Project 3 : "HR Analytics"

```python
# Importing the Dependencies
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
import warnings
warnings.filterwarnings("ignore")

# Importing the dataset
data=pd.read_csv(r"C:\Users\hp\Desktop\FIRST PROJECT\HR-Employee-
Attrition.csv")
pd.set_option('display.max_columns',None)
data.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

   EnvironmentSatisfaction  Gender  HourlyRate  JobInvolvement
JobLevel  \
0                        2  Female          94               3
2
1                        3    Male          61               2
2
```

```
2                        4    Male          92               2
1
3                        4  Female          56               3
1
4                        1    Male          40               3
1

                 JobRole  JobSatisfaction MaritalStatus  MonthlyIncome
\
0        Sales Executive                4        Single           5993

1     Research Scientist                2       Married           5130

2  Laboratory Technician                3        Single           2090

3     Research Scientist                3       Married           2909

4  Laboratory Technician                2       Married           3468


   MonthlyRate  NumCompaniesWorked Over18 OverTime  PercentSalaryHike
\
0        19479                   8      Y      Yes                 11

1        24907                   1      Y       No                 23

2         2396                   6      Y      Yes                 15

3        23159                   1      Y      Yes                 11

4        16632                   9      Y       No                 12


   PerformanceRating  RelationshipSatisfaction  StandardHours  \
0                  3                         1             80
1                  4                         4             80
2                  3                         2             80
3                  3                         3             80
4                  3                         4             80

   StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0                 0                  8                      0
1                 1                 10                      3
2                 0                  7                      3
3                 0                  8                      3
4                 1                  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

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

```
data.tail()
```

```
      Age  Attrition  BusinessTravel  DailyRate  Department
DistanceFromHome  \
1465   36          0             NaN        884           2
23
1466   39          0             2.0        613           2
6
1467   27          0             2.0        155           2
4
1468   49          0             NaN       1023           1
2
1469   34          0             2.0        628           2
8

      Education EducationField  EmployeeCount  EmployeeNumber  \
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 Gender  HourlyRate  JobInvolvement
JobLevel  \
1465                        3   Male          41               4
2
1466                        4   Male          42               2
3
1467                        2   Male          87               4
2
1468                        4   Male          63               2
2
1469                        2   Male          82               4
2

                        JobRole  JobSatisfaction MaritalStatus
MonthlyIncome  \
1465      Laboratory Technician                4       Married
2571
1466  Healthcare Representative                1       Married
```

```
9991
1467     Manufacturing Director                2       Married
6142
1468            Sales Executive                2       Married
5390
1469      Laboratory Technician                3       Married
4404

      MonthlyRate  NumCompaniesWorked Over18 OverTime
PercentSalaryHike  \
1465        12290                   4      Y       No
17
1466        21457                   4      Y       No
15
1467         5174                   1      Y      Yes
20
1468        13243                   2      Y       No
14
1469        10228                   2      Y       No
12

      PerformanceRating  RelationshipSatisfaction  StandardHours  \
1465                  3                         3             80
1466                  3                         1             80
1467                  4                         2             80
1468                  3                         4             80
1469                  3                         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  \
1465                3               5                   2
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
```

```
data.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
```

```
data.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  JobSatisfaction  MonthlyIncome   MonthlyRate  \
count  1470.000000      1470.000000    1470.000000   1470.000000
mean      2.063946         2.728571    6502.931293  14313.103401
std       1.106940         1.102846    4707.956783   7117.786044
min       1.000000         1.000000    1009.000000   2094.000000
25%       1.000000         2.000000    2911.000000   8047.000000
50%       2.000000         3.000000    4919.000000  14235.500000
75%       3.000000         4.000000    8379.000000  20461.500000
max       5.000000         4.000000   19999.000000  26999.000000

       NumCompaniesWorked  PercentSalaryHike  PerformanceRating  \
count         1470.000000        1470.000000        1470.000000
mean             2.693197          15.209524           3.153741
std              2.498009           3.659938           0.360824
min              0.000000          11.000000           3.000000
25%              1.000000          12.000000           3.000000
50%              2.000000          14.000000           3.000000
```

```
75%              4.000000          18.000000           3.000000
max              9.000000          25.000000           4.000000

       RelationshipSatisfaction  StandardHours  StockOptionLevel  \
count               1470.000000         1470.0       1470.000000
mean                   2.712245           80.0          0.793878
std                    1.081209            0.0          0.852077
min                    1.000000           80.0          0.000000
25%                    2.000000           80.0          0.000000
50%                    3.000000           80.0          1.000000
75%                    4.000000           80.0          1.000000
max                    4.000000           80.0          3.000000

       TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  \
count        1470.000000            1470.000000      1470.000000
mean           11.279592               2.799320         2.761224
std             7.780782               1.289271         0.706476
min             0.000000               0.000000         1.000000
25%             6.000000               2.000000         2.000000
50%            10.000000               3.000000         3.000000
75%            15.000000               3.000000         3.000000
max            40.000000               6.000000         4.000000

       YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
count     1470.000000         1470.000000              1470.000000
mean         7.008163            4.229252                 2.187755
std          6.126525            3.623137                 3.222430
min          0.000000            0.000000                 0.000000
25%          3.000000            2.000000                 0.000000
50%          5.000000            3.000000                 1.000000
75%          9.000000            7.000000                 3.000000
max         40.000000           18.000000                15.000000

       YearsWithCurrManager
count           1470.000000
mean               4.123129
std                3.568136
min                0.000000
25%                2.000000
50%                3.000000
75%                7.000000
max               17.000000
```

```
data.describe(include="O")
```

```
       Attrition BusinessTravel              Department EducationField
Gender  \
count       1470           1470                    1470           1470
1470
unique         2              3                       3              6
```

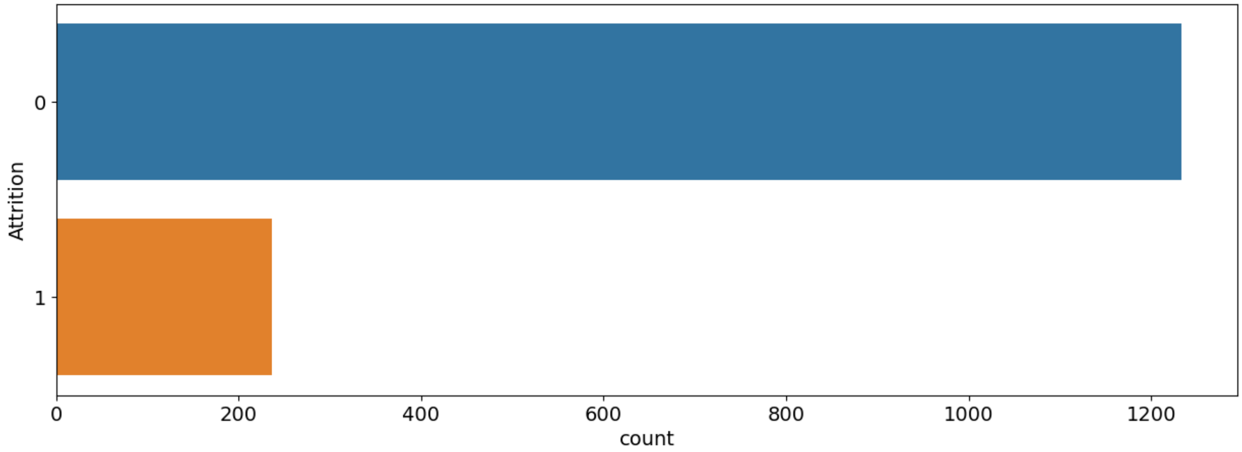
```
2
top           No  Travel_Rarely  Research & Development  Life Sciences
Male
freq        1233           1043                     961            606
882

                JobRole MaritalStatus Over18 OverTime
count              1470          1470   1470     1470
unique                9             3      1        2
top     Sales Executive       Married      Y       No
freq                326           673   1470     1054
```

EDA(EXPLORATORY DATA ANALYSIS)

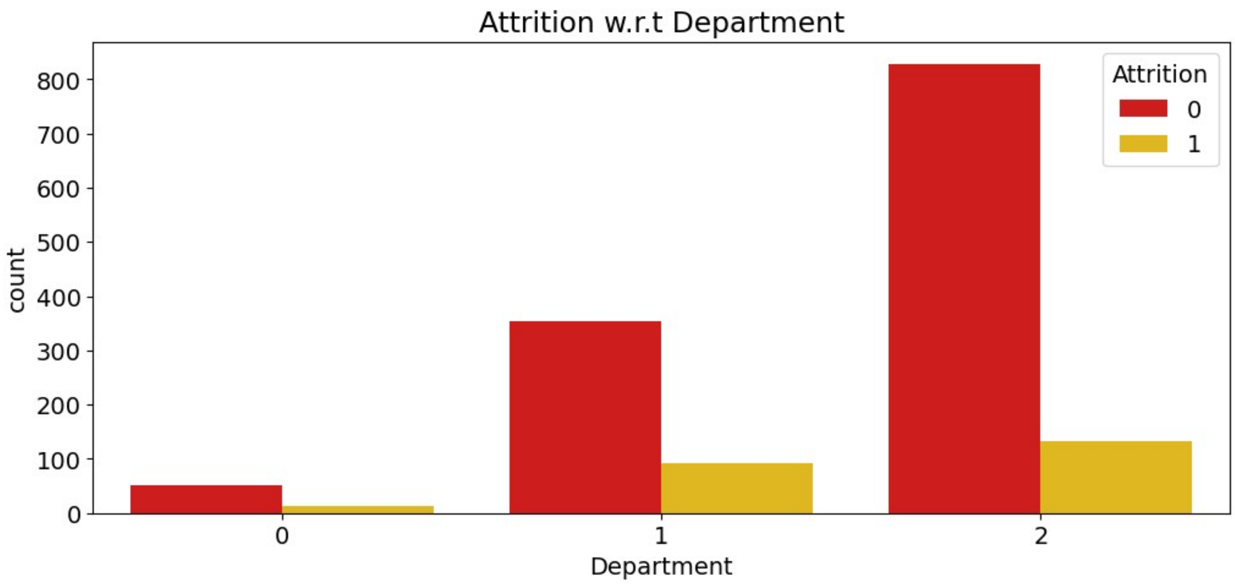
Analysis on categorical columns with respect to the target column(Attrition)

Target column==Attrition

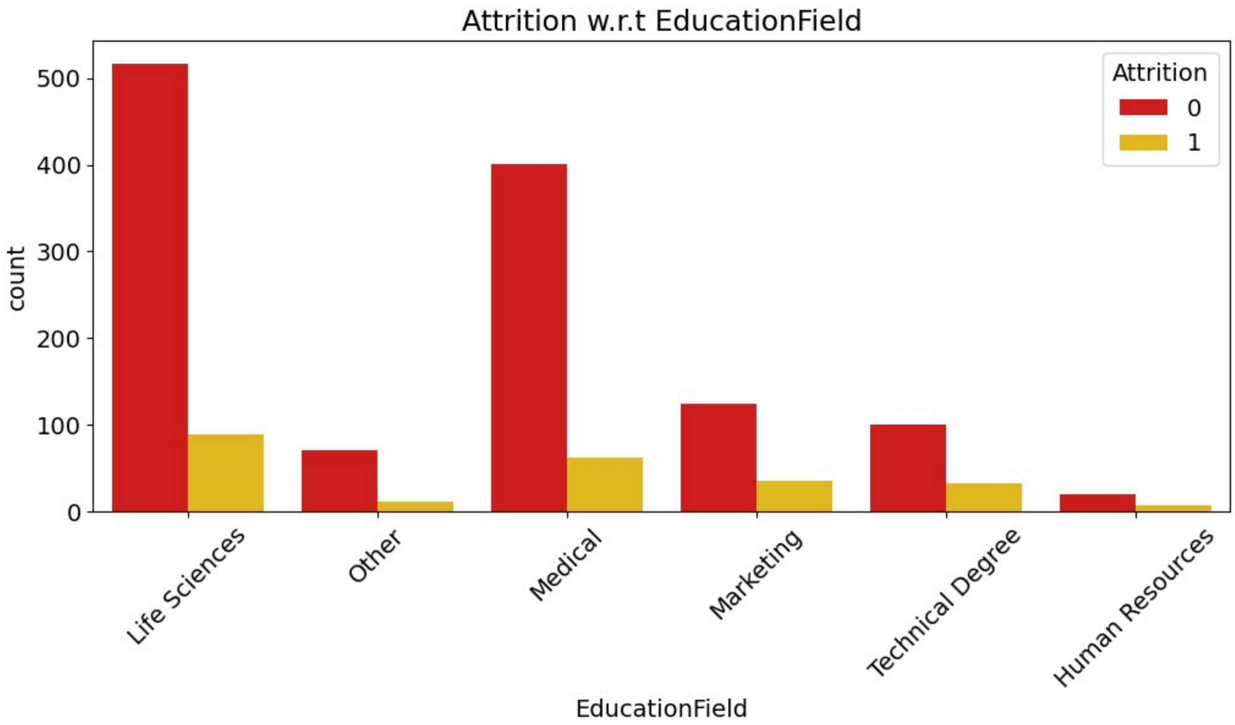
```
plt.figure(figsize=(15,5))
plt.rc("font",size=14)
sns.countplot(y = 'Attrition',data=data)
plt.show()
```



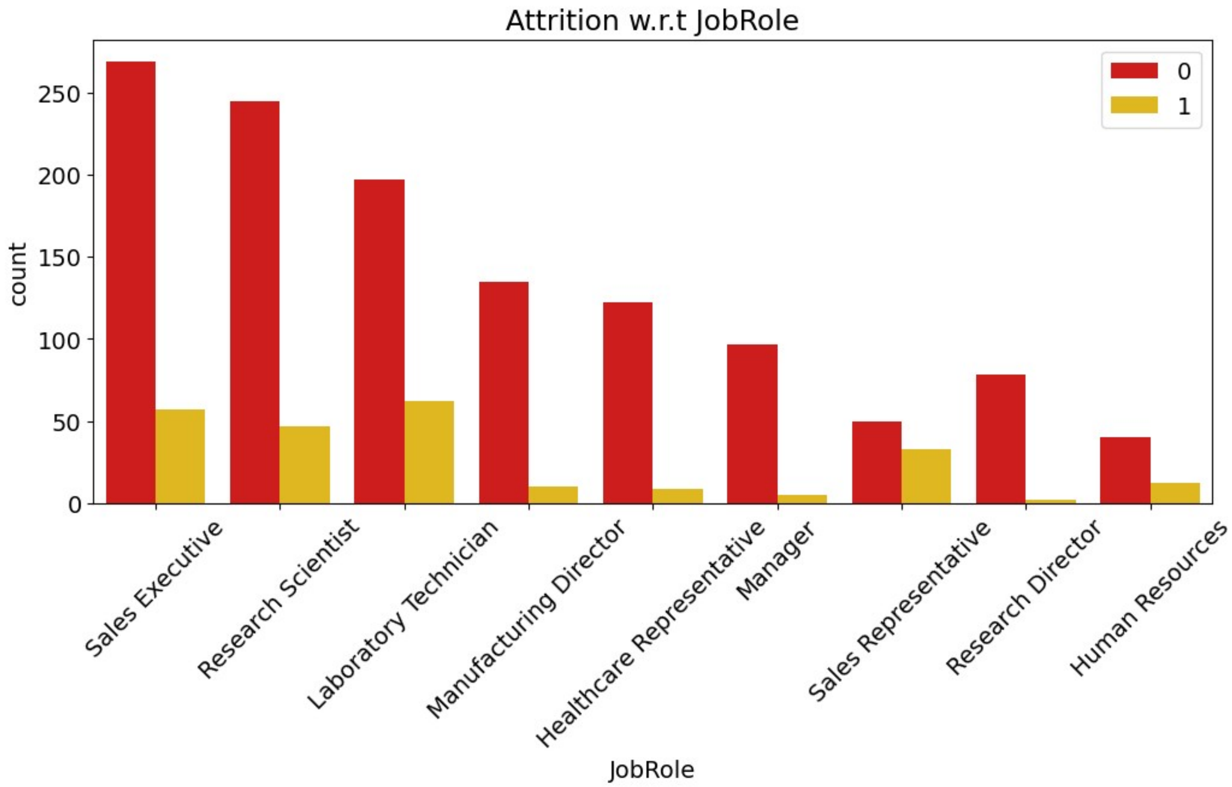
```
#Department w.r.t Attrition
plt.figure(figsize=(12,5))
sns.countplot(x='Department',hue='Attrition',data=data,palette='hot')
plt.title("Attrition w.r.t Department")
plt.show()
```



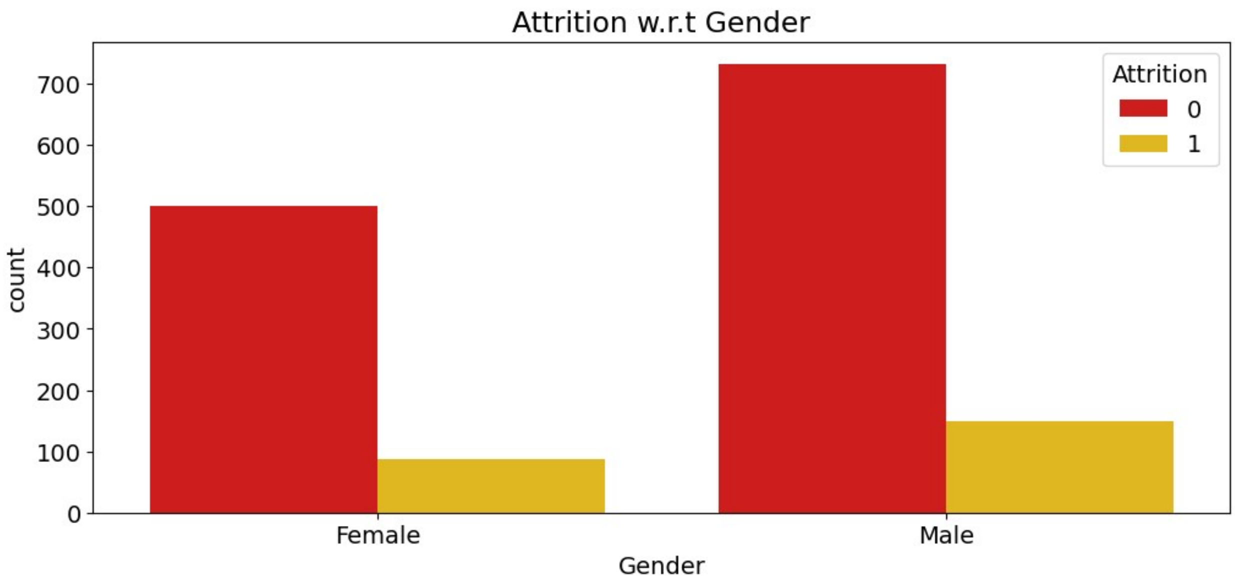
```
#Depatment wrt Attrition
plt.figure(figsize=(12,5))
sns.countplot(x='EducationField',hue='Attrition',data=data,palette='ho
t')
plt.title("Attrition w.r.t EducationField")
plt.xticks(rotation=45)
plt.show()
```



```
#Jobrole
plt.figure(figsize=(12,5))
sns.countplot(x='JobRole',hue='Attrition',data=data,palette='hot')
plt.title("Attrition w.r.t JobRole")
plt.legend(loc='best')
plt.xticks(rotation=45)
plt.show()
```



```
#Gender

plt.figure(figsize=(12,5))
sns.countplot(x='Gender',hue='Attrition',data=data,palette='hot')
plt.title("Attrition w.r.t Gender")
plt.xticks()
plt.show()
```



```
data['Gender'].value_counts()
```

```
Male      882
Female    588
Name: Gender, dtype: int64
```

```
data['count']=1
```

```
data.groupby(["Gender","Attrition"]).agg({"count":"sum"})
```

```
                  count
Gender Attrition
Female 0            501
       1             87
Male   0            732
       1            150
```

```
87/(501+87)
```

```
0.14795918367346939
```

As There is 14% Attrition in Female

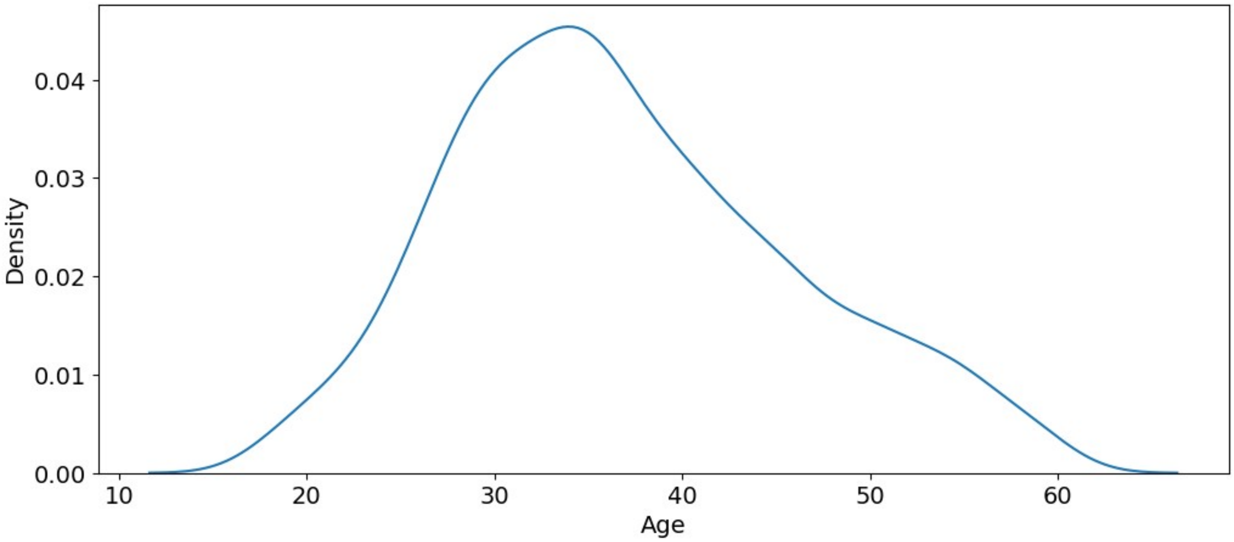
```
150/(732+150)
```

```
0.17006802721088435
```

As there is 17% Attrition in Male

```
#Droping the column which was created for exter analytics part
data.drop(columns = ["count"],inplace =True)
```

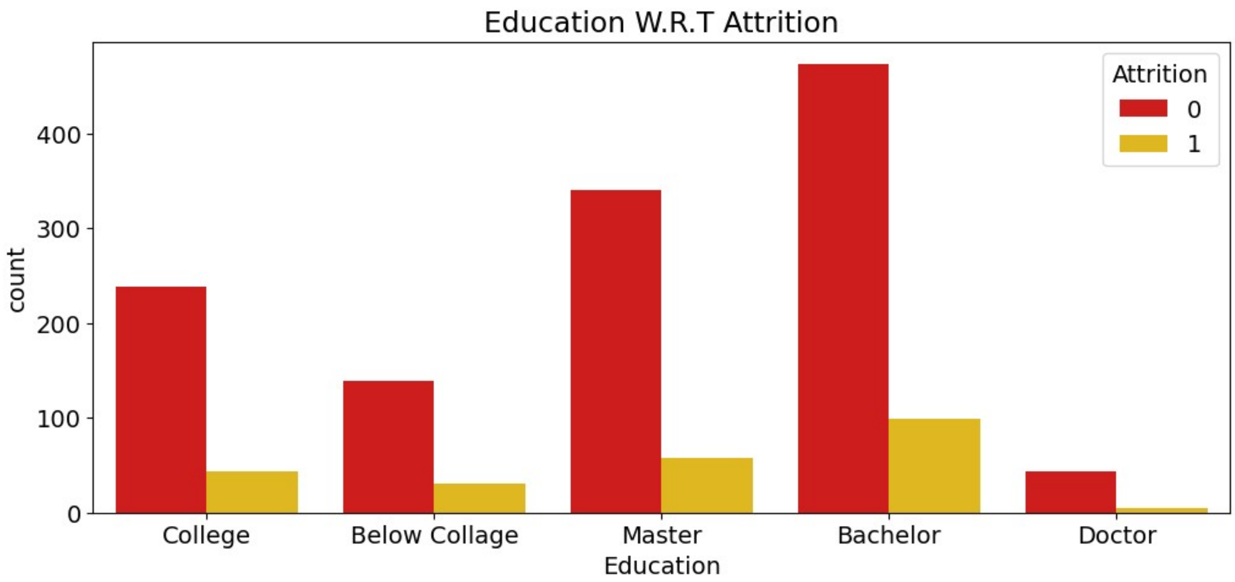
```
#Distribution of age
plt.figure(figsize=(12,5))
sns.distplot(data['Age'],hist=False)
plt.show()
```



```
#This all factor are in numerical form
#Remove leading and trailing whitespaces in column names and correct
typos
ordinal_features = ['Education',
           'EmployeeCount',
           'EnvironmentSatisfaction',
           'JobInvolvement',
           'JobLevel',
           'JobSatisfaction',
           'NumCompaniesWorked',
           'PerformanceRating',
           'RelationshipSatisfaction',
           'StandardHours',
           'StockOptionLevel',
           'TrainingTimesLastYear',
           'WorkLifeBalance']
#Select the columns from the DataFrame
data[ordinal_features].head()
```

```
   Education  EmployeeCount  EnvironmentSatisfaction
JobInvolvement  \
0          2              1                        2                3

1          1              1                        3                2

2          2              1                        4                2
```

```
3          4              1                        4               3

4          1              1                        1               3


   JobLevel  JobSatisfaction  NumCompaniesWorked  PerformanceRating  \
0         2                4                   8                  3
1         2                2                   1                  4
2         1                3                   6                  3
3         1                3                   1                  3
4         1                2                   9                  3

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

   TrainingTimesLastYear  WorkLifeBalance
0                      0                1
1                      3                3
2                      3                3
3                      3                3
4                      3                3

#This all factor are in non-numerical form
edu_map = {1:'Below
Collage',2:'College',3:'Bachelor',4:'Master',5:'Doctor'}
plt.figure(figsize=(12,5))
sns.countplot(x=data['Education'].map(edu_map),hue='Attrition',data=da
ta,palette='hot')
plt.title("Education W.R.T Attrition")
plt.show()
```



```
#List of columns you want to get value counts for
columns_to_count = ['BusinessTravel', 'Department', 'EducationField',
'JobRole', 'MaritalStatus']

#Loop through each column and print its value counts
for column in columns_to_count:
    counts = data[column].value_counts()
    print(f"Value counts for {column}:\{counts}\n")

Value counts for BusinessTravel:\2.0    1043
0.0     150
Name: BusinessTravel, dtype: int64

Value counts for Department:\2    961
1    446
0     63
Name: Department, dtype: int64

Value counts for EducationField:\Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: EducationField, dtype: int64

Value counts for JobRole:\Sales Executive              326
Research Scientist           292
Laboratory Technician        259
Manufacturing Director       145
Healthcare Representative    131
Manager                      102
```

```
Sales Representative          83
Research Director             80
Human Resources               52
Name: JobRole, dtype: int64

Value counts for MaritalStatus:\Married     673
Single      470
Divorced    327
Name: MaritalStatus, dtype: int64
```

```
#Target Variable (Attrition)
data['Attrition'] = data['Attrition'].replace({'No':0,'Yes':1})

#Encode Binary Variable
data['OverTime']=data['OverTime'].map({'No':0,'Yes':1})
#Map 'Male' to 0 and 'Female' to 1
data['Gender'] = data['Gender'].map({'Male': 0, 'Female': 1})

#Encode Categorical Column which are ordinal, use LabelEncoding
#apple Label Encoder to df_categorical

from sklearn.preprocessing import LabelEncoder

encoding_cols = ['BusinessTravel', 'Department', 'EducationField',
'JobRole', 'MaritalStatus']

label_encoders = {}
for column in encoding_cols:
    label_encoders[column] = LabelEncoder()
    data[column] = label_encoders[column].fit_transform(data[column])

data.head(10)
```

```
   Age  Attrition  BusinessTravel  DailyRate  Department
DistanceFromHome  \
0   41          1               1       1102           1
1
1   49          0               2        279           2
8
2   37          1               1       1373           2
2
3   33          0               2       1392           2
3
4   27          0               1        591           2
2
5   32          0               2       1005           2
2
6   59          0               1       1324           2
3
7   30          0               1       1358           2
```

```
24
8   38          0               2        216          2
23
9   36          0               1       1299          2
27

   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
5          2               1              1               8
6          3               3              1              10
7          1               1              1              11
8          3               1              1              12
9          3               3              1              13

   EnvironmentSatisfaction  Gender  HourlyRate  JobInvolvement
JobLevel  \
0                        2       1          94               3
2
1                        3       0          61               2
2
2                        4       0          92               2
1
3                        4       1          56               3
1
4                        1       0          40               3
1
5                        4       0          79               3
1
6                        3       1          81               4
1
7                        4       0          67               3
1
8                        4       0          44               2
3
9                        3       0          94               3
2

   JobRole  JobSatisfaction  MaritalStatus  MonthlyIncome  MonthlyRate
\
0        7                4              2           5993        19479

1        6                2              1           5130        24907

2        2                3              2           2090         2396

3        6                3              1           2909        23159
```

```
4        2                2              1           3468        16632

5        2                4              2           3068        11864

6        2                1              1           2670         9964

7        2                3              0           2693        13335

8        4                3              2           9526         8787

9        0                3              1           5237        16577

   NumCompaniesWorked Over18  OverTime  PercentSalaryHike
PerformanceRating  \
0                   8      Y         1                 11
3
1                   1      Y         0                 23
4
2                   6      Y         1                 15
3
3                   1      Y         1                 11
3
4                   9      Y         0                 12
3
5                   0      Y         0                 13
3
6                   4      Y         1                 20
4
7                   1      Y         0                 22
4
8                   0      Y         0                 21
4
9                   6      Y         0                 13
3

   RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0                         1             80                 0
1                         4             80                 1
2                         2             80                 0
3                         3             80                 0
4                         4             80                 1
5                         3             80                 0
6                         1             80                 3
7                         2             80                 1
8                         2             80                 0
9                         2             80                 2

   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
5                  8                      2                2
7
6                 12                      3                2
1
7                  1                      2                3
1
8                 10                      2                3
9
9                 17                      3                2
7

   YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                   4                        0                     5
1                   7                        1                     7
2                   0                        0                     0
3                   7                        3                     0
4                   2                        2                     2
5                   7                        3                     6
6                   0                        0                     0
7                   0                        0                     0
8                   7                        1                     8
9                   7                        7                     7
```

```
data['Over18']=data['Over18'].map({'Y':1,'No':0})

data.head(10)
```

```
   Age  Attrition  BusinessTravel  DailyRate  Department  
DistanceFromHome  \
0   41          1               1       1102           1
1
1   49          0               2        279           2
8
2   37          1               1       1373           2
2
3   33          0               2       1392           2
3
4   27          0               1        591           2
2
```

```
5   32          0               2       1005           2
2
6   59          0               1       1324           2
3
7   30          0               1       1358           2
24
8   38          0               2        216           2
23
9   36          0               1       1299           2
27

   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
5          2               1              1               8
6          3               3              1              10
7          1               1              1              11
8          3               1              1              12
9          3               3              1              13

   EnvironmentSatisfaction  Gender  HourlyRate  JobInvolvement
JobLevel  \
0                        2       1          94               3
2
1                        3       0          61               2
2
2                        4       0          92               2
1
3                        4       1          56               3
1
4                        1       0          40               3
1
5                        4       0          79               3
1
6                        3       1          81               4
1
7                        4       0          67               3
1
8                        4       0          44               2
3
9                        3       0          94               3
2

   JobRole  JobSatisfaction  MaritalStatus  MonthlyIncome  MonthlyRate
\
0        7                4              2           5993        19479
```

```
1        6                2              1           5130        24907

2        2                3              2           2090         2396

3        6                3              1           2909        23159

4        2                2              1           3468        16632

5        2                4              2           3068        11864

6        2                1              1           2670         9964

7        2                3              0           2693        13335

8        4                3              2           9526         8787

9        0                3              1           5237        16577


   NumCompaniesWorked  Over18  OverTime  PercentSalaryHike
PerformanceRating  \
0                   8       1         1                 11
3
1                   1       1         0                 23
4
2                   6       1         1                 15
3
3                   1       1         1                 11
3
4                   9       1         0                 12
3
5                   0       1         0                 13
3
6                   4       1         1                 20
4
7                   1       1         0                 22
4
8                   0       1         0                 21
4
9                   6       1         0                 13
3

   RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0                         1             80                 0
1                         4             80                 1
2                         2             80                 0
3                         3             80                 0
4                         4             80                 1
5                         3             80                 0
6                         1             80                 3
```

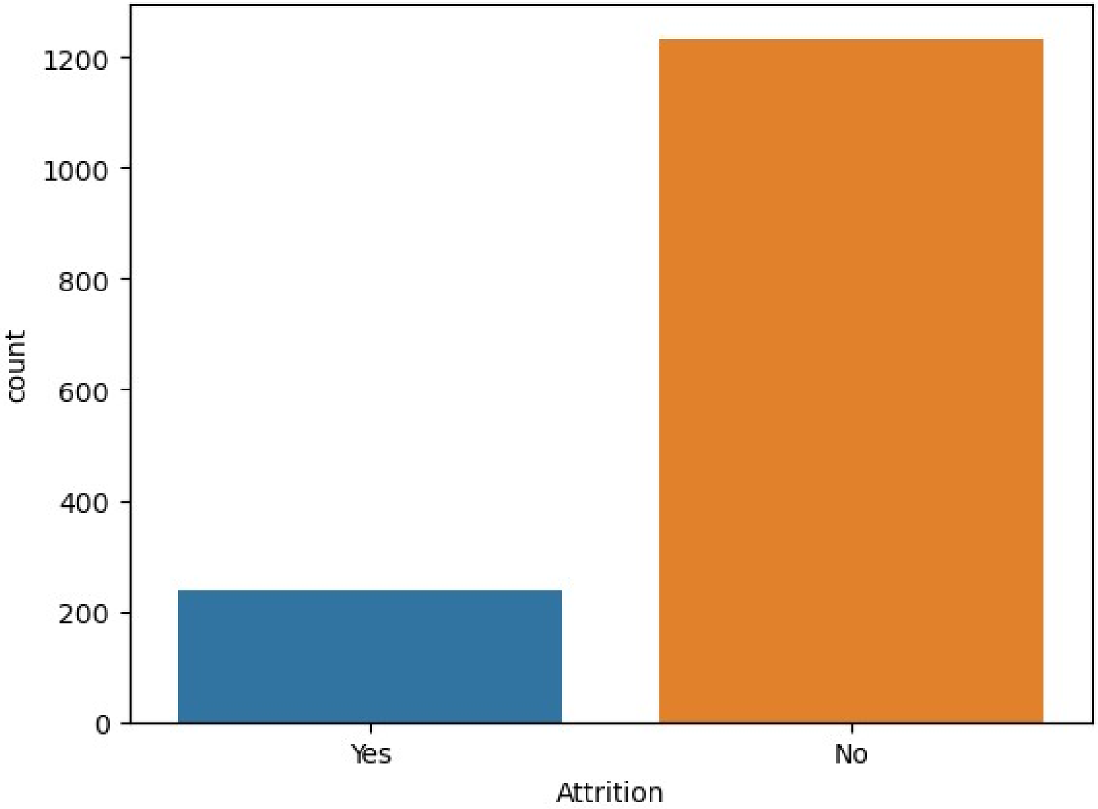
```
7                         2             80                 1
8                         2             80                 0
9                         2             80                 2

   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
5                  8                      2                2
7
6                 12                      3                2
1
7                  1                      2                3
1
8                 10                      2                3
9
9                 17                      3                2
7

   YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                   4                        0                     5
1                   7                        1                     7
2                   0                        0                     0
3                   7                        3                     0
4                   2                        2                     2
5                   7                        3                     6
6                   0                        0                     0
7                   0                        0                     0
8                   7                        1                     8
9                   7                        7                     7
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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   int64
 2   BusinessTravel            1470 non-null   int64
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   int64
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   int32
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   int64
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   int32
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   int32
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   int64
 22  OverTime                  1470 non-null   int64
 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: int32(3), int64(32)
memory usage: 396.2 KB
```

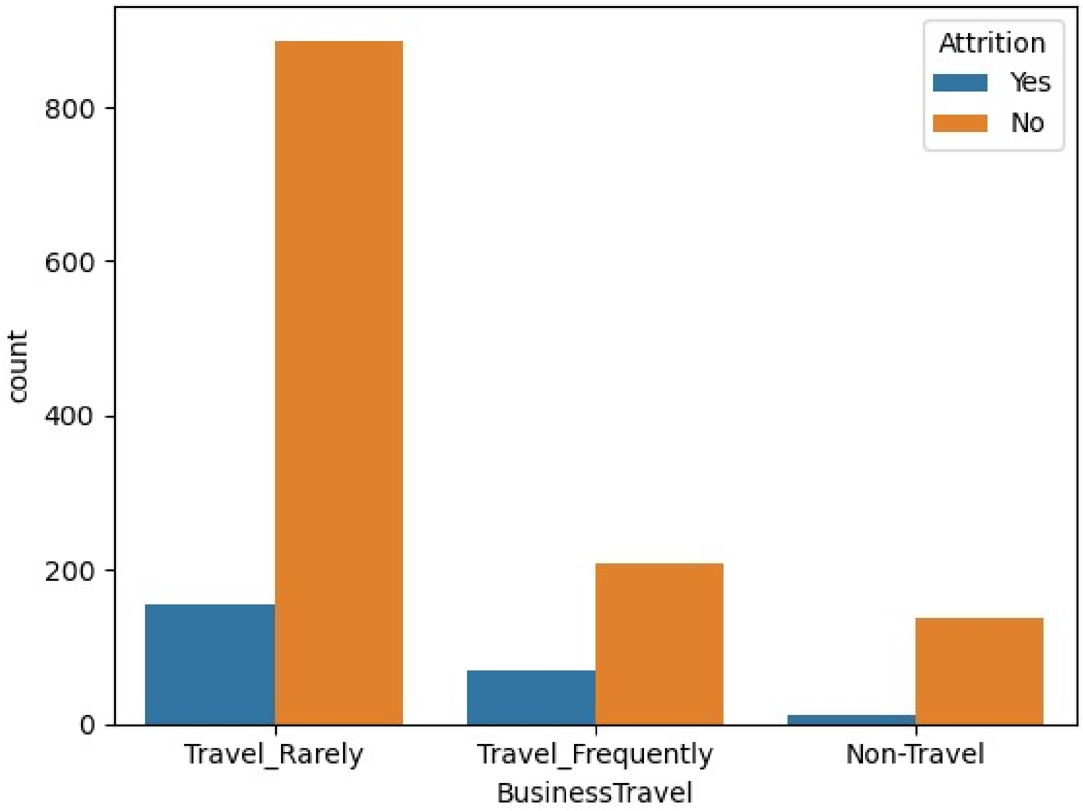
```
x = data.drop(['Attrition','Over18'],axis=1)  #Input Features
y = data['Attrition'].values    #Output Features
```

```
sns.countplot(x=data.Attrition)
plt.show()
```



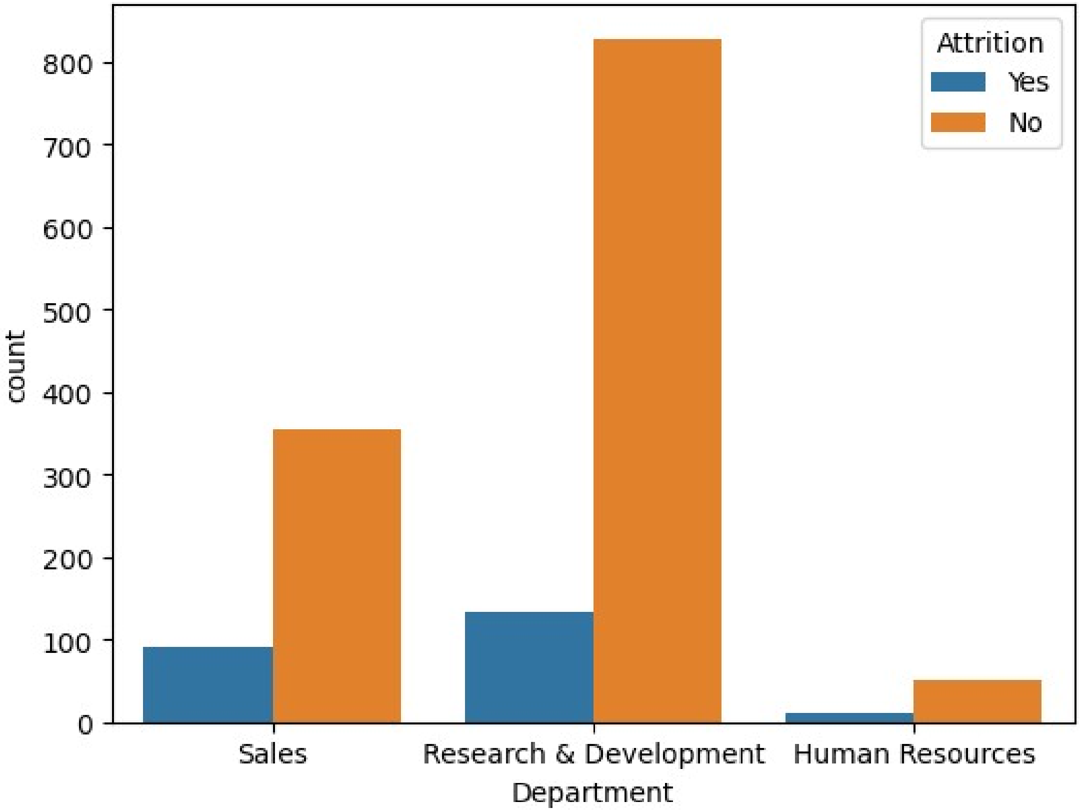
1. Impact Of Business Travel On Attrition

```
sns.countplot(hue=data.Attrition,x=data.BusinessTravel)
plt.show()
```



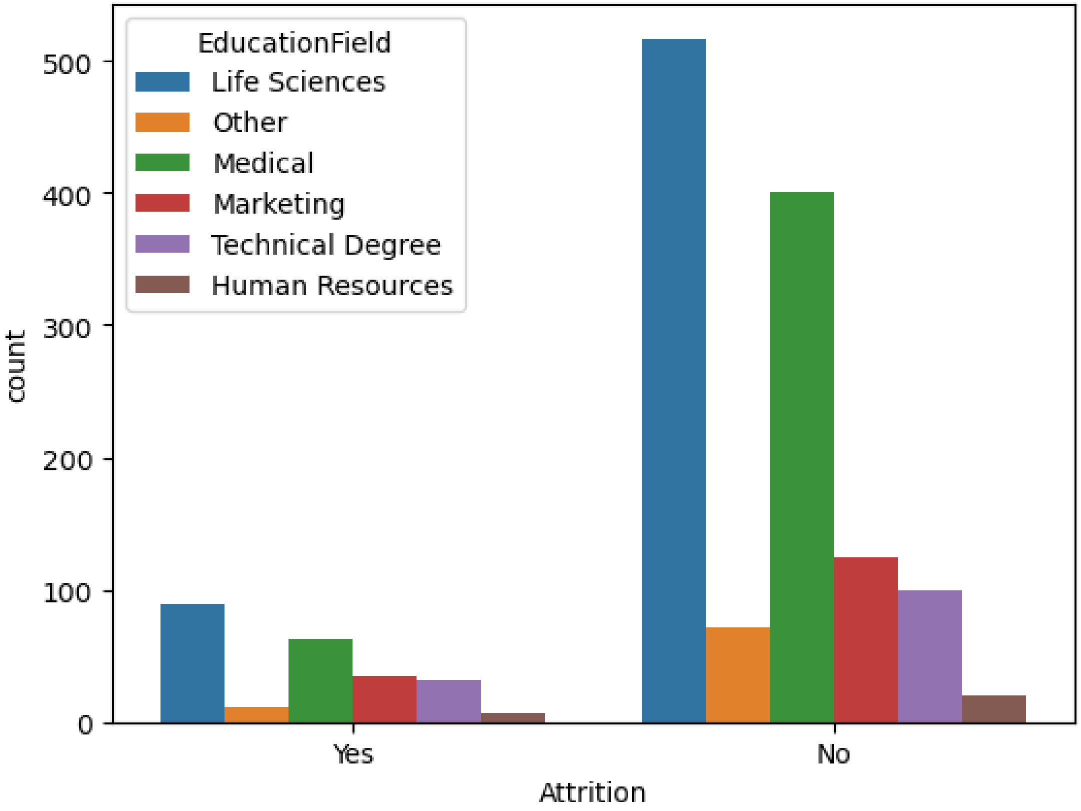
1. Impact Of Department On Attrition

```
sns.countplot(hue=data.Attrition,x=data.Department)
plt.show()
```



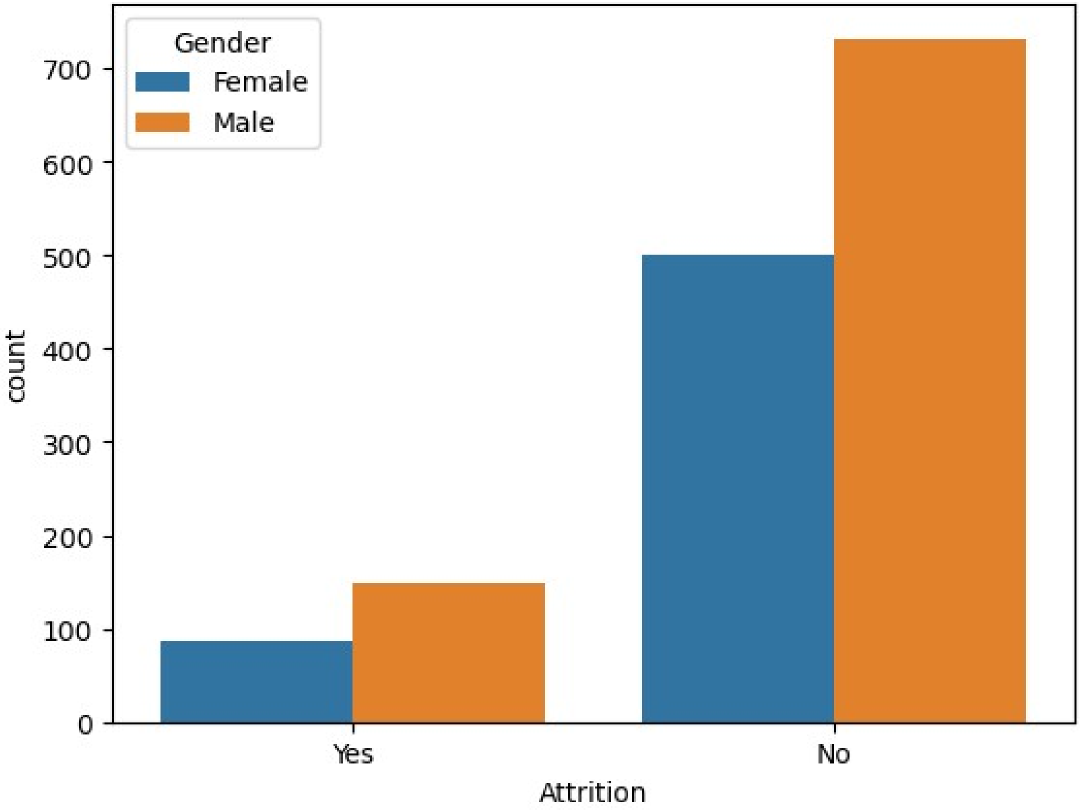
1. Impact Of Education Field On Attrition

```
sns.countplot(x=data.Attrition,hue=data.EducationField)
plt.show()
```

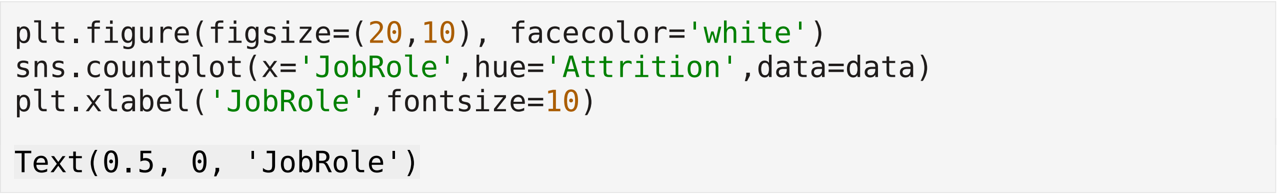


1. Gender And Attrition

```
sns.countplot(x=data.Attrition,hue=data.Gender)
plt.show()
```

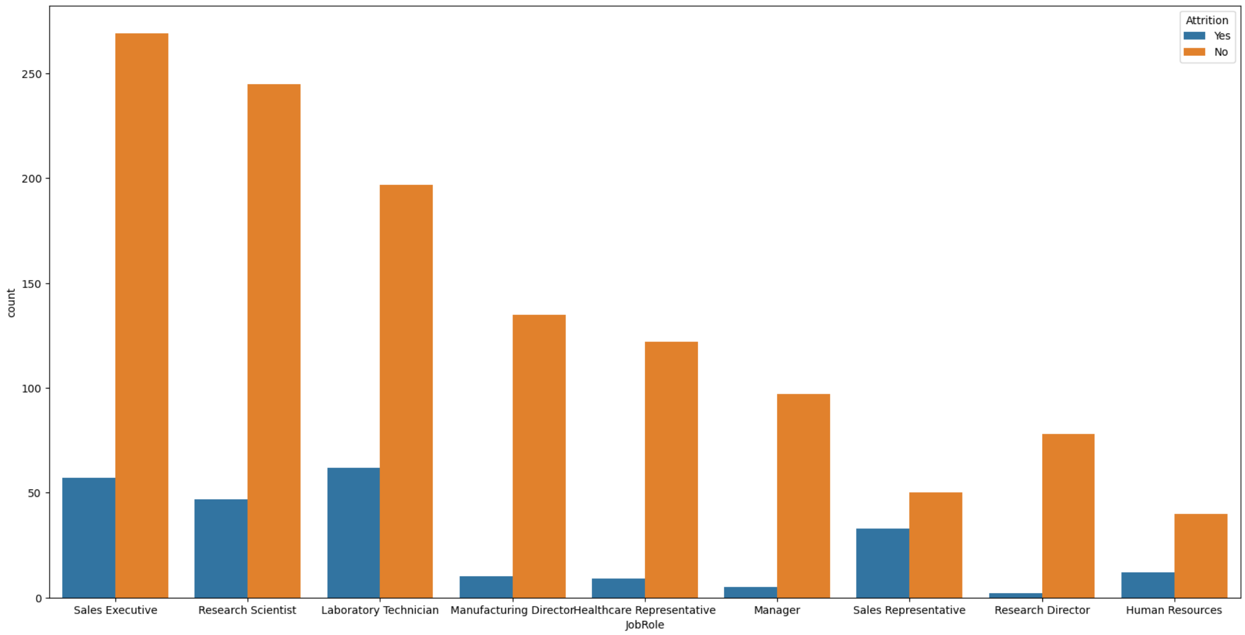


1. Impact Of Job Role On Attrition



```
plt.figure(figsize=(20,10), facecolor='white')
sns.countplot(x='JobRole',hue='Attrition',data=data)
plt.xlabel('JobRole',fontsize=10)
```

```
Text(0.5, 0, 'JobRole')
```



Analysis On Continuos Data With Respect To Target Column

```
numerical_col = []
for column in data.columns:
    if data[column].dtype == "int64" and len(data[column].unique())>=
10:

        numerical_col.append(column)

numerical_col
```

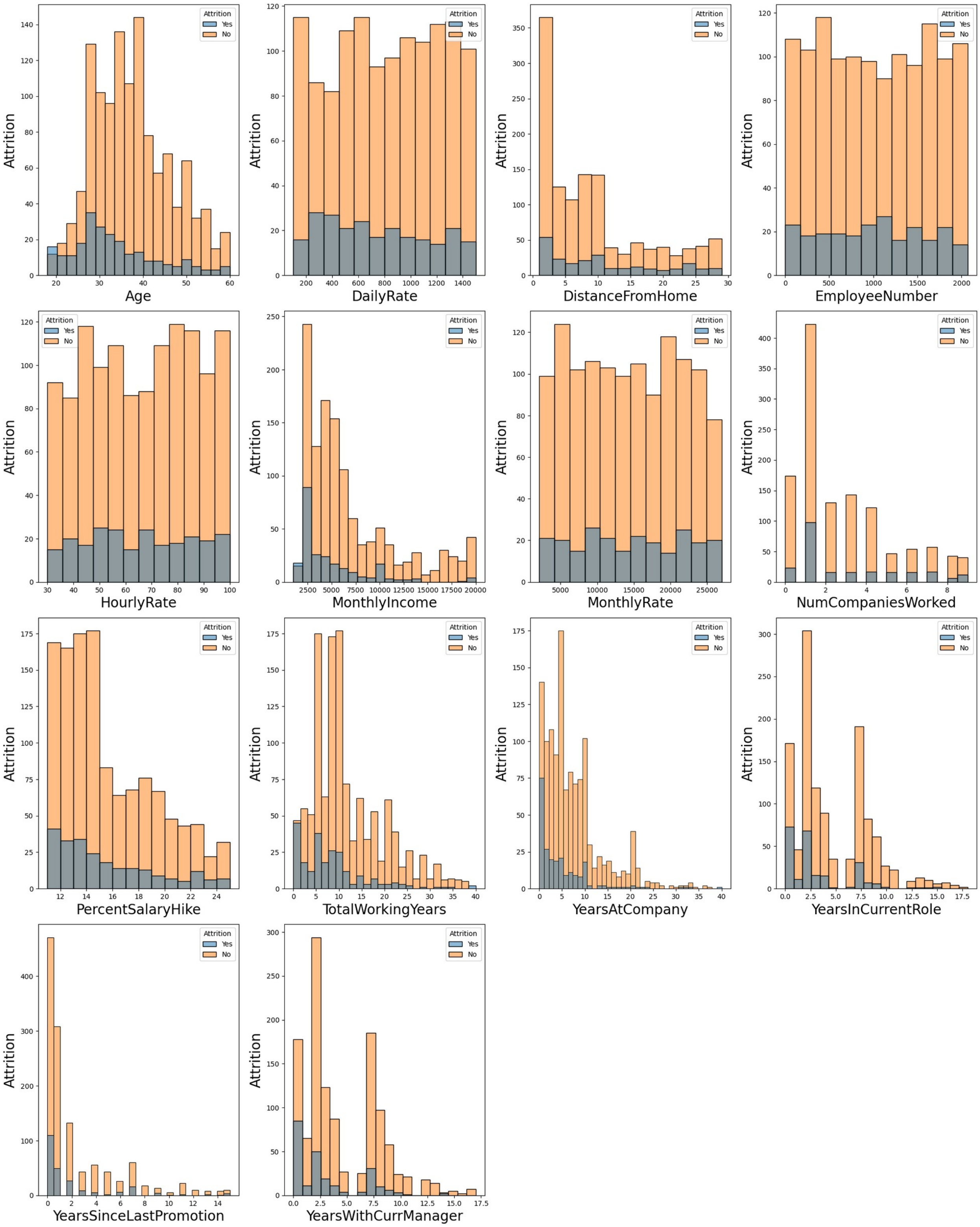
```
['Age',
 'DailyRate',
 'DistanceFromHome',
 'EmployeeNumber',
 'HourlyRate',
 'MonthlyIncome',
 'MonthlyRate',
 'NumCompaniesWorked',
 'PercentSalaryHike',
 'TotalWorkingYears',
 'YearsAtCompany',
 'YearsInCurrentRole',
 'YearsSinceLastPromotion',
 'YearsWithCurrManager']
```

Graphical Representation Of Continuous Data

```
data2=data[['Age',
           'DailyRate',
           'DistanceFromHome',
           'EmployeeNumber',
           'HourlyRate',
           'MonthlyIncome',
           'MonthlyRate',
           'NumCompaniesWorked',
           'PercentSalaryHike',
           'TotalWorkingYears',
           'YearsAtCompany',
           'YearsInCurrentRole',
           'YearsSinceLastPromotion',
           'YearsWithCurrManager']]
```

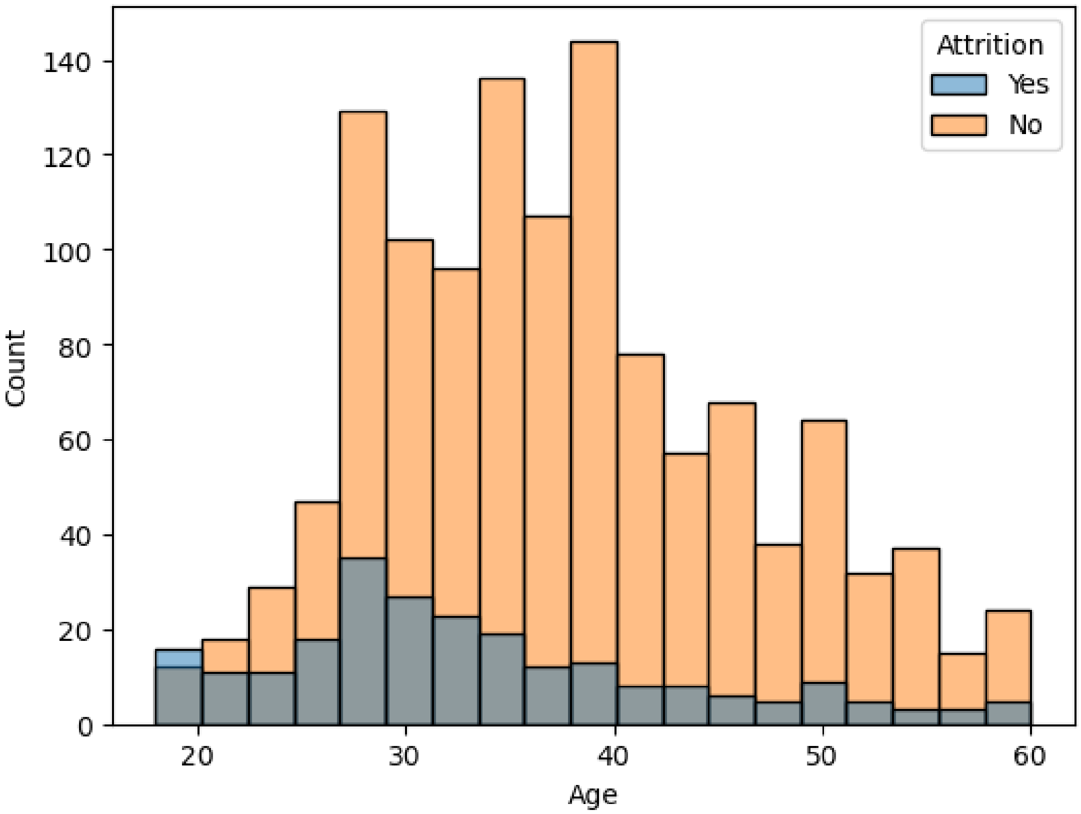
Data Visualization

```
plt.figure(figsize=(20,25), facecolor='white')
plotnumber = 1

for column in data2:
    if plotnumber<=16 :
        ax = plt.subplot(4,4,plotnumber)
        sns.histplot(x=data2[column].dropna(axis=0)
                    ,hue=data.Attrition)
        plt.xlabel(column,fontsize=20)
        plt.ylabel('Attrition',fontsize=20)
    plotnumber+=1
plt.tight_layout()
```



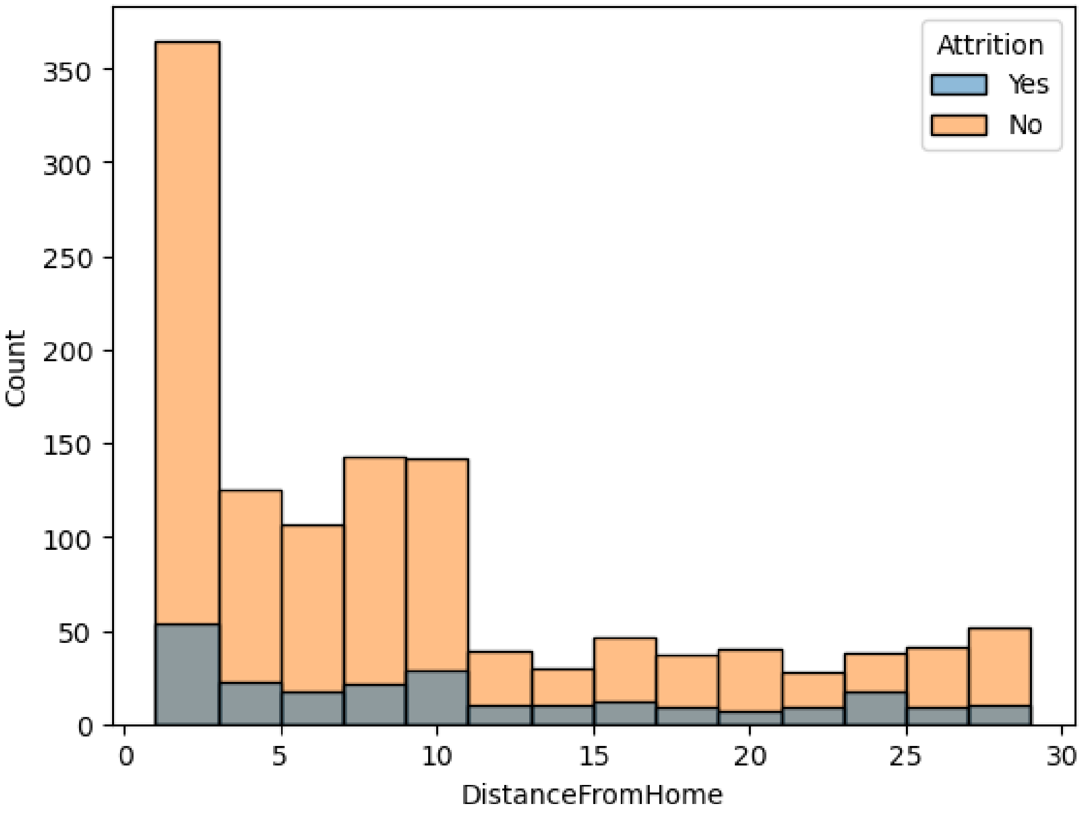
1. Impact Of Age On Attrition

```
sns.histplot(hue=data.Attrition,x=data.Age)
plt.show()
```



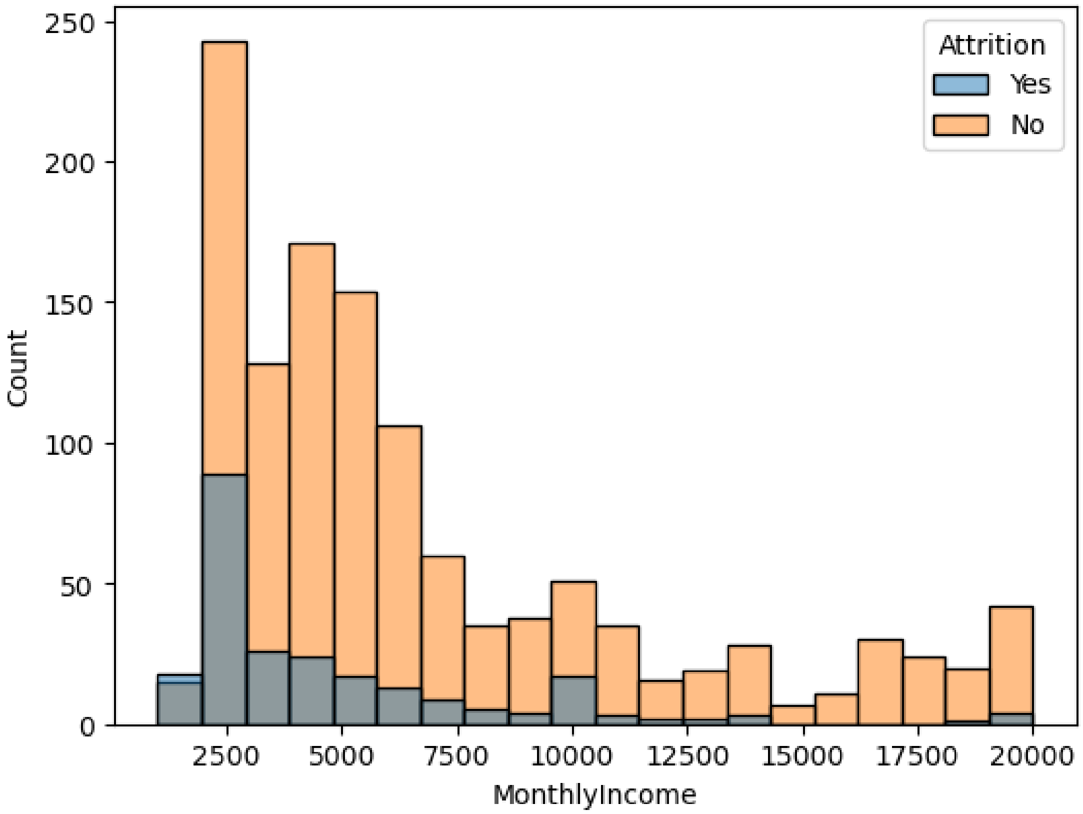
1. Distance From Home And Attrition

```
sns.histplot(hue=data.Attrition,x=data.DistanceFromHome)
plt.show()
```



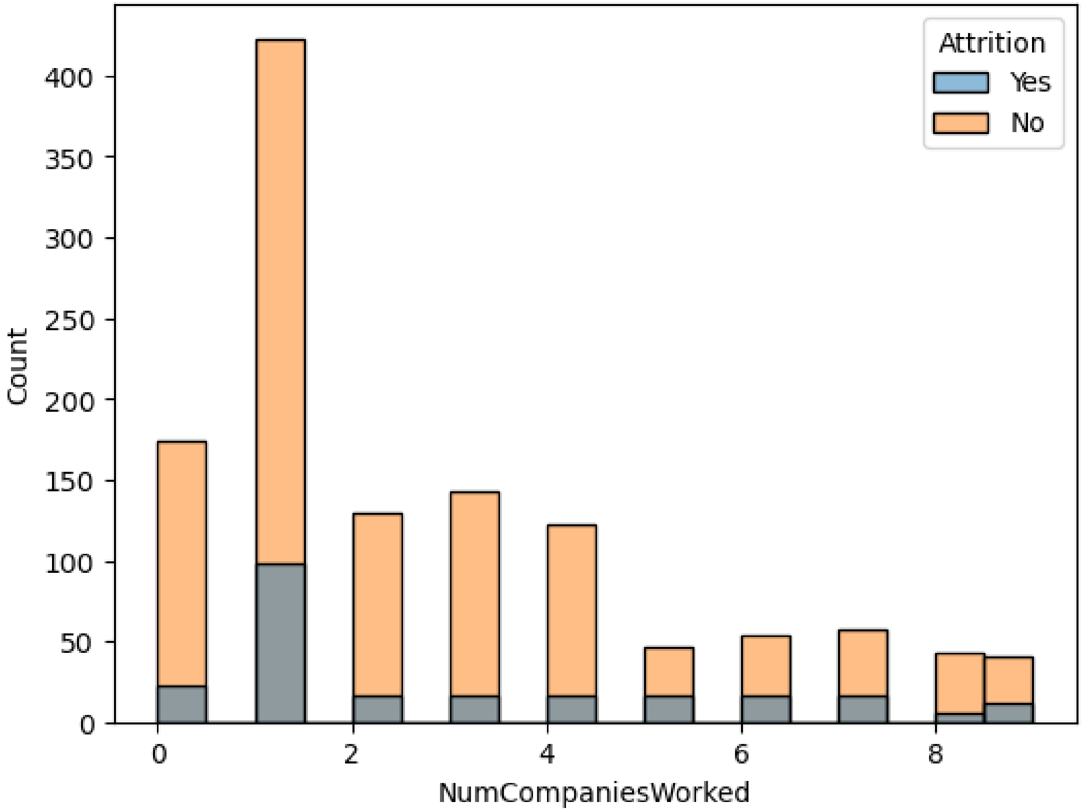
1. How Monthly Income Gives Trends w.r.t Attrition

```
sns.histplot(x=data.MonthlyIncome,hue=data.Attrition)
plt.show()
```



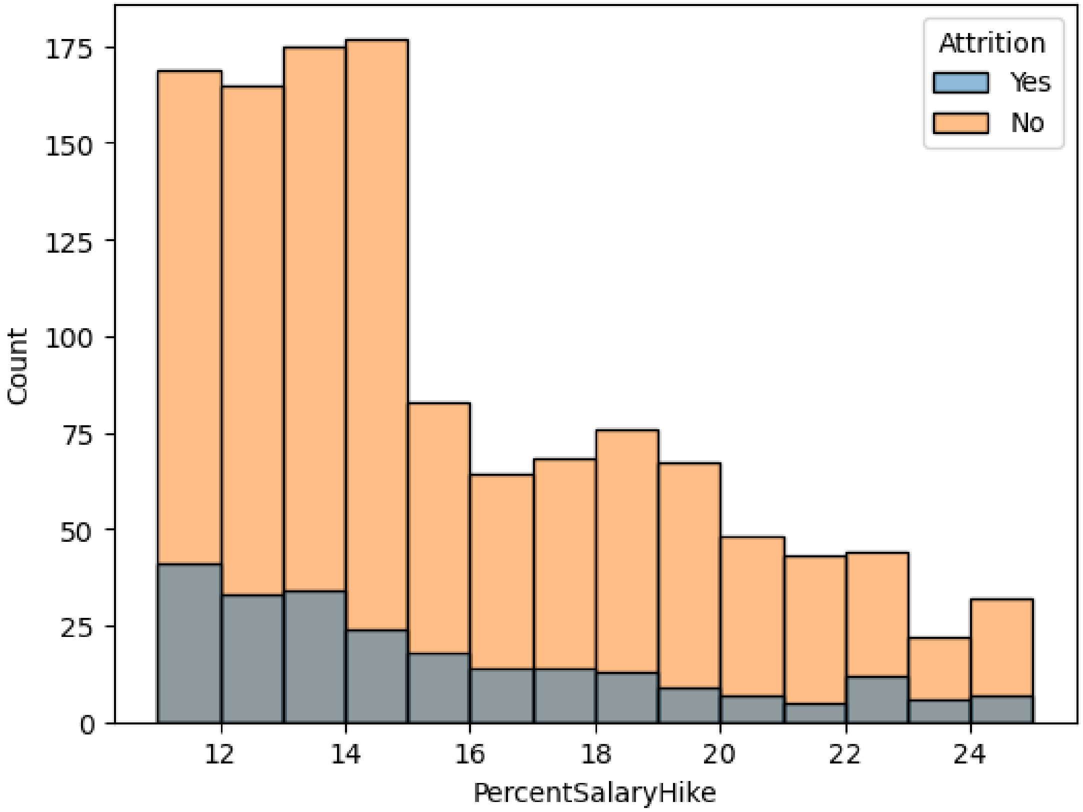
1. Impact Of No. Of Companies Worked

```
sns.histplot(hue=data.Attrition,x=data.NumCompaniesWorked)
plt.show()
```



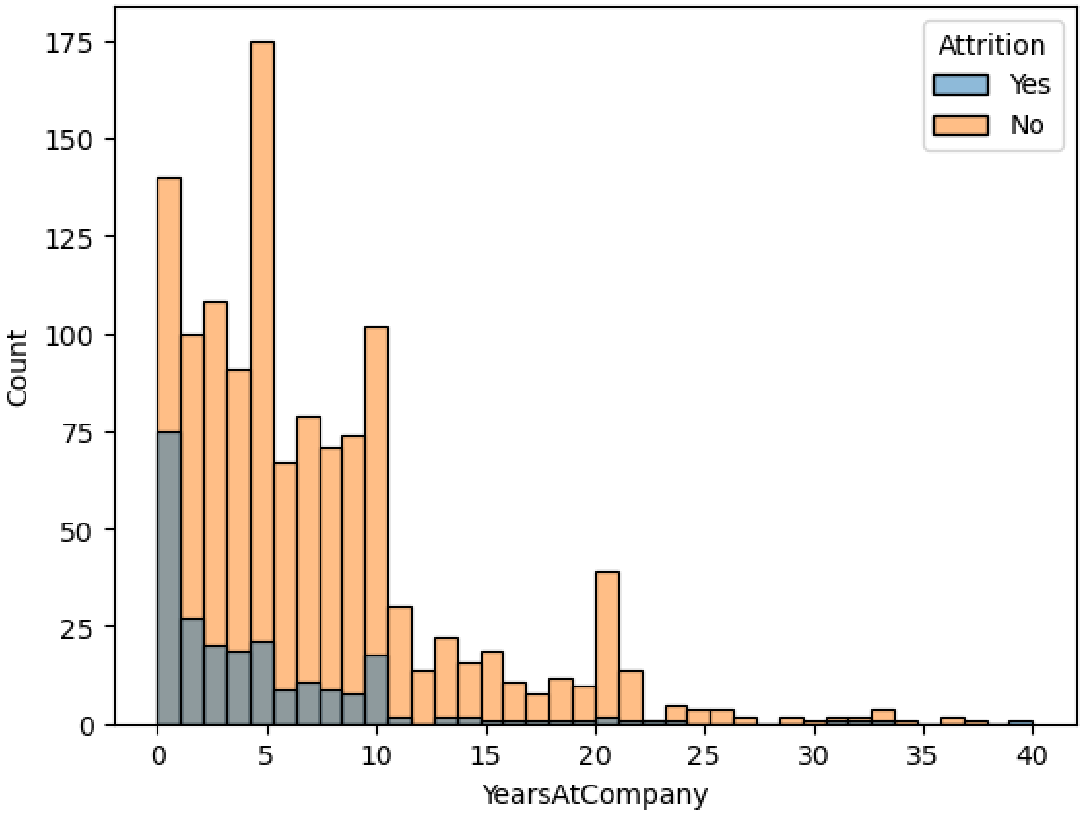
1. How Salary Hike Is Impacting The Attrition

```
sns.histplot(hue=data.Attrition,x=data.PercentSalaryHike)
plt.show()
```



1. Years At The Company

```
sns.histplot(hue=data.Attrition,x=data.YearsAtCompany)
plt.show()
```

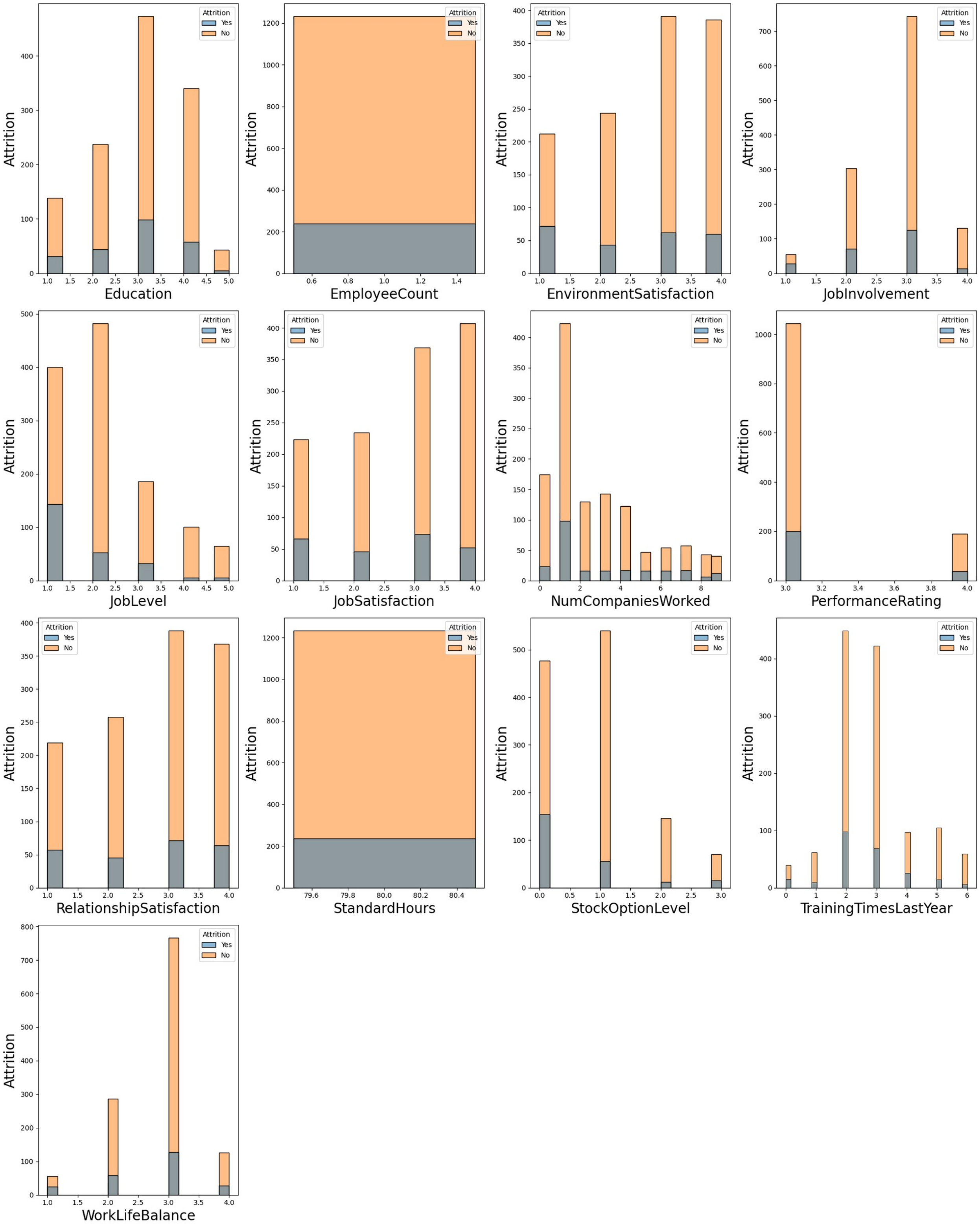


Analysis Of Discrete Data W.R.T Target Column

```
discrete_col = []
for column in data.columns:
    if data[column].dtype == "int64" and len(data[column].unique()) <=
10:
        discrete_col.append(column)

data3=data[['Education',
           'EmployeeCount',
           'EnvironmentSatisfaction',
           'JobInvolvement',
           'JobLevel',
           'JobSatisfaction',
           'NumCompaniesWorked',
           'PerformanceRating',
           'RelationshipSatisfaction',
           'StandardHours',
           'StockOptionLevel',
           'TrainingTimesLastYear',
           'WorkLifeBalance']]
```

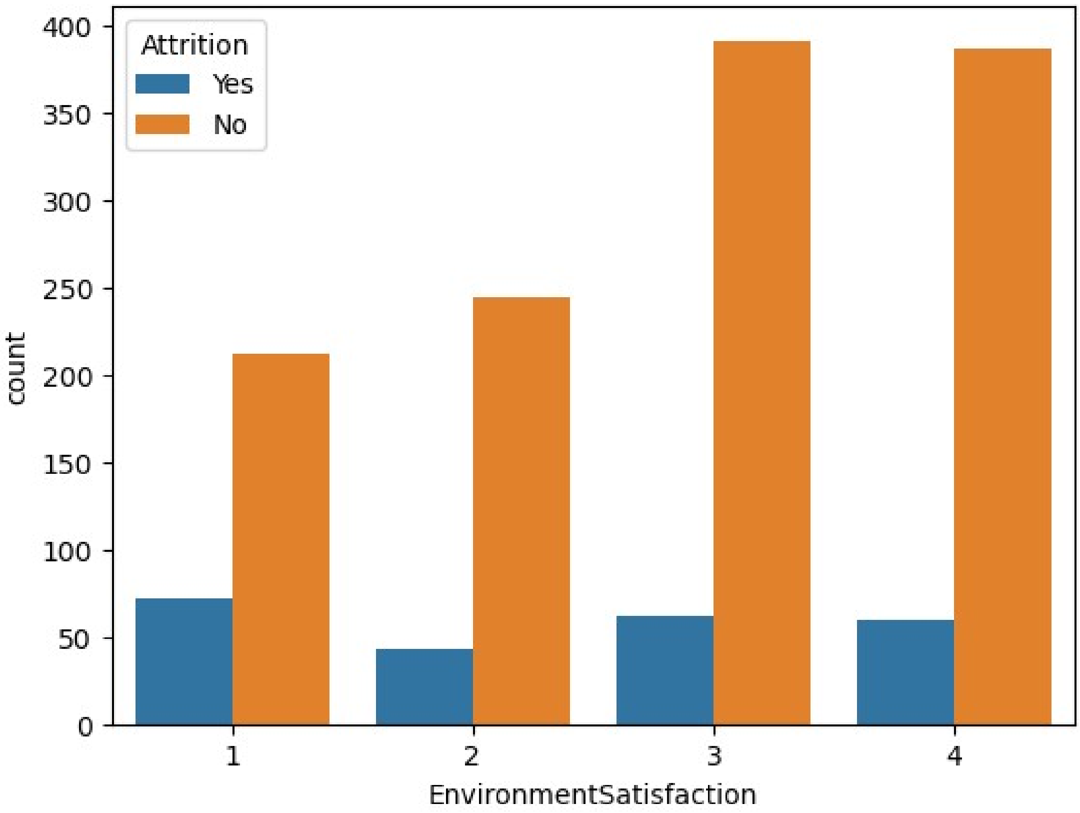
Graphical Representation

```python
plt.figure(figsize=(20,25), facecolor='white')
plotnumber = 1

for column in data3:
    if plotnumber<=16 :
        ax = plt.subplot(4,4,plotnumber)
        sns.histplot(x=data3[column].dropna(axis=0)
                    ,hue=data.Attrition)
        plt.xlabel(column,fontsize=20)
        plt.ylabel('Attrition',fontsize=20)
    plotnumber+=1
plt.tight_layout()
```



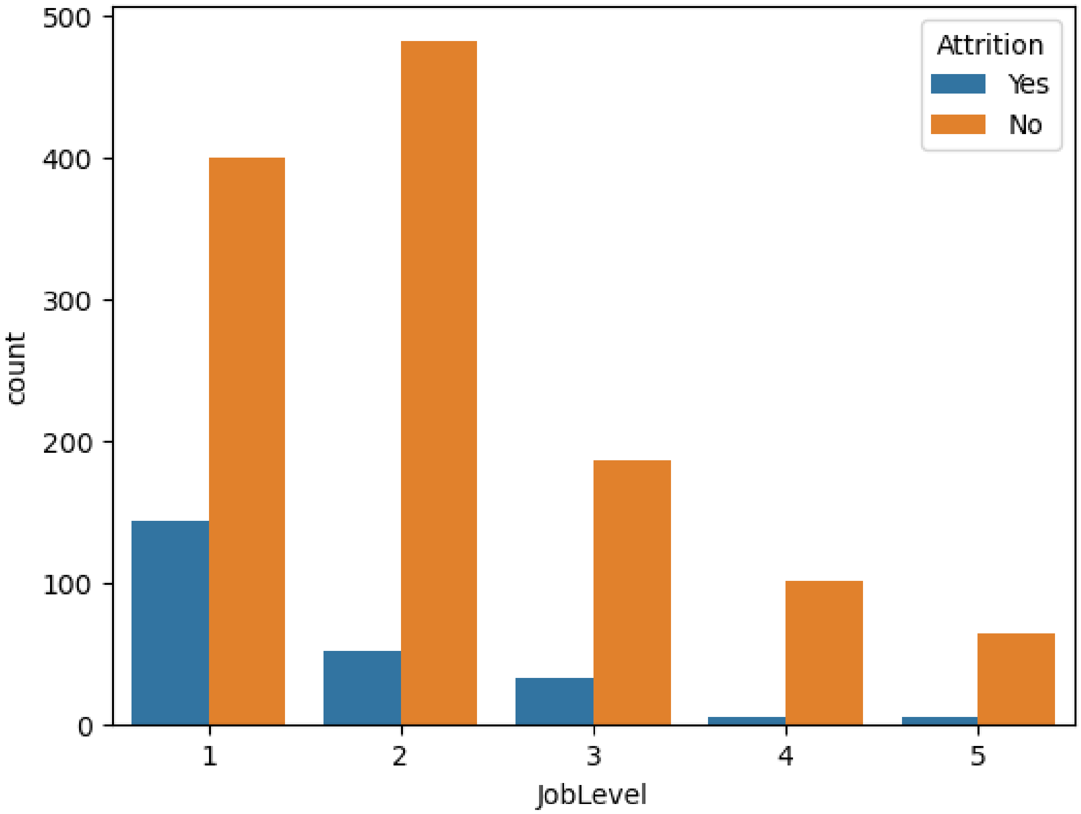
1. Impact Of Environment And Job Satisfaction On Attrition

```
sns.countplot(hue=data.Attrition,x=data.EnvironmentSatisfaction)
plt.show()
```



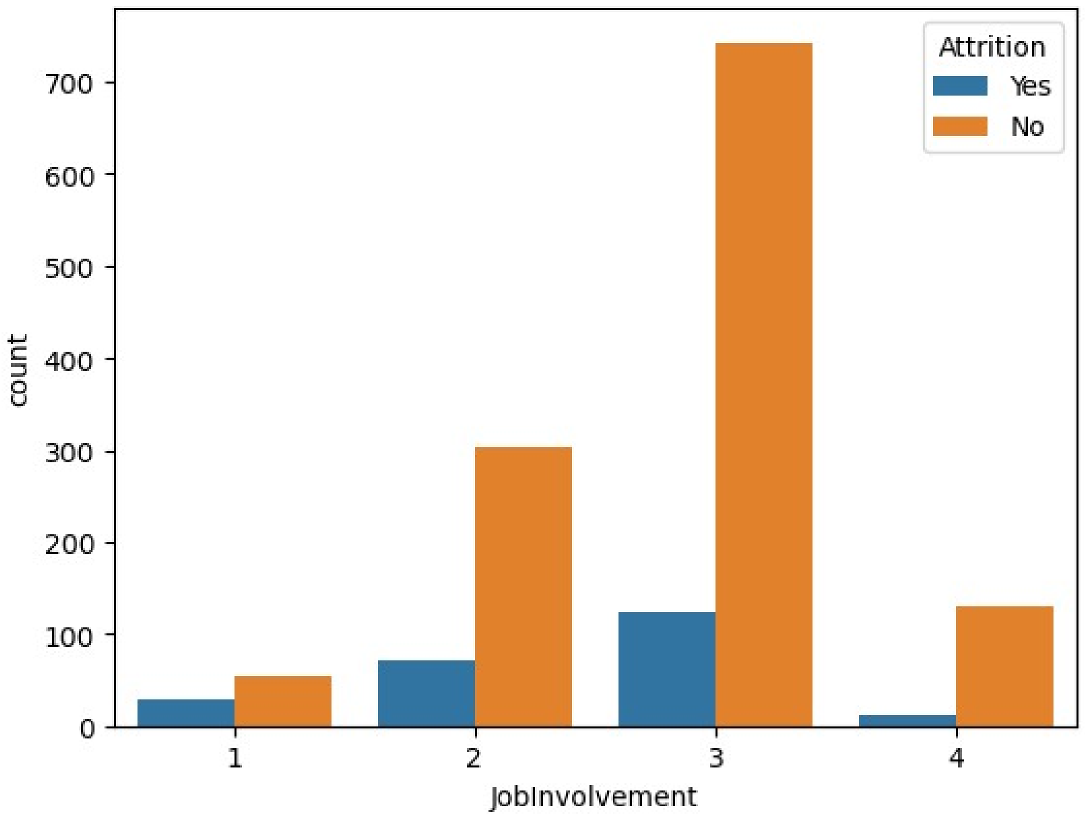
1. How Job Level Is Affecting Attrition

```
sns.countplot(hue=data.Attrition,x=data.JobLevel)
plt.show()
```



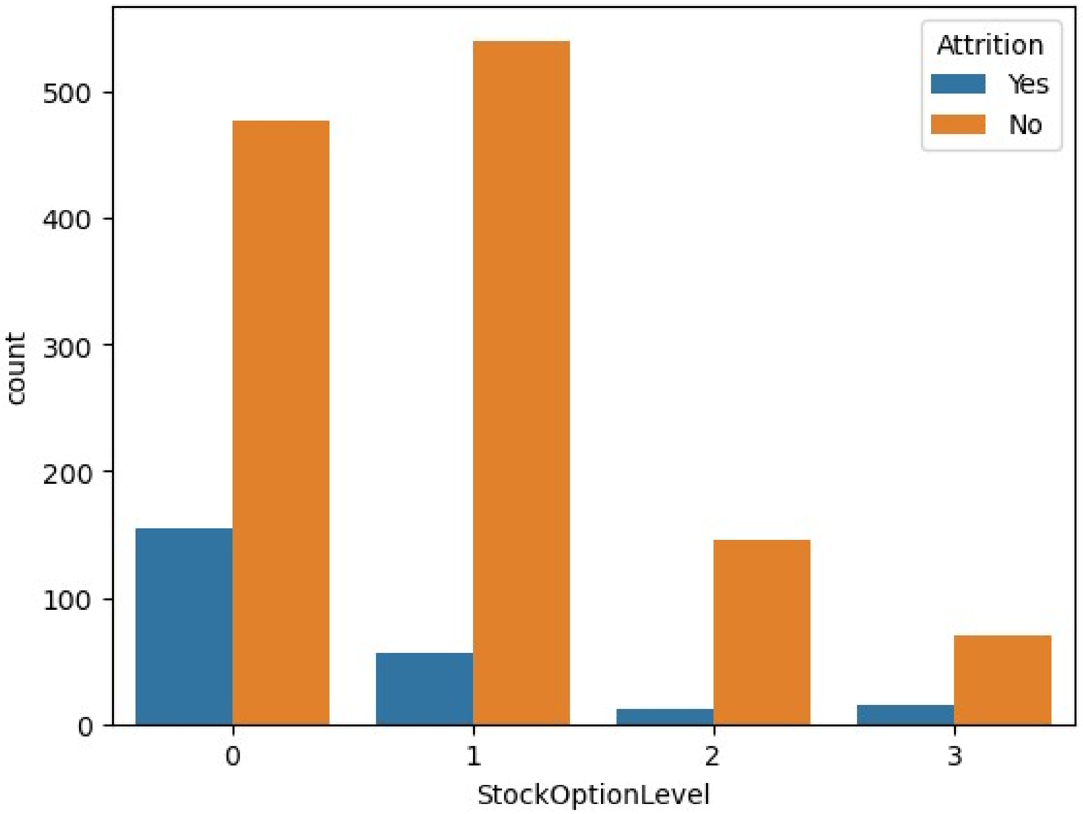
1. Job Involvement Impacting On Attrition

```
sns.countplot(hue=data.Attrition,x=data.JobInvolvement)
plt.show()
```



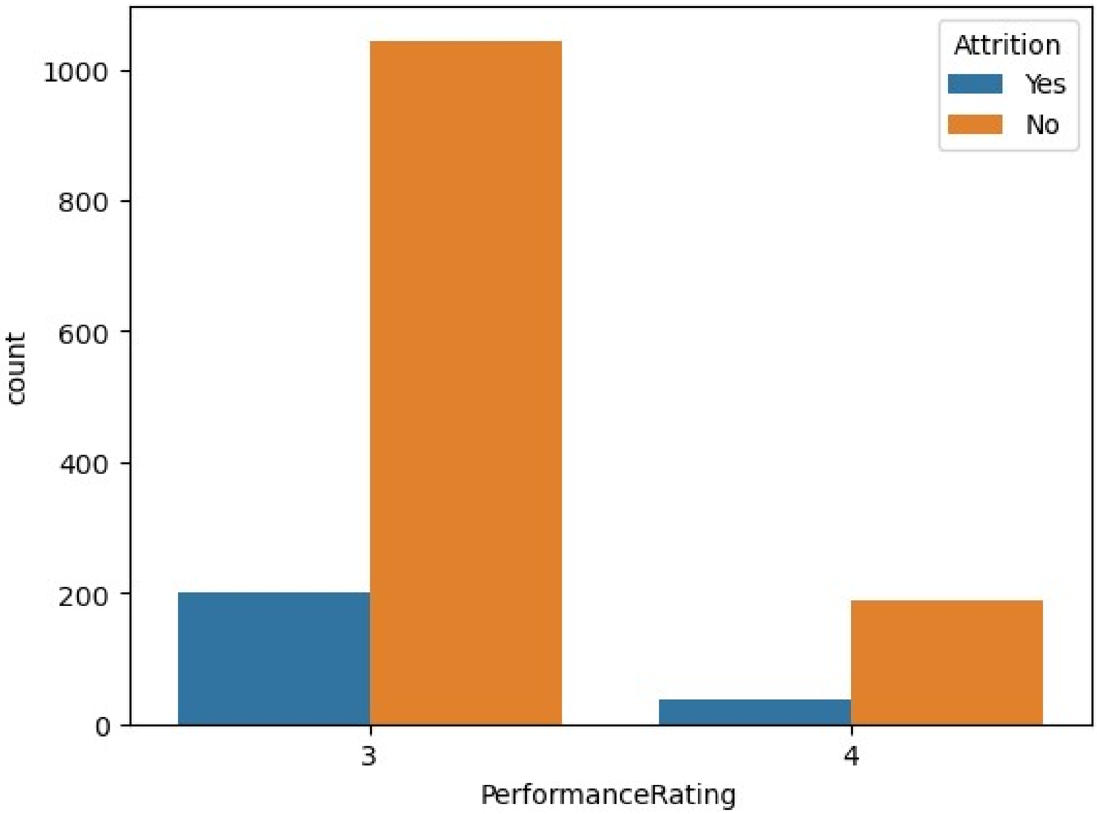
1. Impact Of Stock Option Level On Attrition

```
sns.countplot(hue=data.Attrition,x=data.StockOptionLevel)
plt.show()
```



1. Performance Rating And Attrition

```
sns.countplot(hue=data.Attrition,x=data.PerformanceRating)
plt.show()
```



```
data.isnull().sum()
```

```
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
```

```
print(data.duplicated().value_counts())
data.drop_duplicates(inplace=True)
print(len(data))
```

```
False    1470
dtype: int64
1470
```