hr-data-analysis

March 10, 2024

[1]:
```
import pandas as pd
```

[2]:
```
data = pd.read_csv("C:/Users/Asus/OneDrive/Documents/Projects/Python/HR Data.
↪csv")
```

[3]:
```
data.shape
```

[3]:
```
(1470, 35)
```

[4]:
```
data.columns
```

[4]:
```
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')
```

[5]:
```
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
```

[6]:
```
data.describe()
```

[6]:
```
               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]
```

[7]:
```
data.isnull().sum()
```

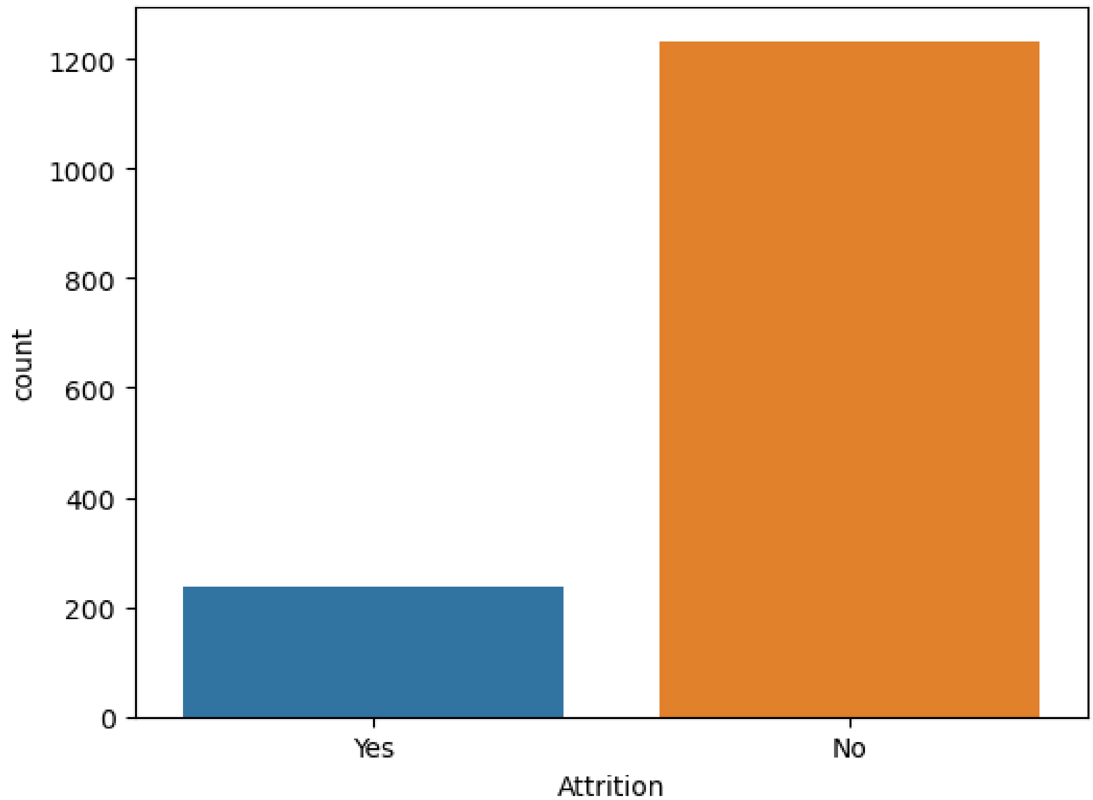
[7]:
```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              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
```

[8]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

1 ATTRITION

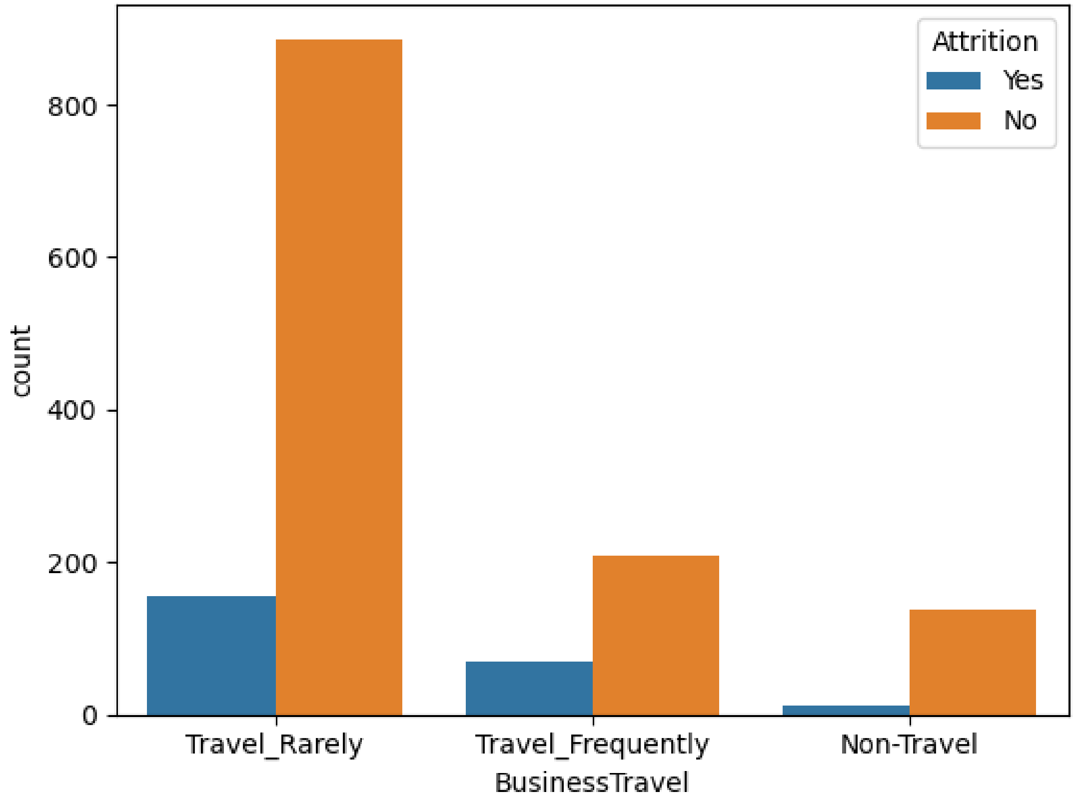
Attrition means the employee wants to leave the company for any reason. Yes -> Means employee wants to leave the company No –> Means employee does not wants to leave the company

[9]:
```
sns.countplot(x=data.Attrition)
plt.show()
```



2 1. Impact on Business Travel of Attrition

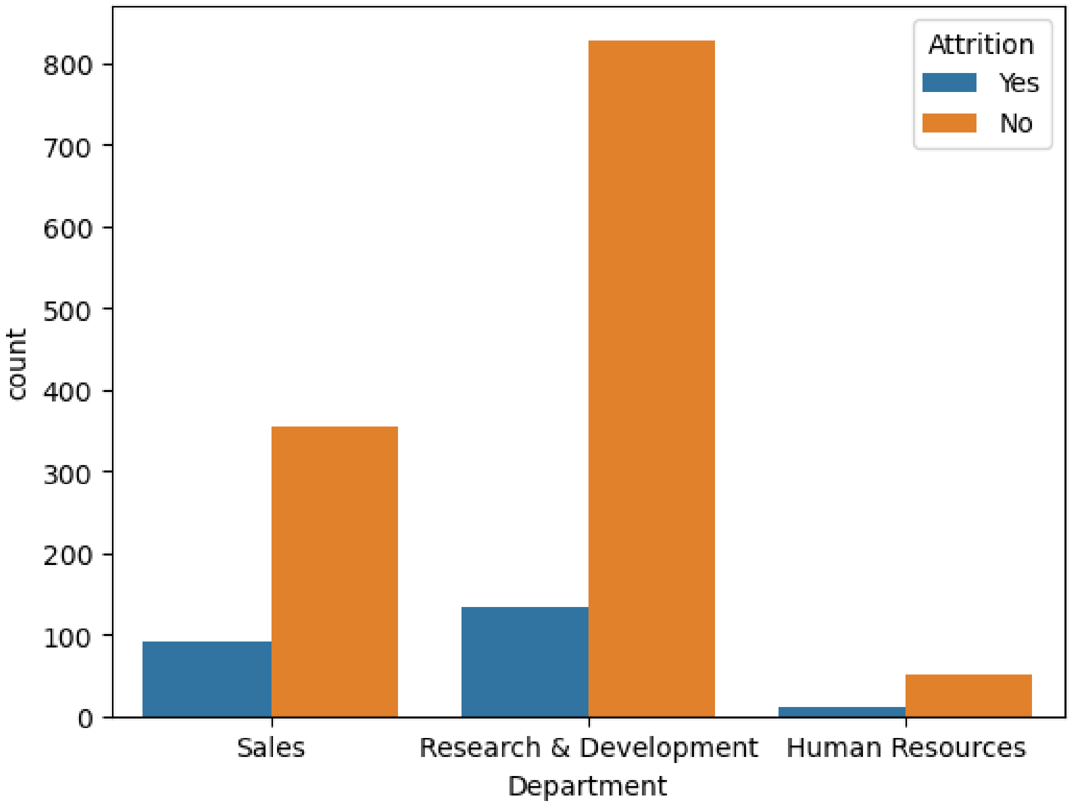
[10]:
```
sns.countplot(hue = data.Attrition, x=data.BusinessTravel)
plt.show()
```



1. Graph tells usthat company has more count or more no. of employees who travels rarely. It means travel rate of company is less.
2. There are more employees which travels rarely and are not satisfied with their job.
3. Non-traveller have least count as well as least attrition.

3 2. IMPACT OF DEPARTMENT ON ATRRITION

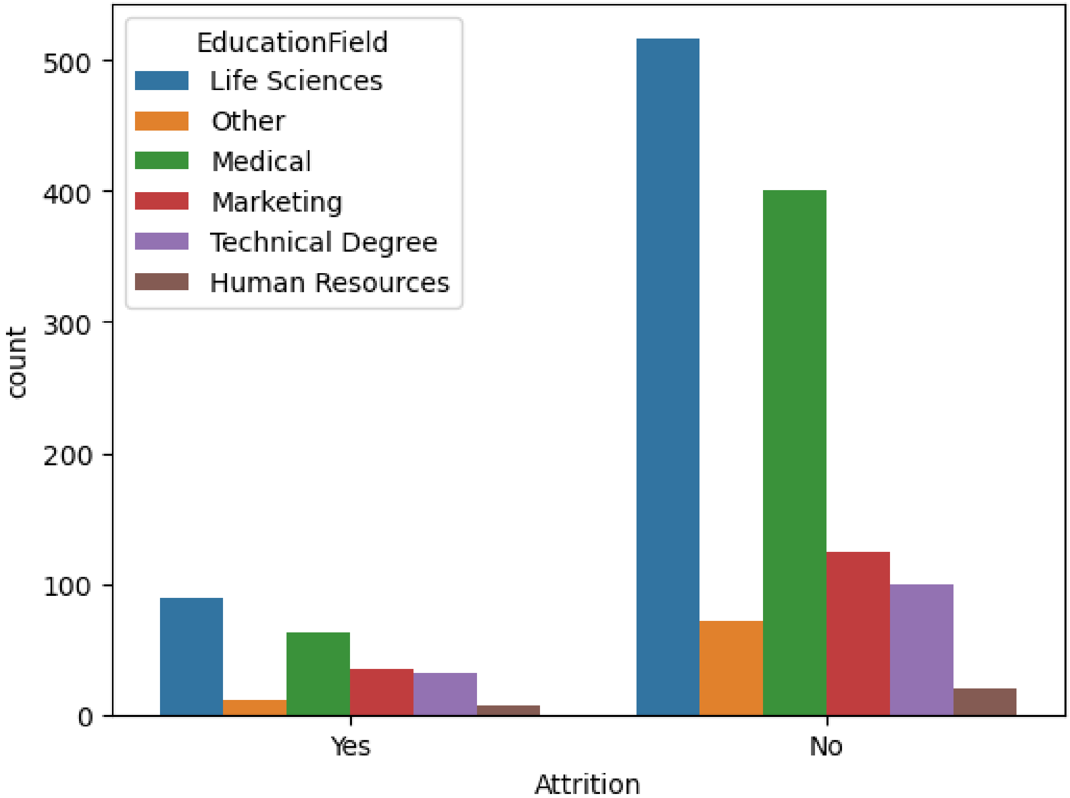
[11]:
```
sns.countplot(hue = data.Attrition, x = data.Department)
plt.show()
```



1. There are 3 number of departments are there –> 1. Sales, 2. Research and Development, 3. HR Department
2. Research and Development Department have more number of Attrition(150 employees) as compared to other two Department
3. Non Traveller have least count as well as least attrition.

4 3. IMPACT ON EDUCATION FIELD ON ATTRITION

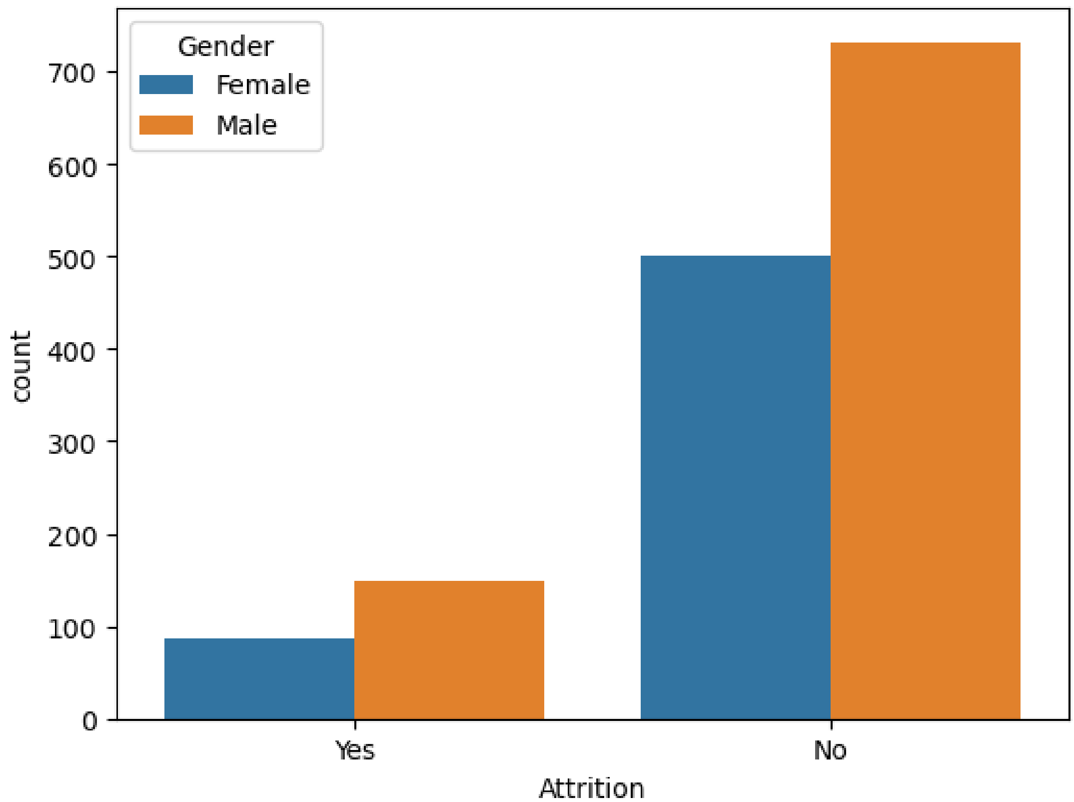
[12]:
```
sns.countplot(x = data.Attrition, hue=data.EducationField)
plt.show()
```



1. First thing is that Employees who are from "Life Science" & "medical" backgrounds are more as compared to other education fields.
2. Nearly 100 number of employees are there who are from "Life Science" education background will leave the company and followed by medical education Employees.
3. As we conclude from analysis of department and attrition, here also HR educational background employees have least attrition.

5 4. GENDER AND ATTRITION

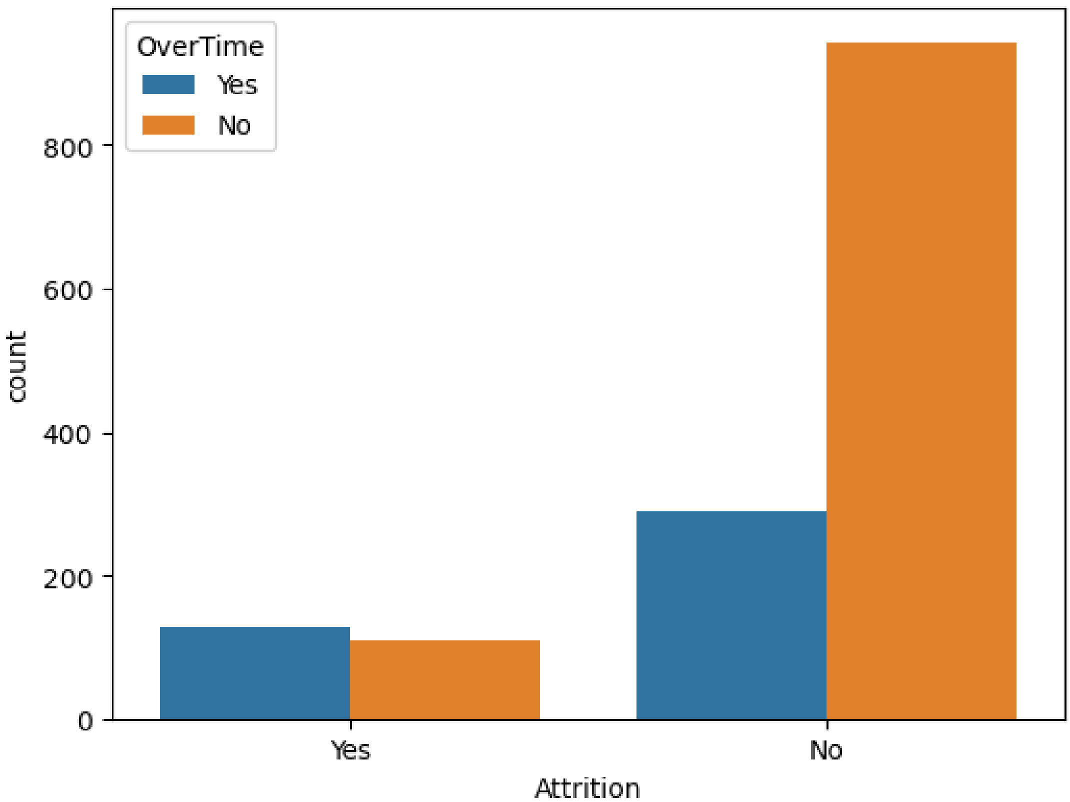
[13]:
```
sns.countplot(x = data.Attrition, hue=data.Gender)
plt.show()
```



1. Male > Female
2. More likely to quit – Male

6 5. OVERTIME AND ATTRITION

[14]:
```
sns.countplot(x = data.Attrition, hue=data.OverTime)
plt.show()
```

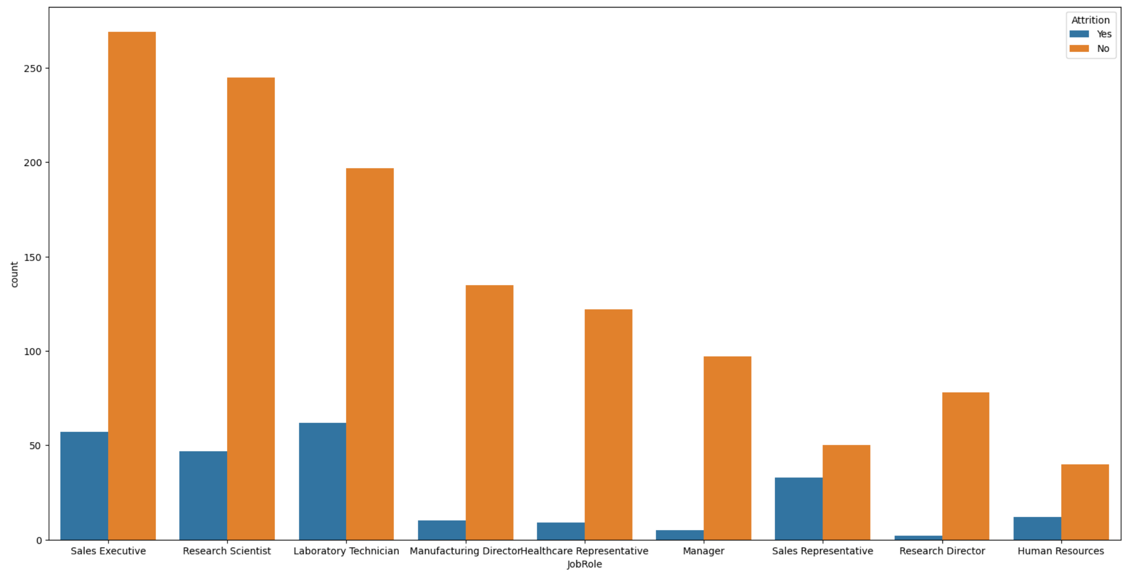


1. As for "Attrition yes", there is minor difference b/w the employees who are doing overtime & who are not.
2. So we can say that Overtime feature is not much affecting Attrition.
3. But we can conclude that most of employees are not doing overtime.

7 6. IMPACT OF JOB ROLE ON ATTRITION

[15]:
```
plt.figure(figsize = (20,10), facecolor = 'white')
sns.countplot(x = "JobRole", hue='Attrition', data=data)
plt.xlabel('JobRole', fontsize=10)
```

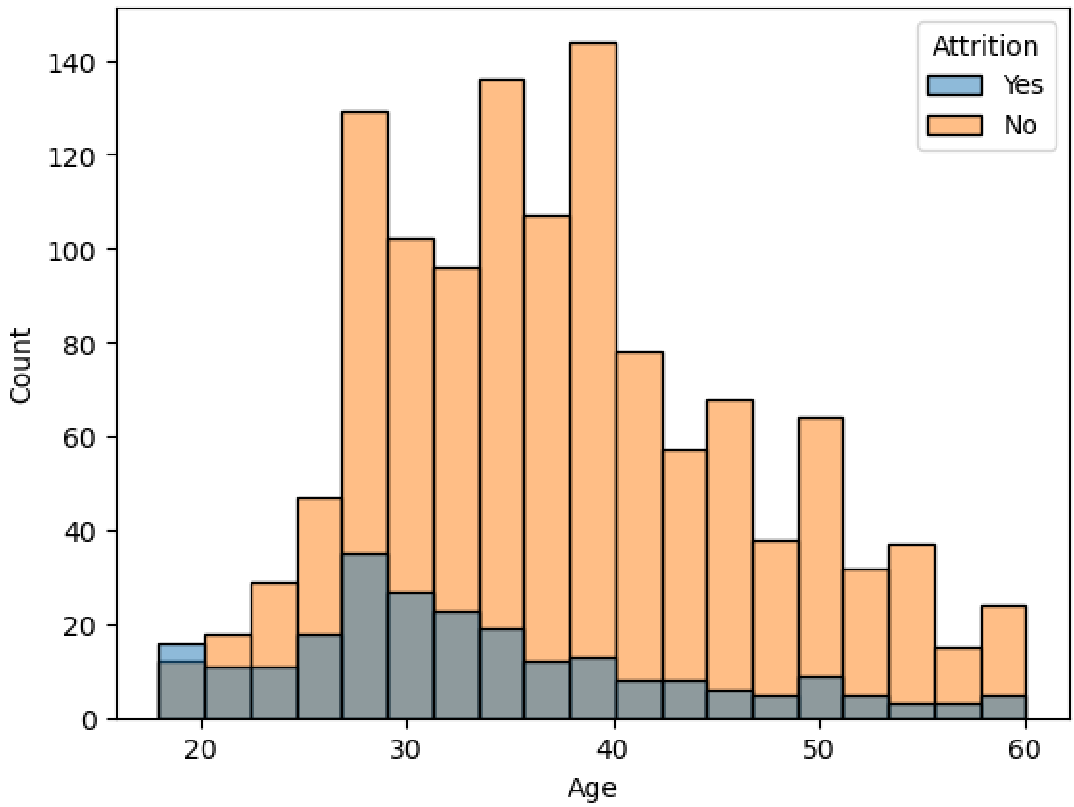
[15]:
```
Text(0.5, 0, 'JobRole')
```



1. There are less number of Research Director who leaves the company.
2. Laboratory technician, sales executive and research scientists are the top three job role in which employees have their attrition yes
3. Apart from these it can also see that there are more number of employees in Sales Executive job role.

8 7. IMPACT OF AGE ON ATTRITION

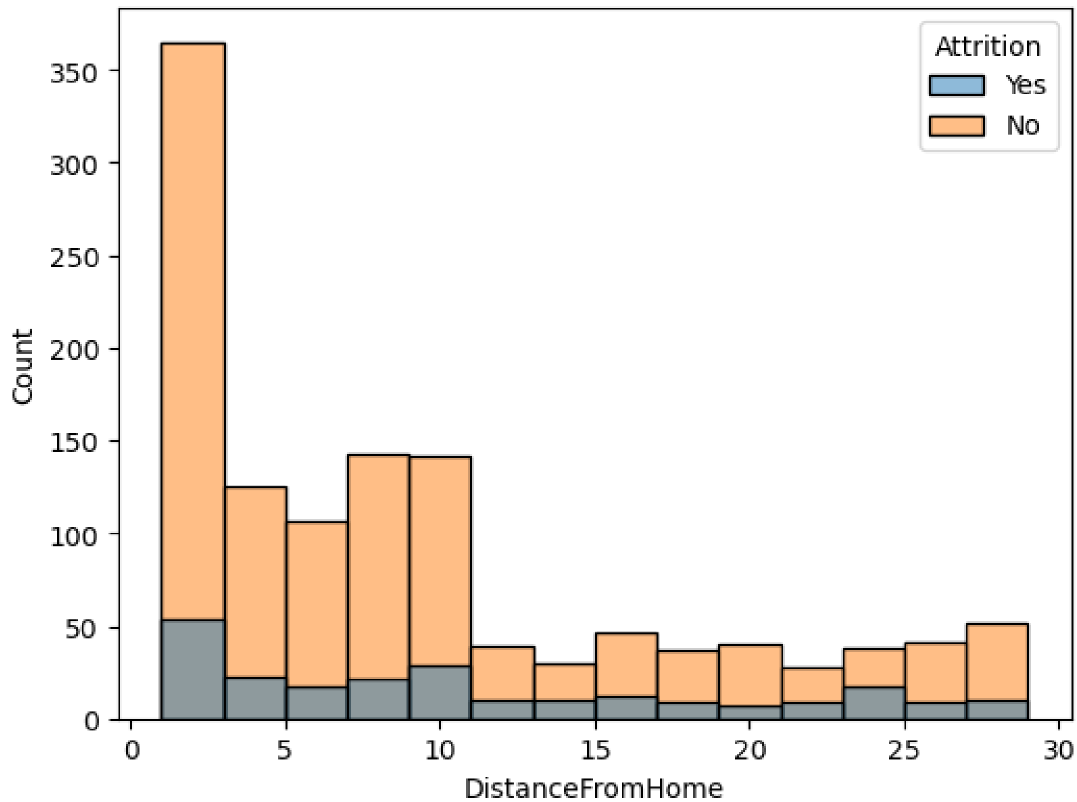
[16]:
```
sns.histplot(hue = data.Attrition, x=data.Age)
plt.show()
```



1. employee in the age of 25 to 35 are more likely to leave the job.
2. After 40 age, the distribution tells us that "Higher the age lesser will be the atrrition".

9 8. DISTANCE FROM HOME AND ATTRITION

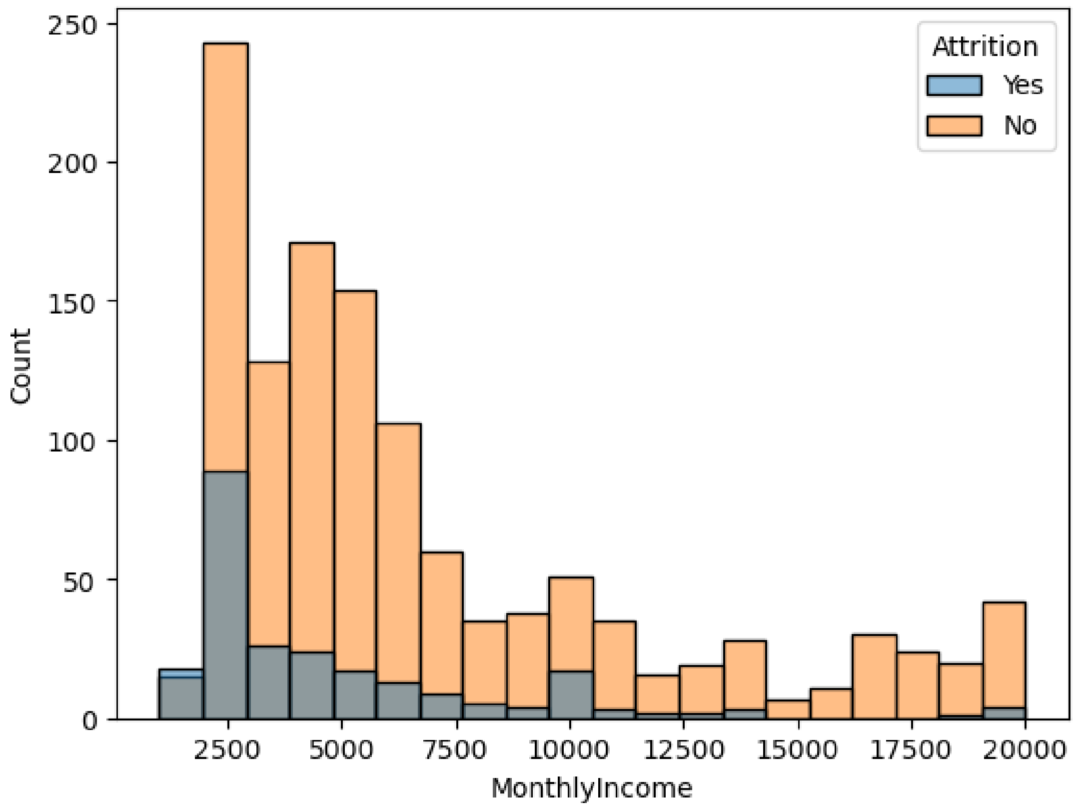
[17]:
```
sns.histplot(hue=data.Attrition, x=data.DistanceFromHome)
plt.show()
```



1. Employees who has distance range of 0-10 km, are more likely to leave the job.
2. We can also conclude that lesser the distance more number of employees are working.

10 9. HOW MONTHLY INCOME GIVES TRENDS W.R.T ATTRITION

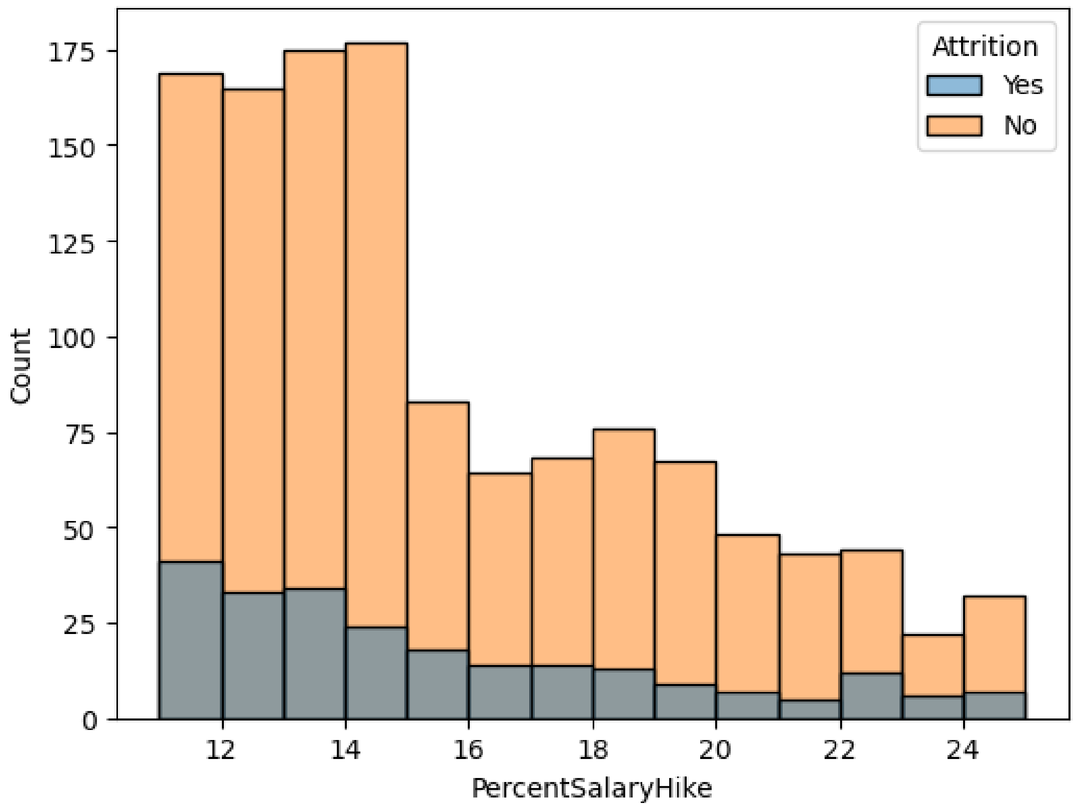
[18]:
```
sns.histplot(x=data.MonthlyIncome, hue=data.Attrition)
plt.show()
```



1. Higher the Monthly Income give rise to less Attrition (means Attrition is "No")
2. Employees who have their Income aprox 2500 are more likely to quit their job, because 2500 is the least range of Income.

11 10. HOW SALARY HIKE IS IMPACTING THE ATRRITION

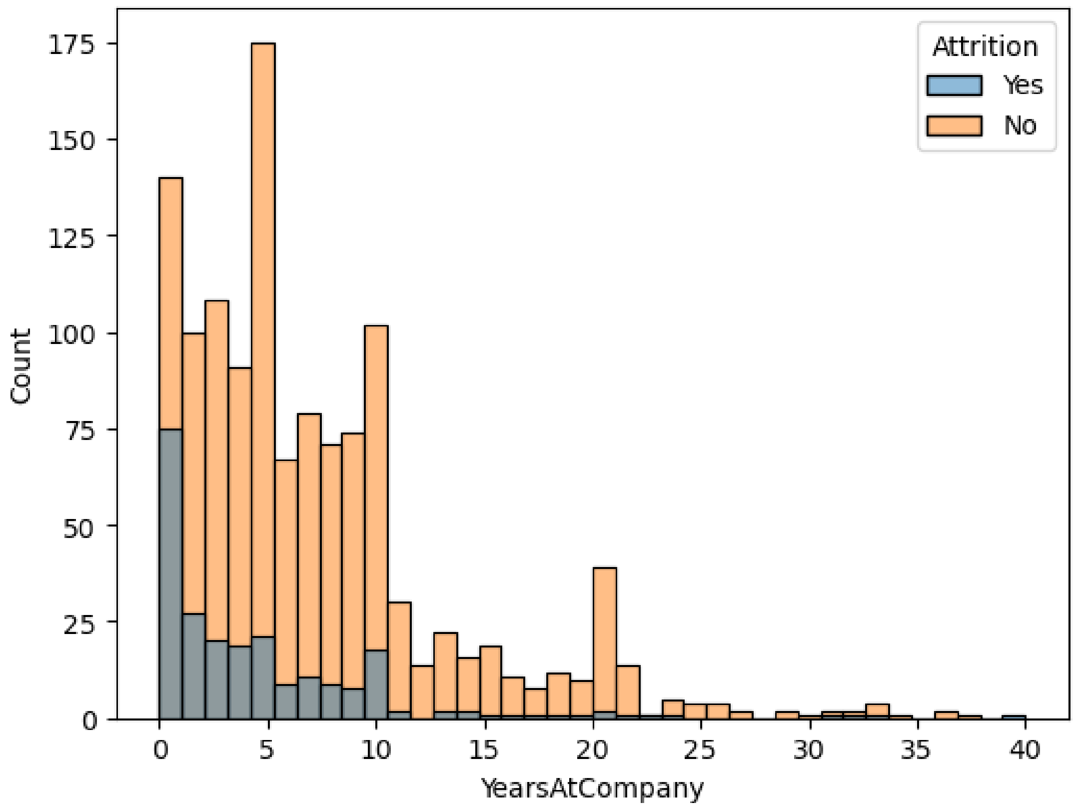
[19]:
```
sns.histplot(hue = data.Attrition, x=data.PercentSalaryHike)
plt.show()
```



1. Higher the salary percentage hike, lesser the Attrition("No")

12 11. YEARS AT THE COMPANY

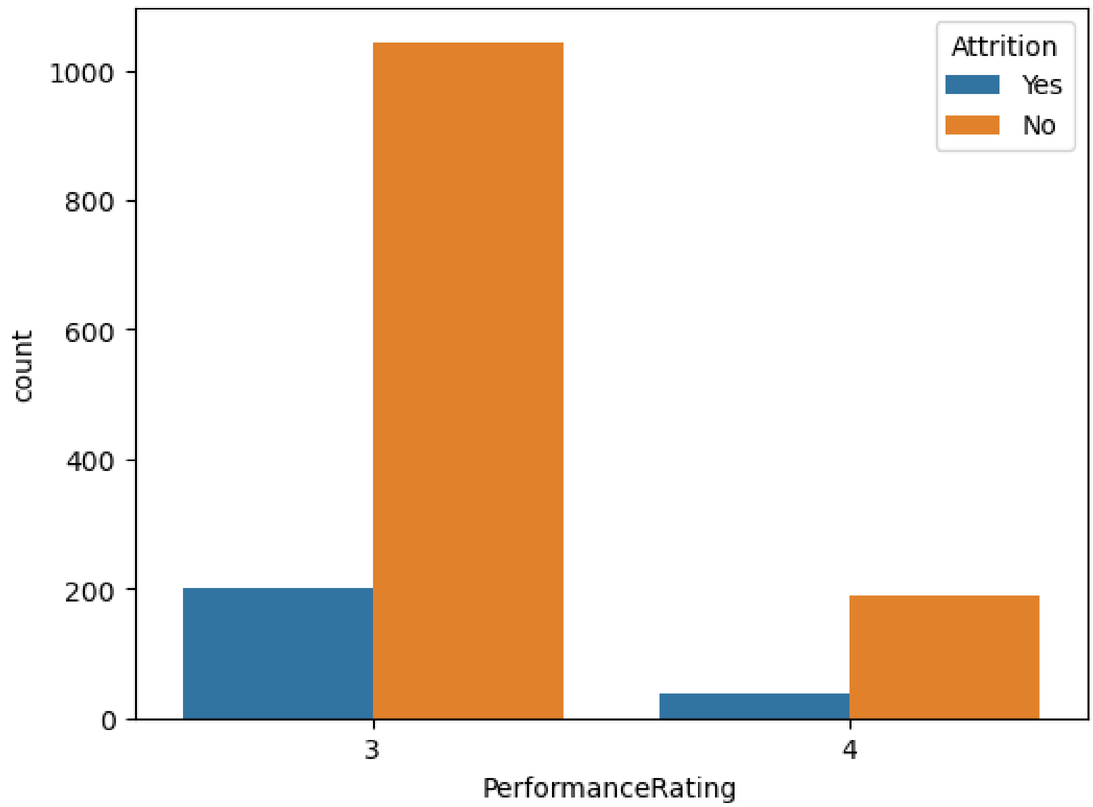
[20]:
```
sns.histplot(x = data.YearsAtCompany, hue=data.Attrition)
plt.show()
```



1. Freshers have higher data of "Attrition Yes" that is of 75 no. of workers or more than half of freshers.
2. Apart from this Employees who ranges from 1 to 10 years working on this company are also likely to quit thier job.

13 12. IMPACT OF PERFORMANCE RATING ON ATTRITION

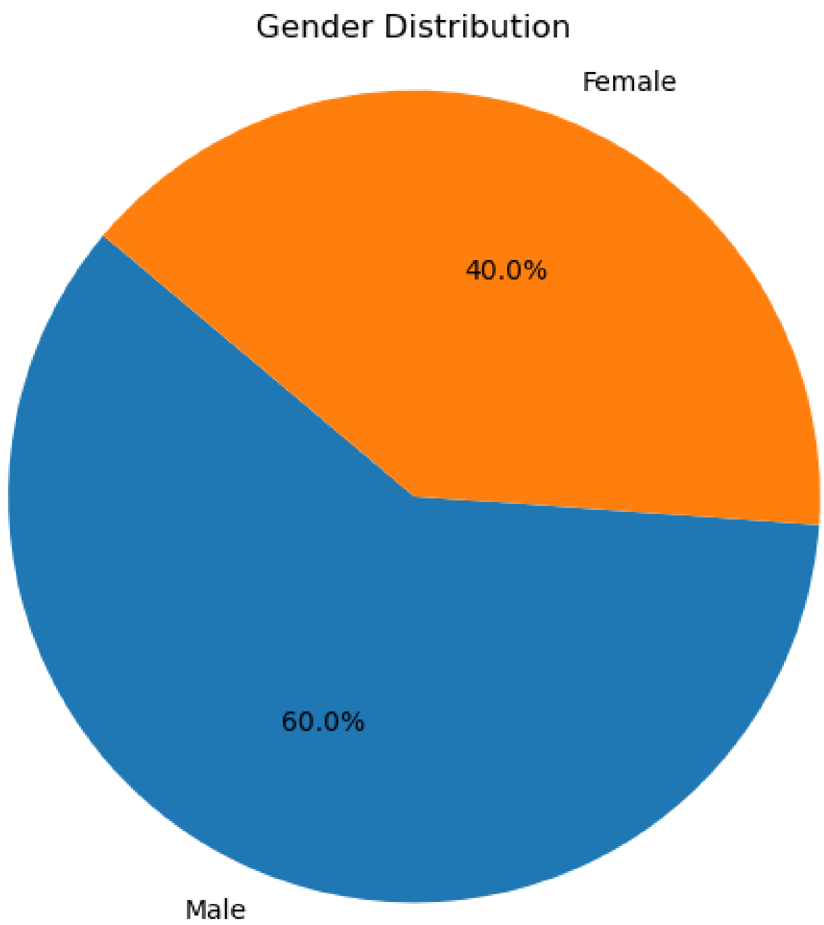
[21]:
```
sns.countplot(x=data.PerformanceRating, hue=data.Attrition)
plt.show()
```



1. On an average, most of employees are moderately performed(because performance rating lies in 3 - 4
2. However employees having less Performance rating are more likely to quit or we can say that company wants to fire that employees.

14 13. GENDER DISTRIBUTION

[22]:
```
gender_counts = data['Gender'].value_counts()

# Create a pie chart
plt.figure(figsize=(6, 6))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%',␣
↪startangle=140)
plt.title('Gender Distribution')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.

# Show the plot
plt.show()
```

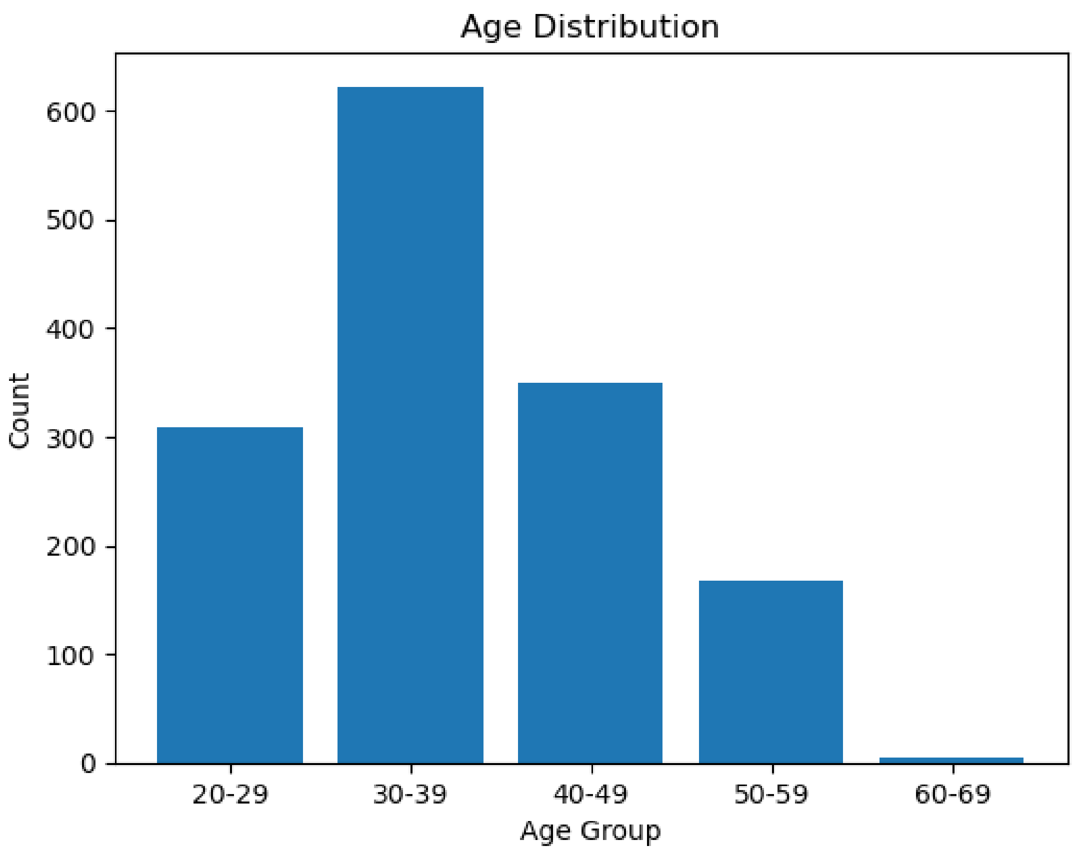


40% are female and 60% are male

15 14. AGE DISTRIBUTION

[23]:
```python
# Define age groups
age_bins = [20, 30, 40, 50, 60, 70]
age_labels = ['20-29', '30-39', '40-49', '50-59', '60-69']

# Bin the ages
age_groups = pd.cut(data['Age'], bins=age_bins, labels=age_labels, right=False)

# Count the occurrences of each age group
age_counts = age_groups.value_counts().sort_index()

# Create a bar chart
plt.bar(age_counts.index, age_counts.values)
```

```
# Add labels and title
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.title('Age Distribution')

# Show the plot
plt.show()
```

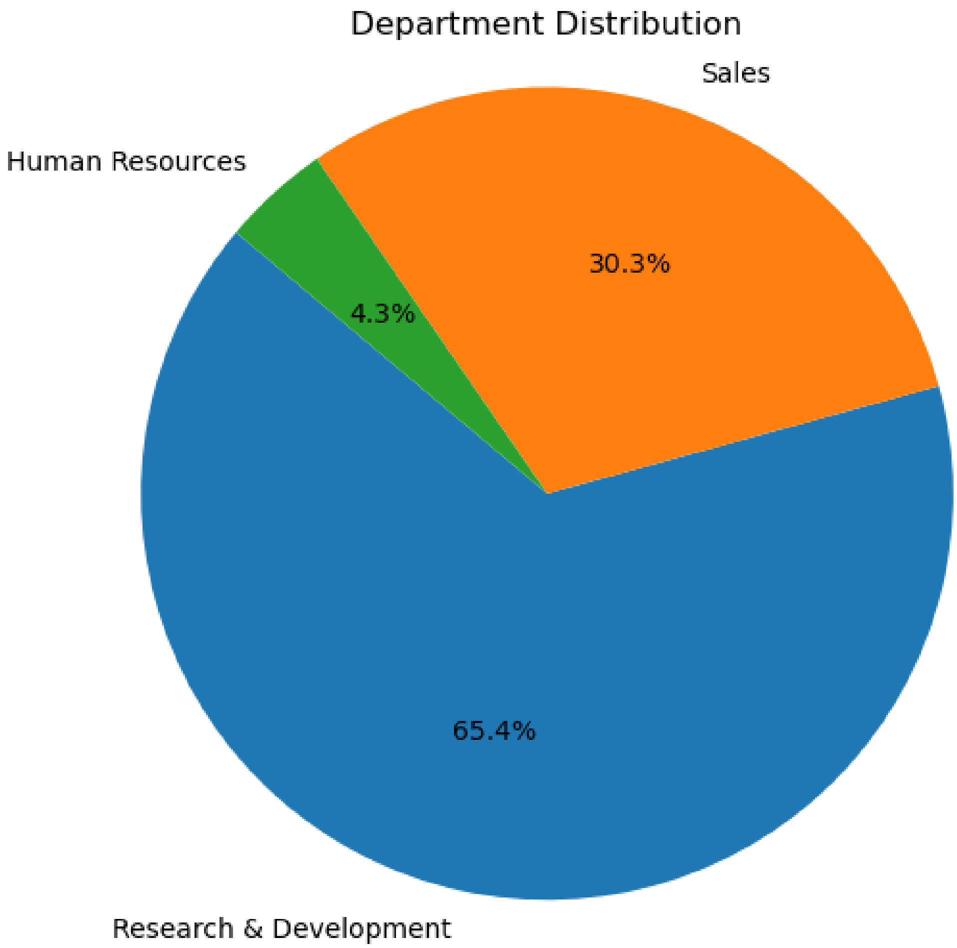


Most of the employees are between the age of 30-39.

16 15. DEPARTMENT DISTRIBUTION

[24]:
```
Department_counts = data['Department'].value_counts()

# Create a pie chart
plt.figure(figsize=(6, 6))
plt.pie(Department_counts, labels=Department_counts.index, autopct='%1.1f%%',␣
 ↪startangle=140)
plt.title('Department Distribution')
```

```
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.

# Show the plot
plt.show()
```



Most of the employees are from Research and Development Department.

[ ]:

```

```