IMPORT LIBRARIES

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

IMPORT DATASET

```python
df=pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")

df
```

```
      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]

df.head()

   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 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

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

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

```
8
4                              6                                 3                      3
2

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

[5 rows x 35 columns]

df.tail()

      Age Attrition        BusinessTravel   DailyRate
Department  \
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  \
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  \
1465            2061   ...                          3            80
1466            2062   ...                          1            80
1467            2064   ...                          2            80
1468            2065   ...                          4            80
1469            2068   ...                          1            80

      StockOptionLevel   TotalWorkingYears   TrainingTimesLastYear  \
1465                 1                  17                        3
1466                 1                   9                        5
1467                 1                   6                        0
1468                 0                  17                        3
1469                 0                   6                        3

     WorkLifeBalance   YearsAtCompany YearsInCurrentRole  \
```

```
          3                 5                 2
1465
1466      3                 7                 7
1467      3                 6                 2
1468      2                 9                 6
1469      4                 4                 3

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

[5 rows x 35 columns]
```

df.shape

(1470, 35)

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

df.describe()
```

	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]

corr=df.corr()
corr
```

```
<ipython-input-11-7d5195e2bf4d>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  corr=df.corr()
```

```
                            Age    DailyRate   DistanceFromHome
Education  \
Age                    1.000000    0.010661          -0.001686
0.208034
DailyRate              0.010661    1.000000          -0.004985  -
0.016806
DistanceFromHome      -0.001686   -0.004985           1.000000
0.021042
Education              0.208034   -0.016806           0.021042
1.000000
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
```

	EnvironmentSatisfaction	HourlyRate	JobInvolvement
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]

plt.subplots(figsize=(22,15))
sns.heatmap(corr,annot=True,cmap="coolwarm")

<Axes: >
```
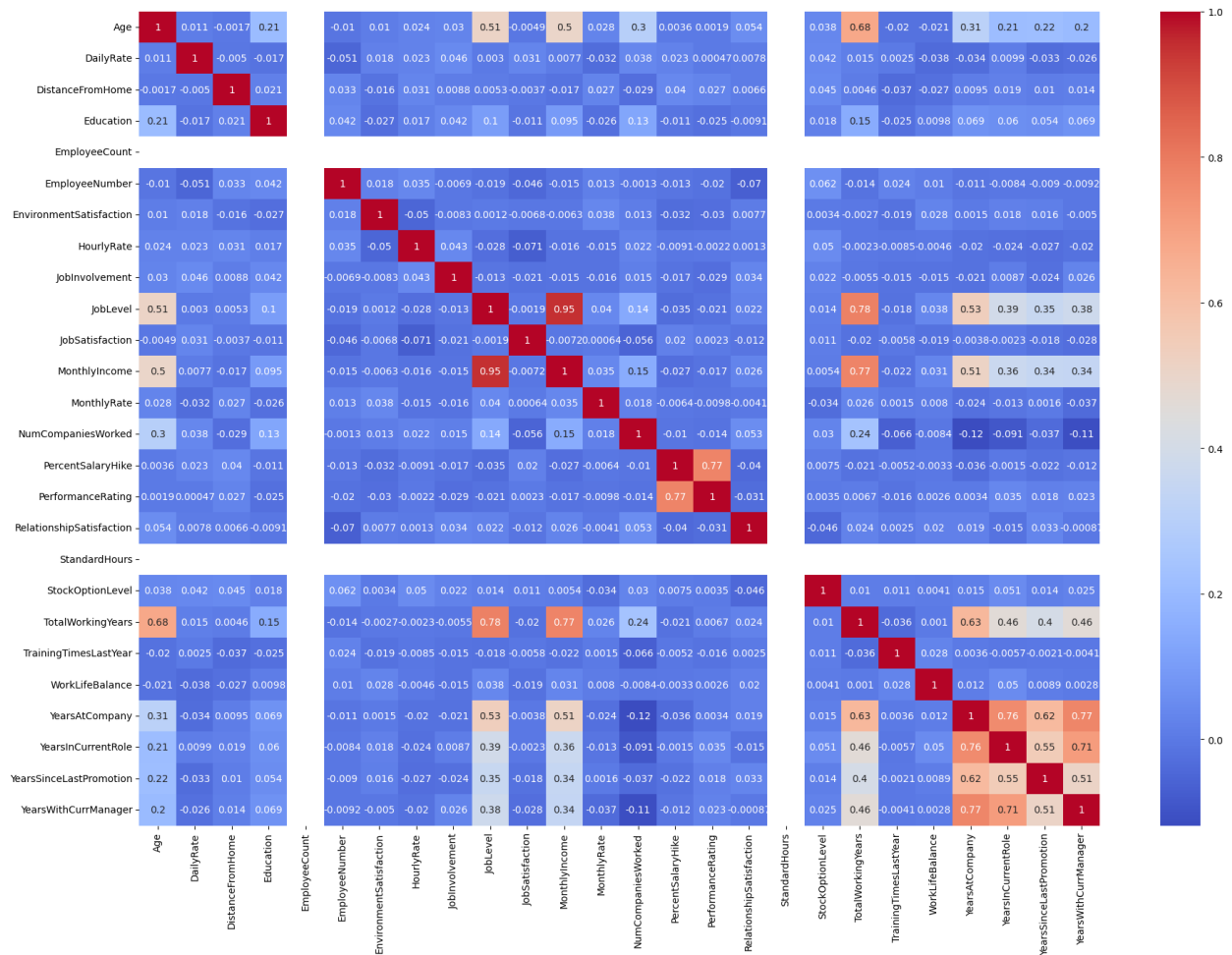
```
df.Attrition.value_counts()

No      1233
Yes      237
Name: Attrition, dtype: int64
```

Checking for NULL Values

```
df.isnull().any()

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
```

Data Visualization

```python
attrition_counts = df['Attrition'].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(attrition_counts, labels=attrition_counts.index,
autopct='%1.1f%%', startangle=90)
plt.title('Attrition Distribution')
plt.axis('equal')

plt.show()
```

Attrition Distribution

Yes

16.1%

83.9%

No
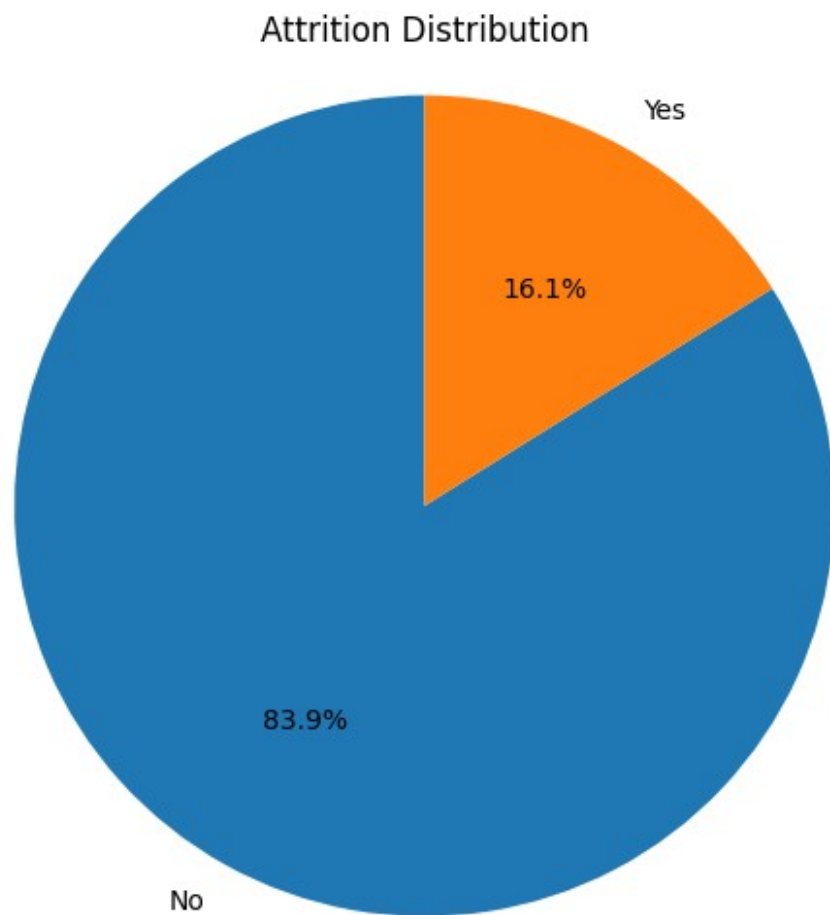
```python
plt.figure(figsize=(8, 6))
sns.countplot(x="Attrition", data=df)
plt.title("Attrition Count")
plt.show()
```

Attrition Count
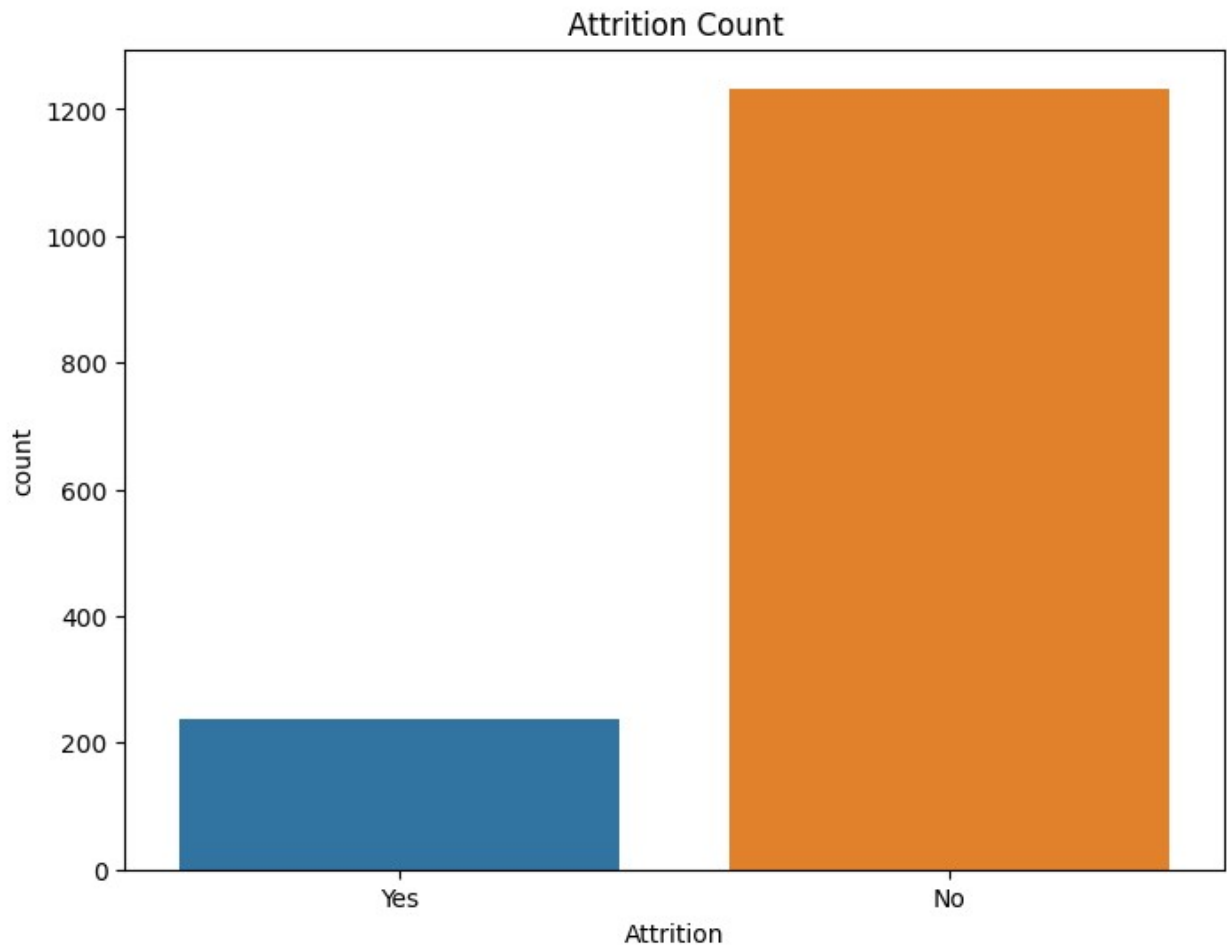
```
plt.figure(figsize=(8, 6))
sns.histplot(data=df, x="Age", kde=True)
plt.title("Distribution of Age")
plt.show()
```

Distribution of Age

Outlier Detection

```python
plt.figure(figsize=(35, 8))
sns.boxplot(data=df)
plt.title('Box Plots for all the attributes')
plt.show()
```



```python
sns.boxplot(data=df, x='YearsInCurrentRole')
plt.title('Years In Current Role')
```

```
plt.xlabel('Fare')
plt.show()
```


Years In Current Role
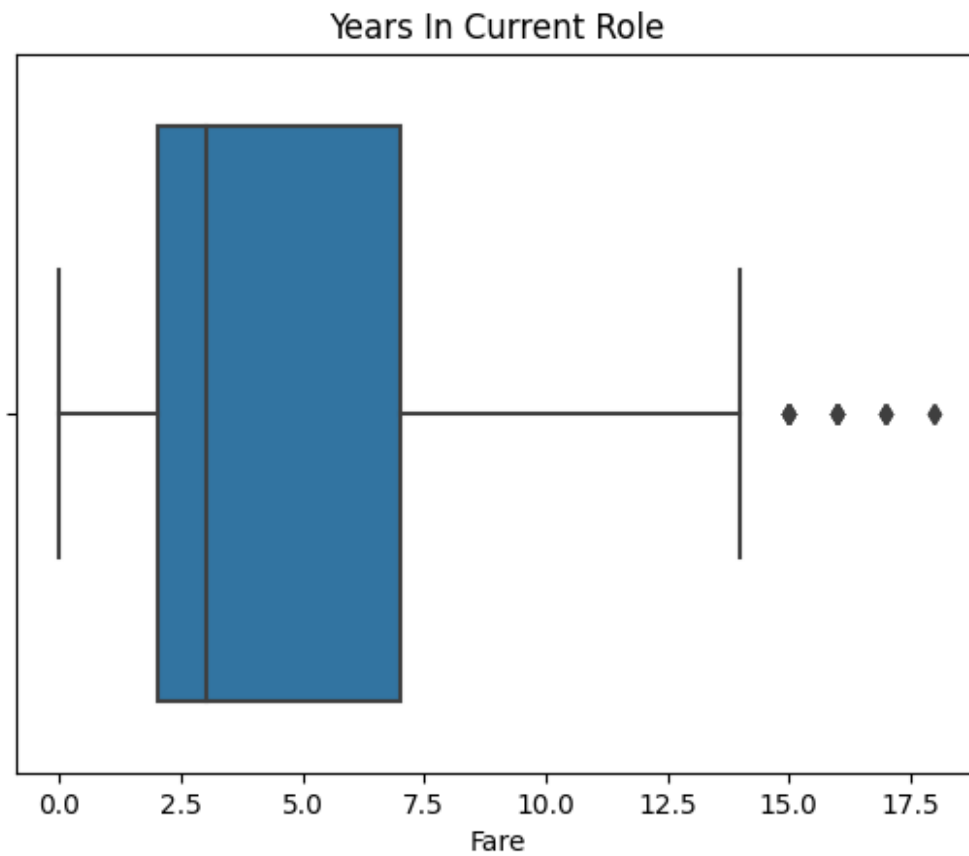
```
sns.boxplot(data=df, x='MonthlyIncome')
plt.title('Monthly Income')
plt.xlabel('Fare')
plt.show()
```

Monthly Income
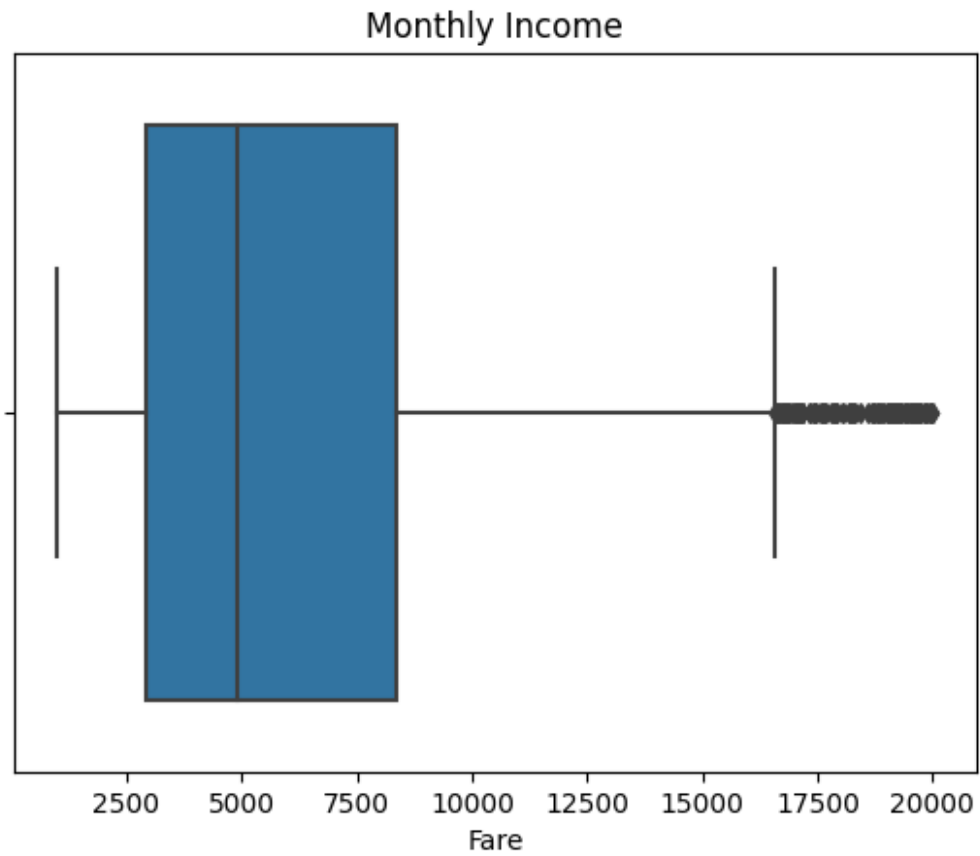

```python
from scipy import stats

z_scores = stats.zscore(df['MonthlyIncome'])
z_score_threshold = 3
df_cleaned = df[(np.abs(z_scores) <= z_score_threshold)]

sns.boxplot(data=df_cleaned, x='MonthlyIncome')
plt.title('Monthly Income')
plt.xlabel('Fare')
plt.show()
```

Monthly Income
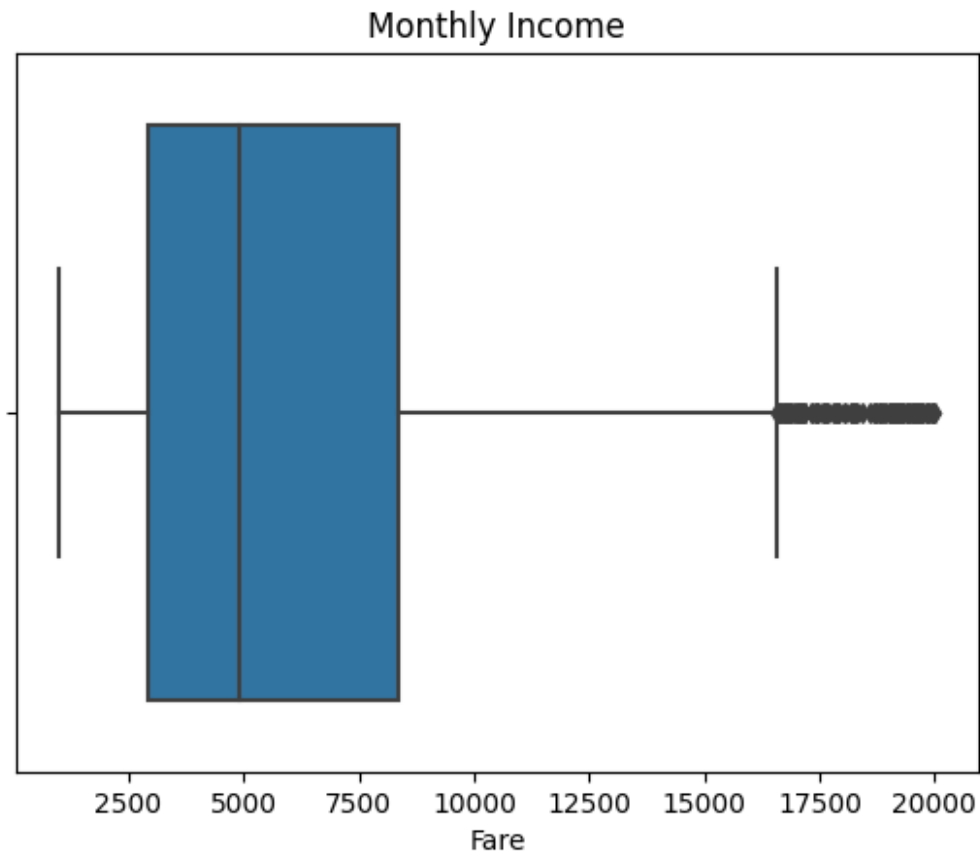
So the outliers are in large quantity, and they are inside the threshold, so let us not remove the outliers

SPLITTING INDEPENDENT AND DEPENDENT VARIABLES

```
x= df.drop(columns=["Attrition"])
y = df["Attrition"]

x.head()
```

```
   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]

y.head()
```

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

ENCODING

```
categorical_features =
x.select_dtypes(include=['object']).columns.tolist()
x_encoded = pd.get_dummies(x, columns=categorical_features,
drop_first=True)

x_encoded.head()

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

   EnvironmentSatisfaction  HourlyRate  JobInvolvement  JobLevel  ...
\
0                        2          94               3         2  ...

1                        3          61               2         2  ...

2                        4          92               2         1  ...

3                        4          56               3         1  ...

4                        1          40               3         1  ...


   JobRole_Laboratory Technician  JobRole_Manager  \
0                              0                0
1                              0                0
2                              1                0
3                              0                0
4                              1                0

   JobRole_Manufacturing Director  JobRole_Research Director  \
```

```
0                                           0                           0
1                                           0                           0
2                                           0                           0
3                                           0                           0
4                                           0                           0

    JobRole_Research Scientist  JobRole_Sales Executive  \
0                             0                        1
1                             1                        0
2                             0                        0
3                             1                        0
4                             0                        0

    JobRole_Sales Representative  MaritalStatus_Married  \
MaritalStatus_Single  \
0                              0                      0
1
1                              0                      1
0
2                              0                      0
1
3                              0                      1
0
4                              0                      1
0

    OverTime_Yes
0              1
1              0
2              1
3              1
4              0

[5 rows x 47 columns]
```

FEATURE SCALING

```python
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
x_scaled = pd.DataFrame(scaler.fit_transform(x_encoded),
columns=x_encoded.columns)

x_scaled.head()

        Age  DailyRate  DistanceFromHome  Education  EmployeeCount  \
0  0.446350   0.742527         -1.010909  -0.891688            0.0
1  1.322365  -1.297775         -0.147150  -1.868426            0.0
2  0.008343   1.414363         -0.887515  -0.891688            0.0
3 -0.429664   1.461466         -0.764121   1.061787            0.0
```

```
4 -1.086676   -0.524295           -0.887515 -1.868426              0.0

   EmployeeNumber  EnvironmentSatisfaction  HourlyRate  JobInvolvement
\
0       -1.701283                -0.660531    1.383138        0.379672

1       -1.699621                 0.254625   -0.240677       -1.026167

2       -1.696298                 1.169781    1.284725       -1.026167

3       -1.694636                 1.169781   -0.486709        0.379672

4       -1.691313                -1.575686   -1.274014        0.379672


   JobLevel  ...  JobRole_Laboratory Technician  JobRole_Manager  \
0 -0.057788  ...                      -0.462464        -0.273059
1 -0.057788  ...                      -0.462464        -0.273059
2 -0.961486  ...                       2.162331        -0.273059
3 -0.961486  ...                      -0.462464        -0.273059
4 -0.961486  ...                       2.162331        -0.273059

   JobRole_Manufacturing Director  JobRole_Research Director  \
0                       -0.330808                  -0.239904
1                       -0.330808                  -0.239904
2                       -0.330808                  -0.239904
3                       -0.330808                  -0.239904
4                       -0.330808                  -0.239904

   JobRole_Research Scientist  JobRole_Sales Executive  \
0                   -0.497873                 1.873287
1                    2.008543                -0.533821
2                   -0.497873                -0.533821
3                    2.008543                -0.533821
4                   -0.497873                -0.533821

   JobRole_Sales Representative  MaritalStatus_Married
MaritalStatus_Single  \
0                     -0.244625              -0.918921
1.458650
1                     -0.244625               1.088232             -
0.685565
2                     -0.244625              -0.918921
1.458650
3                     -0.244625               1.088232             -
0.685565
4                     -0.244625               1.088232             -
0.685565

   OverTime_Yes
```

```
0         1.591746
1        -0.628241
2         1.591746
3         1.591746
4        -0.628241

[5 rows x 47 columns]

x=x_scaled
```

Train and test split

```
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)
```

MODEL BUILDING

```python
# Import the necessary libraries
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix
from joblib import dump

logreg_model = LogisticRegression(random_state=42)
dt_model = DecisionTreeClassifier(random_state=42)

logreg_model.fit(x_train, y_train)
dt_model.fit(x_train, y_train)

DecisionTreeClassifier(random_state=42)

logreg_predictions = logreg_model.predict(x_test)

dt_predictions = dt_model.predict(x_test)

logreg_accuracy = accuracy_score(y_test, logreg_predictions)
print("Logistic Regression Accuracy:", logreg_accuracy)

dt_accuracy = accuracy_score(y_test, dt_predictions)
print("Decision Tree Accuracy:", dt_accuracy)

logreg_report = classification_report(y_test, logreg_predictions)
print("Classification Report for Logistic Regression:\n",
logreg_report)

dt_report = classification_report(y_test, dt_predictions)
print("Classification Report for Decision Tree Classifier:\n",
dt_report)
```

```python
logreg_conf_matrix = confusion_matrix(y_test, logreg_predictions)
print("Confusion Matrix for Logistic Regression:\n",
logreg_conf_matrix)

dt_conf_matrix = confusion_matrix(y_test, dt_predictions)
print("Confusion Matrix for Decision Tree Classifier:\n",
dt_conf_matrix)
```

```
Logistic Regression Accuracy: 0.8809523809523809
Decision Tree Accuracy: 0.7721088435374149
Classification Report for Logistic Regression:
              precision    recall  f1-score   support

          No       0.92      0.95      0.93       255
         Yes       0.56      0.46      0.51        39

    accuracy                           0.88       294
   macro avg       0.74      0.70      0.72       294
weighted avg       0.87      0.88      0.88       294

Classification Report for Decision Tree Classifier:
              precision    recall  f1-score   support

          No       0.87      0.86      0.87       255
         Yes       0.17      0.18      0.17        39

    accuracy                           0.77       294
   macro avg       0.52      0.52      0.52       294
weighted avg       0.78      0.77      0.78       294

Confusion Matrix for Logistic Regression:
 [[241  14]
 [ 21  18]]
Confusion Matrix for Decision Tree Classifier:
 [[220  35]
 [ 32   7]]
```