emp-attrition-assignment

September 27, 2023

0.1 Data Collection and Data Preprocessing

0.1.1 Importing Libraries and Dataset

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

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

[3]:
```
a
```

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

0.1.2 Reading the Data Types

[4]:
```
a.dtypes
```

[4]:
```
Age                          int64
Attrition                   object
BusinessTravel              object
DailyRate                    int64
Department                  object
DistanceFromHome             int64
Education                    int64
EducationField              object
EmployeeCount                int64
EmployeeNumber               int64
EnvironmentSatisfaction      int64
Gender                      object
HourlyRate                   int64
JobInvolvement               int64
JobLevel                     int64
JobRole                     object
JobSatisfaction              int64
MaritalStatus               object
MonthlyIncome                int64
MonthlyRate                  int64
NumCompaniesWorked           int64
Over18                      object
OverTime                    object
PercentSalaryHike            int64
PerformanceRating            int64
RelationshipSatisfaction     int64
StandardHours                int64
StockOptionLevel             int64
TotalWorkingYears            int64
TrainingTimesLastYear        int64
WorkLifeBalance              int64
YearsAtCompany               int64
YearsInCurrentRole           int64
YearsSinceLastPromotion      int64
YearsWithCurrManager         int64
```

```
dtype: object
```

0.1.3 Shape of the Dataset

[5]:
```
a.shape
```

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

0.1.4 Information about the Dataset

[6]:
```
a.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
```

0.1.5 Statistics about the Dataset

[7]:
```
a.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]
```

0.1.6 Identifying Null Values

[8]:
```
a.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]:
```
a.isnull().sum()
```

[9]:
```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
OverTime                    0
PercentSalaryHike           0
```

```
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

0.2 Data Visualization

[10]:
```
d=a.corr()
d
```

```
C:\Users\sbkom\AppData\Local\Temp\ipykernel_10068\554136312.py:1: FutureWarning:
The default value of numeric_only in DataFrame.corr is deprecated. In a future
version, it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  d=a.corr()
```

[10]:

	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	

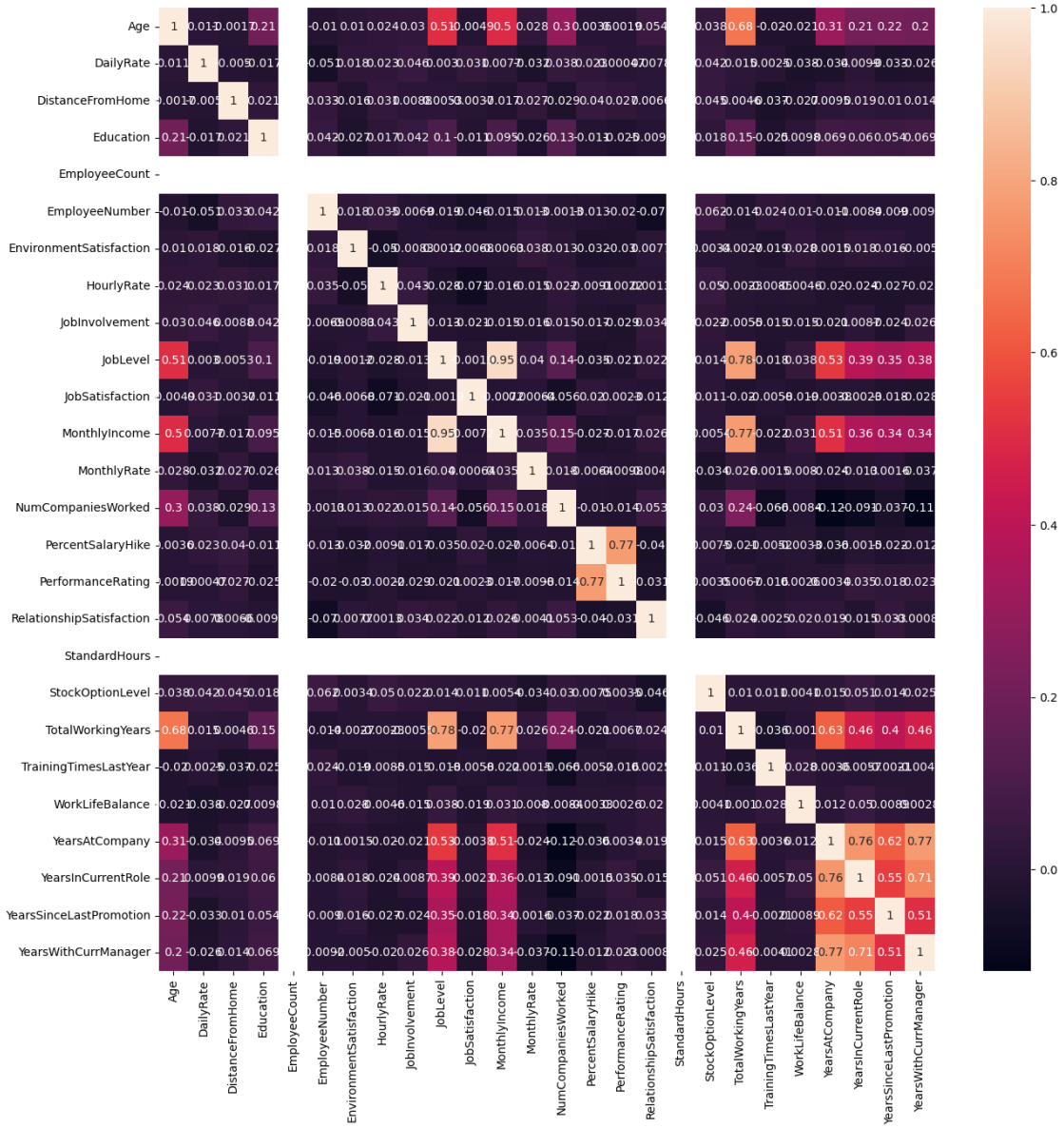
	YearsSinceLastPromotion	YearsWithCurrManager
Age	0.216513	0.202089
DailyRate	-0.033229	-0.026363
DistanceFromHome	0.010029	0.014406
Education	0.054254	0.069065
EmployeeCount	NaN	NaN
EmployeeNumber	-0.009019	-0.009197
EnvironmentSatisfaction	0.016194	-0.004999
HourlyRate	-0.026716	-0.020123
JobInvolvement	-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]
```

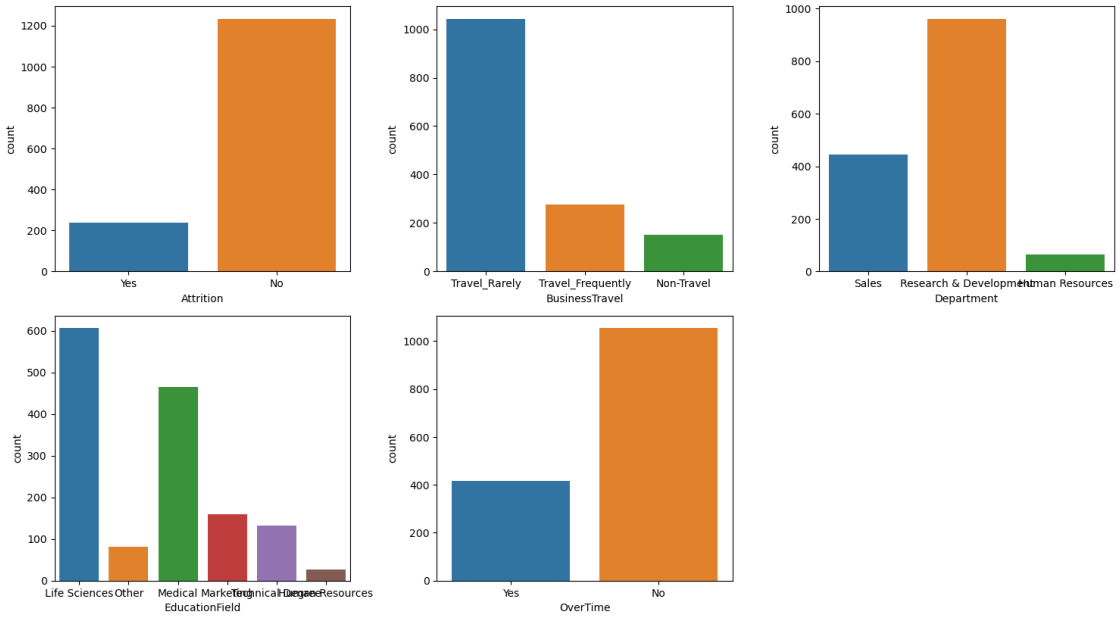
[11]:
```
plt.subplots(figsize=(15,15))
sns.heatmap(d,annot=True)
```

[11]: <Axes: >



[12]:
```
f = plt.figure()
f.set_figwidth(15)
f.set_figheight(12)
#Plot 1
plt.subplot(3, 3, 1)
sns.countplot(x="Attrition", data=a)
#Plot 2
plt.subplot(3, 3, 2)
sns.countplot(x="BusinessTravel", data=a)
#Plot 5
plt.subplot(3, 3, 3)
```

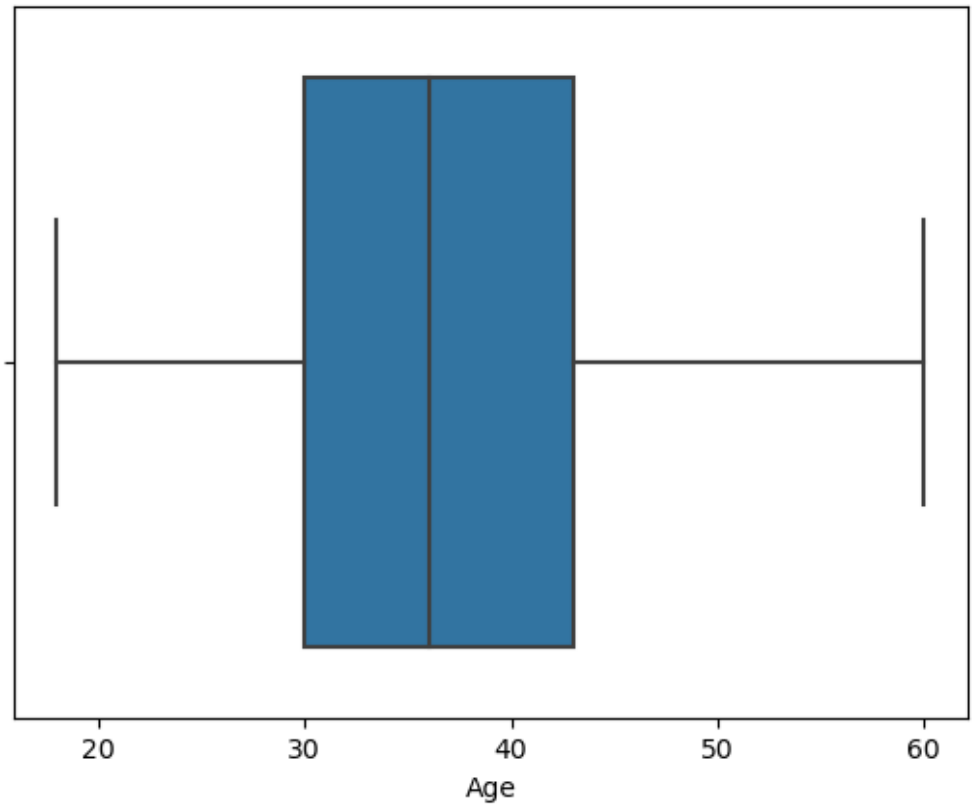
```
sns.countplot(x="Department", data=a)
#Plot 8
plt.subplot(3, 3, 4)
sns.countplot(x="EducationField", data=a)
#Plot 9
plt.subplot(3, 3, 5)
sns.countplot(x="OverTime", data=a)
# Adjust layout
plt.tight_layout()
# Show Plots
plt.show()
```



0.2.1 Outlier Detection

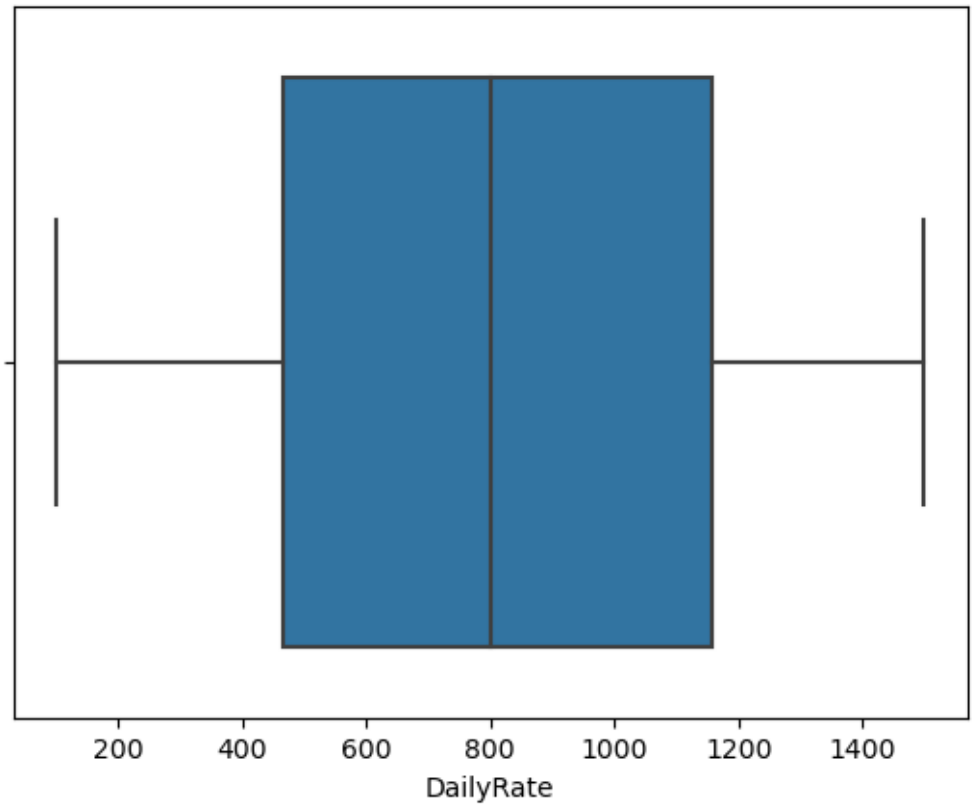
[13]:
```
sns.boxplot(x="Age",data=a)
```

[13]:
```
<Axes: xlabel='Age'>
```



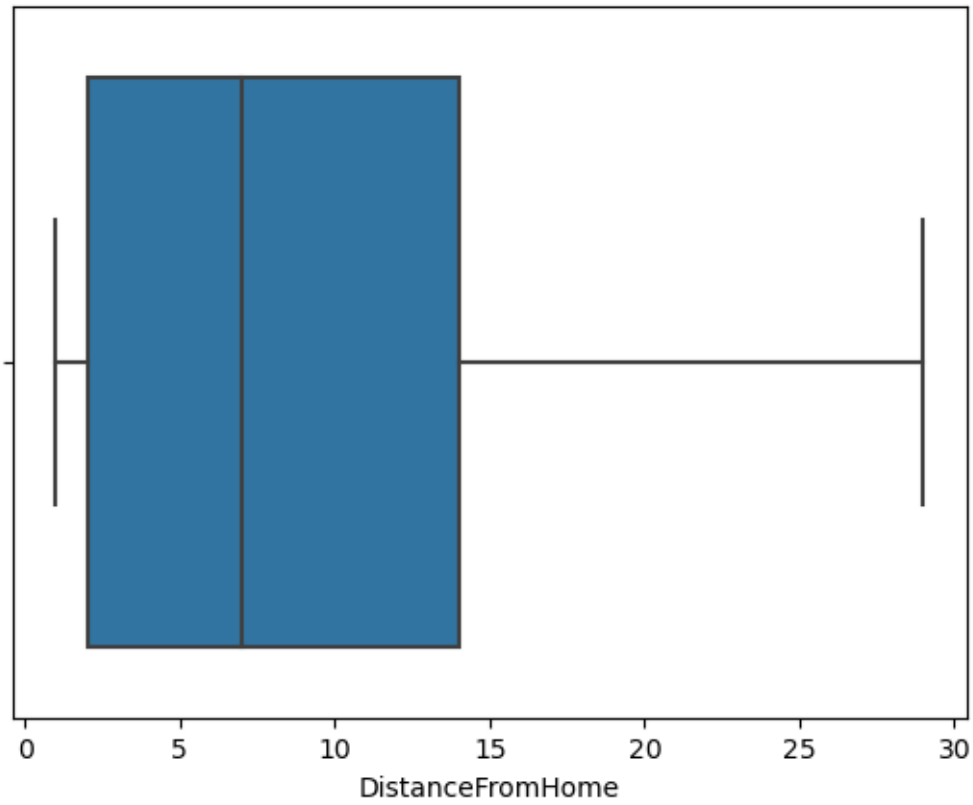
[14]:
```
sns.boxplot(x="DailyRate",data=a)
```

[14]:
```
<Axes: xlabel='DailyRate'>
```



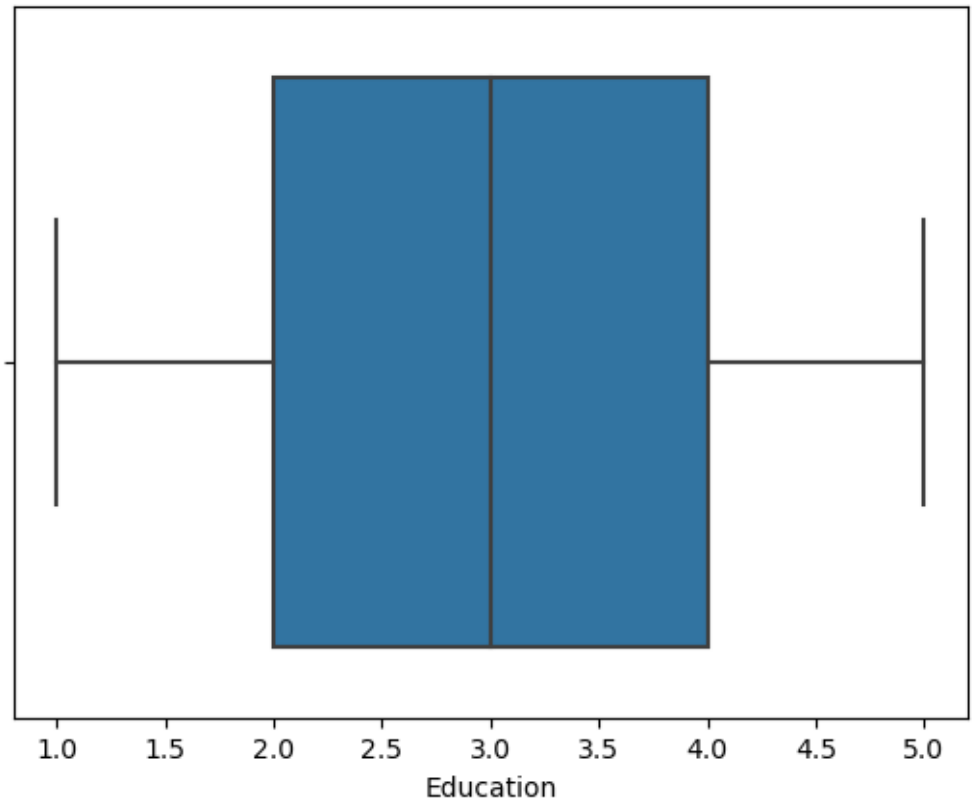
[15]:
```
sns.boxplot(x="DistanceFromHome",data=a)
```

[15]: <Axes: xlabel='DistanceFromHome'>



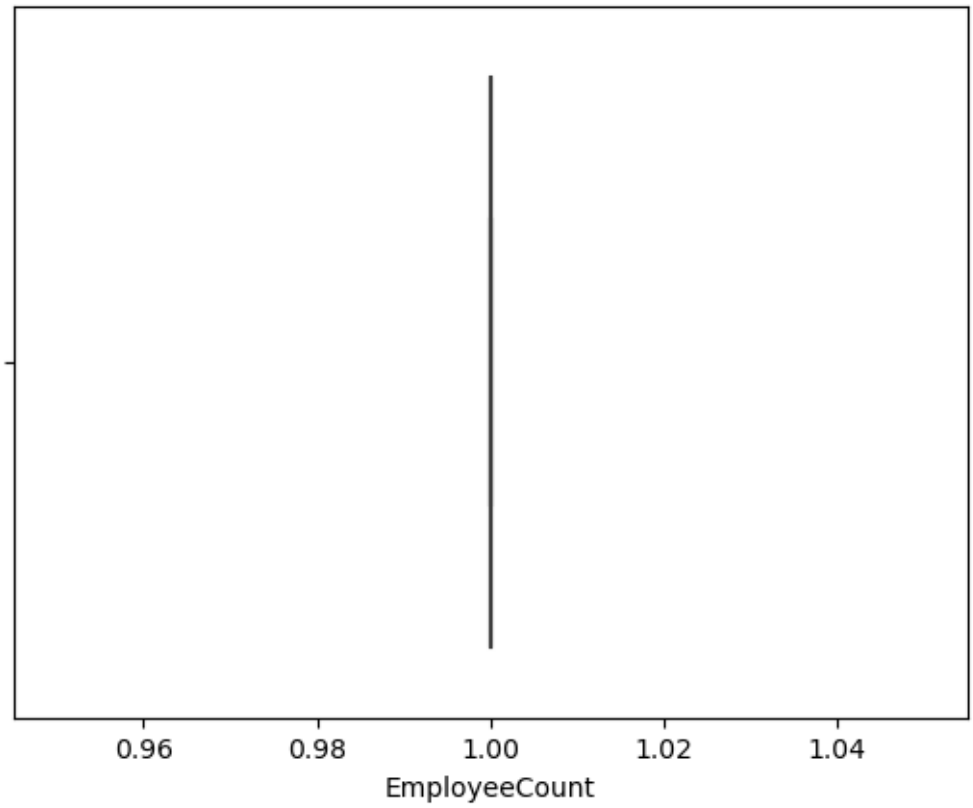
[16]:
```
sns.boxplot(x="Education",data=a)
```

[16]:
```
<Axes: xlabel='Education'>
```



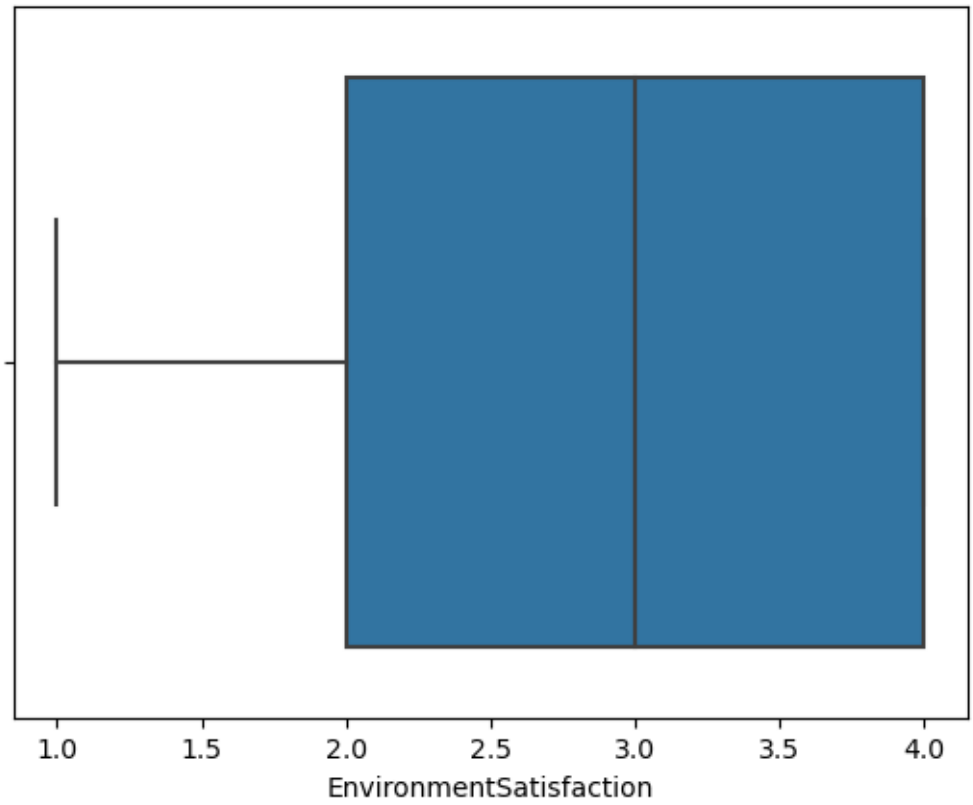
[17]:
```
sns.boxplot(x="EmployeeCount",data=a)
```

[17]:
```
<Axes: xlabel='EmployeeCount'>
```



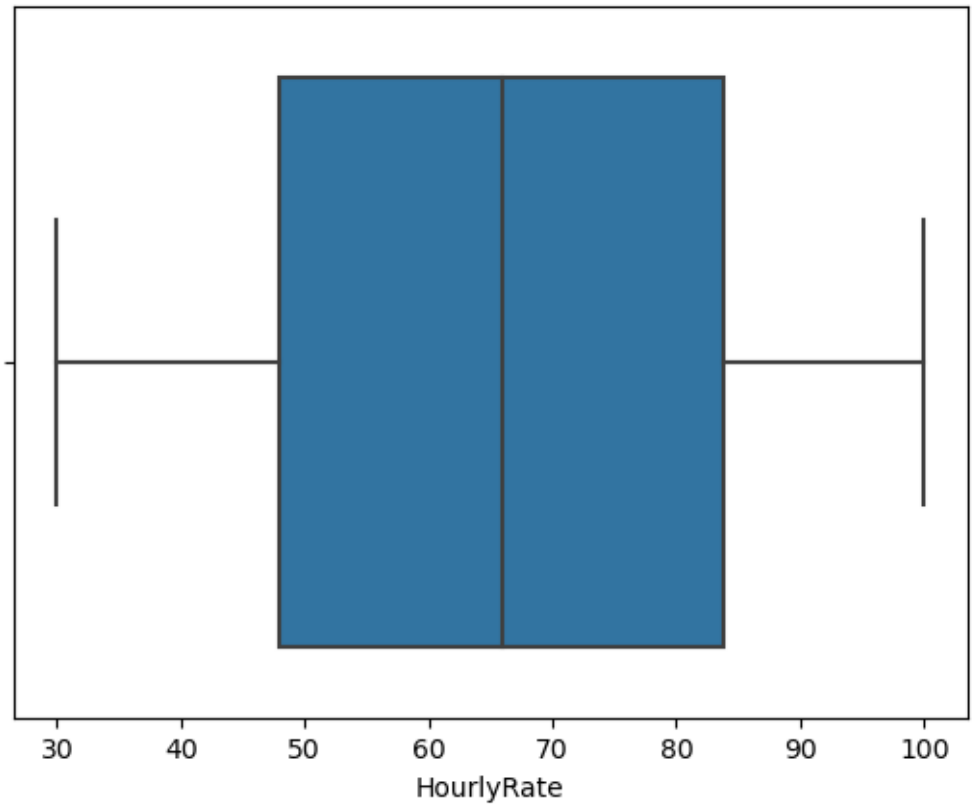
[18]:
```
sns.boxplot(x="EnvironmentSatisfaction",data=a)
```

[18]:
```
<Axes: xlabel='EnvironmentSatisfaction'>
```



[19]:
```
sns.boxplot(x="HourlyRate",data=a)
```

[19]:
```
<Axes: xlabel='HourlyRate'>
```



0.2.2 Splitting dependent and independent variables

[20]:
```
x=a.drop(columns=["Attrition"],axis=1)
x.head()
```

[20]:

	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	
4	27	Travel_Rarely	591	Research & Development	

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

	EnvironmentSatisfaction	…	RelationshipSatisfaction	StandardHours	\
0	2	…	1	80	

```
1                        3  …                        4              80
2                        4  …                        2              80
3                        4  …                        3              80
4                        1  …                        4              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
4                 1                  6                     3                3

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
4                     2

[5 rows x 34 columns]
```

[21]:
```
x.shape
```

[21]:
```
(1470, 34)
```

[22]:
```
y=a["Attrition"]
y.head()
```

[22]:
```
0    Yes
1     No
2    Yes
3     No
4     No
Name: Attrition, dtype: object
```

[23]:
```
y.shape
```

[23]:
```
(1470,)
```

0.3 Encoding

[24]:
```
from sklearn.preprocessing import LabelEncoder
```

[25]:
```
l=LabelEncoder()
```

[26]:
```
x["Gender"]=l.fit_transform(x["Gender"])
x['Gender']
```

[26]:
```
0       0
1       1
2       1
3       0
4       1
       ..
1465    1
1466    1
1467    1
1468    1
1469    1
Name: Gender, Length: 1470, dtype: int32
```

[27]:
```
x['Gender'].value_counts()
```

[27]:
```
1    882
0    588
Name: Gender, dtype: int64
```

[28]:
```
x['Gender'].nunique()
```

[28]:
```
2
```

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

[29]:

	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	
4	27	Travel_Rarely	591	Research & Development	

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

```
   EnvironmentSatisfaction  …  RelationshipSatisfaction  StandardHours  \
0                        2  …                         1             80
1                        3  …                         4             80
2                        4  …                         2             80
3                        4  …                         3             80
4                        1  …                         4             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
4                 1                  6                      3                3

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
4                     2

[5 rows x 34 columns]
```

[30]:
```
Dept = pd.get_dummies(a, columns=["Department"])
print(Dept)
```

```
      Age Attrition     BusinessTravel  DailyRate  DistanceFromHome  \
0      41       Yes      Travel_Rarely       1102                 1
1      49        No  Travel_Frequently        279                 8
2      37       Yes      Travel_Rarely       1373                 2
3      33        No  Travel_Frequently       1392                 3
4      27        No      Travel_Rarely        591                 2
…    …       …                  …        …                 …
1465   36        No  Travel_Frequently        884                23
1466   39        No      Travel_Rarely        613                 6
1467   27        No      Travel_Rarely        155                 4
1468   49        No  Travel_Frequently       1023                 2
1469   34        No      Travel_Rarely        628                 8
```

```
      Education EducationField  EmployeeCount  EmployeeNumber  \
0             2  Life Sciences              1               1
1             1  Life Sciences              1               2
2             2          Other              1               4
3             4  Life Sciences              1               5
4             1        Medical              1               7
…         …             …            …             …
1465          2        Medical              1            2061
1466          1        Medical              1            2062
1467          3  Life Sciences              1            2064
1468          3        Medical              1            2065
1469          3        Medical              1            2068

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

      Department_Human Resources  Department_Research & Development  \
0                              0                                   0
1                              0                                   1
2                              0                                   1
3                              0                                   1
4                              0                                   1
…                            …                                   …
1465                           0                                   1
1466                           0                                   1
1467                           0                                   1
1468                           0                                   0
1469                           0                                   1

      Department_Sales
0                    1
1                    0
2                    0
3                    0
4                    0
…                  …
1465                 0
1466                 0
1467                 0
1468                 1
1469                 0

[1470 rows x 37 columns]
```

[31]:
```
print(x)
```

```
      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
4      27      Travel_Rarely        591  Research & Development
…   …                  …        …                      …
1465   36  Travel_Frequently        884  Research & Development
1466   39      Travel_Rarely        613  Research & Development
1467   27      Travel_Rarely        155  Research & Development
1468   49  Travel_Frequently       1023                   Sales
1469   34      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  EnvironmentSatisfaction  …  RelationshipSatisfaction  \
0                  1                        2  …                         1
1                  2                        3  …                         4
2                  4                        4  …                         2
3                  5                        4  …                         3
4                  7                        1  …                         4
…                …                        …  …                       …
1465            2061                        3  …                         3
1466            2062                        4  …                         1
1467            2064                        2  …                         2
1468            2065                        4  …                         4
1469            2068                        2  …                         1

      StandardHours  StockOptionLevel  TotalWorkingYears  \
0                80                 0                  8
1                80                 1                 10
2                80                 0                  7
3                80                 0                  8
4                80                 1                  6
…              …                  …                  …
1465             80                 1                 17
1466             80                 1                  9
1467             80                 1                  6
1468             80                 0                 17
1469             80                 0                  6

     TrainingTimesLastYear  WorkLifeBalance YearsAtCompany  \
0                        0                1              6
1                        3                3             10
2                        3                3              0
3                        3                3              8
4                        3                3              2
…                      …                …            …
1465                     3                3              5
1466                     5                3              7
1467                     0                3              6
```

```
1468                      3                 2              9
1469                      3                 4              4

      YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                      4                        0                     5
1                      7                        1                     7
2                      0                        0                     0
3                      7                        3                     0
4                      2                        2                     2
…                    …                        …                     …
1465                   2                        0                     3
1466                   7                        1                     7
1467                   2                        0                     3
1468                   6                        0                     8
1469                   3                        1                     2

[1470 rows x 34 columns]
```

[32]:
```
a=pd.get_dummies(a,columns=["EducationField"])
a.head()
```

[32]:
```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education  EmployeeCount  EmployeeNumber  \
0                 1          2              1               1
1                 8          1              1               2
2                 2          2              1               4
3                 3          4              1               5
4                 2          1              1               7

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

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

```
4                         2                         2

   EducationField_Human Resources  EducationField_Life Sciences  \
0                               0                             1
1                               0                             1
2                               0                             0
3                               0                             1
4                               0                             0

   EducationField_Marketing  EducationField_Medical  EducationField_Other  \
0                         0                       0                     0
1                         0                       0                     0
2                         0                       0                     1
3                         0                       0                     0
4                         0                       1                     0

    EducationField_Technical Degree
0                                 0
1                                 0
2                                 0
3                                 0
4                                 0

[5 rows x 40 columns]
```

[33]:
```
x.head()
```

[33]:
```
    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
4   27      Travel_Rarely        591  Research & Development

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

    EnvironmentSatisfaction  …  RelationshipSatisfaction  StandardHours  \
0                         2  …                         1             80
1                         3  …                         4             80
2                         4  …                         2             80
3                         4  …                         3             80
4                         1  …                         4             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
4                 1                  6                     3                3

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
4                     2

[5 rows x 34 columns]
```

[34]:
```
Dept=pd.get_dummies(x["Department"],drop_first=True)
Dept
```

[34]:
```
      Research & Development  Sales
0                          0      1
1                          1      0
2                          1      0
3                          1      0
4                          1      0
…                        …    …
1465                       1      0
1466                       1      0
1467                       1      0
1468                       0      1
1469                       1      0

[1470 rows x 2 columns]
```

[35]:
```
x=pd.concat([x,Dept],axis=1)
x.head()
```

[35]:
```
   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
4   27      Travel_Rarely        591  Research & Development

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

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

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

  YearsSinceLastPromotion  YearsWithCurrManager  Research & Development  Sales
0                       0                     5                       0      1
1                       1                     7                       1      0
2                       0                     0                       1      0
3                       3                     0                       1      0
4                       2                     2                       1      0

[5 rows x 36 columns]
```

0.4 Feature Scaling

[36]:
```
from sklearn.preprocessing import StandardScaler
```

[37]:
```
scaler = StandardScaler()
```

[38]:
```
X = a[['Age', 'MonthlyIncome', 'YearsAtCompany', 'JobSatisfaction',␣
 ↪'EnvironmentSatisfaction','YearsWithCurrManager','WorkLifeBalance']]
Y = a['Attrition']
```

[39]:
```
X.head()
```

[39]:
```
   Age  MonthlyIncome  YearsAtCompany  JobSatisfaction  \
0   41           5993               6                4
1   49           5130              10                2
2   37           2090               0                3
3   33           2909               8                3
4   27           3468               2                2

   EnvironmentSatisfaction  YearsWithCurrManager  WorkLifeBalance
0                        2                     5                1
1                        3                     7                3
2                        4                     0                3
3                        4                     0                3
4                        1                     2                3
```

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

[40]:
```
      Age     BusinessTravel  DailyRate              Department  \
1465   36  Travel_Frequently        884  Research & Development
1466   39      Travel_Rarely        613  Research & Development
1467   27      Travel_Rarely        155  Research & Development
1468   49  Travel_Frequently       1023                   Sales
1469   34      Travel_Rarely        628  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount  \
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  EnvironmentSatisfaction  …  StockOptionLevel  \
1465            2061                        3  …                 1
1466            2062                        4  …                 1
1467            2064                        2  …                 1
1468            2065                        4  …                 0
1469            2068                        2  …                 0

      TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  \
1465                 17                      3                3
1466                  9                      5                3
1467                  6                      0                3
1468                 17                      3                2
1469                  6                      3                4

     YearsAtCompany  YearsInCurrentRole YearsSinceLastPromotion  \
1465              5                   2                       0
1466              7                   7                       1
```

```
1467              6                   2                       0
1468              9                   6                       0
1469              4                   3                       1

      YearsWithCurrManager  Research & Development  Sales
1465                     3                       1      0
1466                     7                       1      0
1467                     3                       1      0
1468                     8                       0      1
1469                     2                       1      0

[5 rows x 36 columns]
```

[41]:
```
x
```

[41]:
```
       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
4       27      Travel_Rarely        591  Research & Development
…    …                  …        …                       …
1465    36  Travel_Frequently        884  Research & Development
1466    39      Travel_Rarely        613  Research & Development
1467    27      Travel_Rarely        155  Research & Development
1468    49  Travel_Frequently       1023                   Sales
1469    34      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  EnvironmentSatisfaction  …  StockOptionLevel  \
0                  1                        2  …                 0
1                  2                        3  …                 1
2                  4                        4  …                 0
3                  5                        4  …                 0
4                  7                        1  …                 1
```

```
…              …                        …  …                 …
1465           2061                     3  …                 1
1466           2062                     4  …                 1
1467           2064                     2  …                 1
1468           2065                     4  …                 0
1469           2068                     2  …                 0

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

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

      YearsWithCurrManager  Research & Development  Sales
0                        5                       0      1
1                        7                       1      0
2                        0                       1      0
3                        0                       1      0
4                        2                       1      0
…                      …                       …    …
1465                     3                       1      0
1466                     7                       1      0
1467                     3                       1      0
1468                     8                       0      1
1469                     2                       1      0

[1470 rows x 36 columns]
```

0.4.1 Splitting Data into Test and Train Sets

[42]:
```
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=42)
```

[43]:
```
X_train,X_test,Y_train,Y_test.shape
```

[43]:
```
(      Age  MonthlyIncome  YearsAtCompany  JobSatisfaction  \
1097   24           2296               1                1
727    18           1051               0                4
254    29           6931               3                4
1175   39           5295               5                2
1341   31           4197              10                3
…   …            …              …              …
1130   35           3407              10                3
1294   41           6870               3                2
860    22           2853               0                4
1459   29           4025               4                2
1126   50          19331               1                3

      EnvironmentSatisfaction  YearsWithCurrManager  WorkLifeBalance
1097                        3                     0                3
727                         2                     0                3
254                         4                     2                3
1175                        4                     0                3
1341                        2                     2                3
…                         …                     …              …
1130                        2                     8                2
1294                        2                     2                1
860                         3                     0                3
1459                        4                     3                3
1126                        3                     0                3

[1176 rows x 7 columns],
      Age  MonthlyIncome  YearsAtCompany  JobSatisfaction  \
1041   28           8463               5                1
184    53           4450               4                1
1222   24           1555               1                3
67     45           9724               1                1
220    36           5914              13                2
…   …            …              …              …
567    34           6274               6                4
560    34           5121               0                1
945    50          16880               3                1
522    37           4680               1                4
651    47           4537               7                4
```

```
      EnvironmentSatisfaction  YearsWithCurrManager  WorkLifeBalance
1041                        4                     3                3
184                         4                     3                3
1222                        4                     0                3
67                          2                     0                3
220                         4                     7                4
…                         …                     …              …
567                         4                     4                3
560                         2                     0                3
945                         4                     2                3
522                         4                     0                3
651                         3                     7                3

[294 rows x 7 columns],
1097     No
727      No
254      No
1175     No
1341     No
       …
1130     No
1294     No
860     Yes
1459     No
1126     No
Name: Attrition, Length: 1176, dtype: object,
(294,))
```

0.5 Logistic Regression

0.5.1 Model Building and Importing the Model Building Libraries

[44]:
```
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

[45]:
```
model.fit(X_train, Y_train)
```

[45]:
```
LogisticRegression()
```

[46]:
```
pred=model.predict(X_test)
```

[47]:
```
pred
```

[47]:
```
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', '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', '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',
'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', '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', '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'], dtype=object)
```

[48]:
```
Y_test
```

[48]:
```
1041     No
184      No
1222    Yes
67       No
220      No
       …
567      No
560      No
945      No
522      No
651      No
Name: Attrition, Length: 294, dtype: object
```

[49]:
```
a
```

[49]:
```
      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  EmployeeCount  EmployeeNumber  \
0                    1          2              1               1
1                    8          1              1               2
2                    2          2              1               4
3                    3          4              1               5
4                    2          1              1               7
…                  …        …            …              …
1465                23          2              1            2061
1466                 6          1              1            2062
1467                 4          3              1            2064
1468                 2          3              1            2065
1469                 8          3              1            2068

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

```
      EducationField_Human Resources  EducationField_Life Sciences  \
0                                  0                             1
1                                  0                             1
2                                  0                             0
3                                  0                             1
4                                  0                             0
…                                …                             …
1465                               0                             0
1466                               0                             0
1467                               0                             1
1468                               0                             0
1469                               0                             0

      EducationField_Marketing  EducationField_Medical  EducationField_Other  \
0                            0                       0                     0
1                            0                       0                     0
2                            0                       0                     1
3                            0                       0                     0
4                            0                       1                     0
…                          …                       …                   …
1465                         0                       1                     0
1466                         0                       1                     0
1467                         0                       0                     0
1468                         0                       1                     0
1469                         0                       1                     0

      EducationField_Technical Degree
0                                    0
1                                    0
2                                    0
3                                    0
4                                    0
…                                  …
1465                                 0
1466                                 0
1467                                 0
1468                                 0
1469                                 0

[1470 rows x 40 columns]
```

0.6 Evaluation of Classification Model

[50]:
```
from sklearn.metrics import␣
↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[51]:
```
accuracy = accuracy_score(Y_test, pred)
```

[52]:
```
report = classification_report(Y_test, pred, zero_division=1)
```

[53]:
```
print(f'Accuracy: {accuracy}')
print(f'Classification Report:\n{report}')
```

```
Accuracy: 0.8673469387755102
Classification Report:
              precision    recall  f1-score   support

          No       0.87      1.00      0.93       255
         Yes       1.00      0.00      0.00        39

    accuracy                           0.87       294
   macro avg       0.93      0.50      0.46       294
weighted avg       0.88      0.87      0.81       294
```

[54]:
```
confusion_matrix(Y_test,pred)
```

[54]:
```
array([[255,   0],
       [ 39,   0]], dtype=int64)
```

[55]:
```
pd.crosstab(Y_test,pred)
```

[55]:
```
col_0       No
Attrition
No         255
Yes         39
```

0.7 ROC-AUC Curve

[56]:
```
probability=model.predict_proba(X_test)[:,1]
```

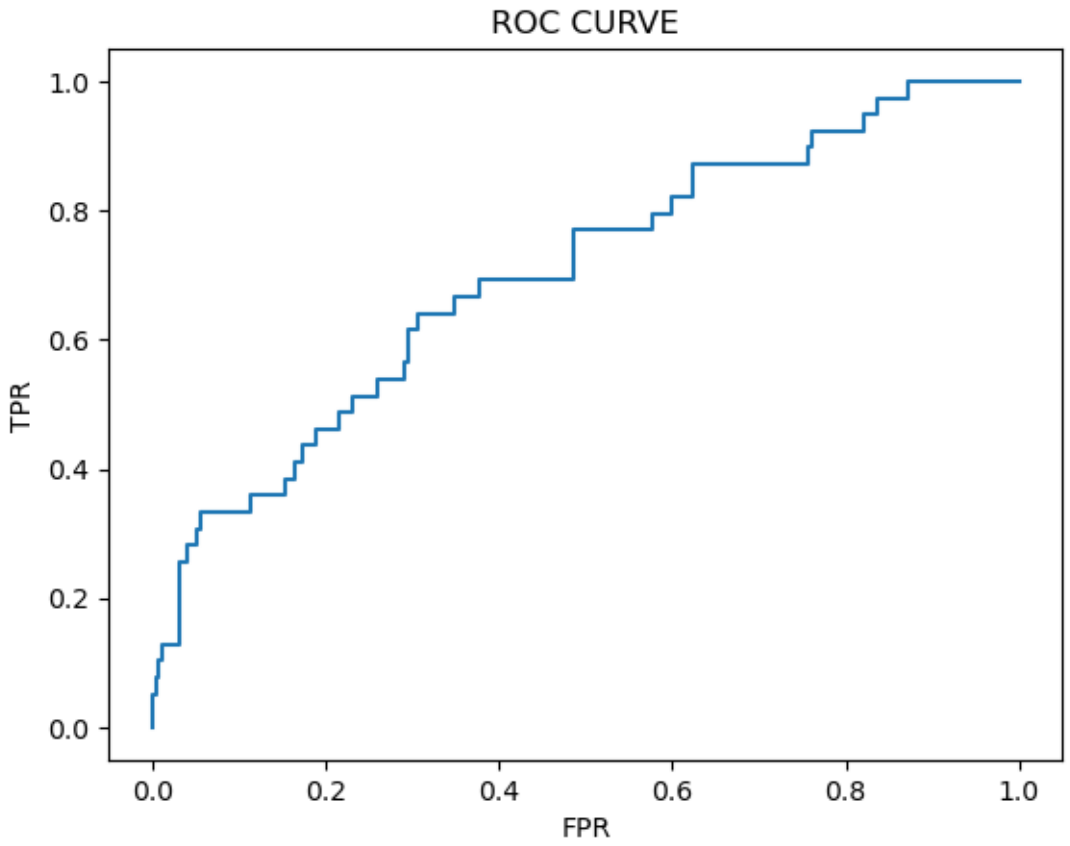
[57]:
```
probability
```

[57]:
```
array([0.14873939, 0.17373604, 0.25084589, 0.1865791 , 0.11911736,
       0.14963007, 0.15969356, 0.20644099, 0.08193936, 0.18537088,
       0.16096129, 0.02189805, 0.15660552, 0.11782876, 0.18248771,
       0.13287268, 0.14334387, 0.0892007 , 0.06858367, 0.05708062,
       0.1753651 , 0.14395111, 0.10012064, 0.15057687, 0.2329628 ,
       0.03338823, 0.27116899, 0.15771848, 0.18762417, 0.10029771,
       0.10548668, 0.15048832, 0.12644387, 0.14778903, 0.2030313 ,
       0.06737083, 0.04935137, 0.35253675, 0.19926438, 0.23846212,
       0.08198467, 0.28864725, 0.23955634, 0.19282516, 0.22246873,
       0.11288909, 0.17545014, 0.24051176, 0.14059822, 0.32377579,
       0.08977525, 0.15148043, 0.01896052, 0.14635136, 0.20158982,
       0.10191406, 0.10573264, 0.08537077, 0.1631479 , 0.12443613,
```

```
0.10510977, 0.33623452, 0.11027653, 0.05493965, 0.28005007,
0.18450874, 0.12499532, 0.17197795, 0.17873294, 0.06110176,
0.18127058, 0.08791989, 0.15005295, 0.15959692, 0.19866202,
0.07388538, 0.19341696, 0.19100387, 0.08712656, 0.08033949,
0.02928375, 0.13253218, 0.05956382, 0.16844954, 0.08753921,
0.17957673, 0.12899389, 0.16872069, 0.16947305, 0.12397644,
0.10991471, 0.24576674, 0.07821105, 0.2716565 , 0.12140547,
0.06524951, 0.1337184 , 0.14536957, 0.18726004, 0.10915274,
0.04570312, 0.10169758, 0.07390408, 0.22704117, 0.07208355,
0.08035364, 0.18593691, 0.16647288, 0.10818369, 0.05315879,
0.17696614, 0.18973955, 0.22476227, 0.17342537, 0.21403334,
0.16943373, 0.16771766, 0.09747364, 0.11387728, 0.2559594 ,
0.32393512, 0.08431327, 0.13118746, 0.10751731, 0.09837009,
0.25991497, 0.18954525, 0.11954205, 0.10534474, 0.09694665,
0.07268098, 0.30507638, 0.06501248, 0.14080365, 0.1255734 ,
0.11537899, 0.23299235, 0.17264787, 0.24765337, 0.06927027,
0.21512755, 0.09901074, 0.16646941, 0.08047622, 0.03233445,
0.1536394 , 0.14131117, 0.25851265, 0.26761484, 0.1665985 ,
0.10685997, 0.11549038, 0.19827263, 0.19076354, 0.13247131,
0.26173972, 0.17180386, 0.21324175, 0.04115976, 0.15054569,
0.16012435, 0.09434315, 0.09921354, 0.22000675, 0.06421677,
0.16643204, 0.12016003, 0.14827189, 0.08450615, 0.05725373,
0.12102272, 0.02681568, 0.18300015, 0.21076054, 0.11715199,
0.16127828, 0.18483891, 0.09043029, 0.14086669, 0.20253644,
0.0594472 , 0.10383826, 0.01617733, 0.15428555, 0.08595315,
0.22434066, 0.11577714, 0.07998958, 0.07811109, 0.12006352,
0.12845942, 0.14824842, 0.10405812, 0.19816497, 0.1162661 ,
0.21477996, 0.24395257, 0.04972863, 0.2156586 , 0.16831872,
0.17867722, 0.15398516, 0.21871738, 0.03416769, 0.07072713,
0.22242289, 0.10244091, 0.10919764, 0.12517809, 0.0706504 ,
0.07399615, 0.24438034, 0.17159597, 0.17617076, 0.10663942,
0.13898632, 0.15178098, 0.10545547, 0.2723432 , 0.07462743,
0.23465253, 0.26405406, 0.10124306, 0.30280889, 0.12410107,
0.1909214 , 0.20302625, 0.13276688, 0.0401135 , 0.18943046,
0.23129362, 0.25951761, 0.08630086, 0.21347439, 0.20469075,
0.13330949, 0.08581729, 0.10996843, 0.06690194, 0.04616928,
0.18853288, 0.11542819, 0.21231547, 0.03597583, 0.07176025,
0.17130681, 0.11593175, 0.23407496, 0.1533375 , 0.09696206,
0.16256038, 0.06366454, 0.04689748, 0.0855508 , 0.23703024,
0.07106702, 0.18067447, 0.2069784 , 0.22648723, 0.02715875,
0.17170263, 0.14167866, 0.27663201, 0.10463943, 0.12037205,
0.21133882, 0.02933273, 0.0973697 , 0.23466029, 0.23184944,
0.1882965 , 0.04906958, 0.19036583, 0.13999651, 0.11412922,
0.22223015, 0.12517666, 0.24824295, 0.07113102, 0.07508479,
0.14609486, 0.15491467, 0.18318556, 0.09382192, 0.04811606,
0.20893659, 0.20088061, 0.23217748, 0.10747859, 0.11268901,
0.25784861, 0.07464244, 0.1744561 , 0.09272658])
```

[58]:
```
from sklearn.preprocessing import LabelBinarizer
lb = LabelBinarizer()
Y_test_bin = lb.fit_transform(Y_test)
fpr, tpr, thresholds = roc_curve(Y_test_bin, probability)
```

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



0.8 Decision Tree

[60]:
```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report
```

[61]:
```
dt_model = DecisionTreeClassifier(random_state=50)
```

[62]:
```
dt_model.fit(X_train, Y_train)
```

[62]: DecisionTreeClassifier(random_state=50)

[63]:
```
dt_predictions = dt_model.predict(X_test)
```

[64]:
```
dt_accuracy = accuracy_score(Y_test, dt_predictions)
```

[65]:
```
dt_report = classification_report(Y_test, dt_predictions)
```

[66]:
```
print(f'Decision Tree Accuracy: {dt_accuracy}')
```

```
Decision Tree Accuracy: 0.7789115646258503
```

[67]:
```
print(f'Decision Tree Classification Report:\n{dt_report}')
```

```
Decision Tree Classification Report:
              precision    recall  f1-score   support

          No       0.90      0.84      0.87       255
         Yes       0.28      0.41      0.33        39

    accuracy                           0.78       294
   macro avg       0.59      0.62      0.60       294
weighted avg       0.82      0.78      0.80       294
```

0.9 Random Forest Classifier

[68]:
```
from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(random_state=50)
rf_model.fit(X_train, Y_train)
```

[68]: RandomForestClassifier(random_state=50)

[69]:
```
rf_predictions = rf_model.predict(X_test)
rf_accuracy = accuracy_score(Y_test, rf_predictions)
rf_report = classification_report(Y_test, rf_predictions)
print(f'Random Forest Accuracy: {rf_accuracy}')
```

```
Random Forest Accuracy: 0.8435374149659864
```

[70]:
```
print(f'Random Forest Classification Report:\n{rf_report}')
```

```
Random Forest Classification Report:
              precision    recall  f1-score   support

          No       0.88      0.95      0.91       255
         Yes       0.33      0.18      0.23        39
```

```
              precision    recall  f1-score   support

    accuracy                       0.84       294
   macro avg       0.61      0.56      0.57       294
weighted avg       0.81      0.84      0.82       294
```