assignment-4-22-09-2023

September 27, 2023

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

[2]:
```
#Importing the dataset.
df=pd.read_csv("HR-Employee-Attrition.csv")
```

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

[3]:

	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]
```

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

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

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
```

```
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

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

[7]:

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

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

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

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	\
count	1470.000000	1470.000000	1470.000000	
mean	0.793878	11.279592	2.799320	

```
std            0.852077           7.780782              1.289271
min            0.000000           0.000000              0.000000
25%            0.000000           6.000000              2.000000
50%            1.000000          10.000000              3.000000
75%            1.000000          15.000000              3.000000
max            3.000000          40.000000              6.000000

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

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

[8 rows x 26 columns]
```

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

[8]:
```
Age                         False
Attrition                   False
BusinessTravel              False
DailyRate                   False
Department                  False
DistanceFromHome            False
Education                   False
EducationField              False
EmployeeCount               False
EmployeeNumber              False
EnvironmentSatisfaction     False
Gender                      False
HourlyRate                  False
JobInvolvement              False
JobLevel                    False
JobRole                     False
```

```
JobSatisfaction             False
MaritalStatus               False
MonthlyIncome               False
MonthlyRate                 False
NumCompaniesWorked          False
Over18                      False
OverTime                    False
PercentSalaryHike           False
PerformanceRating           False
RelationshipSatisfaction    False
StandardHours               False
StockOptionLevel            False
TotalWorkingYears           False
TrainingTimesLastYear       False
WorkLifeBalance             False
YearsAtCompany              False
YearsInCurrentRole          False
YearsSinceLastPromotion     False
YearsWithCurrManager        False
dtype: bool
```

[9]:
```
df.isnull().sum()
```

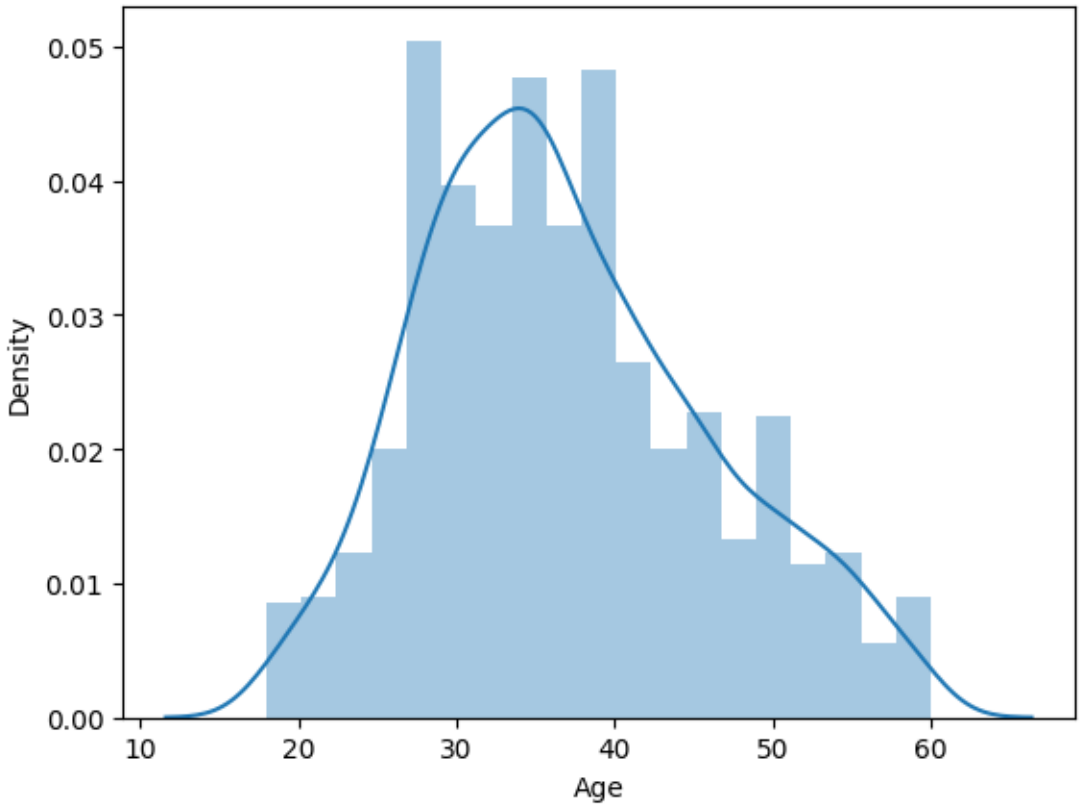
[9]:
```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              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
```

[11]:
```
#Data Visualization
sns.distplot(df["Age"])
```

```
<ipython-input-11-40d61bfc4c0b>:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["Age"])
```

[11]: <Axes: xlabel='Age', ylabel='Density'>



[12]:
```
df.corr()
```

```
<ipython-input-12-2f6f6606aa2c>: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()
```

[12]:

	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

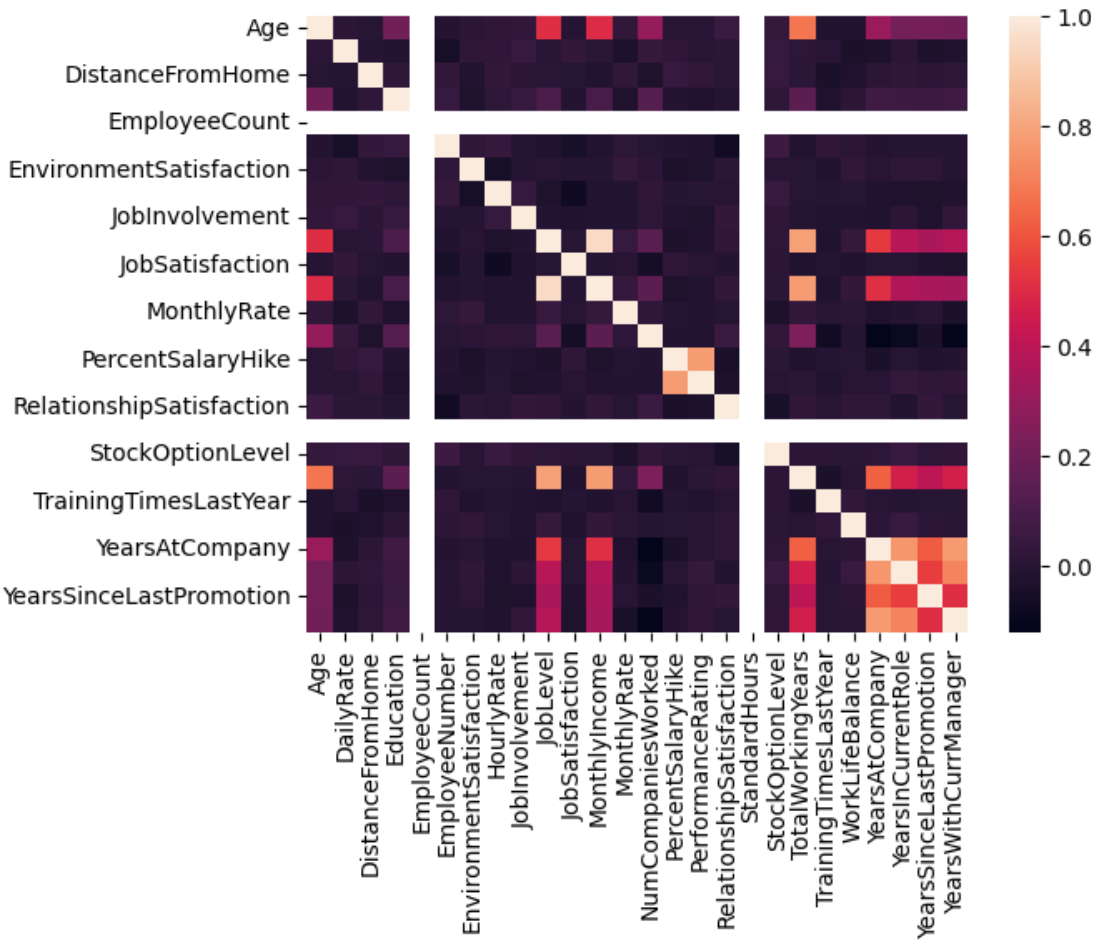
EmployeeCount	NaN	NaN
EmployeeNumber	-0.009019	-0.009197
EnvironmentSatisfaction	0.016194	-0.004999
HourlyRate	-0.026716	-0.020123
JobInvolvement	-0.024184	0.025976
JobLevel	0.353885	0.375281
JobSatisfaction	-0.018214	-0.027656
MonthlyIncome	0.344978	0.344079
MonthlyRate	0.001567	-0.036746
NumCompaniesWorked	-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]

[15]:
```
sns.heatmap(df.corr(),annot=False)
```

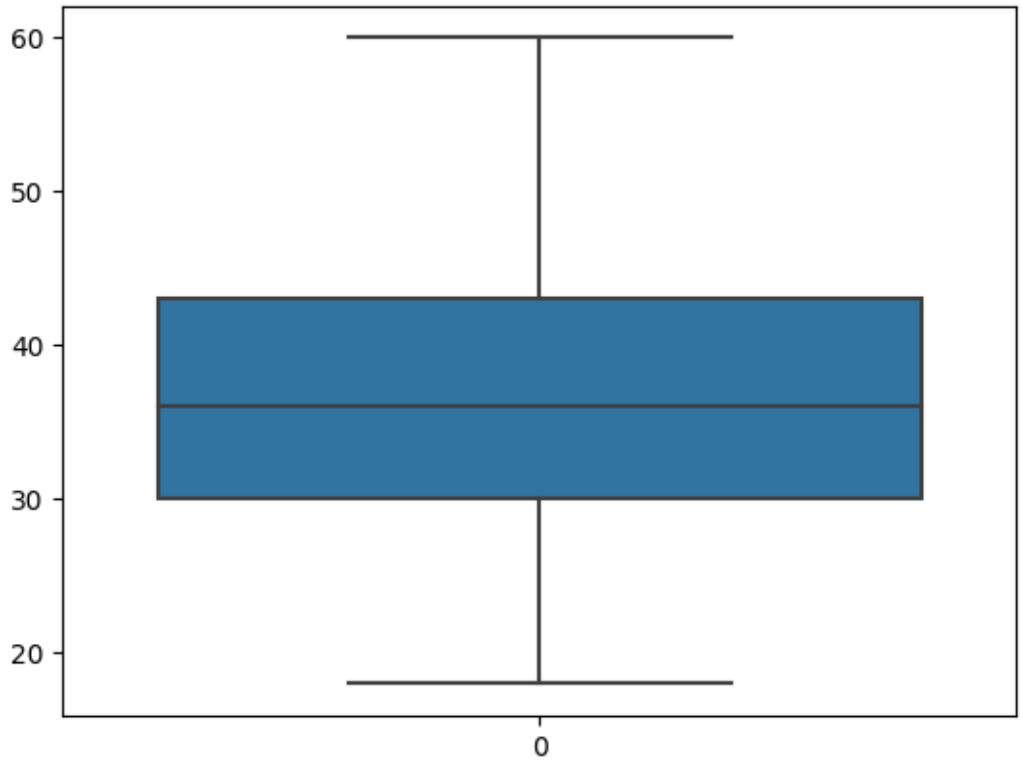
```
<ipython-input-15-e7f8babea3a9>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr(),annot=False)
```

[15]: <Axes: >



[16]:
```
sns.boxplot(df.Age)
```

[16]: <Axes: >



[26]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

[27]:
```
df["Attrition"]=le.fit_transform(df["Attrition"])
df["Attrition"]
```

[27]:
```
0       1
1       0
2       1
3       0
4       0
       ..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Attrition, Length: 1470, dtype: int64
```

[34]:
```
categorical_columns = df.select_dtypes(include=['object']).columns.tolist()
for col in categorical_columns:
```

```
    df[col] = le.fit_transform(df[col])
```

[35]:
```
#Splitting Depenndent and Independent variables
x=df.drop(columns=['Attrition'])
x.head()
```

[35]:

	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	
4	27	2	591	1	2	1	

	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	\
0	1	1	1	2	
1	1	1	2	3	
2	4	1	4	4	
3	1	1	5	4	
4	3	1	7	1	

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

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

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

[5 rows x 34 columns]

[36]:
```
y=df['Attrition']
y.head()
```

[36]:
```
0    1
1    0
2    1
3    0
4    0
Name: Attrition, dtype: int64
```

[38]:
```
#feature scaling
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled=pd.DataFrame(ms.fit_transform(x),columns=x.columns)
```

[39]:
```
x_scaled
```

[39]:

	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	0.214286	1.0	0.350036	0.5	0.035714	
…	…	…	…	…	…	
1465	0.428571	0.5	0.559771	0.5	0.785714	
1466	0.500000	1.0	0.365784	0.5	0.178571	
1467	0.214286	1.0	0.037938	0.5	0.107143	
1468	0.738095	0.5	0.659270	1.0	0.035714	
1469	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	0.00	0.6	0.0	0.002903	
…	…	…	…	…	
1465	0.25	0.6	0.0	0.996613	
1466	0.00	0.6	0.0	0.997097	
1467	0.50	0.2	0.0	0.998065	
1468	0.50	0.6	0.0	0.998549	
1469	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	0.000000	…	1.000000	0.0	
…	…	…	…	…	

```
1465                0.666667  …                 0.666667           0.0
1466                1.000000  …                 0.000000           0.0
1467                0.333333  …                 0.333333           0.0
1468                1.000000  …                 1.000000           0.0
1469                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             0.333333              0.150               0.500000
…                ….                 …                  …
1465          0.333333              0.425               0.500000
1466          0.333333              0.225               0.833333
1467          0.333333              0.150               0.000000
1468          0.000000              0.425               0.500000
1469          0.000000              0.150               0.500000

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0            0.000000           0.150            0.222222
1            0.666667           0.250            0.388889
2            0.666667           0.000            0.000000
3            0.666667           0.200            0.388889
4            0.666667           0.050            0.111111
…               …              …                  …
1465         0.666667           0.125            0.111111
1466         0.666667           0.175            0.388889
1467         0.666667           0.150            0.111111
1468         0.333333           0.225            0.333333
1469         1.000000           0.100            0.166667

      YearsSinceLastPromotion  YearsWithCurrManager
0                    0.000000              0.294118
1                    0.066667              0.411765
2                    0.000000              0.000000
3                    0.200000              0.000000
4                    0.133333              0.117647
…                       …                   …
1465                 0.000000              0.176471
1466                 0.066667              0.411765
1467                 0.000000              0.176471
1468                 0.000000              0.470588
1469                 0.066667              0.117647

[1470 rows x 34 columns]
```

[40]:
```
#Splitting Data into Train and Test.
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)
```

[41]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[41]:
```
((1176, 34), (294, 34), (1176,), (294,))
```

[42]:
```
x_train.head()
```

[42]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	\
1374	0.952381	1.0	0.360057	1.0	0.714286	
1092	0.642857	1.0	0.607015	0.5	0.964286	
768	0.523810	1.0	0.141732	1.0	0.892857	
569	0.428571	0.0	0.953472	1.0	0.250000	
911	0.166667	0.5	0.355762	1.0	0.821429	

	Education	EducationField	EmployeeCount	EmployeeNumber	\
1374	0.50	0.2	0.0	0.937107	
1092	0.50	1.0	0.0	0.747460	
768	0.50	0.4	0.0	0.515239	
569	0.75	0.2	0.0	0.381229	
911	0.00	0.2	0.0	0.615385	

	EnvironmentSatisfaction	…	RelationshipSatisfaction	StandardHours	\
1374	1.000000	…	0.666667	0.0	
1092	1.000000	…	1.000000	0.0	
768	0.666667	…	0.333333	0.0	
569	0.000000	…	0.333333	0.0	
911	0.666667	…	1.000000	0.0	

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

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

	YearsSinceLastPromotion	YearsWithCurrManager

```
1374                0.000000              0.000000
1092                0.000000              0.176471
768                 0.466667              0.294118
569                 0.000000              0.529412
911                 0.066667              0.000000

[5 rows x 34 columns]
```

[43]:
```
#Model Building using Logistic Regression
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

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

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

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

[46]:
```
pred
```

[46]:
```
array([0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0])
```

[47]:
```
y_test
```

[47]:
```
442     0
1091    0
981     1
785     0
1332    1
       ..
1439    0
481     0
124     1
198     0
```

```
1229    0
Name: Attrition, Length: 294, dtype: int64
```

[48]:
```
df
```

[48]:
```
      Age  Attrition  BusinessTravel  DailyRate  Department  DistanceFromHome  \
0      41          1               2       1102           2                 1
1      49          0               1        279           1                 8
2      37          1               2       1373           1                 2
3      33          0               1       1392           1                 3
4      27          0               2        591           1                 2
…   …        …              …         …          …                 …
1465   36          0               1        884           1                23
1466   39          0               2        613           1                 6
1467   27          0               2        155           1                 4
1468   49          0               1       1023           2                 2
1469   34          0               2        628           1                 8

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

      RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0                            1             80                 0
1                            4             80                 1
2                            2             80                 0
3                            3             80                 0
4                            4             80                 1
…                         …           …              …
1465                         3             80                 1
1466                         1             80                 1
1467                         2             80                 1
1468                         4             80                 0
1469                         1             80                 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
0                        5
1                        7
2                        0
3                        0
4                        2
…                      …
1465                     3
1466                     7
1467                     3
1468                     8
1469                     2

[1470 rows x 35 columns]
```

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

[51]:
```
accuracy_score(y_test,pred)
```

[51]:
```
0.8843537414965986
```

[52]:
```
confusion_matrix(y_test,pred)
```

[52]:
```
array([[242,   3],
       [ 31,  18]])
```

[53]:
```
pd.crosstab(y_test,pred)
```

[53]:
```
col_0        0   1
Attrition
0          242   3
1           31  18
```

[54]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.89      0.99      0.93       245
           1       0.86      0.37      0.51        49

    accuracy                           0.88       294
   macro avg       0.87      0.68      0.72       294
weighted avg       0.88      0.88      0.86       294
```

[56]:
```
#Model Building using DecisionTree Classifier
from sklearn.tree import DecisionTreeClassifier
dc=DecisionTreeClassifier()
```

[57]:
```
dc.fit(x_train,y_train)
```

[57]:
```
DecisionTreeClassifier()
```

[58]:
```
pred=dc.predict(x_test)
```

[59]:
```
pred
```

[59]:
```
array([0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,
       1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1,
```

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

[60]:
```
y_test
```

[60]:
```
442     0
1091    0
981     1
785     0
1332    1
       ..
1439    0
481     0
124     1
198     0
1229    0
Name: Attrition, Length: 294, dtype: int64
```

[61]:
```
df
```

[61]:
```
      Age  Attrition  BusinessTravel  DailyRate  Department  DistanceFromHome  \
0      41          1               2       1102           2                 1
1      49          0               1        279           1                 8
2      37          1               2       1373           1                 2
3      33          0               1       1392           1                 3
4      27          0               2        591           1                 2
…   …        …              …         …          …                 …
1465   36          0               1        884           1                23
1466   39          0               2        613           1                 6
1467   27          0               2        155           1                 4
1468   49          0               1       1023           2                 2
1469   34          0               2        628           1                 8

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

      RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0                            1             80                 0
```

```
1                           4              80                  1
2                           2              80                  0
3                           3              80                  0
4                           4              80                  1
…                         …              …                   …
1465                        3              80                  1
1466                        1              80                  1
1467                        2              80                  1
1468                        4              80                  0
1469                        1              80                  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
0                        5
1                        7
2                        0
3                        0
4                        2
…                      …
1465                     3
1466                     7
1467                     3
```

```
1468                    8
1469                    2

[1470 rows x 35 columns]
```

[62]:
```
#Accuracy score
from sklearn.metrics import␣
 ↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[63]:
```
accuracy_score(y_test,pred)
```

[63]:
```
0.7653061224489796
```

[64]:
```
confusion_matrix(y_test,pred)
```

[64]:
```
array([[208,  37],
       [ 32,  17]])
```

[65]:
```
pd.crosstab(y_test,pred)
```

[65]:
```
col_0        0   1
Attrition
0          208  37
1           32  17
```

[80]:
```
#Model Building using Random Forest
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier()
```

[81]:
```
rf.fit(x_train,y_train)
```

[81]:
```
RandomForestClassifier()
```

[82]:
```
pred=rf.predict(x_test)
```

[83]:
```
accuracy_score(y_test,pred)
```

[83]:
```
0.8537414965986394
```

[84]:
```
confusion_matrix(y_test, pred)
```

[84]:
```
array([[243,   2],
       [ 41,   8]])
```

[85]:
```
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.86	0.99	0.92	245
1	0.80	0.16	0.27	49
accuracy			0.85	294
macro avg	0.83	0.58	0.59	294
weighted avg	0.85	0.85	0.81	294