assignment-22-sep

September 27, 2023

[71]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[72]:
```
df=pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
```

[73]:
```
df.head()
```

[73]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	…	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	…	1	80	0	8	0	1	6
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	…	4	80	1	10	3	3	10
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	…	2	80	0	7	3	3	0
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	…	3	80	0	8	3	3	8
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	…	4	80	1	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]
```

[74]:
```
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
```

[75]:
```
df.shape
```

[75]:
```
(1470, 35)
```

[76]:
```
df.        Attrition.value_counts()
```

[76]:
```
No     1233
Yes     237
Name: Attrition, dtype: int64
```

[77]:
```
df.corr()
```

```
<ipython-input-77-2f6f6606aa2c>: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()
```

[77]:

	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]
```

[78]:
```
df.isnull().any()
```

[78]:
```
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
```

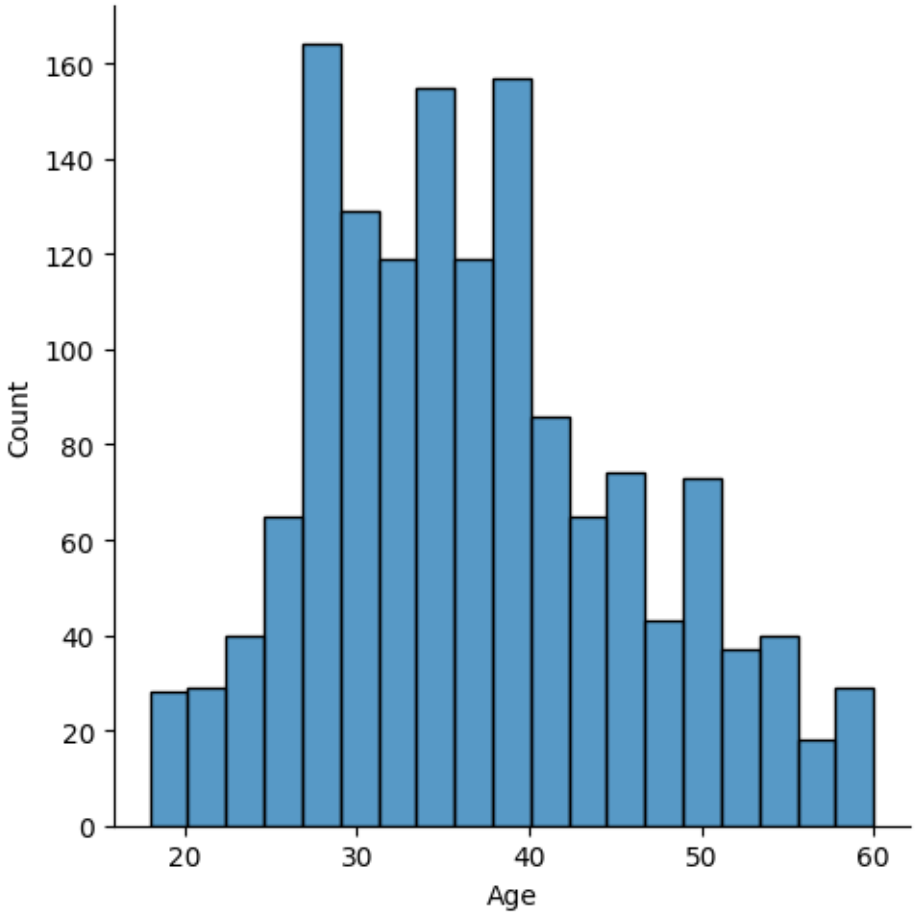
```
WorkLifeBalance             False
YearsAtCompany              False
YearsInCurrentRole          False
YearsSinceLastPromotion     False
YearsWithCurrManager        False
dtype: bool
```

[79]:
```
df.isnull().sum()
```

[79]:
```
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
```

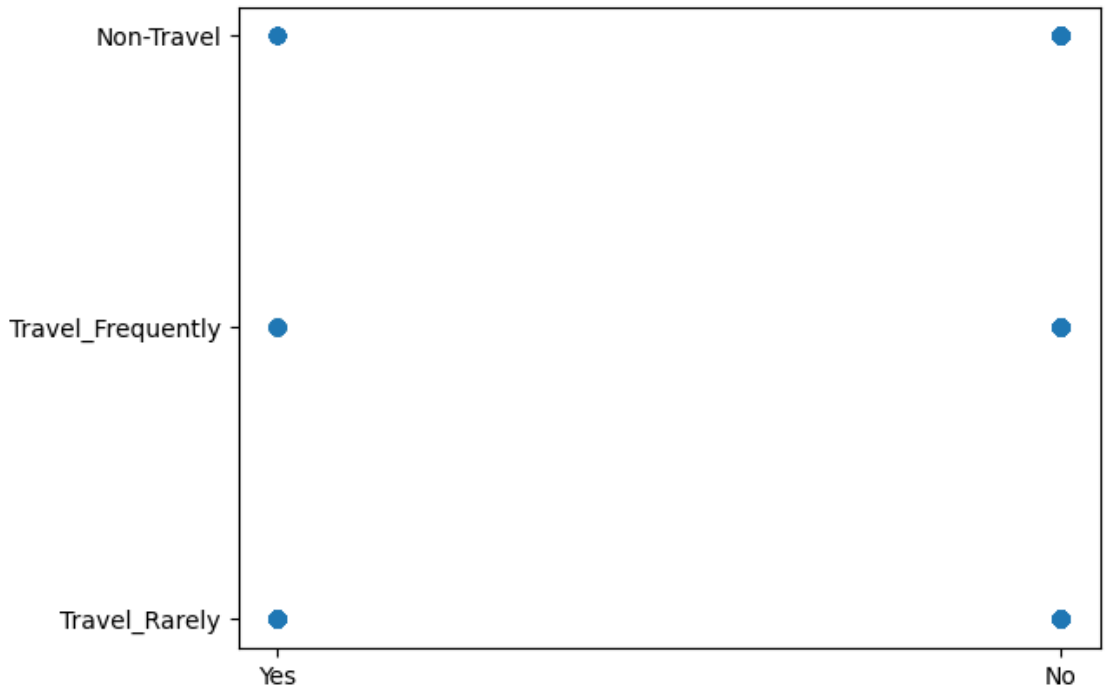
[80]:
```
sns.displot(df["Age"])
```

[80]: <seaborn.axisgrid.FacetGrid at 0x7ec537403ee0>



[81]:
```
plt.scatter(df['Attrition'],df['BusinessTravel'])
```

[81]: <matplotlib.collections.PathCollection at 0x7ec5338caad0>



[82]:
```
plt.scatter(df['Attrition'],df['DistanceFromHome'])
```

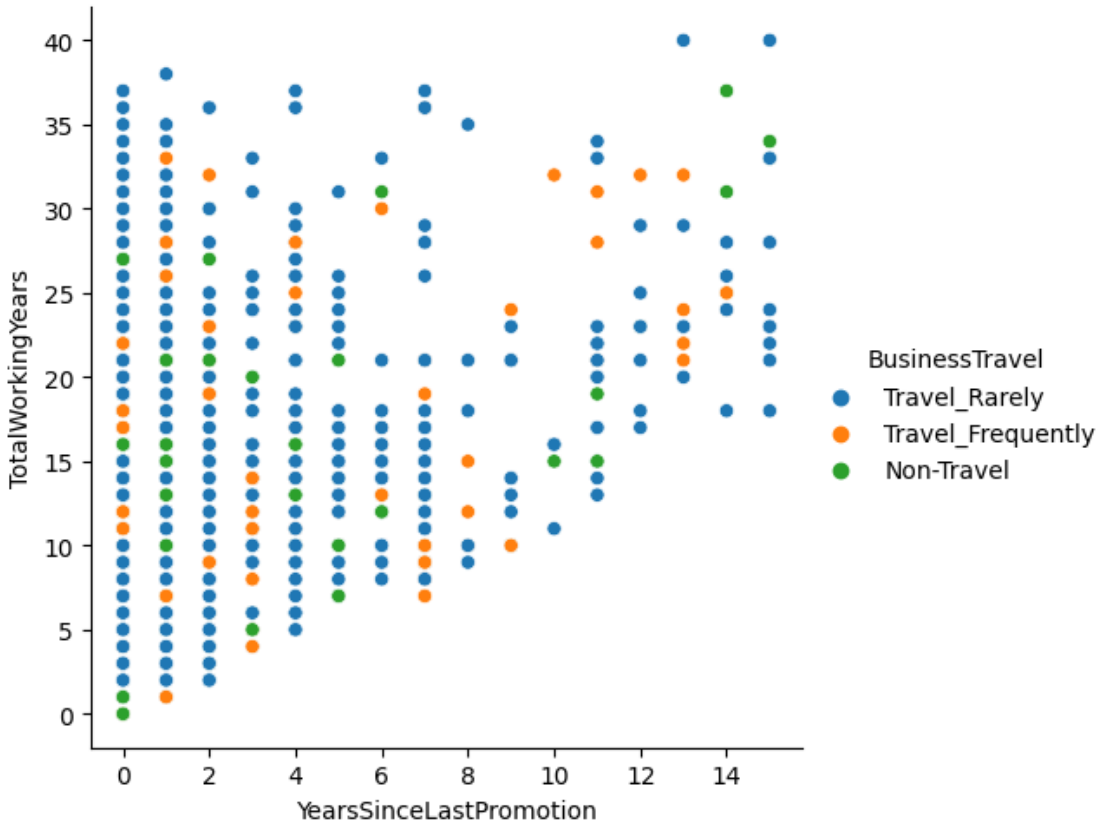
[82]: <matplotlib.collections.PathCollection at 0x7ec533915810>

[83]:
```
plt.scatter(df['Attrition'],df['StandardHours'])
```

[83]: <matplotlib.collections.PathCollection at 0x7ec533947550>

[84]:
```
sns.
↪relplot(x="YearsSinceLastPromotion",y="TotalWorkingYears",data=df,hue="BusinessTravel")
```

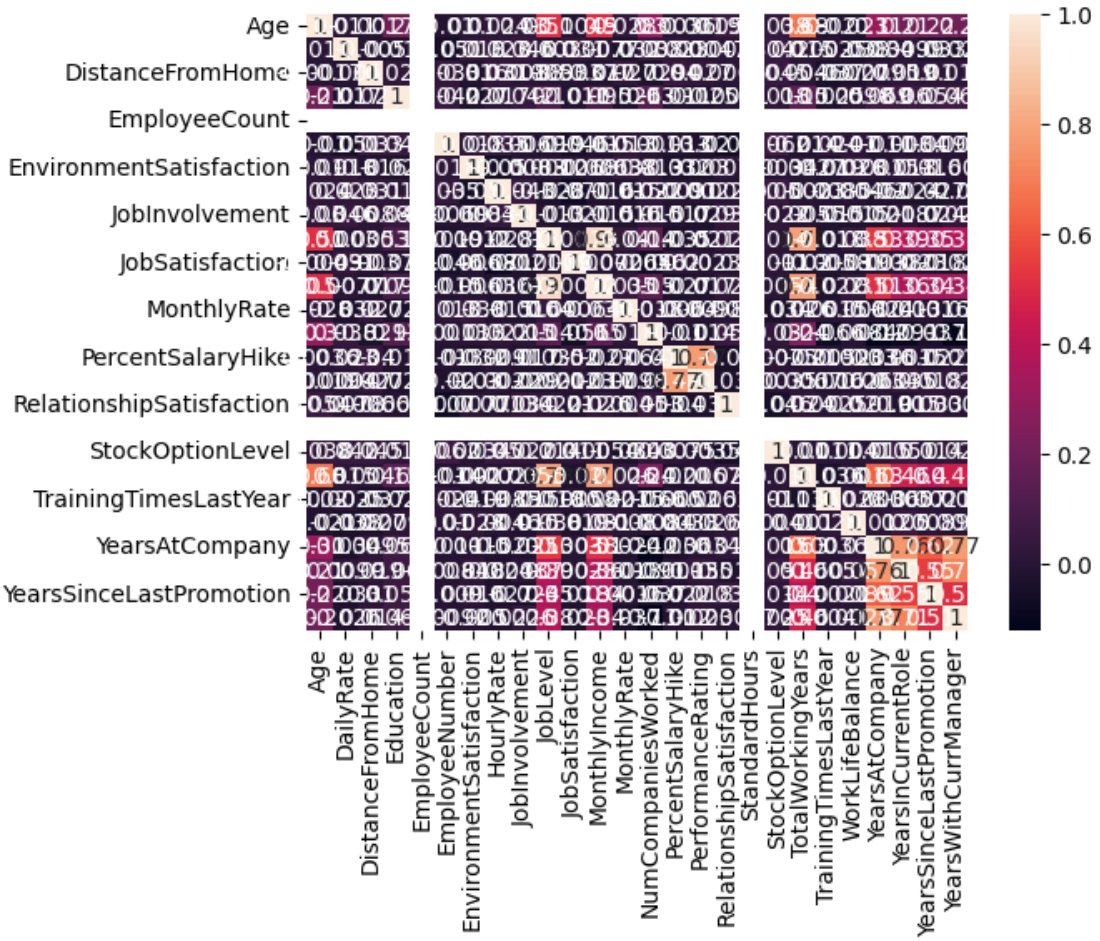
[84]: <seaborn.axisgrid.FacetGrid at 0x7ec533a46fe0>



[85]:
```
sns.heatmap(df.corr(),annot=True)
```

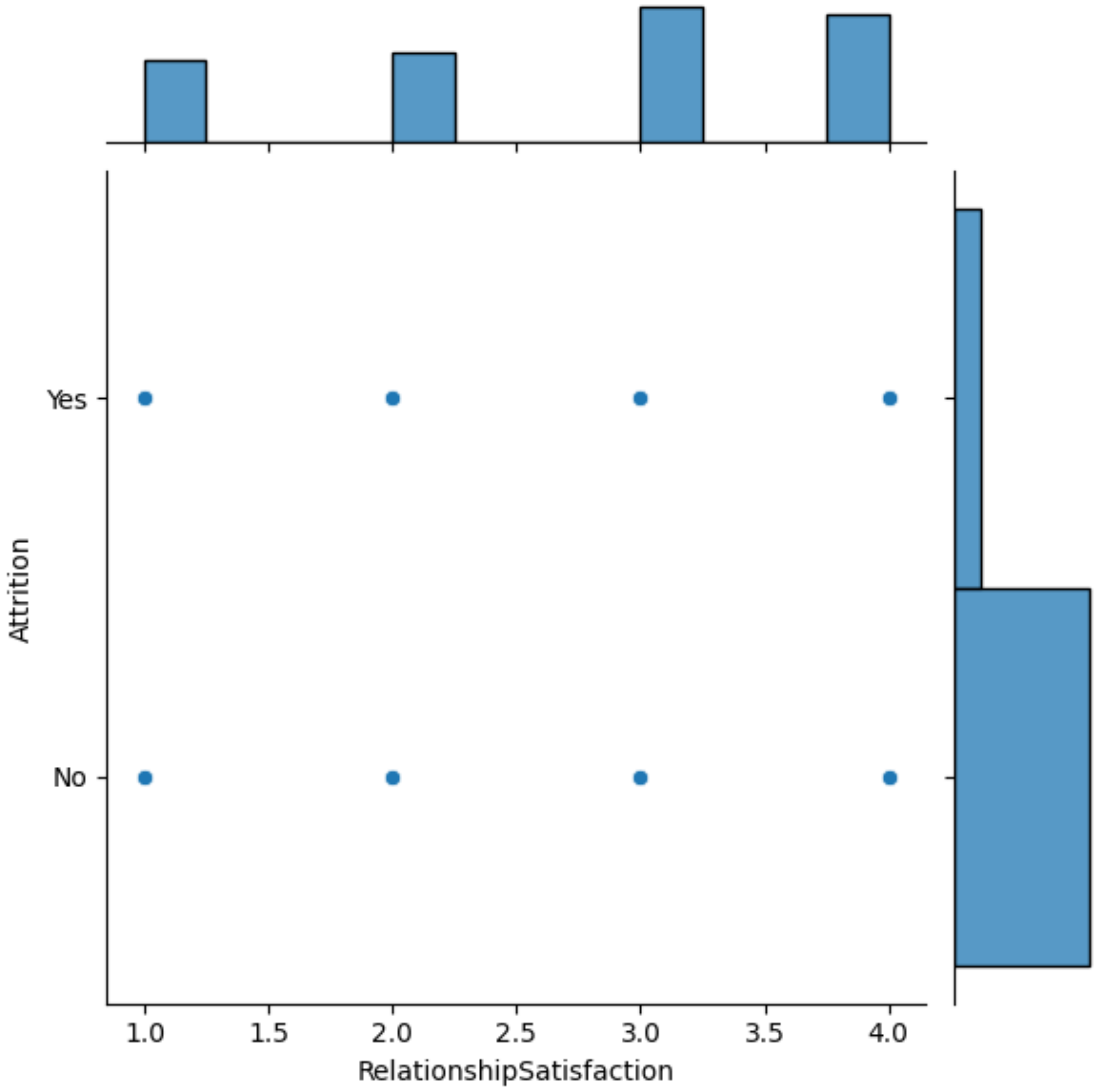
```
<ipython-input-85-8df7bcac526d>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr(),annot=True)
```

[85]: <Axes: >



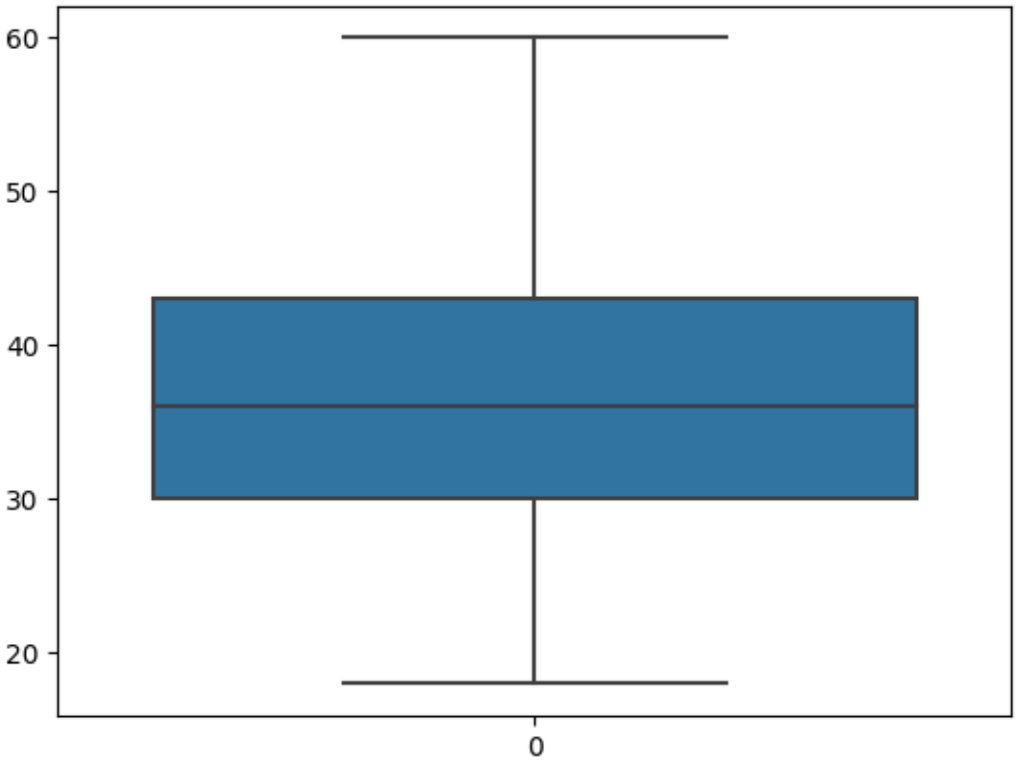
[86]:
```
sns.jointplot(x="RelationshipSatisfaction",y="Attrition",data=df)
```

[86]: <seaborn.axisgrid.JointGrid at 0x7ec53379d8a0>



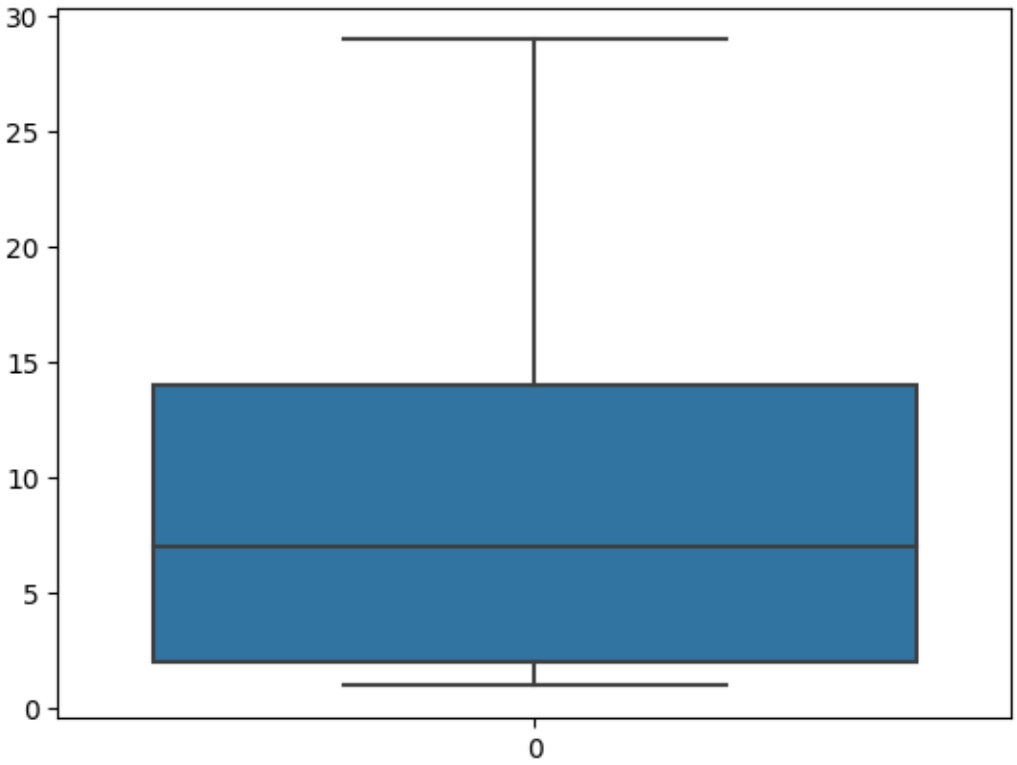
[87]:
```
sns.boxplot(df.Age)
```

[87]: <Axes: >



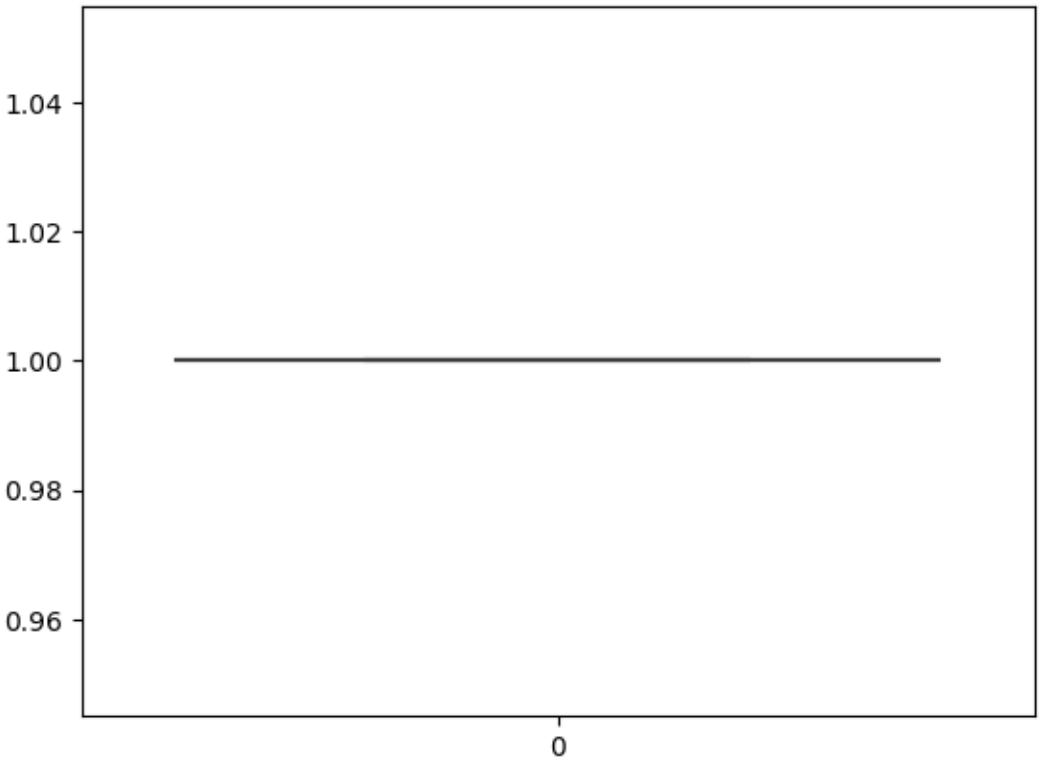
[88]:
```
sns.boxplot(df.DistanceFromHome)
```

[88]: <Axes: >



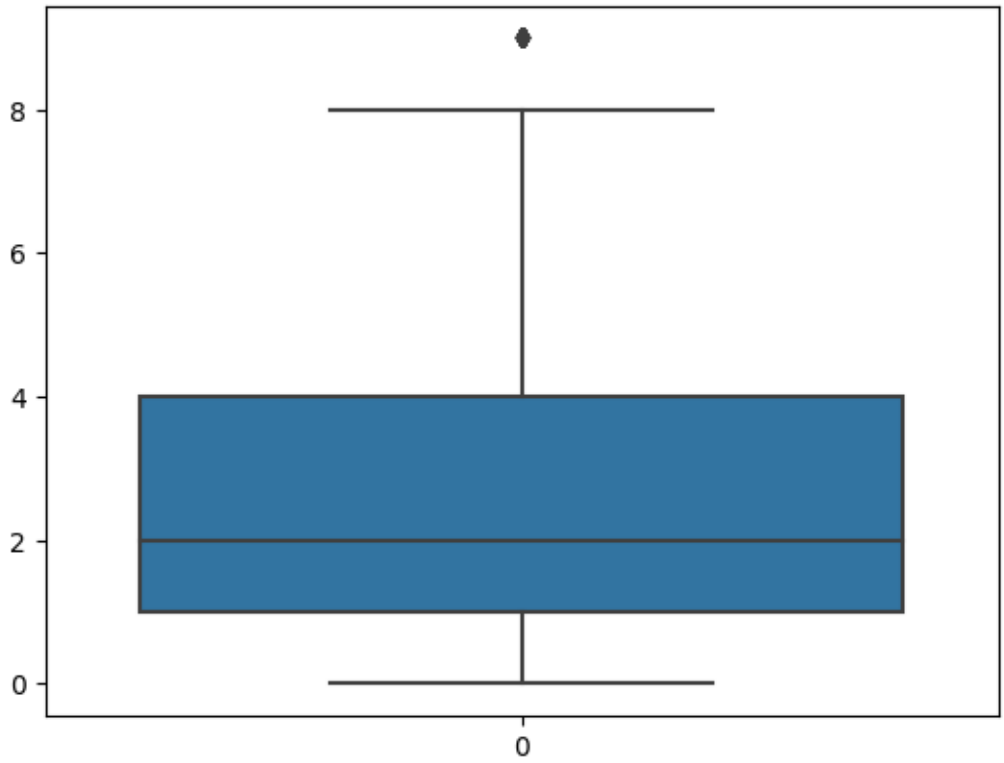
[89]:
```
sns.boxplot(df.EmployeeCount )
```

[89]: <Axes: >



[90]:
```
sns.boxplot(df.NumCompaniesWorked)
```

[90]: <Axes: >



[91]:
```
q1=df.NumCompaniesWorked.quantile(0.25)
q3=df.NumCompaniesWorked.quantile(0.75)
```

[92]:
```
print(q1)
print(q3)
```

```
1.0
4.0
```

[93]:
```
IQR=q3-q1
```

[94]:
```
IQR
```

[94]: 3.0

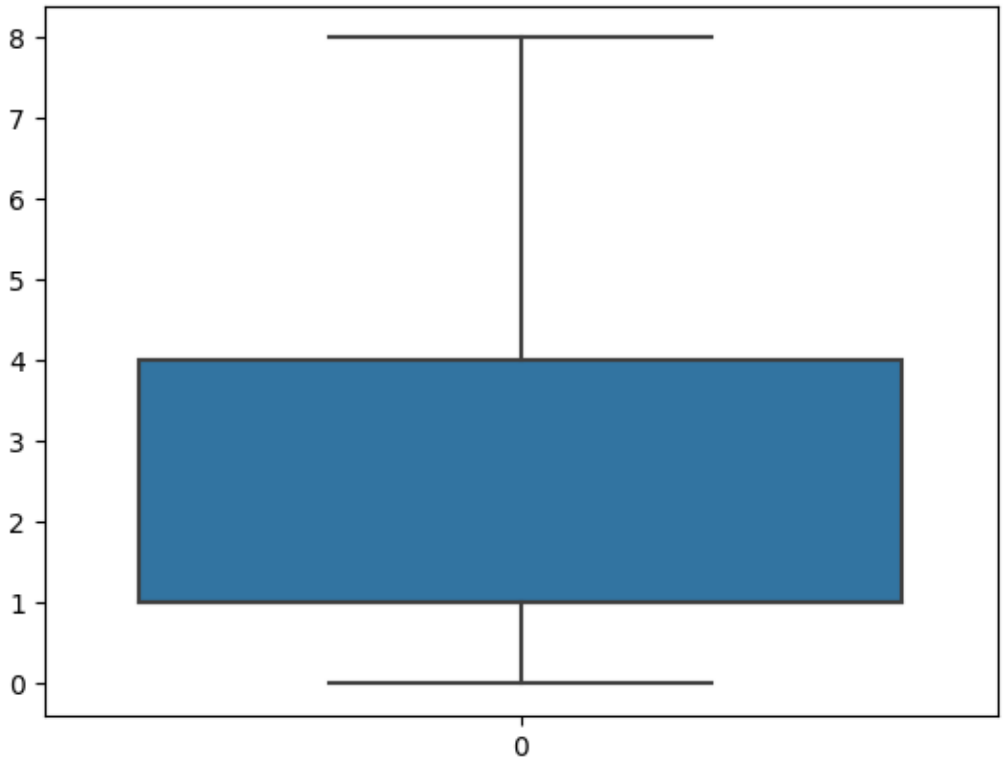
[95]:
```
upper_limit=q3+1.5*IQR
upper_limit
```

[95]: 8.5

[96]:
```
df=df[df.NumCompaniesWorked<upper_limit]
```

[97]:
```
sns.boxplot(df.NumCompaniesWorked)
```

[97]: <Axes: >



[145]:
```
#dependent variable
y=df.Attrition
```

[146]:
```
y.head()
```

[146]:
```
0    Yes
1     No
2    Yes
3     No
5     No
Name: Attrition, dtype: object
```

[147]:
```
#independent varible
x=df.drop(["Attrition"],axis=1)
```

[148]:
```
x.head()
```

[148]:
```
   Age     BusinessTravel  DailyRate              Department  \
0   41      Travel_Rarely       1102                   Sales
1   49  Travel_Frequently        279  Research & Development
2   37      Travel_Rarely       1373  Research & Development
3   33  Travel_Frequently       1392  Research & Development
5   32  Travel_Frequently       1005  Research & Development

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

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

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

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
5                     6

[5 rows x 34 columns]
```

[149]:
```
x.shape
```

[149]:
```
(1418, 34)
```

[150]:
```
y.shape
```

[150]:
```
(1418,)
```

[151]:
```
df.head()
```

[151]:

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

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

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

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

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

[5 rows x 35 columns]

[152]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x["BusinessTravel"]=le.fit_transform(x["BusinessTravel"])
```

[153]:
```
x["BusinessTravel"]
```

[153]:
```
0       2
1       1
2       2
3       1
5       1
       ..
1465    1
1466    2
1467    2
1468    1
1469    2
Name: BusinessTravel, Length: 1418, dtype: int64
```

[154]:
```
x.head()
```

[154]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	…	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion
0	41	2	1102	Sales	1	2	Life Sciences	1	1	2	…	1	80	0	8	0	1	6	4	0
1	49	1	279	Research & Development	8	1	Life Sciences	1	2	3	…	4	80	1	10	3	3			
2	37	2	1373	Research & Development	2	2	Other	1	4	4	…	2	80	0	7	3	3			
3	33	1	1392	Research & Development	3	4	Life Sciences	1	5	4	…	3	80	0	8	3	3			
5	32	1	1005	Research & Development	2	2	Life Sciences	1	8	4	…	3	80	0	8	2	2			

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
5                     6

[5 rows x 34 columns]
```

[155]:
```
x["Department"]=le.fit_transform(x["Department"])
```

[156]:
```
x["Department"]
```

[156]:
```
0       2
1       1
2       1
3       1
5       1
       ..
1465    1
1466    1
1467    1
1468    2
1469    1
Name: Department, Length: 1418, dtype: int64
```

[157]:
```
x.head()
```

[157]:
```
   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41               2       1102           2                 1          2
1   49               1        279           1                 8          1
2   37               2       1373           1                 2          2
3   33               1       1392           1                 3          4
5   32               1       1005           1                 2          2

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

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

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

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

[5 rows x 34 columns]
```

[158]:
```
x["EducationField"]=le.fit_transform(x["EducationField"])
```

[159]:
```
x["EducationField"]
```

[159]:
```
0       1
1       1
2       4
3       1
5       1
       ..
1465    3
1466    3
1467    1
1468    3
1469    3
Name: EducationField, Length: 1418, dtype: int64
```

[160]:
```
x.head()
```

[160]:
```
   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41               2       1102           2                 1          2
1   49               1        279           1                 8          1
2   37               2       1373           1                 2          2
3   33               1       1392           1                 3          4
```

```
5   32               1       1005           1                  2           2

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

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

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

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

[5 rows x 34 columns]
```

[162]:
```
non_numeric_columns = x.select_dtypes(exclude=['number']).columns
```

[161]:
```
print(non_numeric_columns)
```

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

[163]:
```
x["Gender"]=le.fit_transform(x["Gender"])
```

[164]:
```
x["Gender"]
```

[164]:
```
0       0
1       1
2       1
3       0
```

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

[165]:
```
x.head()
```

[165]:

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

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

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

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

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

[5 rows x 34 columns]

[166]:
```
x["JobRole"]=le.fit_transform(x["JobRole"])
```

[167]:
```
x["JobRole"]
```

[167]:
```
0       7
1       6
2       2
3       6
5       2
       ..
1465    2
1466    0
1467    4
1468    7
1469    2
Name: JobRole, Length: 1418, dtype: int64
```

[168]:
```
x.head()
```

[168]:

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

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

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

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

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

[5 rows x 34 columns]
```

[169]:
```
x["MaritalStatus"]=le.fit_transform(x["MaritalStatus"])
```

[170]:
```
x["MaritalStatus"]
```

[170]:
```
0       2
1       1
2       2
3       1
5       2
       ..
1465    1
1466    1
1467    1
1468    1
1469    1
Name: MaritalStatus, Length: 1418, dtype: int64
```

[171]:
```
x.head()
```

[171]:
```
   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41               2       1102           2                 1          2
1   49               1        279           1                 8          1
2   37               2       1373           1                 2          2
3   33               1       1392           1                 3          4
5   32               1       1005           1                 2          2

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

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

```
5  …                        3             80                  0

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

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

[5 rows x 34 columns]
```

[172]:
```
x["Over18"]=le.fit_transform(x["Over18"])
```

[173]:
```
x["Over18"]
```

[173]:
```
0       0
1       0
2       0
3       0
5       0
       ..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Over18, Length: 1418, dtype: int64
```

[174]:
```
x.head()
```

[174]:
```
   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41               2       1102           2                 1          2
1   49               1        279           1                 8          1
2   37               2       1373           1                 2          2
3   33               1       1392           1                 3          4
5   32               1       1005           1                 2          2

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

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

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

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

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

[5 rows x 34 columns]
```

[175]:
```
x["OverTime"]=le.fit_transform(x["OverTime"])
```

[176]:
```
x["OverTime"]
```

[176]:
```
0       1
1       0
2       1
3       1
5       0
       ..
1465    0
1466    0
1467    1
1468    0
1469    0
Name: OverTime, Length: 1418, dtype: int64
```

[177]:
```
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled = pd.DataFrame(ms.fit_transform(x), columns=x.columns)
```

[178]:
```
x_scaled
```

[178]:

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

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

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

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

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

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

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

[1418 rows x 34 columns]
```

[179]:
```
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)
```

[180]:
```
print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)
```

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

[181]:
```
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

[182]:
```
model.fit(x_train,y_train)
```

```
[182]: LogisticRegression()
```

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

```
[184]: pred
```

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

```
[185]: y_test
```

```
[185]: 451     No
       639     No
       832     No
       1287    No
       1277    No
               ..
       521     No
       550     No
       1113    No
       335     No
       917     No
       Name: Attrition, Length: 284, dtype: object
```

[186]:
```
df
```

[186]:
```
      Age Attrition     BusinessTravel  DailyRate              Department  \
0      41       Yes      Travel_Rarely       1102                   Sales
1      49        No  Travel_Frequently        279  Research & Development
2      37       Yes      Travel_Rarely       1373  Research & Development
3      33        No  Travel_Frequently       1392  Research & Development
5      32        No  Travel_Frequently       1005  Research & Development
…   …       …                   …        …                       …
1465   36        No  Travel_Frequently        884  Research & Development
1466   39        No      Travel_Rarely        613  Research & Development
1467   27        No      Travel_Rarely        155  Research & Development
1468   49        No  Travel_Frequently       1023                   Sales
1469   34        No      Travel_Rarely        628  Research & Development

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

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

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

```
…                …                …                      …
1465               1               17                     3
1466               1                9                     5
1467               1                6                     0
1468               0               17                     3
1469               0                6                     3

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

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

[1418 rows x 35 columns]
```

[190]:
```
from sklearn.metrics import␣
↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[191]:
```
accuracy_score(y_test,pred)
```

[191]:
```
0.8697183098591549
```

[192]:
```
confusion_matrix(y_test,pred)
```

[192]:
```
array([[237,   3],
       [ 34,  10]])
```

[193]:
```
pd.crosstab(y_test,pred)
```

[193]:
```
col_0       No  Yes
Attrition
No         237    3
Yes         34   10
```

[194]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

          No       0.87      0.99      0.93       240
         Yes       0.77      0.23      0.35        44

    accuracy                           0.87       284
   macro avg       0.82      0.61      0.64       284
weighted avg       0.86      0.87      0.84       284
```

[195]:
```
#ROC_AUC Curve

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

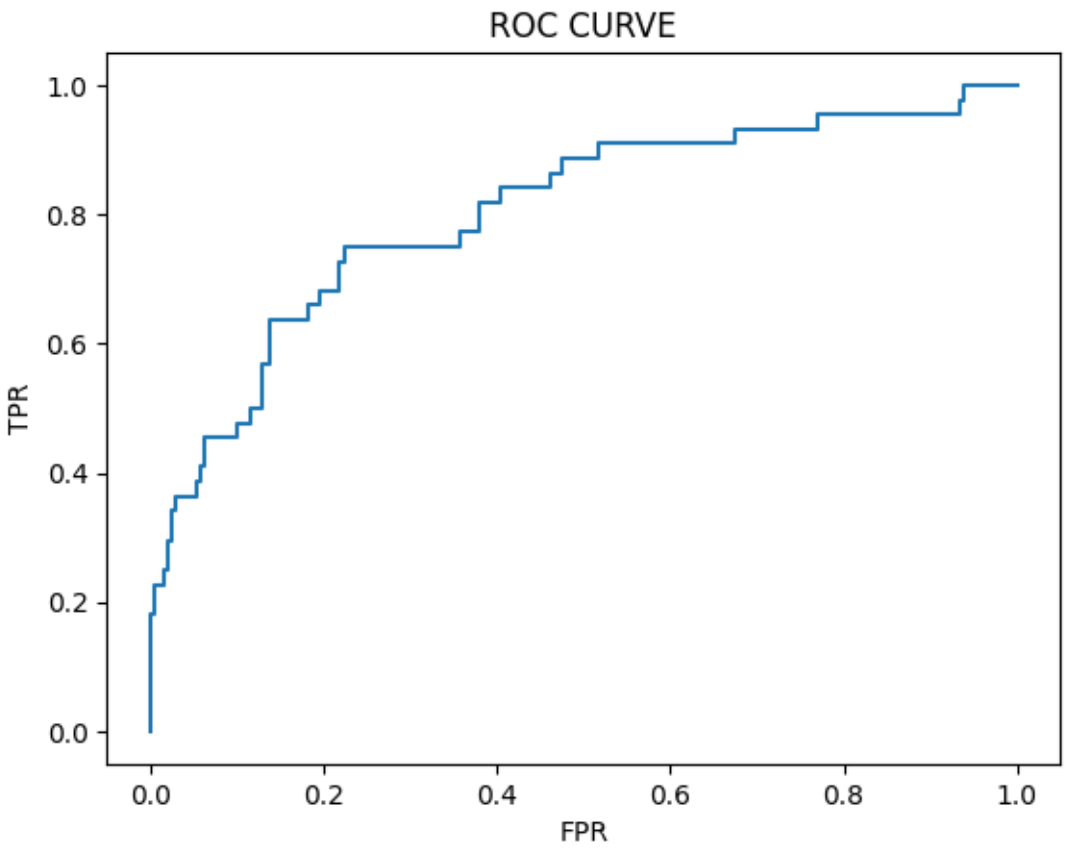
[195]:
```
array([0.13657895, 0.03742004, 0.08053736, 0.08659374, 0.023358  ,
       0.10563069, 0.13815154, 0.00229225, 0.06771379, 0.12744425,
       0.08172802, 0.05965762, 0.0638561 , 0.12855128, 0.2275486 ,
       0.08936636, 0.06240484, 0.09603478, 0.21199145, 0.05717384,
       0.01180209, 0.00367791, 0.07898725, 0.02473968, 0.11962886,
       0.12904799, 0.0184306 , 0.0365714 , 0.02049336, 0.10008116,
       0.16143025, 0.03099261, 0.05571065, 0.04469354, 0.21600549,
       0.42230677, 0.2197372 , 0.5227653 , 0.18101958, 0.10182865,
       0.03088844, 0.18054679, 0.08248226, 0.01733578, 0.19733818,
       0.06725397, 0.01197982, 0.01366601, 0.02702768, 0.18659878,
       0.04323244, 0.00445696, 0.05192806, 0.1866853 , 0.1632088 ,
       0.27853238, 0.07437663, 0.09816652, 0.00573849, 0.00449716,
       0.0059488 , 0.03111943, 0.00839901, 0.00669404, 0.04253402,
       0.18695255, 0.19941885, 0.03278527, 0.00238087, 0.01663221,
       0.58136087, 0.1578733 , 0.21711936, 0.03898385, 0.04521495,
       0.03220001, 0.06616953, 0.19809653, 0.10991992, 0.22934288,
       0.05904098, 0.02037218, 0.66970453, 0.26829173, 0.08216447,
       0.04010601, 0.11590138, 0.27057603, 0.22694055, 0.20450222,
       0.56793147, 0.22053355, 0.36393157, 0.01755166, 0.01233427,
       0.01492107, 0.2081514 , 0.12205625, 0.40315397, 0.04856193,
       0.07330096, 0.25379683, 0.14516211, 0.28647266, 0.02781388,
       0.18391223, 0.26396952, 0.01946723, 0.28598072, 0.04347479,
       0.15563751, 0.13357455, 0.00963796, 0.02116195, 0.07528362,
```

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

[200]:
```
y_test_encoded = le.fit_transform(y_test)
```

[201]:
```
fpr, tpr,threshsholds = roc_curve (y_test_encoded,probability)
```

[202]:
```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show ()
```


ROC CURVE
TPR
FPR
0.0
0.2
0.4
0.6
0.8
1.0

[ ]: