# EDA - HR ATTRITION DATASET

November 26, 2022

## 0.1 EXPLORATORY DATA ANALYSIS OF EMPLOYEE ATTRITION; THE HR ATTRITION DATASET

**BY**

### 0.1.1 MINA KUDZAWU

The capacity of an organisation to survive in today's fiercely competitive global economic environment depends in significant part on a number of elements. One of such aspects is an enthusiastic workforce. That is why an extremely high attrition rate among a company's workforce must be cause for concern.

Employee attrition appears to be caused by a variety of factors, including dissatisfaction with compensation or benefits, a lack of opportunity for professional growth, and even unfavourable working circumstances. What are the main elements that affect employee churn, though, is the question.

In order to help executives take the necessary actions to retain the employee, my purpose is to analyse the data set and study the many links that exist between all the different components.

The Dataset In this project, I perform an analysis on a publicly available dataset (https://www.kaggle.com/datasets/itssuru/hr-employee-attrition?select=HR-Employee-Attrition.csv)

The key steps covered in this project are as follows; 1. Importing the relevant libraries 2. Cleaning the data 3. Visualizing relationships

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from IPython import get_ipython
import warnings
warnings.filterwarnings("ignore")
```

[2]:
```
HR_df = pd.read_csv('HR-Employee-Attrition.csv')
HR_df.head()
```

[2]:
```
   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]
```

[3]:
```
HR_df.tail()
```

[3]:
```
      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  \
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

[5 rows x 35 columns]
```

[4]:
```
HR_df.shape
```

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

[9]:
```
HR_df.columns
```

[9]:
```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
```

```
       'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

[11]:
```
HR_df.duplicated().sum()
```

[11]:
```
0
```

[12]:
```
HR_df.isnull().sum()
```

[12]:
```
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
```

[13]:
```
HR_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
```

[15]:
```
HR_df.describe().transpose()
```

[15]:

| | count | mean | std | min | 25% | \ |
|---|---|---|---|---|---|---|
| Age | 1470.0 | 36.923810 | 9.135373 | 18.0 | 30.00 | |
| DailyRate | 1470.0 | 802.485714 | 403.509100 | 102.0 | 465.00 | |
| DistanceFromHome | 1470.0 | 9.192517 | 8.106864 | 1.0 | 2.00 | |
| Education | 1470.0 | 2.912925 | 1.024165 | 1.0 | 2.00 | |
| EmployeeCount | 1470.0 | 1.000000 | 0.000000 | 1.0 | 1.00 | |
| EmployeeNumber | 1470.0 | 1024.865306 | 602.024335 | 1.0 | 491.25 | |
| EnvironmentSatisfaction | 1470.0 | 2.721769 | 1.093082 | 1.0 | 2.00 | |
| HourlyRate | 1470.0 | 65.891156 | 20.329428 | 30.0 | 48.00 | |
| JobInvolvement | 1470.0 | 2.729932 | 0.711561 | 1.0 | 2.00 | |
| JobLevel | 1470.0 | 2.063946 | 1.106940 | 1.0 | 1.00 | |
| JobSatisfaction | 1470.0 | 2.728571 | 1.102846 | 1.0 | 2.00 | |
| MonthlyIncome | 1470.0 | 6502.931293 | 4707.956783 | 1009.0 | 2911.00 | |
| MonthlyRate | 1470.0 | 14313.103401 | 7117.786044 | 2094.0 | 8047.00 | |
| NumCompaniesWorked | 1470.0 | 2.693197 | 2.498009 | 0.0 | 1.00 | |
| PercentSalaryHike | 1470.0 | 15.209524 | 3.659938 | 11.0 | 12.00 | |
| PerformanceRating | 1470.0 | 3.153741 | 0.360824 | 3.0 | 3.00 | |
| RelationshipSatisfaction | 1470.0 | 2.712245 | 1.081209 | 1.0 | 2.00 | |
| StandardHours | 1470.0 | 80.000000 | 0.000000 | 80.0 | 80.00 | |
| StockOptionLevel | 1470.0 | 0.793878 | 0.852077 | 0.0 | 0.00 | |
| TotalWorkingYears | 1470.0 | 11.279592 | 7.780782 | 0.0 | 6.00 | |
| TrainingTimesLastYear | 1470.0 | 2.799320 | 1.289271 | 0.0 | 2.00 | |
| WorkLifeBalance | 1470.0 | 2.761224 | 0.706476 | 1.0 | 2.00 | |
| YearsAtCompany | 1470.0 | 7.008163 | 6.126525 | 0.0 | 3.00 | |
| YearsInCurrentRole | 1470.0 | 4.229252 | 3.623137 | 0.0 | 2.00 | |
| YearsSinceLastPromotion | 1470.0 | 2.187755 | 3.222430 | 0.0 | 0.00 | |
| YearsWithCurrManager | 1470.0 | 4.123129 | 3.568136 | 0.0 | 2.00 | |

| | 50% | 75% | max |
|---|---|---|---|
| Age | 36.0 | 43.00 | 60.0 |
| DailyRate | 802.0 | 1157.00 | 1499.0 |
| DistanceFromHome | 7.0 | 14.00 | 29.0 |
| Education | 3.0 | 4.00 | 5.0 |
| EmployeeCount | 1.0 | 1.00 | 1.0 |
| EmployeeNumber | 1020.5 | 1555.75 | 2068.0 |
| EnvironmentSatisfaction | 3.0 | 4.00 | 4.0 |
| HourlyRate | 66.0 | 83.75 | 100.0 |
| JobInvolvement | 3.0 | 3.00 | 4.0 |
| JobLevel | 2.0 | 3.00 | 5.0 |
| JobSatisfaction | 3.0 | 4.00 | 4.0 |
| MonthlyIncome | 4919.0 | 8379.00 | 19999.0 |
| MonthlyRate | 14235.5 | 20461.50 | 26999.0 |
| NumCompaniesWorked | 2.0 | 4.00 | 9.0 |
| PercentSalaryHike | 14.0 | 18.00 | 25.0 |
| PerformanceRating | 3.0 | 3.00 | 4.0 |
| RelationshipSatisfaction | 3.0 | 4.00 | 4.0 |
| StandardHours | 80.0 | 80.00 | 80.0 |

```
StockOptionLevel              1.0      1.00      3.0
TotalWorkingYears            10.0     15.00     40.0
TrainingTimesLastYear         3.0      3.00      6.0
WorkLifeBalance               3.0      3.00      4.0
YearsAtCompany                5.0      9.00     40.0
YearsInCurrentRole            3.0      7.00     18.0
YearsSinceLastPromotion       1.0      3.00     15.0
YearsWithCurrManager          3.0      7.00     17.0
```

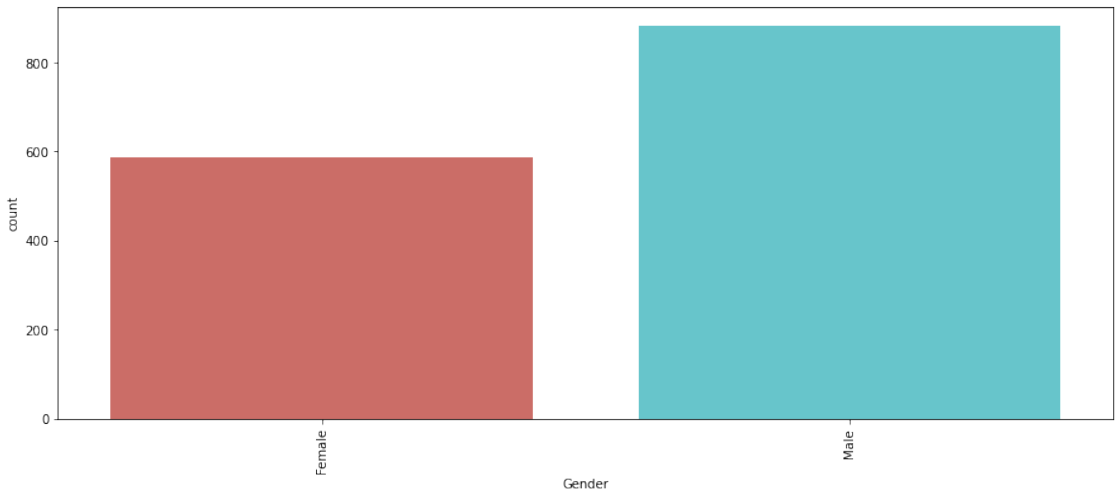
[20]:
```
HR_df['Gender'].unique()
```

[20]:
```
array(['Female', 'Male'], dtype=object)
```

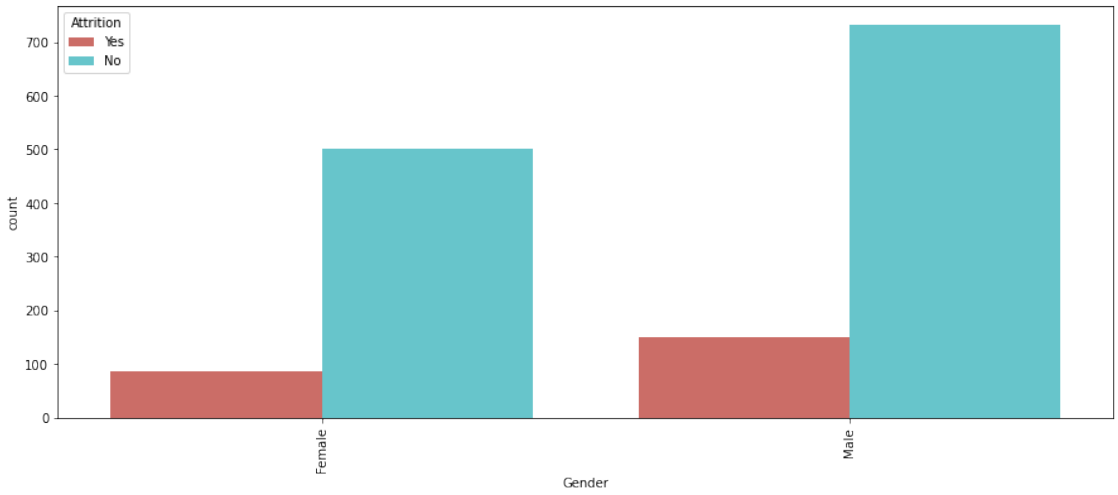
[21]:
```
HR_df['Gender'].value_counts()
```

[21]:
```
Male      882
Female    588
Name: Gender, dtype: int64
```

[23]:
```
plt.figure(figsize=(15,6))
sns.countplot('Gender', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[24]:
```
plt.figure(figsize=(15,6))
sns.countplot('Gender', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[25]:
```
HR_df['Age'].unique()
```

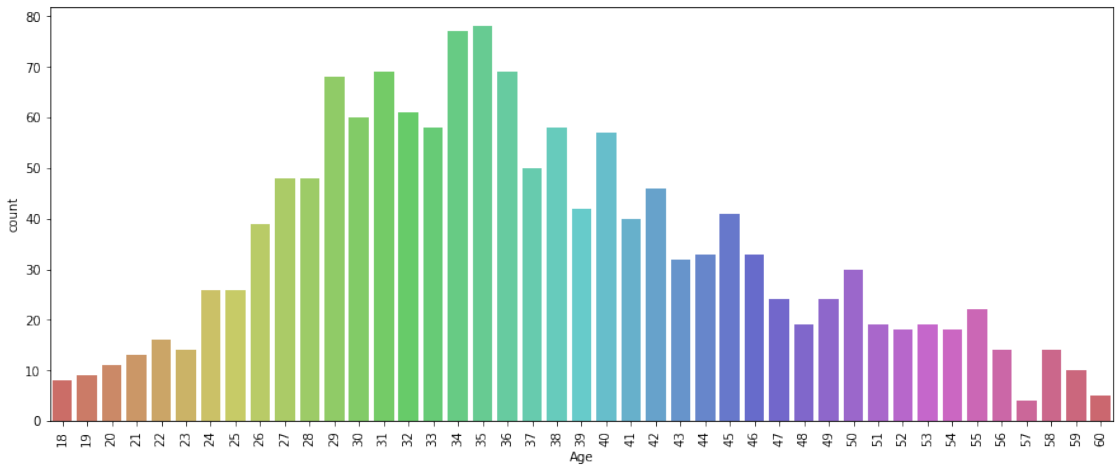
[25]:
```
array([41, 49, 37, 33, 27, 32, 59, 30, 38, 36, 35, 29, 31, 34, 28, 22, 53,
       24, 21, 42, 44, 46, 39, 43, 50, 26, 48, 55, 45, 56, 23, 51, 40, 54,
       58, 20, 25, 19, 57, 52, 47, 18, 60])
```

[26]:
```
HR_df['Age'].value_counts()
```

[26]:
```
35    78
34    77
36    69
31    69
29    68
32    61
30    60
33    58
38    58
40    57
37    50
27    48
28    48
42    46
39    42
45    41
41    40
26    39
44    33
46    33
43    32
50    30
```

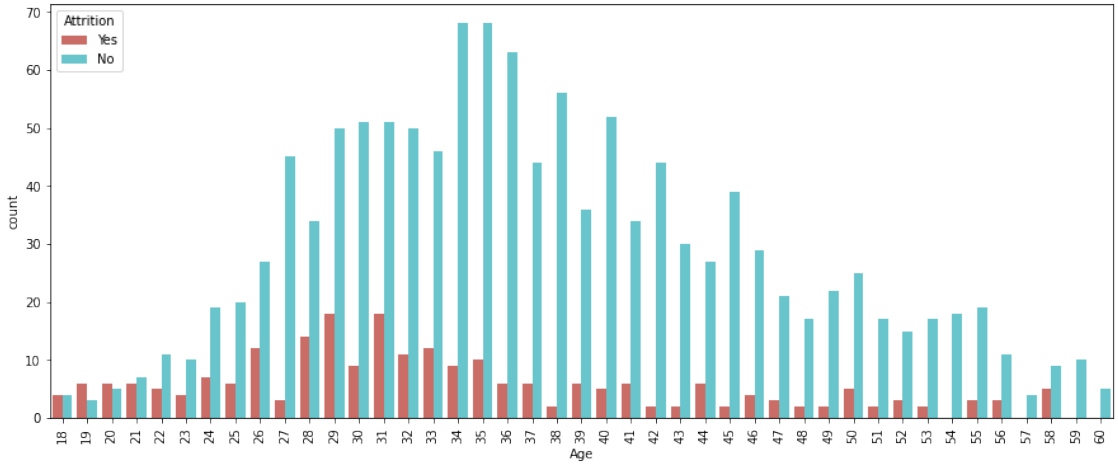
```
25    26
24    26
49    24
47    24
55    22
51    19
53    19
48    19
54    18
52    18
22    16
56    14
23    14
58    14
21    13
20    11
59    10
19     9
18     8
60     5
57     4
Name: Age, dtype: int64
```

[27]:
```
plt.figure(figsize=(15,6))
sns.countplot('Age', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[30]:
```
plt.figure(figsize=(15,6))
sns.countplot('Age', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
```

```
plt.xticks(rotation = 90)
plt.show()
```



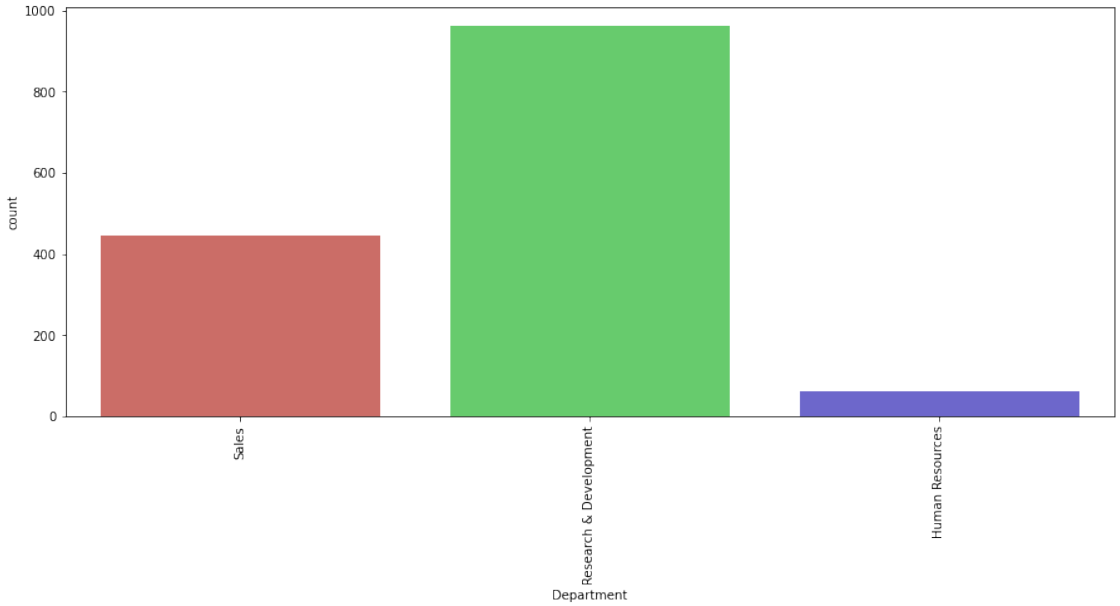

[31]:
```
HR_df['Department'].unique()
```

[31]:
```
array(['Sales', 'Research & Development', 'Human Resources'], dtype=object)
```

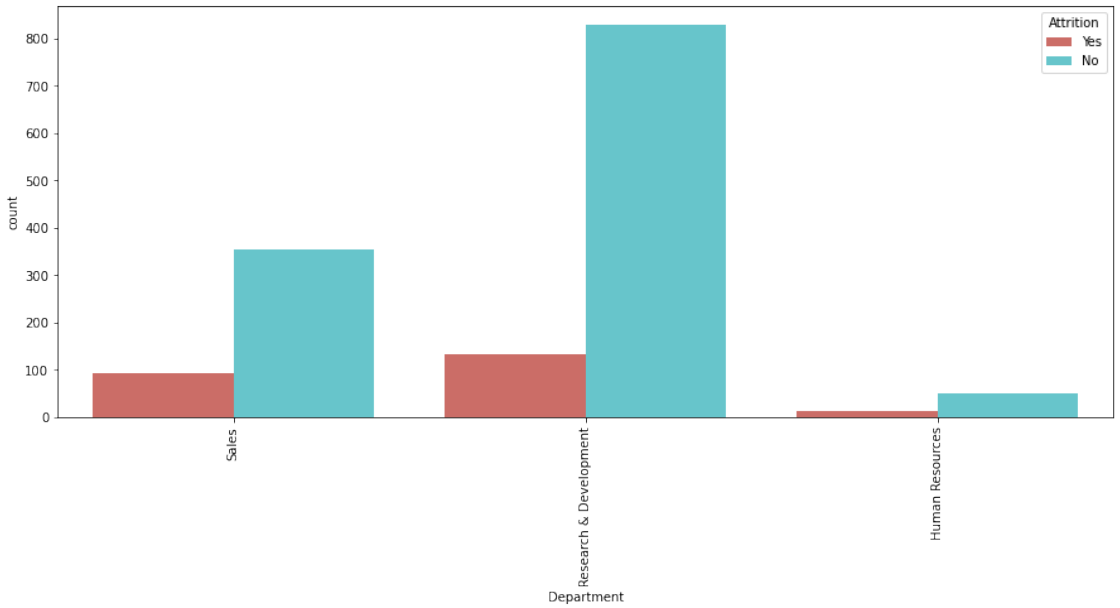
[35]:
```
HR_df['Department'].value_counts()
```

[35]:
```
Research & Development    961
Sales                     446
Human Resources            63
Name: Department, dtype: int64
```

[36]:
```
plt.figure(figsize=(15,6))
sns.countplot('Department', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[41]:
```
plt.figure(figsize=(15,6))
sns.countplot('Department', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



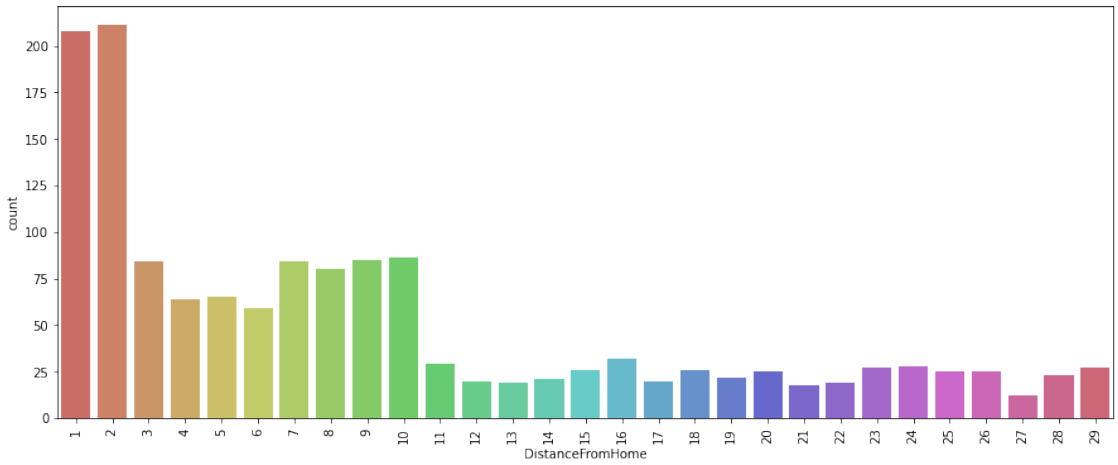
[42]:
```
HR_df['DistanceFromHome'].unique()
```

[42]:
```
array([ 1,  8,  2,  3, 24, 23, 27, 16, 15, 26, 19, 21,  5, 11,  9,  7,  6,
       10,  4, 25, 12, 18, 29, 22, 14, 20, 28, 17, 13])
```

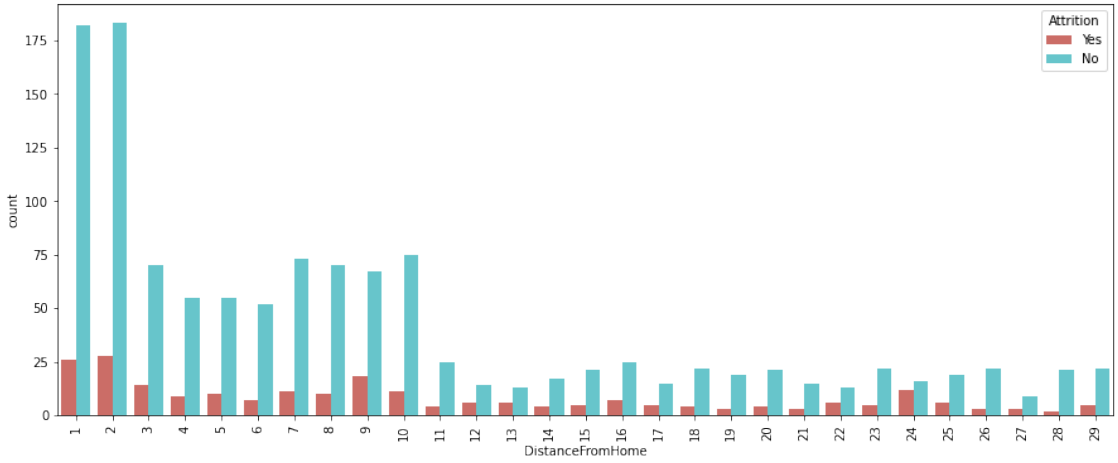
[43]:
```
HR_df['DistanceFromHome'].value_counts()
```

[43]:
```
2     211
1     208
10     86
9      85
3      84
7      84
8      80
5      65
4      64
6      59
16     32
11     29
24     28
23     27
29     27
15     26
18     26
26     25
25     25
20     25
28     23
19     22
14     21
12     20
17     20
22     19
13     19
21     18
27     12
Name: DistanceFromHome, dtype: int64
```

[44]:
```
plt.figure(figsize=(15,6))
sns.countplot('DistanceFromHome', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[45]:
```
plt.figure(figsize=(15,6))
sns.countplot('DistanceFromHome', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



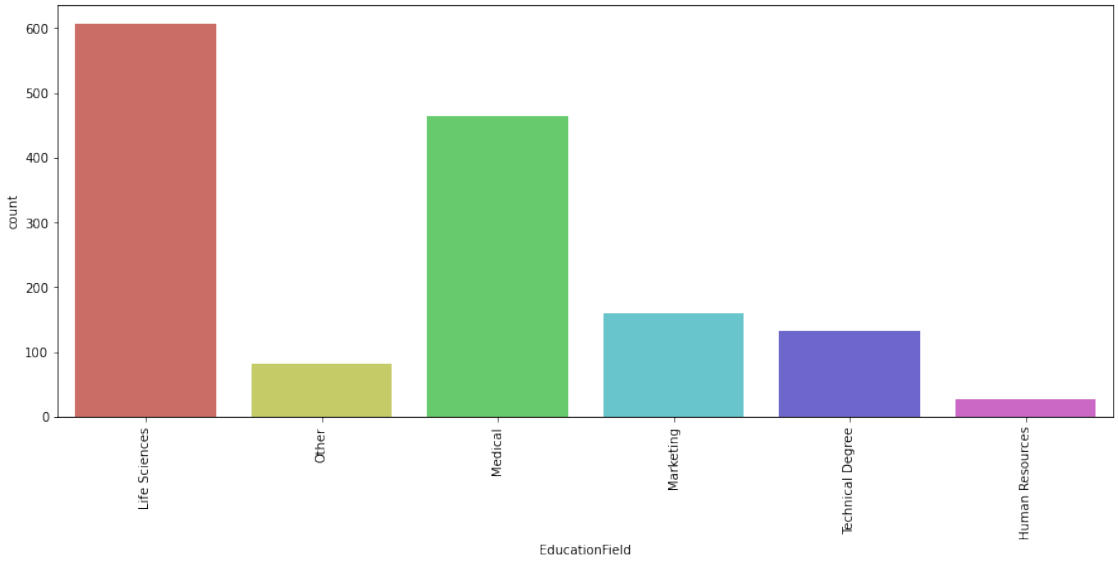

[46]:
```
HR_df['EducationField'].unique()
```

[46]:
```
array(['Life Sciences', 'Other', 'Medical', 'Marketing',
       'Technical Degree', 'Human Resources'], dtype=object)
```

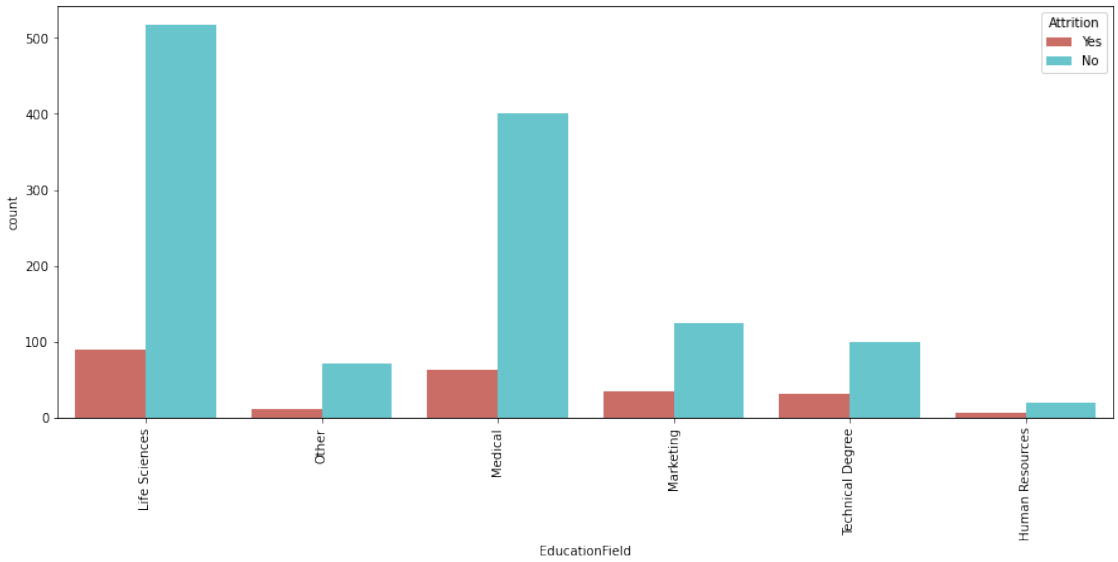
[47]:
```
HR_df['EducationField'].value_counts()
```

[47]:
```
Life Sciences       606
Medical             464
Marketing           159
Technical Degree    132
Other                82
Human Resources      27
Name: EducationField, dtype: int64
```

[48]:
```
plt.figure(figsize=(15,6))
sns.countplot('EducationField', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[49]:
```
plt.figure(figsize=(15,6))
sns.countplot('EducationField', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



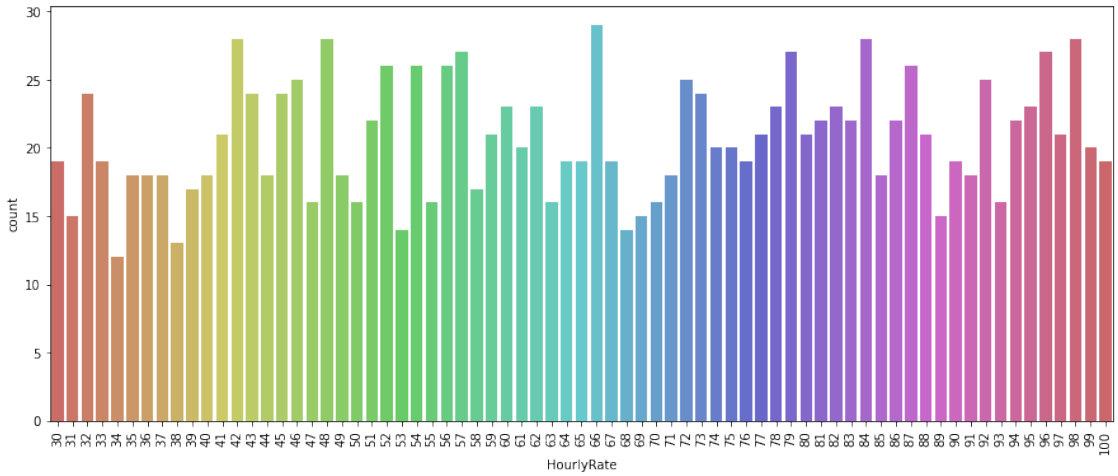

[50]:
```
HR_df['HourlyRate'].unique()
```

[50]:
```
array([ 94,  61,  92,  56,  40,  79,  81,  67,  44,  84,  49,  31,  93,
        50,  51,  80,  96,  78,  45,  82,  53,  83,  58,  72,  48,  42,
        41,  86,  97,  75,  33,  37,  73,  98,  36,  47,  71,  30,  43,
        99,  59,  95,  57,  76,  87,  66,  55,  32,  52,  70,  62,  64,
        63,  60, 100,  46,  39,  77,  35,  91,  54,  34,  90,  65,  88,
        85,  89,  68,  69,  74,  38])
```

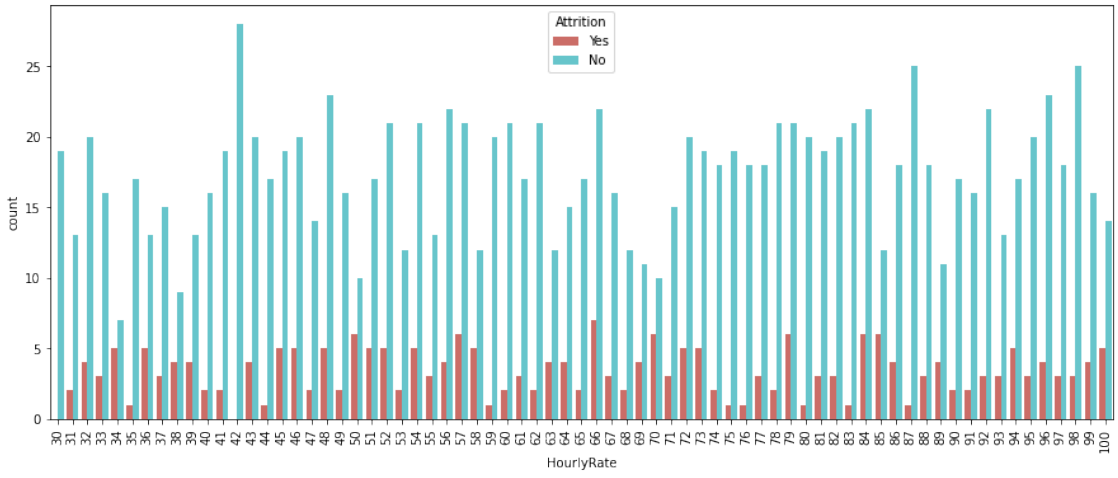
[51]:
```
HR_df['HourlyRate'].value_counts()
```

[51]:
```
66    29
98    28
42    28
48    28
84    28
      ..
31    15
53    14
68    14
38    13
34    12
Name: HourlyRate, Length: 71, dtype: int64
```

[52]:
```
plt.figure(figsize=(15,6))
sns.countplot('HourlyRate', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[53]:
```
plt.figure(figsize=(15,6))
sns.countplot('HourlyRate', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[54]:
```
HR_df['JobInvolvement'].unique()
```

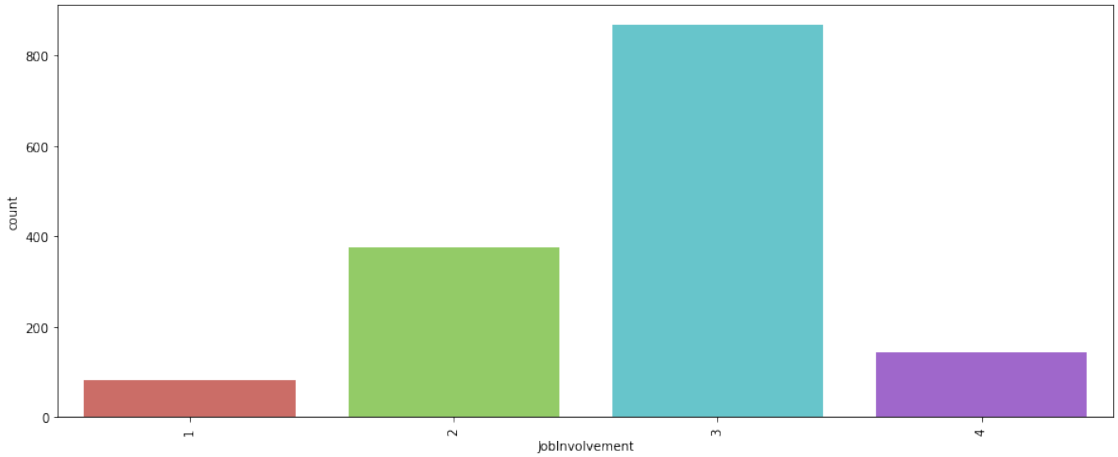
[54]:
```
array([3, 2, 4, 1])
```

[55]:
```
HR_df['JobInvolvement'].value_counts()
```

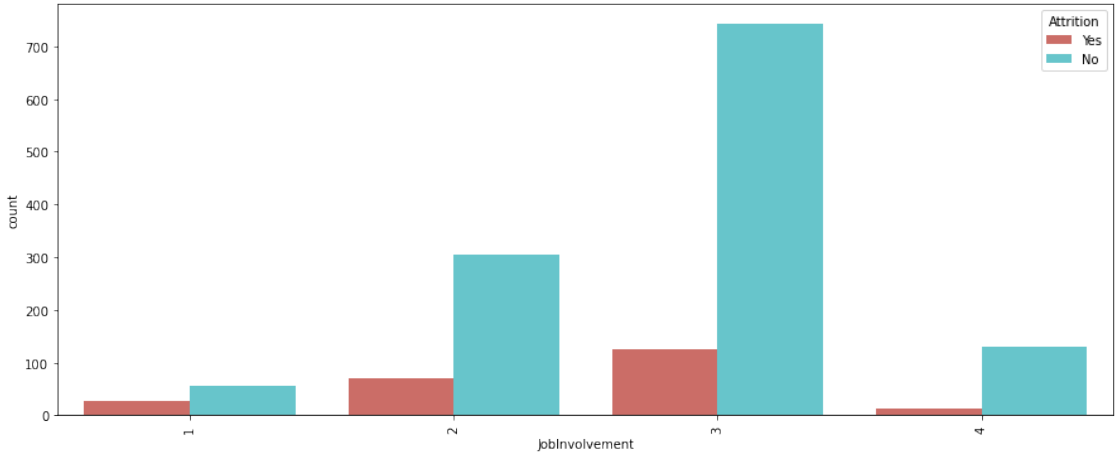
[55]:
```
3    868
2    375
```

```
4    144
1     83
Name: JobInvolvement, dtype: int64
```

[56]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobInvolvement', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[57]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobInvolvement', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



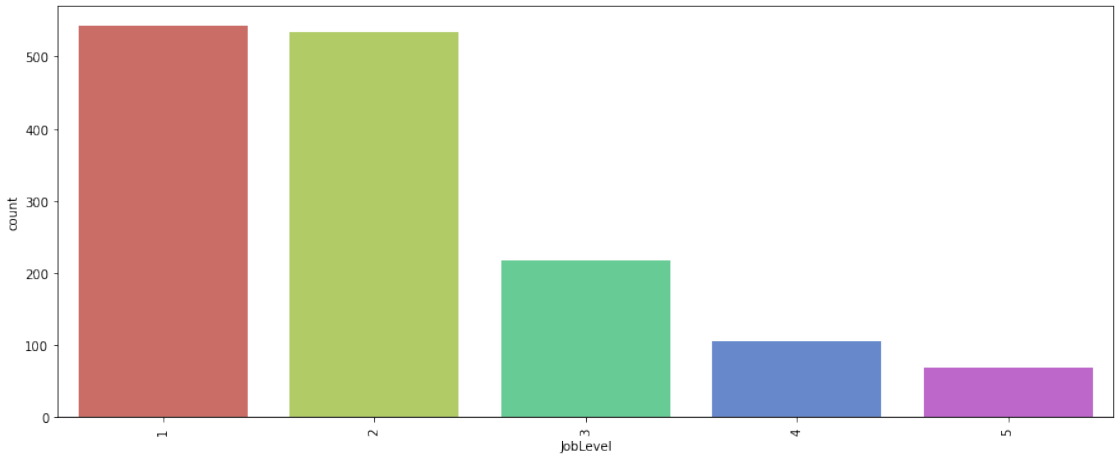
[58]:
```
HR_df['JobLevel'].unique()
```

[58]:
```
array([2, 1, 3, 4, 5])
```

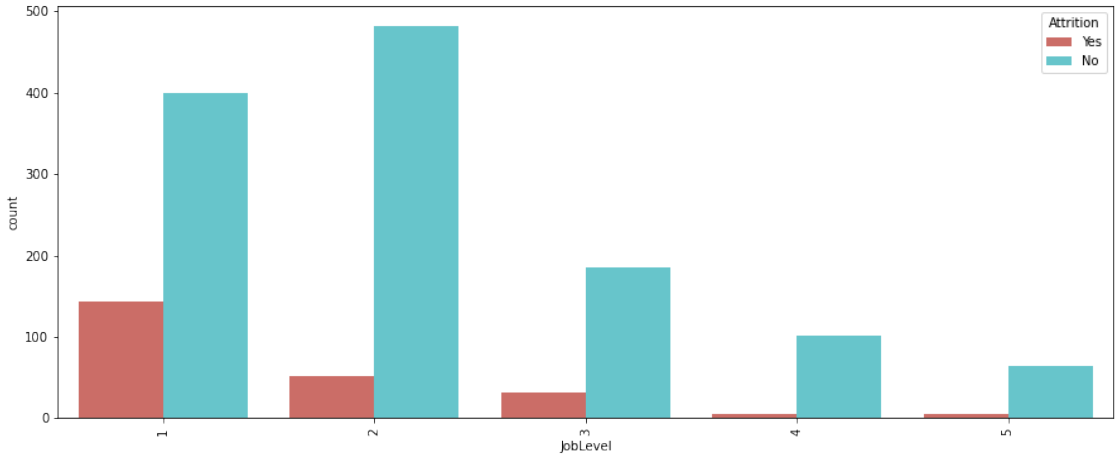
[59]:
```
HR_df['JobLevel'].value_counts()
```

[59]:
```
1    543
2    534
3    218
4    106
5     69
Name: JobLevel, dtype: int64
```

[60]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobLevel', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[61]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobLevel', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



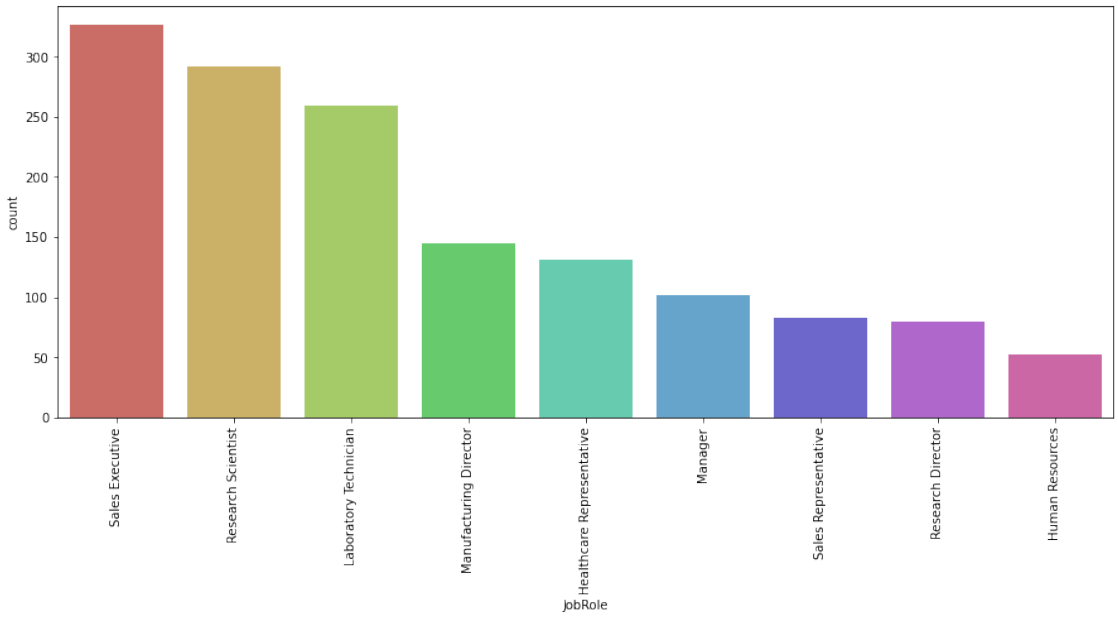

[62]:
```
HR_df['JobRole'].unique()
```

[62]:
```
array(['Sales Executive', 'Research Scientist', 'Laboratory Technician',
       'Manufacturing Director', 'Healthcare Representative', 'Manager',
       'Sales Representative', 'Research Director', 'Human Resources'],
      dtype=object)
```

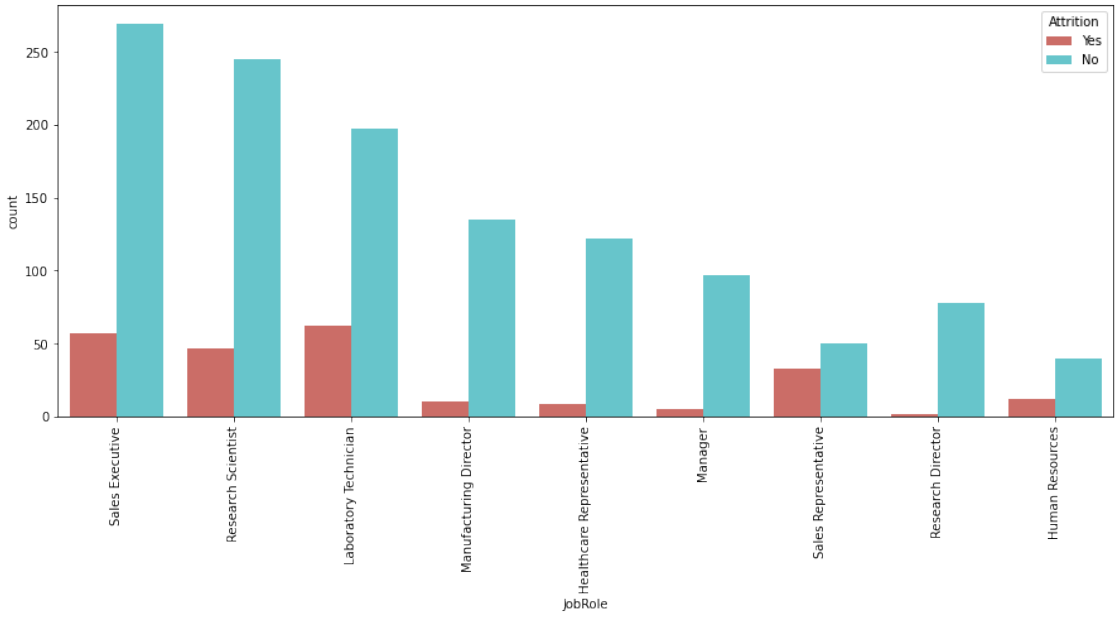
[63]:
```
HR_df['JobRole'].value_counts()
```

[63]:
```
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
```

[64]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobRole', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[65]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobRole', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



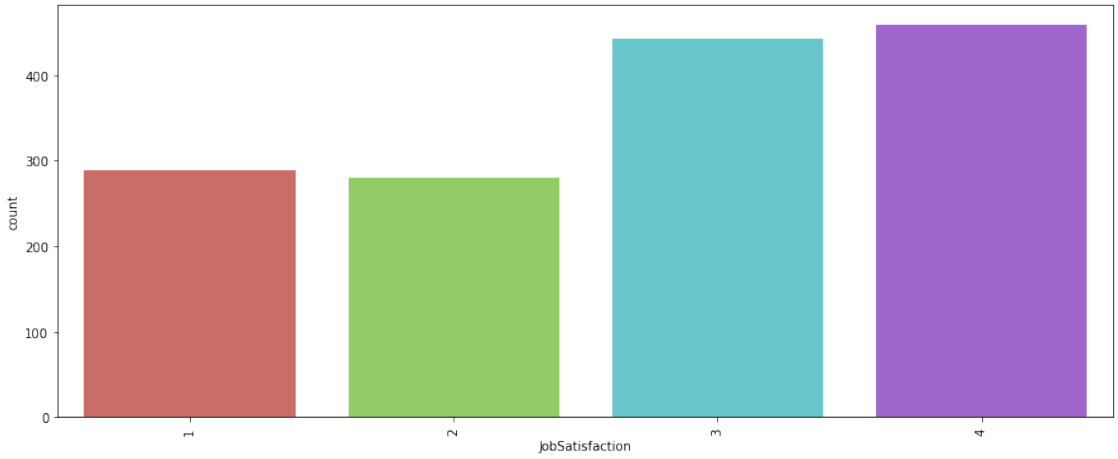
[66]:
```
HR_df['JobSatisfaction'].unique()
```

[66]:
```
array([4, 2, 3, 1])
```

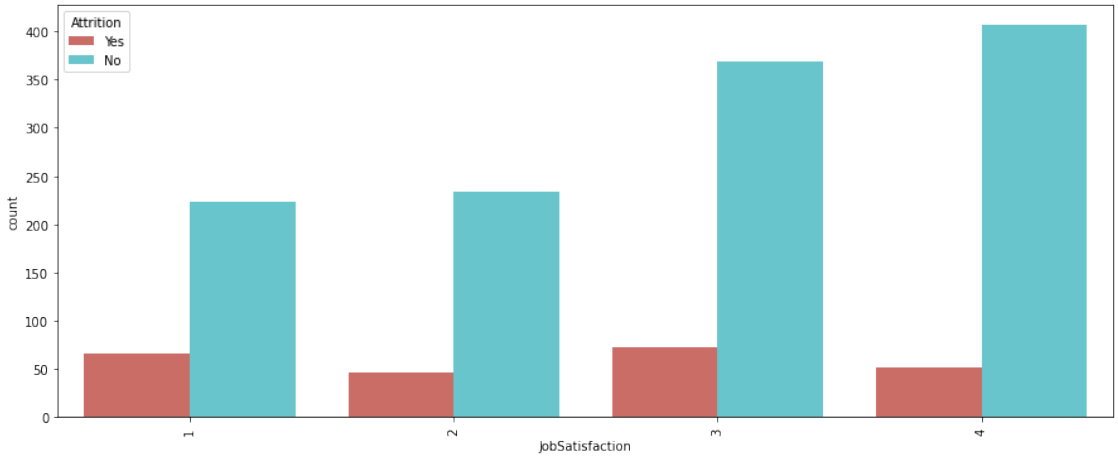
[67]:
```
HR_df['JobSatisfaction'].value_counts()
```

[67]:
```
4    459
3    442
1    289
2    280
Name: JobSatisfaction, dtype: int64
```

[68]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobSatisfaction', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[69]:
```
plt.figure(figsize=(15,6))
sns.countplot('JobSatisfaction', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



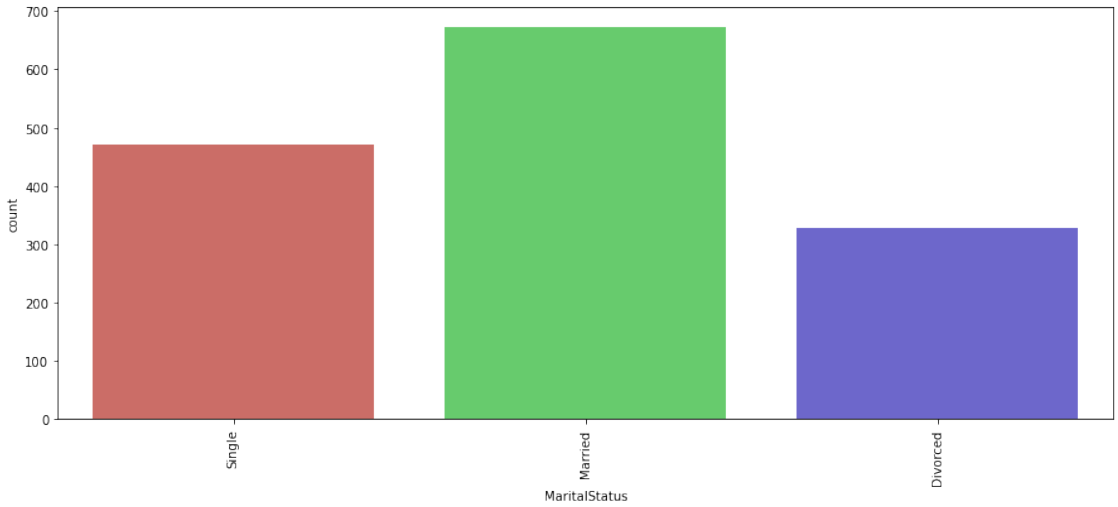

[70]:
```
HR_df['MaritalStatus'].unique()
```

[70]:
```
array(['Single', 'Married', 'Divorced'], dtype=object)
```

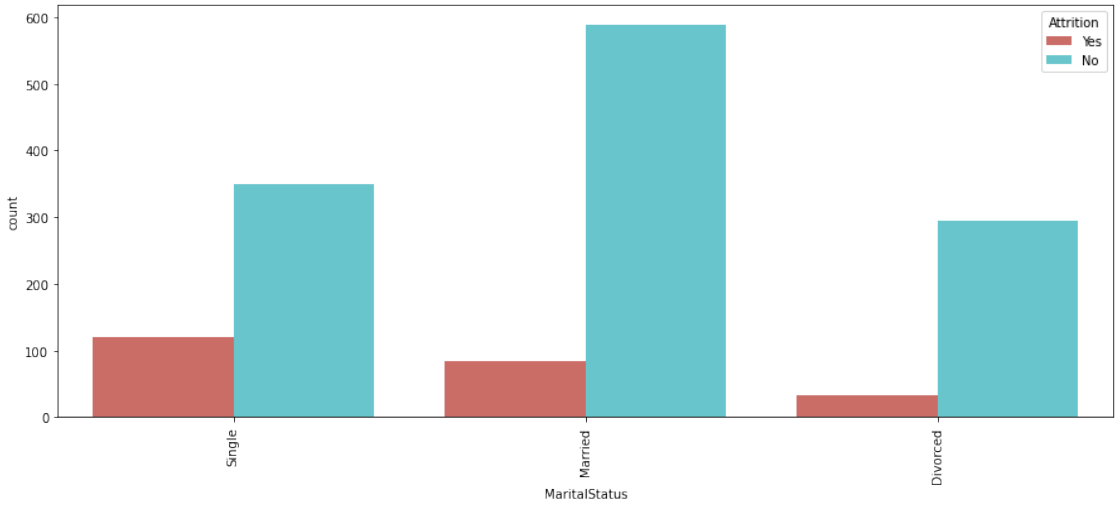
[71]:
```
HR_df['MaritalStatus'].value_counts()
```

[71]:
```
Married     673
Single      470
Divorced    327
Name: MaritalStatus, dtype: int64
```

[72]:
```
plt.figure(figsize=(15,6))
sns.countplot('MaritalStatus', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



[73]:
```
plt.figure(figsize=(15,6))
sns.countplot('MaritalStatus', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



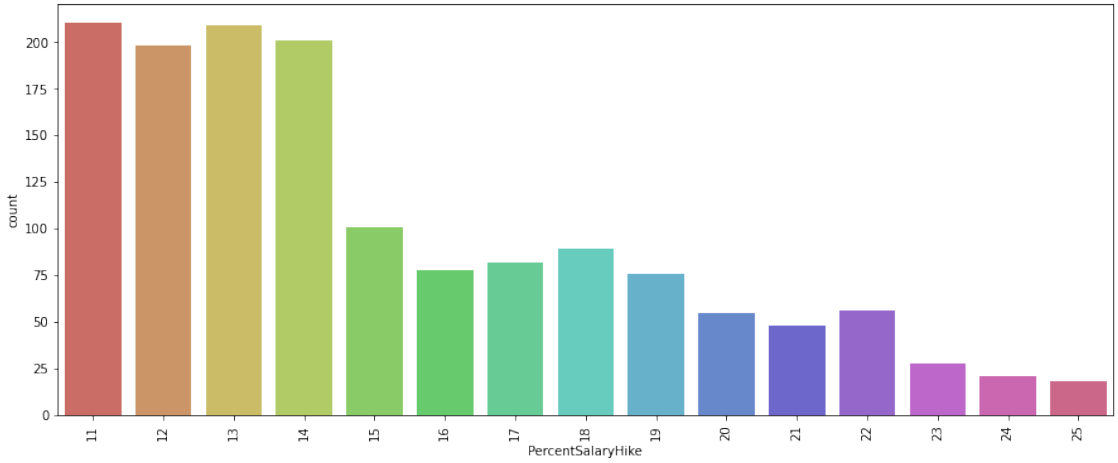

[74]:
```
HR_df['PercentSalaryHike'].unique()
```

[74]:
```
array([11, 23, 15, 12, 13, 20, 22, 21, 17, 14, 16, 18, 19, 24, 25])
```

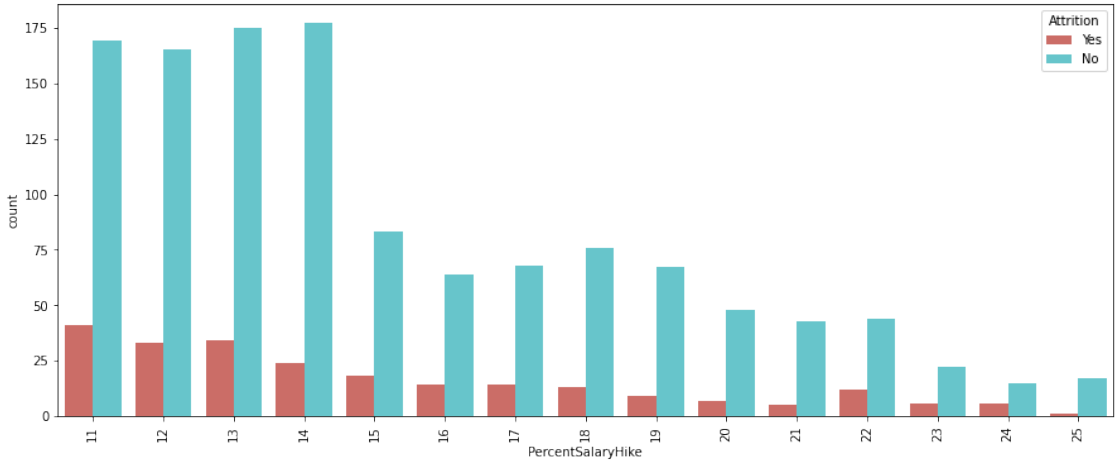
[75]:
```
HR_df['PercentSalaryHike'].value_counts()
```

[75]:
```
11    210
13    209
14    201
12    198
15    101
18     89
17     82
16     78
19     76
22     56
20     55
21     48
23     28
24     21
25     18
Name: PercentSalaryHike, dtype: int64
```

[76]:
```
plt.figure(figsize=(15,6))
sns.countplot('PercentSalaryHike', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```




[77]:
```
plt.figure(figsize=(15,6))
sns.countplot('PercentSalaryHike', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



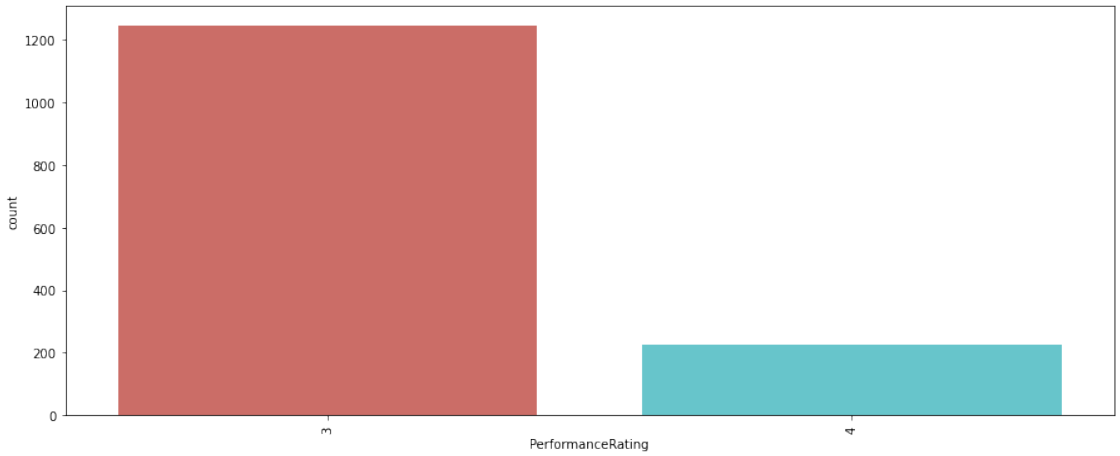

[78]:
```
HR_df['PerformanceRating'].unique()
```

[78]:
```
array([3, 4])
```

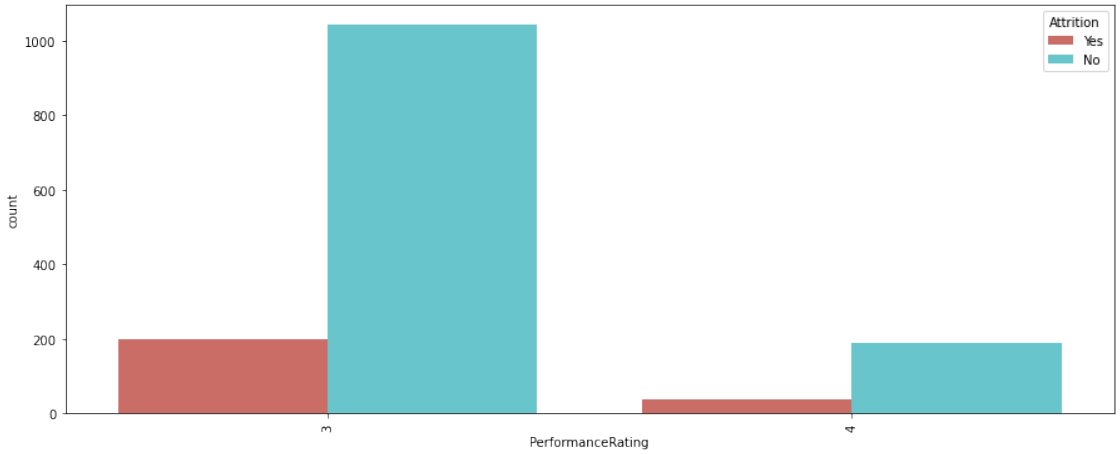
[79]:
```
HR_df['PerformanceRating'].value_counts()
```

[79]:
```
3    1244
4     226
Name: PerformanceRating, dtype: int64
```

[80]:
```
plt.figure(figsize=(15,6))
sns.countplot('PerformanceRating', data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```

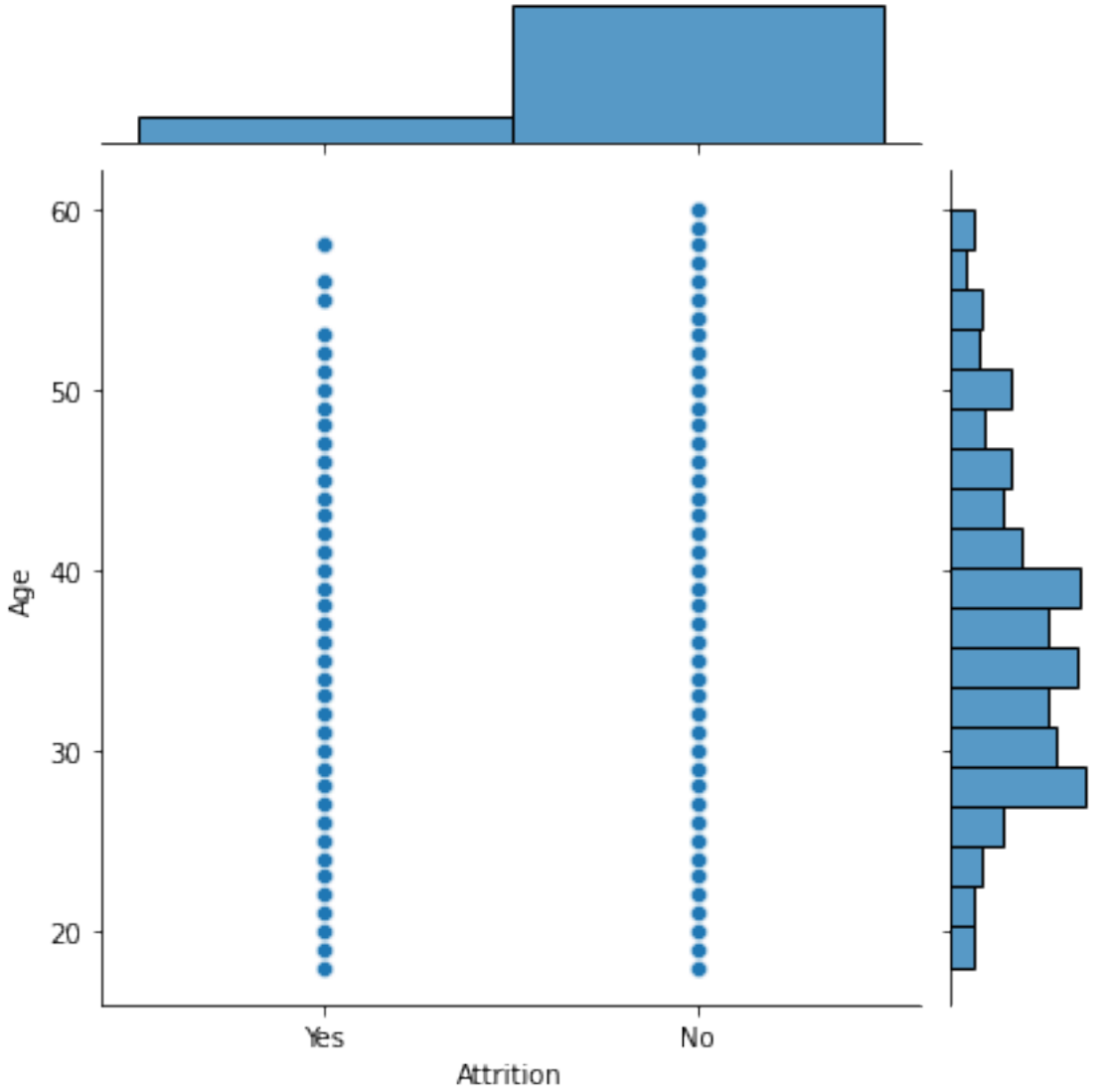



[81]:
```
plt.figure(figsize=(15,6))
sns.countplot('PerformanceRating', hue = HR_df['Attrition'],
              data = HR_df, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



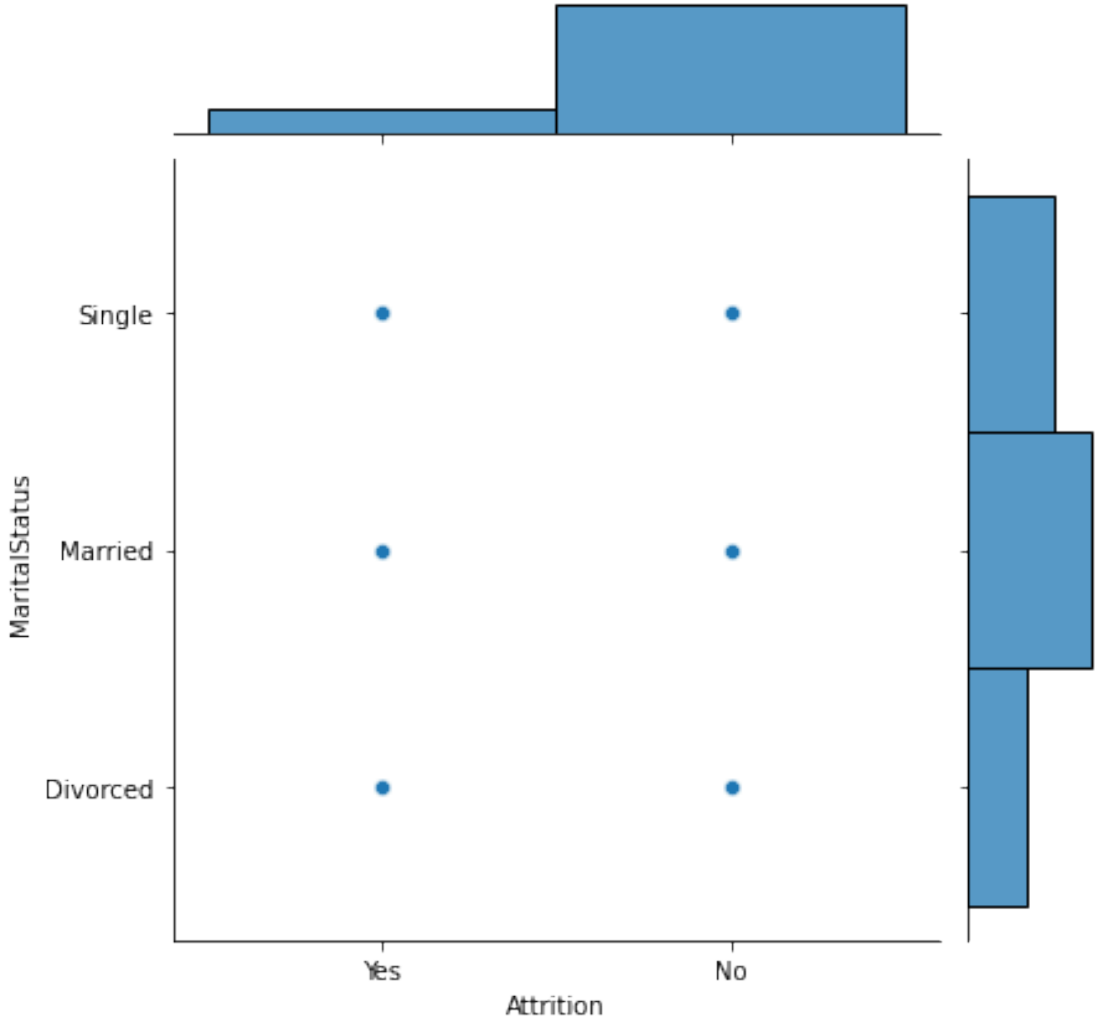
[82]:
```
plt.figure(figsize=(15,6))
sns.jointplot(x='Attrition',y='Age',data = data)
plt.xticks(rotation = 90)
plt.show()
```

```
<Figure size 1080x432 with 0 Axes>
```




[83]:
```
plt.figure(figsize=(15,6))
sns.jointplot(x='Attrition',y='MaritalStatus',data = data)
plt.xticks(rotation = 90)
plt.show()
```

<Figure size 1080x432 with 0 Axes>




[84]:
```
corr_matrix=HR_df.corr()
corr_matrix
```

[84]:

| | 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 |

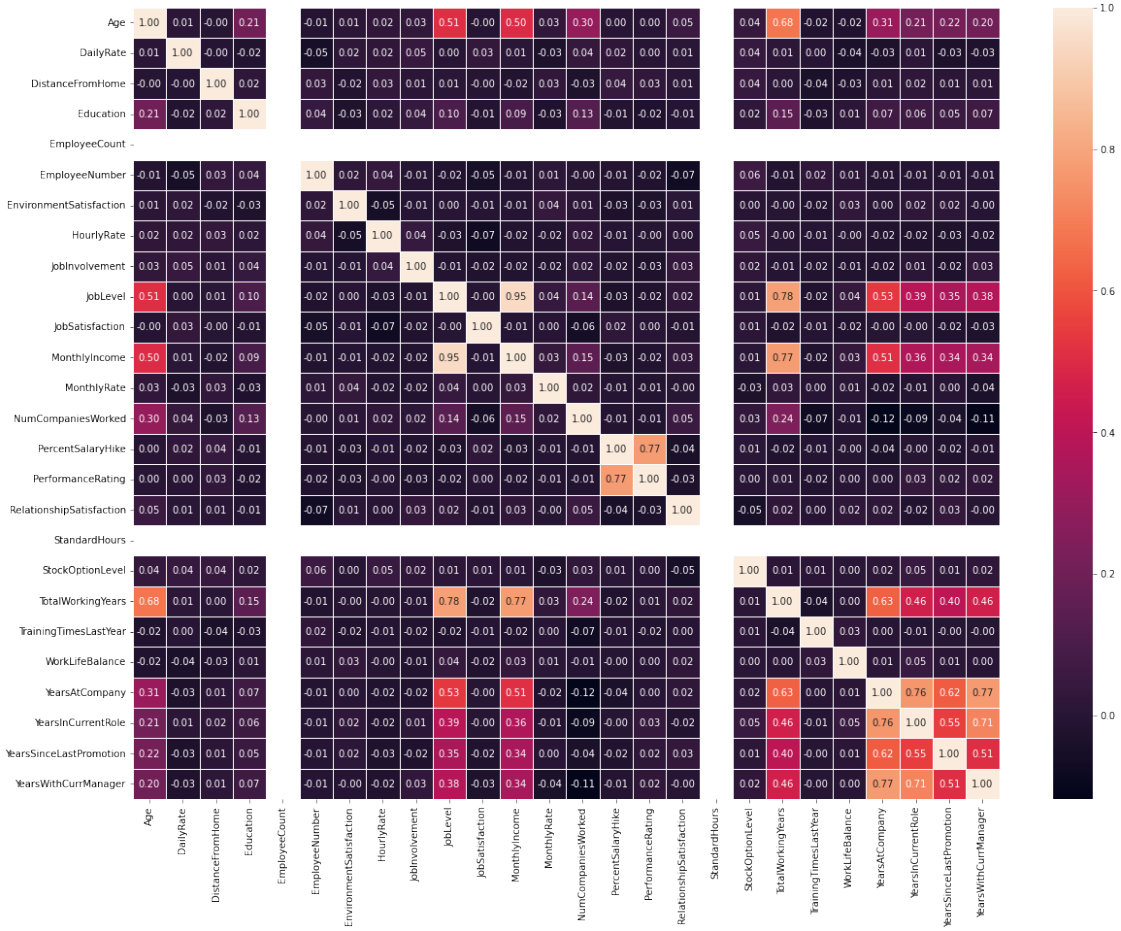
| | YearsAtCompany | YearsInCurrentRole | \ |
|---|---|---|---|
| Age | 0.311309 | 0.212901 | |
| DailyRate | -0.034055 | 0.009932 | |
| DistanceFromHome | 0.009508 | 0.018845 | |
| Education | 0.069114 | 0.060236 | |
| EmployeeCount | NaN | NaN | |
| EmployeeNumber | -0.011240 | -0.008416 | |
| EnvironmentSatisfaction | 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]

[86]:
```
fig,ax=plt.subplots(figsize=(20,15))
ax=sns.heatmap(corr_matrix,
               annot=True,
               linewidths=0.5,
               fmt=".2f"
              )
```


Age
DailyRate
DistanceFromHome
Education
EmployeeCount
EmployeeNumber
EnvironmentSatisfaction
HourlyRate
JobInvolvement
JobLevel
JobSatisfaction
MonthlyIncome
MonthlyRate
NumCompaniesWorked
PercentSalaryHike
PerformanceRating
RelationshipSatisfaction
StandardHours
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
WorkLifeBalance
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager
1.0
0.8
0.6
0.4
0.2
0.0

[ ]: