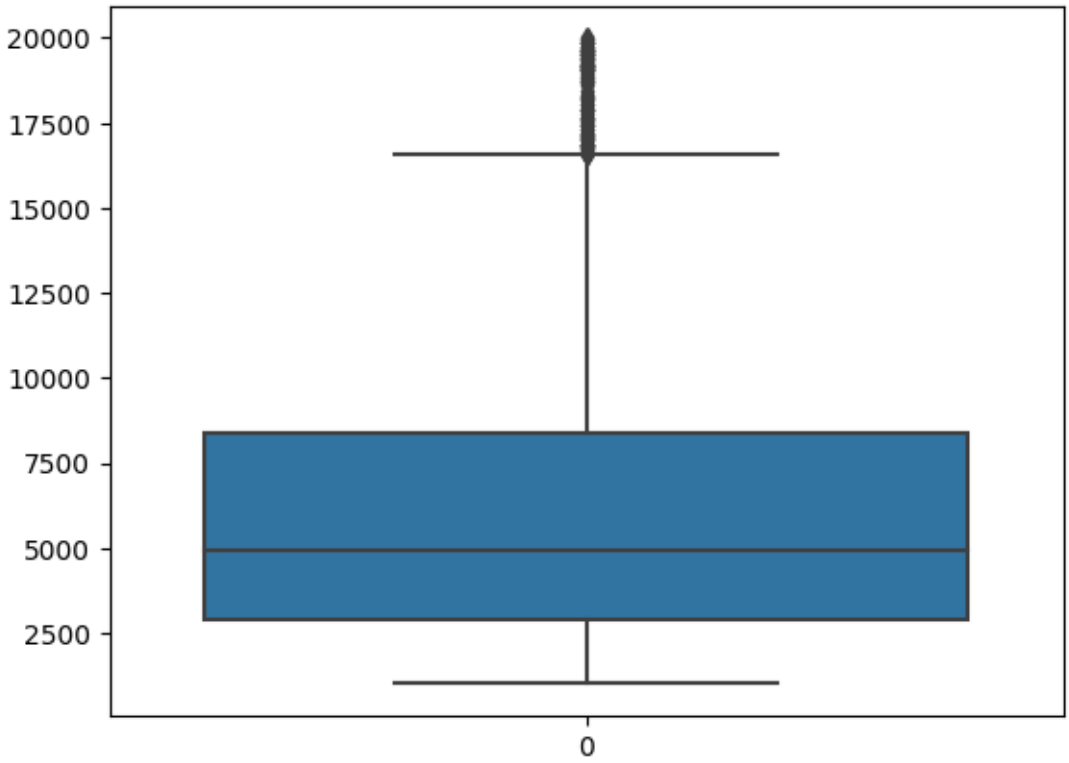
[13]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	\
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	
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						

```
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

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

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

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

[5 rows x 35 columns]
```

[14]:
```
sns.boxplot(df.MonthlyIncome)
```

[14]:
```
<Axes: >
```



[15]:
```
q1 = df.MonthlyIncome.quantile(0.25)
q3 = df.MonthlyIncome.quantile(0.75)
```

[16]:
```
IQR =q3-q1
```

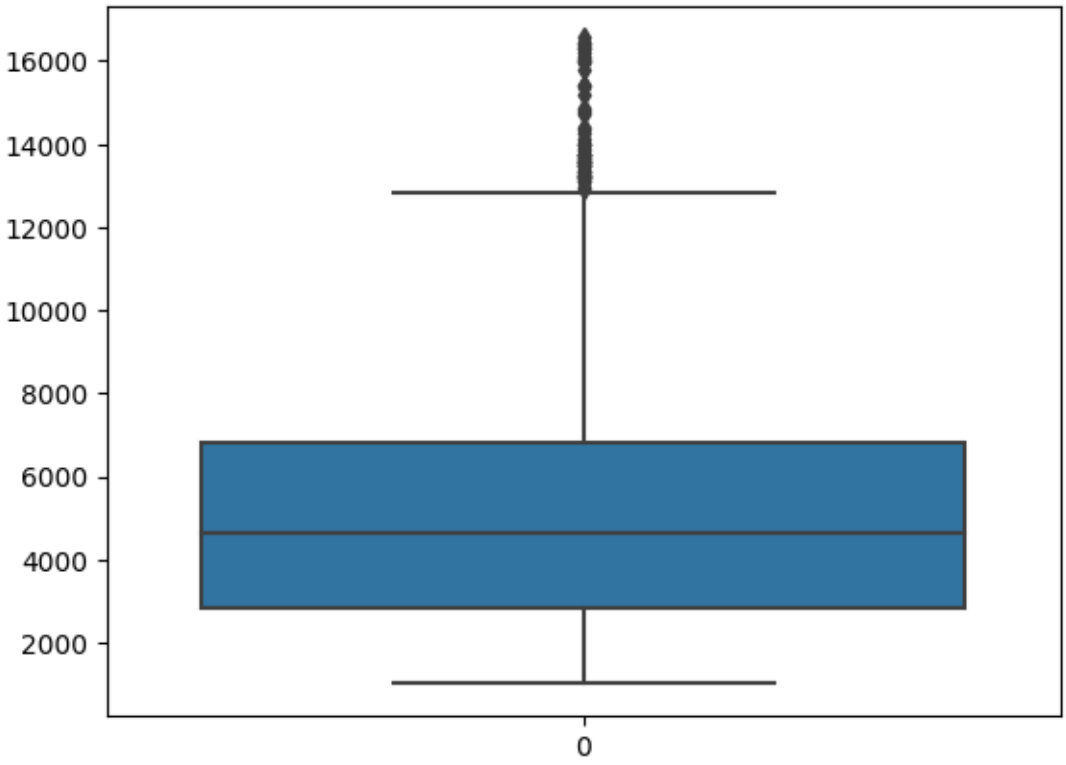
[17]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[17]: 16581.0

[18]:
```
df = df[df.MonthlyIncome<upper_limit]
```

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

[19]: <Axes: >



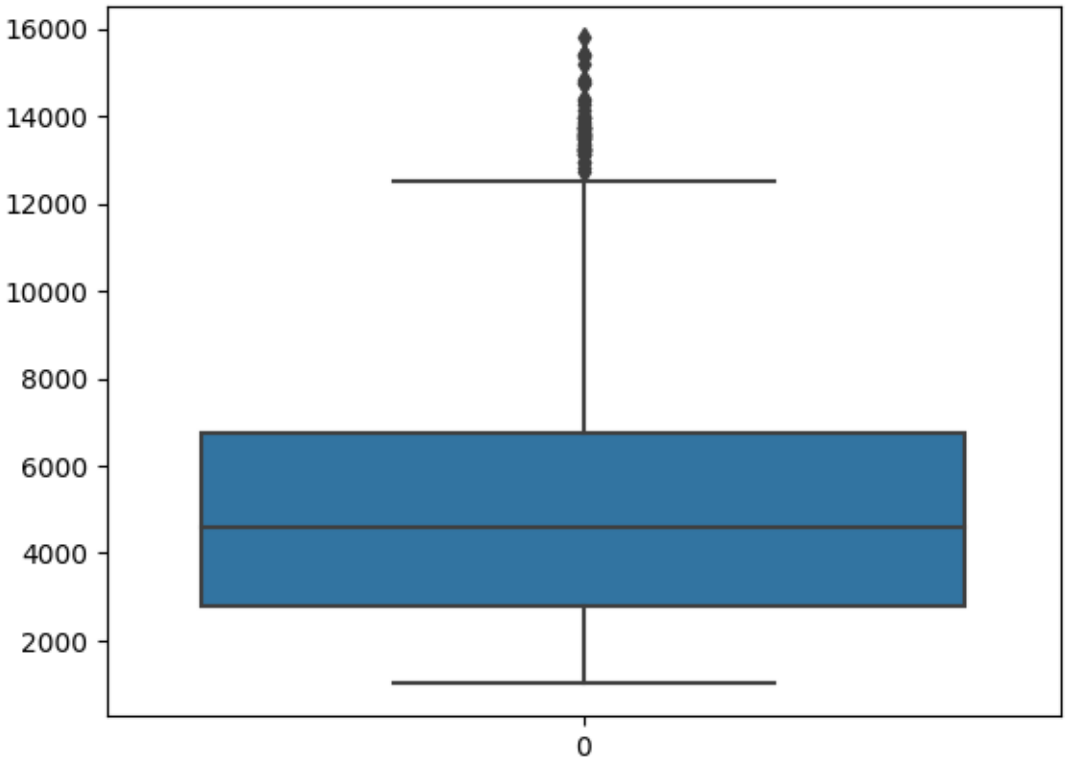
[20]:
```
p99 = df.MonthlyIncome.quantile(0.99)
p99
```

[20]: 15870.25000000001

[21]:
```
df = df[df.MonthlyIncome<=p99]
```

[22]:
```
sns.boxplot(df.MonthlyIncome)
```

[22]: <Axes: >



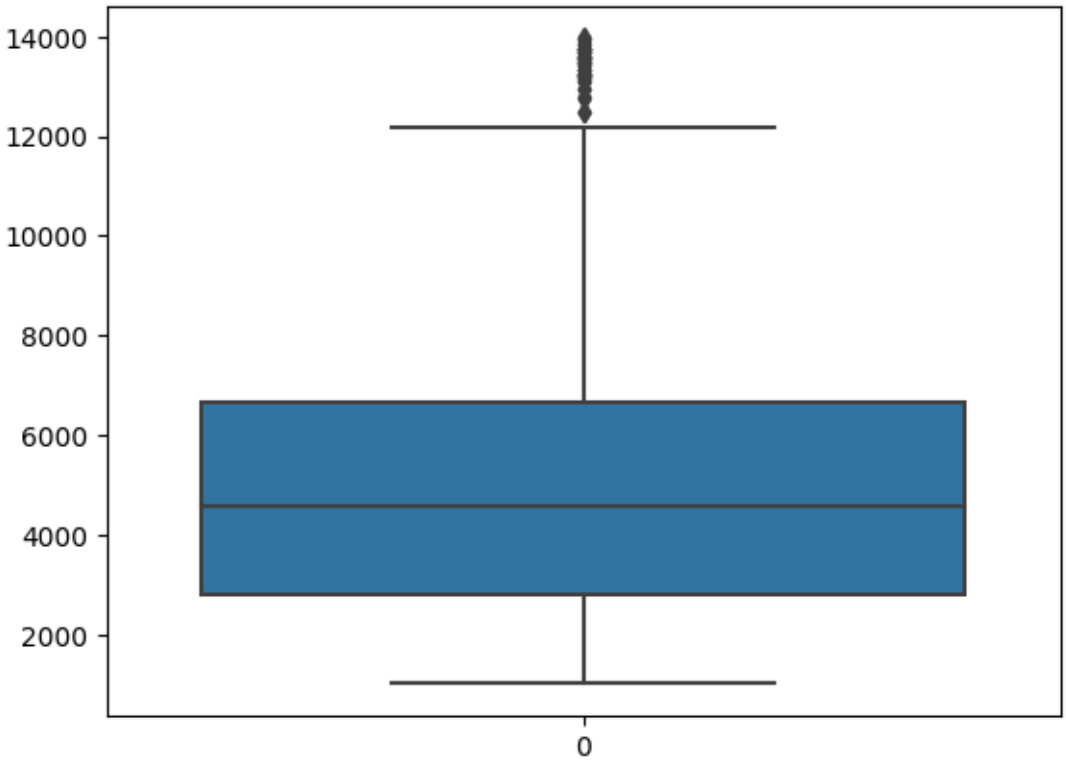
[23]:
```
p99 = df.MonthlyIncome.quantile(0.99)
p99
```

[23]: 14004.269999999995

[24]:
```
df = df[df.MonthlyIncome<=p99]
```

[25]:
```
sns.boxplot(df.MonthlyIncome)
```

[25]: <Axes: >



[26]:
```
q1 = df.MonthlyIncome.quantile(0.25)
q3 = df.MonthlyIncome.quantile(0.75)
```

[27]:
```
IQR =q3-q1
```

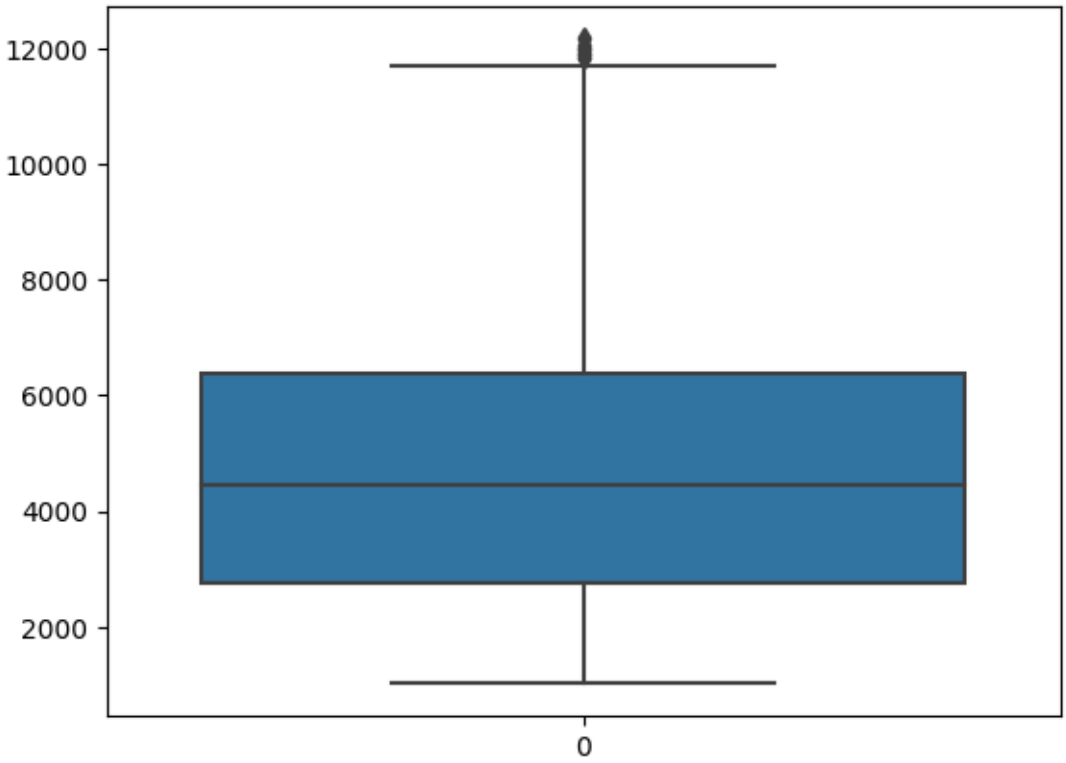
[28]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[28]: 12436.0

[29]:
```
df = df[df.MonthlyIncome<upper_limit]
```

[30]:
```
sns.boxplot(df.MonthlyIncome)
```

[30]: <Axes: >



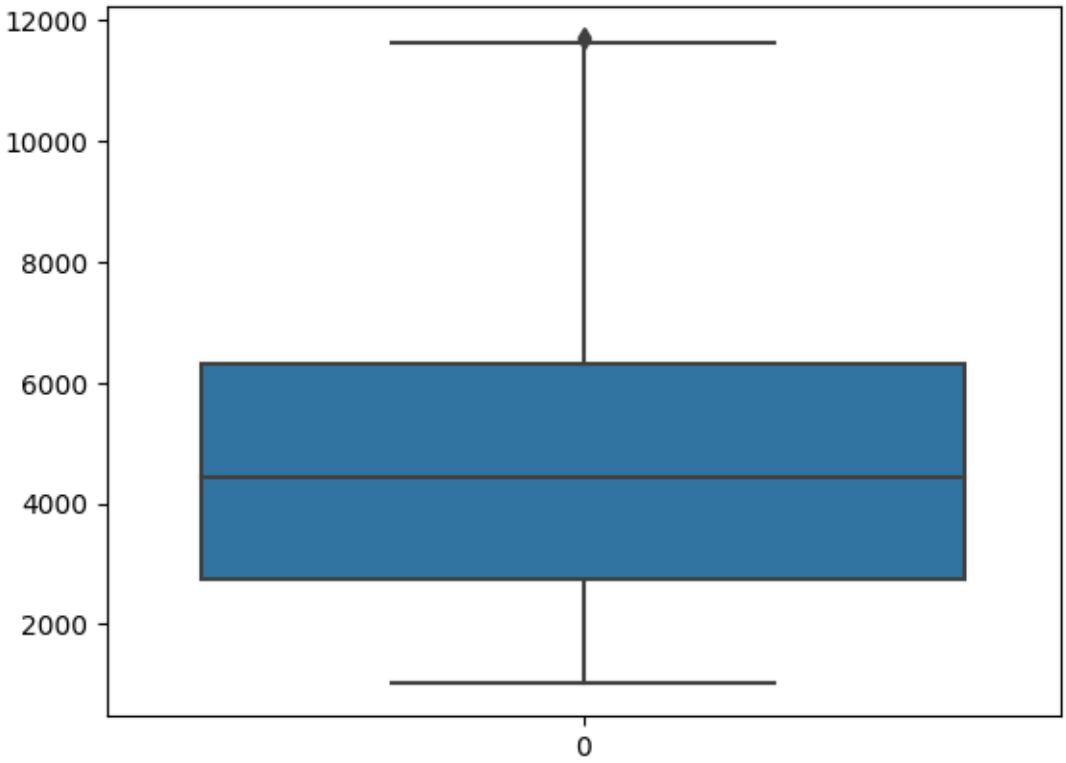
[31]:
```
p99 = df.MonthlyIncome.quantile(0.99)
p99
```

[31]: 11740.060000000003

[32]:
```
df = df[df.MonthlyIncome<=p99]
```

[33]:
```
sns.boxplot(df.MonthlyIncome)
```

[33]: <Axes: >



[34]:
```
q1 = df.MonthlyIncome.quantile(0.25)
q3 = df.MonthlyIncome.quantile(0.75)
```

[35]:
```
IQR =q3-q1
```

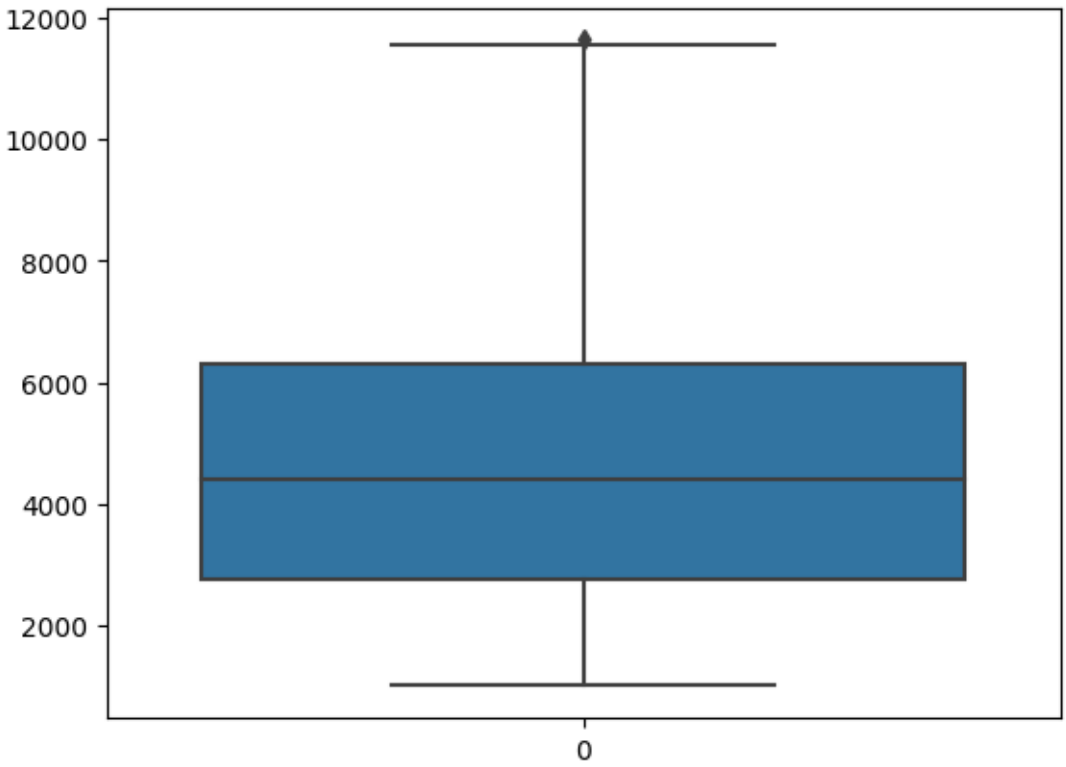
[36]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[36]: 11656.125

[37]:
```
df = df[df.MonthlyIncome<upper_limit]
```

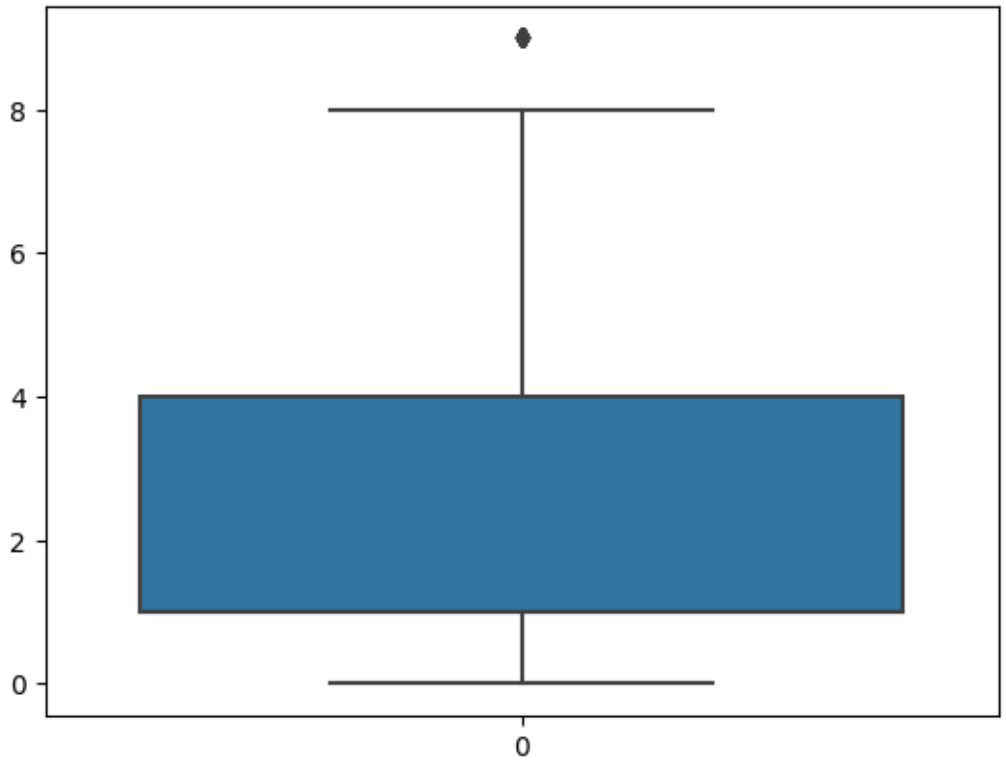
[38]:
```
sns.boxplot(df.MonthlyIncome)
```

[38]: <Axes: >



[39]:
```
sns.boxplot(df.NumCompaniesWorked)
```

[39]: <Axes: >



[40]:
```
q1 = df.NumCompaniesWorked.quantile(0.25)
q3 = df.NumCompaniesWorked.quantile(0.75)
```

[41]:
```
IQR =q3-q1
```

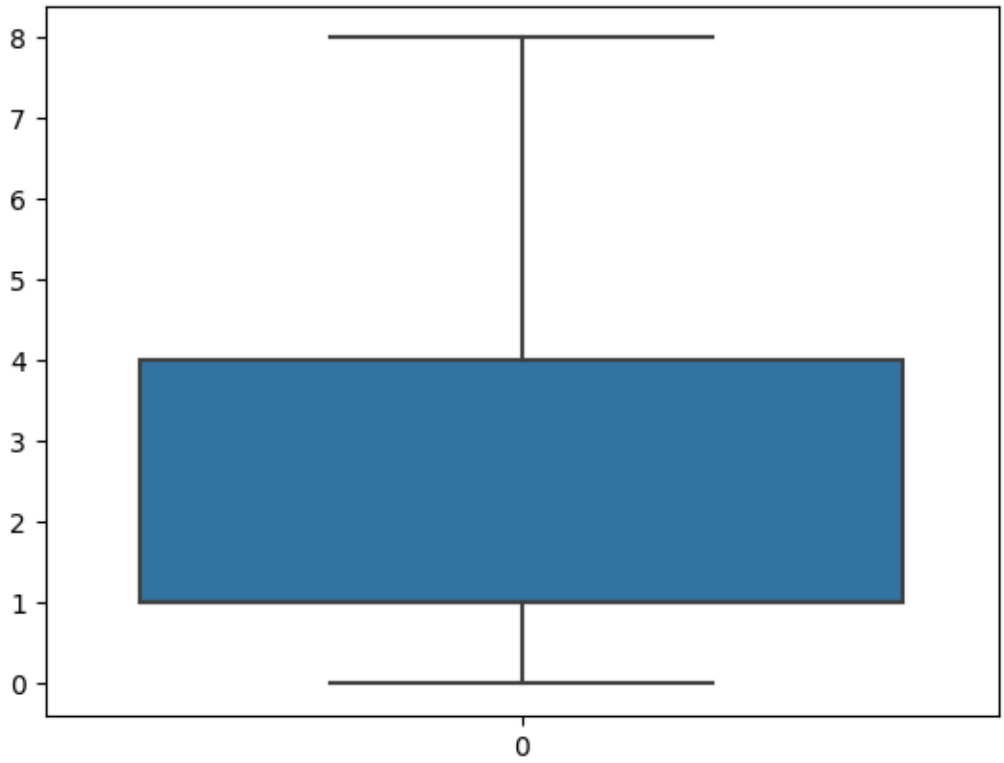
[42]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[42]: 8.5

[43]:
```
df = df[df.NumCompaniesWorked<upper_limit]
```

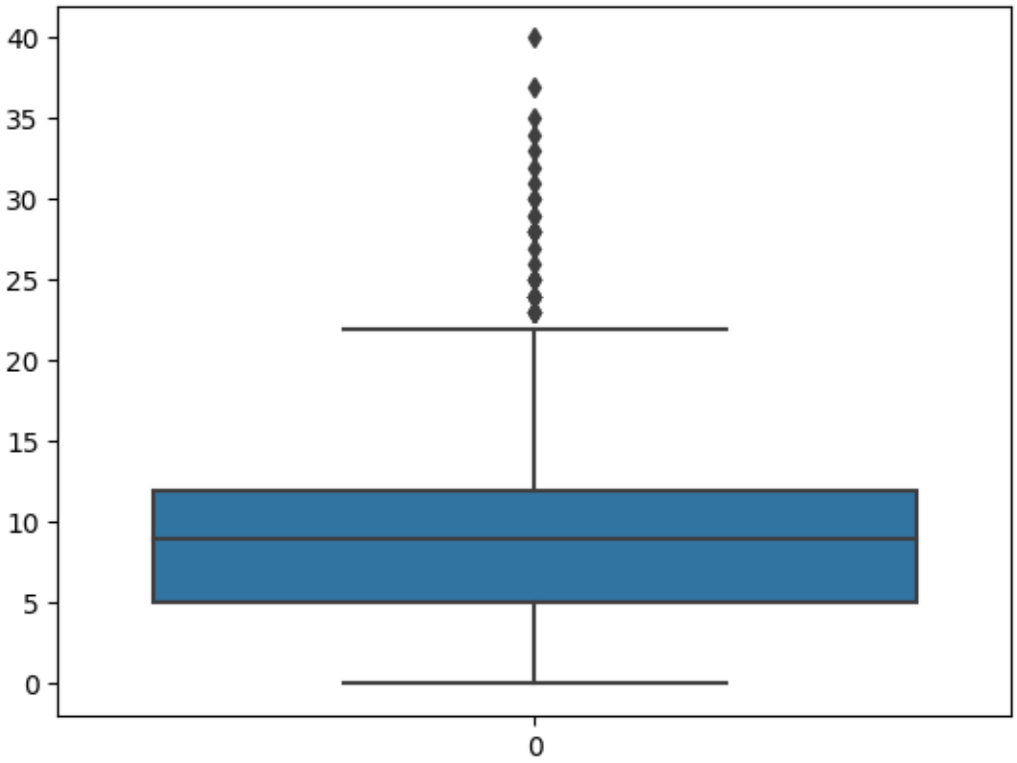
[44]:
```
sns.boxplot(df.NumCompaniesWorked)
```

[44]: <Axes: >



[45]:
```
sns.boxplot(df.TotalWorkingYears)
```

[45]: <Axes: >



[46]:
```
q1 = df.TotalWorkingYears.quantile(0.25)
q3 = df.TotalWorkingYears.quantile(0.75)
```

[47]:
```
IQR =q3-q1
```

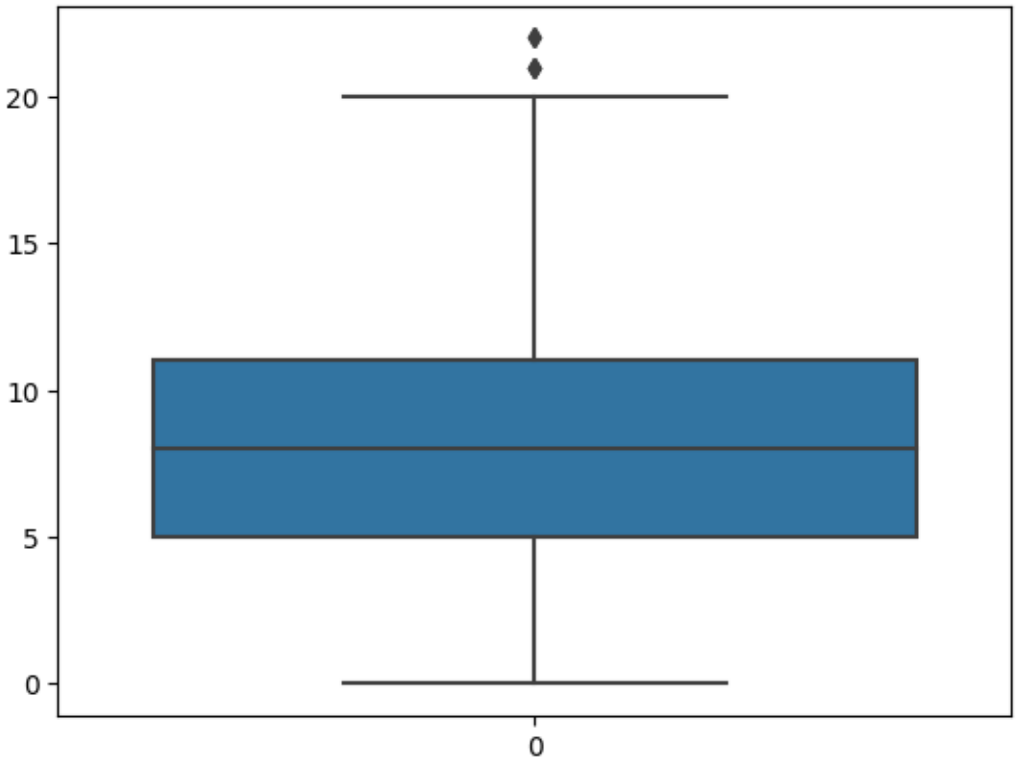
[48]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[48]: 22.5

[49]:
```
df = df[df.TotalWorkingYears<upper_limit]
```

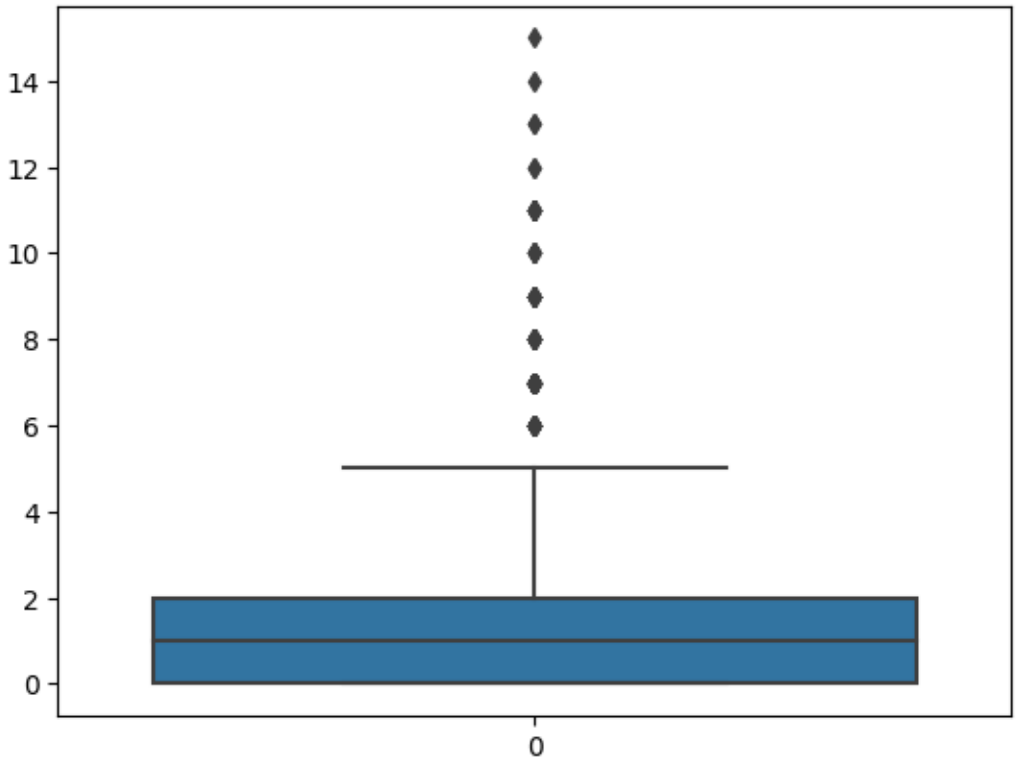
[50]:
```
sns.boxplot(df.TotalWorkingYears)
```

[50]: <Axes: >



[51]:
```
sns.boxplot(df.YearsSinceLastPromotion)
```

[51]: <Axes: >



[52]:
```
q1 = df.YearsSinceLastPromotion.quantile(0.25)
q3 = df.YearsSinceLastPromotion.quantile(0.75)
```

[53]:
```
IQR =q3-q1
```

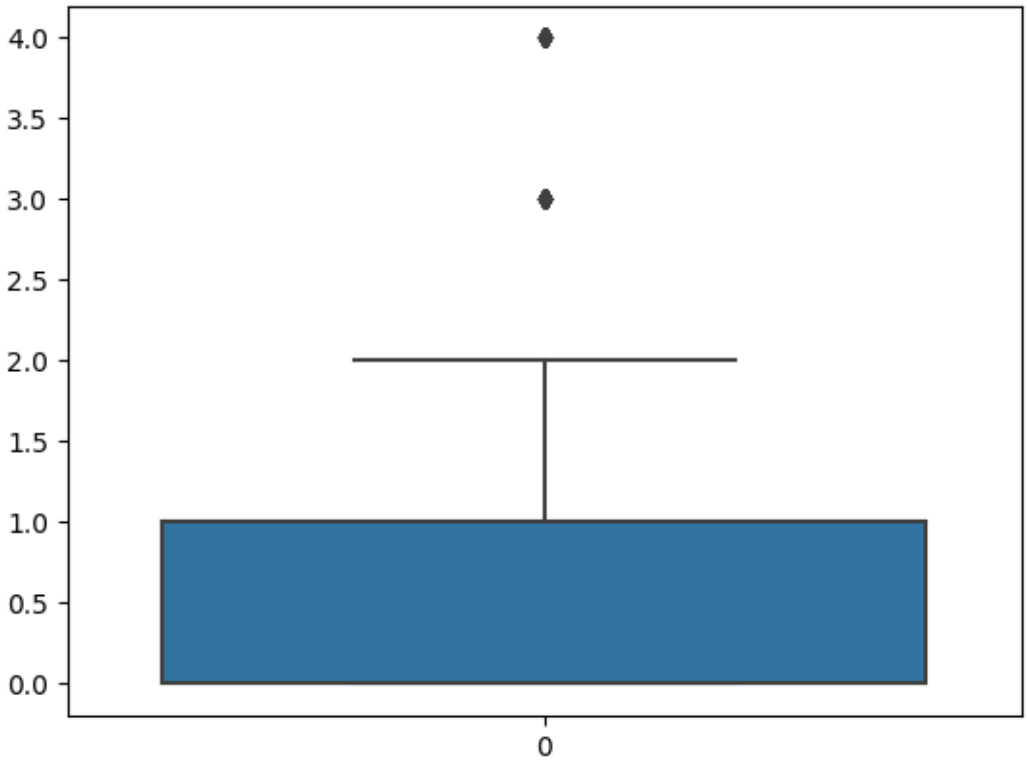
[54]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[54]: 5.0

[55]:
```
df = df[df.YearsSinceLastPromotion<upper_limit]
```

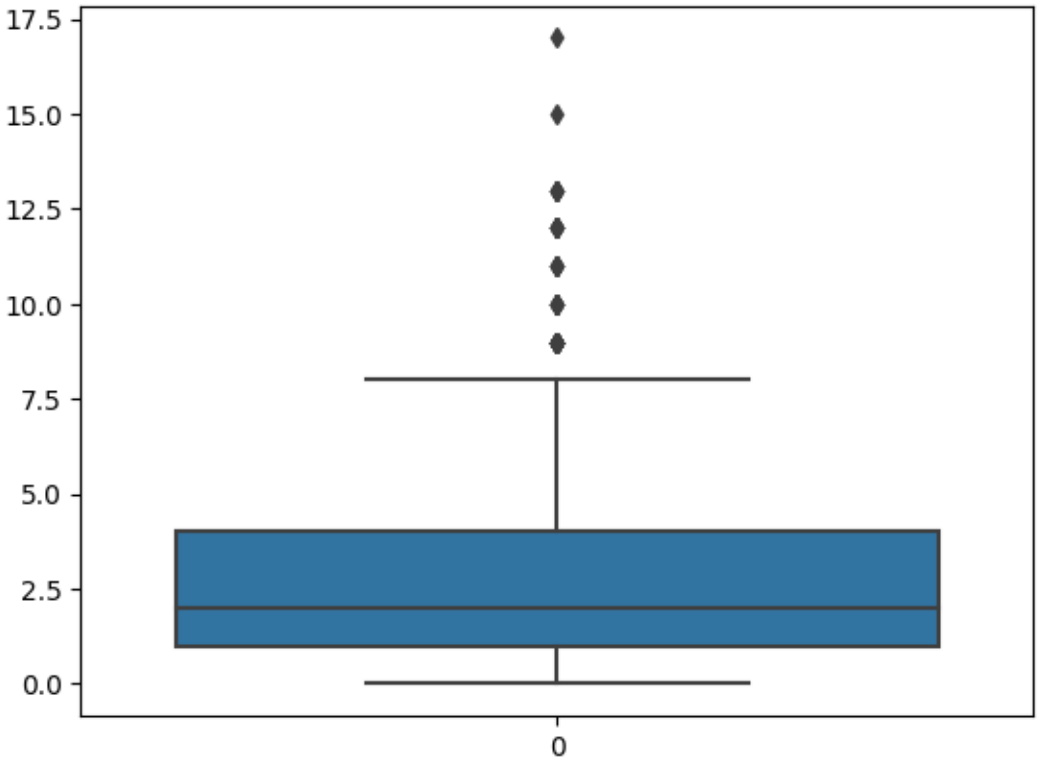
[56]:
```
sns.boxplot(df.YearsSinceLastPromotion)
```

[56]: <Axes: >



[57]:
```
sns.boxplot(df.YearsWithCurrManager)
```

[57]: <Axes: >



[58]:
```
q1 = df.YearsWithCurrManager.quantile(0.25)
q3 = df.YearsWithCurrManager.quantile(0.75)
```

[59]:
```
IQR =q3-q1
```

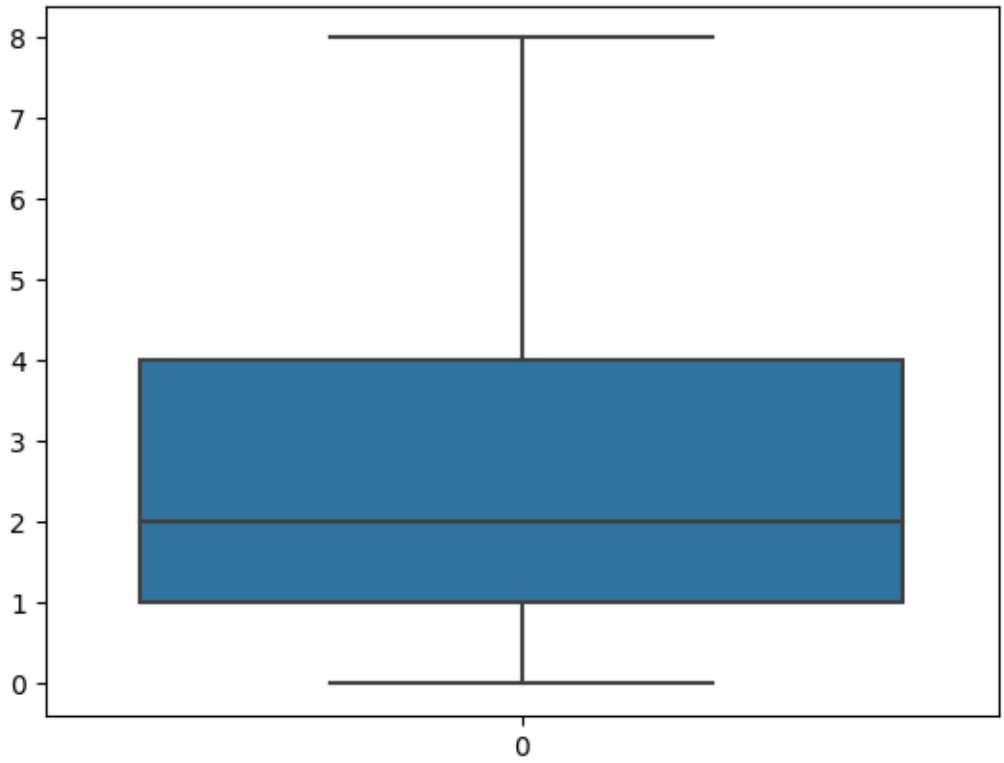
[60]:
```
upper_limit =q3+1.5*IQR
upper_limit
```

[60]: 8.5

[61]:
```
df = df[df.YearsWithCurrManager<upper_limit]
```

[62]:
```
sns.boxplot(df.YearsWithCurrManager)
```

[62]: <Axes: >



[67]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

[68]:
```
columns=["Attrition","Over18"]
df[columns]=df[columns].apply(le.fit_transform)
```

[70]:
```
df.head()
```

[70]:

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

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

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

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

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

[5 rows x 35 columns]
```

[79]:
```
y = df.iloc[:, 1]
X = df
X.drop('Attrition',axis = 1, inplace = True)
```

[81]:
```
x.head()
```

[81]:
```
   Age  Attrition  DailyRate  DistanceFromHome  Education  EmployeeCount  \
0   41          1       1102                 1          2              1
1   49          0        279                 8          1              1
2   37          1       1373                 2          2              1
3   33          0       1392                 3          4              1
5   32          0       1005                 2          2              1

   EmployeeNumber  EnvironmentSatisfaction  HourlyRate  JobInvolvement  …  \
0               1                        2          94               3  …
1               2                        3          61               2  …
2               4                        4          92               2  …
3               5                        4          56               3  …
5               8                        4          79               3  …

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

```
3                        3              80                 0
5                        3              80                 0

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

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

[5 rows x 28 columns]
```

[80]:
```
y.head()
```

[80]:
```
0    1
1    0
2    1
3    0
5    0
Name: Attrition, dtype: int64
```

[84]:
```
x.shape
```

[84]:
```
(970, 28)
```

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

[85]:
```
(970,)
```

[86]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.
 ↪2,random_state = 47)
```

[87]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[87]:
```
((776, 28), (194, 28), (776,), (194,))
```

[88]:
```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
```

[89]:
```
x_train = sc.fit_transform(x_train)
test = sc.fit_transform(x_test)
x_train
```

[89]:
```
array([[-1.04011092, -0.46508165,  1.4281238 , …,  1.58657549,
        -0.7843004 ,  1.72232723],
       [-0.68558995, -0.46508165, -0.01338251, …,  1.58657549,
         1.30000477,  2.13356598],
       [ 0.96884124, -0.46508165, -0.62938954, …,  1.58657549,
        -0.7843004 , -1.15634401],
       …,
       [-0.68558995, -0.46508165,  0.90938104, …, -1.16774796,
         0.25785219, -0.33386651],
       [ 0.37797296, -0.46508165,  0.98419971, …, -0.3807984 ,
        -0.7843004 ,  0.07737224],
       [-0.92193726, -0.46508165, -0.84636367, …, -0.3807984 ,
         1.30000477, -1.15634401]])
```

- Model Building

o Import the model building Libraries
o Initializing the model
o Training and testing the model
o Evaluation of Model
o Save the Model

#Using Logistic Regression

[90]:
```
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

[91]:
```
model.fit(x_train,y_train)
```

[91]:
```
LogisticRegression()
```

[92]:
```
pred = model.predict(x_test)
pred
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:432: UserWarning: X has
feature names, but LogisticRegression was fitted without feature names
  warnings.warn(
```

[92]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0,
       1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1,
```

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

[93]:
```
y_test
```

[93]:
```
1019    0
825     0
1112    1
463     1
364     0
       ..
509     0
591     1
1306    0
1285    0
1213    1
Name: Attrition, Length: 194, dtype: int64
```

[94]:
```
from sklearn.metrics import␣
↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[95]:
```
accuracy_score(y_test,pred)
```

[95]:
```
0.30927835051546393
```

[96]:
```
confusion_matrix(y_test,pred)
```

[96]:
```
array([[ 22, 129],
       [  5,  38]])
```

[97]:
```
pd.crosstab(y_test,pred)
```

[97]:
```
col_0       0    1
Attrition
0          22  129
1           5   38
```

[98]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.81      0.15      0.25       151
           1       0.23      0.88      0.36        43

    accuracy                           0.31       194
   macro avg       0.52      0.51      0.30       194
weighted avg       0.68      0.31      0.27       194
```

[99]:
```
probability=model.predict_proba(x_test)[:,1]
probability
```

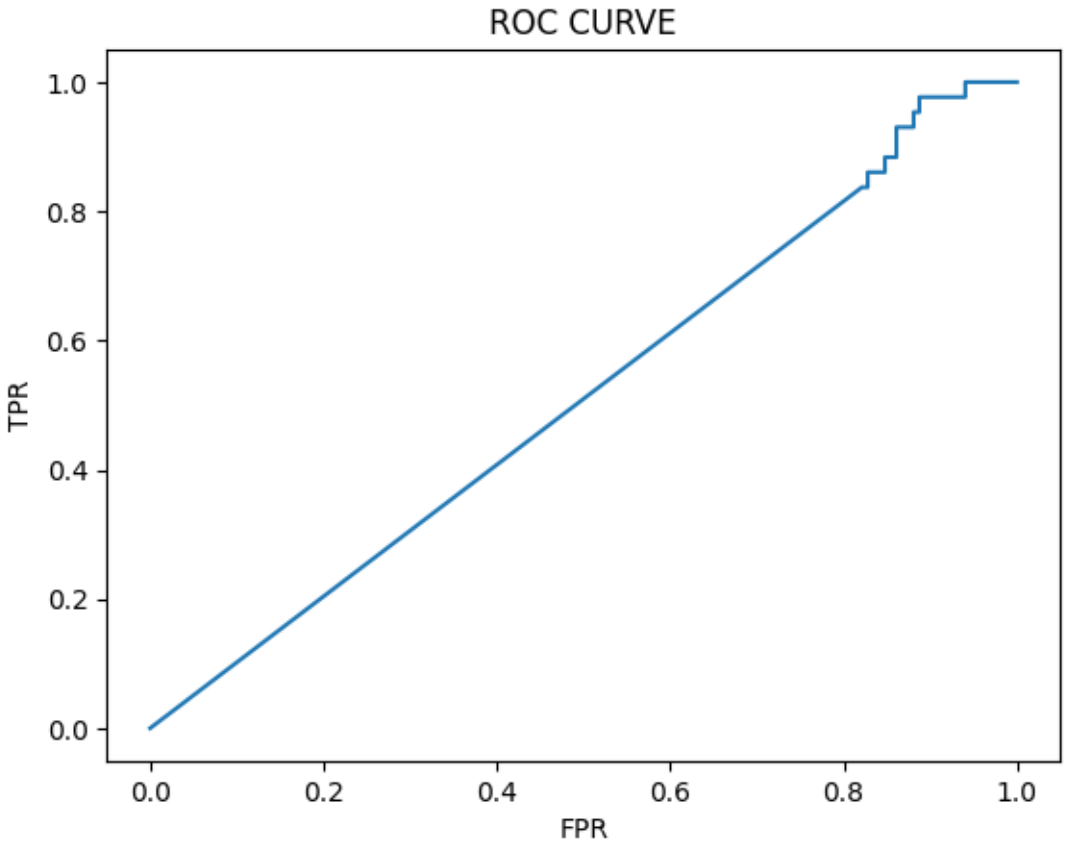
```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:432: UserWarning: X has
feature names, but LogisticRegression was fitted without feature names
  warnings.warn(
```

[99]:
```
array([1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 9.99423305e-001, 1.00000000e+000,
       3.23901917e-017, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.30163466e-069, 9.40582222e-001, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 5.36591599e-021, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 5.28165788e-075, 1.00000000e+000, 6.56757672e-024,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 9.54377873e-110, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.11827772e-065, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 5.87110155e-009,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       4.83667648e-118, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.09650752e-012, 1.00000000e+000, 3.12240891e-012, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 3.93635858e-041,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 2.48714063e-048, 8.06423554e-023,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 3.36346413e-121, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 1.00000000e+000, 4.02505017e-007, 1.00000000e+000,
       1.00000000e+000, 2.74788366e-048, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 2.66763897e-036, 1.00000000e+000, 1.00000000e+000,
       1.00000000e+000, 4.15229391e-034, 1.00000000e+000, 1.00000000e+000,
```

```
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 9.99998187e-001,
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
1.00000000e+000, 4.34236410e-045, 8.02391227e-016, 1.00000000e+000,
1.00000000e+000, 3.59117236e-041, 1.00000000e+000, 1.00000000e+000,
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 8.94221403e-106,
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
9.99775187e-001, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 2.97313538e-009,
1.00000000e+000, 1.00000000e+000, 1.00000000e+000, 1.00000000e+000,
5.17962948e-031, 6.70618985e-065, 1.00000000e+000, 1.91126784e-053,
1.00000000e+000, 1.00000000e+000])
```

[100]:
```
fpr,tpr,threshsholds = roc_curve(y_test,probability)
```

[101]:
```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show()
```



#Using Decision Tree

[106]:
```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier()
```

[107]:
```
dtc.fit(x_train,y_train)
```

[107]:
```
DecisionTreeClassifier()
```

[108]:
```
pred1 = dtc.predict(x_test)
pred1
```

[108]:
```
array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1,
       0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0])
```

[109]:
```
y_test
```

[109]:
```
1419    0
434     0
17      0
1056    1
700     1
       ..
885     0
744     1
23      0
644     0
765     0
Name: Attrition, Length: 194, dtype: int64
```

[110]:
```
from sklearn.metrics import␣
↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[111]:
```
accuracy_score(y_test,pred)
```

[111]:
```
0.29896907216494845
```

[112]:
```
confusion_matrix(y_test,pred)
```

[112]:
```
array([[ 23, 132],
       [  4,  35]])
```

[113]:
```
pd.crosstab(y_test,pred)
```

[113]:
```
col_0       0    1
Attrition
0          23  132
1           4   35
```

[114]:
```
print(classification_report(y_test,pred))
```

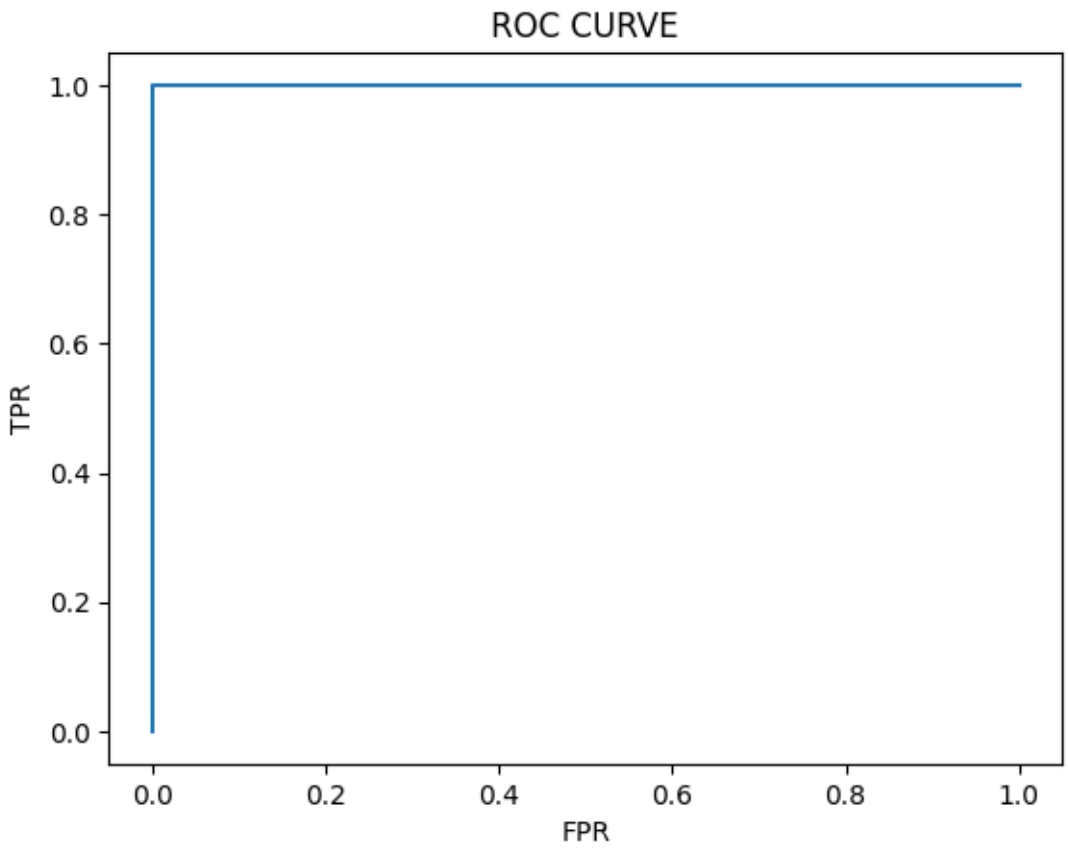
```
              precision    recall  f1-score   support

           0       0.85      0.15      0.25       155
           1       0.21      0.90      0.34        39

    accuracy                           0.30       194
   macro avg       0.53      0.52      0.30       194
weighted avg       0.72      0.30      0.27       194
```

[115]:
```
probability=dtc.predict_proba(x_test)[:,1]
probability
```

[115]:
```
array([0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 1., 0.,
       0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0.,
       0., 0., 0., 0., 1., 0., 0., 1., 1., 0., 0., 0., 1., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 1., 0., 0., 1., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 1., 0., 1., 0., 1., 0., 0., 0., 1., 0., 1., 0., 0., 0.,
       1., 0., 0., 0., 0., 1., 0., 1., 1., 0., 1., 1., 0., 0., 0., 1., 0.,
       1., 0., 0., 0., 0., 1., 0., 0., 0., 1., 0., 1., 1., 0., 0., 0., 1.,
       1., 0., 0., 1., 0., 0., 0.])
```

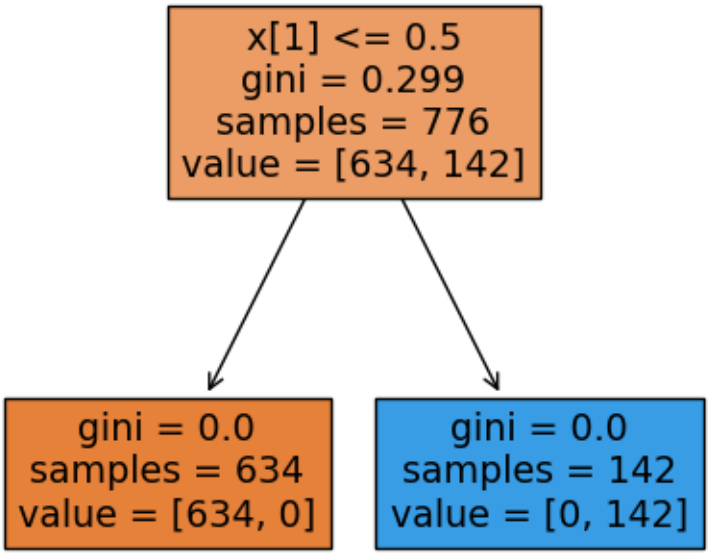
[116]:
```
fpr,tpr,threshsholds = roc_curve(y_test,probability)
```

[117]:
```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show()
```



[120]:
```
from sklearn import tree
plt.figure(figsize=(5,5))
tree.plot_tree(dtc,filled=True)
```

[120]:
```
[Text(0.5, 0.75, 'x[1] <= 0.5\ngini = 0.299\nsamples = 776\nvalue = [634,
142]'),
 Text(0.25, 0.25, 'gini = 0.0\nsamples = 634\nvalue = [634, 0]'),
 Text(0.75, 0.25, 'gini = 0.0\nsamples = 142\nvalue = [0, 142]')]
```



[121]:
```
from sklearn.model_selection import GridSearchCV
parameter={
 'criterion':['gini','entropy'],
  'splitter':['best','random'],
  'max_depth':[1,2,3,4,5],
  'max_features':['auto', 'sqrt', 'log2']
}
```

[122]:
```
grid_search=GridSearchCV(estimator=dtc,param_grid=parameter,cv=5,scoring="accuracy")
```

[123]:
```
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(
/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(
```

[123]:
```
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')
```

[124]:
```
grid_search.best_params_
```

[124]:
```
{'criterion': 'entropy',
 'max_depth': 4,
 'max_features': 'auto',
 'splitter': 'random'}
```

[125]:
```
dtc_cv=DecisionTreeClassifier(criterion= 'entropy',
 max_depth=3,
 max_features='sqrt',
 splitter='best')
dtc_cv.fit(x_train,y_train)
```

[125]:
```
DecisionTreeClassifier(criterion='entropy', max_depth=3, max_features='sqrt')
```

[126]:
```
pred = dtc_cv.predict(x_test)
```

[127]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.89      1.00      0.94       155
           1       1.00      0.51      0.68        39

    accuracy                           0.90       194
   macro avg       0.95      0.76      0.81       194
```

```
weighted avg       0.91      0.90      0.89       194
```

#Using Random Forest

[128]:
```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
```

[129]:
```
forest_params = [{'max_depth': list(range(10, 15)), 'max_features':␣
 ↪list(range(0,14))}]
```

[130]:
```
rfc_cv = GridSearchCV(rfc,param_grid=forest_params,cv=10,scoring="accuracy")
```

[131]:
```
rfc_cv.fit(x_train,y_train)
```

```
/usr/local/lib/python3.10/dist-
packages/sklearn/model_selection/_validation.py:378: FitFailedWarning:
50 fits failed out of a total of 700.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
50 fits failed with the following error:
Traceback (most recent call last):
  File "/usr/local/lib/python3.10/dist-
packages/sklearn/model_selection/_validation.py", line 686, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py",
line 340, in fit
    self._validate_params()
  File "/usr/local/lib/python3.10/dist-packages/sklearn/base.py", line 600, in
_validate_params
    validate_parameter_constraints(
  File "/usr/local/lib/python3.10/dist-
packages/sklearn/utils/_param_validation.py", line 97, in
validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features'
parameter of RandomForestClassifier must be an int in the range [1, inf), a
float in the range (0.0, 1.0], a str among {'sqrt', 'auto' (deprecated), 'log2'}
or None. Got 0 instead.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_search.py:952:
UserWarning: One or more of the test scores are non-finite: [       nan
```

```
0.88403263 0.98196803 0.9987013  1.         1.
 1.         1.         1.         1.         1.         1.
 1.         1.                nan 0.89302364 0.96648352 1.
 1.         1.         1.         1.         1.         1.
 1.         1.         1.         1.                nan 0.90331335
 0.97805528 1.         1.         1.         1.         1.
 1.         1.         1.         1.         1.         1.
        nan 0.88919414 0.97552448 1.         1.         1.
 1.         1.         1.         1.         1.         1.
 1.         1.                nan 0.9032634  0.98198468 1.
 1.         1.         1.         1.         1.         1.
 1.         1.         1.         1.        ]
  warnings.warn(
```

[131]:
```
GridSearchCV(cv=10, estimator=RandomForestClassifier(),
             param_grid=[{'max_depth': [10, 11, 12, 13, 14],
                          'max_features': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
                                           12, 13]}],
             scoring='accuracy')
```

[132]:
```
pred = rfc_cv.predict(x_test)
```

[134]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       155
           1       1.00      1.00      1.00        39

    accuracy                           1.00       194
   macro avg       1.00      1.00      1.00       194
weighted avg       1.00      1.00      1.00       194
```

[ ]:
```

```