```python
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import plotly.express as px
        import seaborn as sns
```

```python
In [3]: df = pd.read_csv("HR-Employee-Attrition.csv")
```

```python
In [4]: df.columns
```

```
Out[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')
```

```python
In [7]: df.head()
```

Out[7]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Emplo
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	

5 rows × 35 columns

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

```
Out[9]: Age                        0
        Attrition                  0
        BusinessTravel             0
        DailyRate                  0
        Department                 0
        DistanceFromHome           0
        Education                  0
        EducationField             0
        EmployeeCount              0
        EmployeeNumber             0
        EnvironmentSatisfaction    0
        Gender                     0
        HourlyRate                 0
        JobInvolvement             0
        JobLevel                   0
        JobRole                    0
        JobSatisfaction            0
        MaritalStatus              0
        MonthlyIncome              0
        MonthlyRate                0
        NumCompaniesWorked         0
        Over18                     0
        OverTime                   0
        PercentSalaryHike          0
        PerformanceRating          0
        RelationshipSatisfaction   0
        StandardHours              0
        StockOptionLevel           0
        TotalWorkingYears          0
        TrainingTimesLastYear      0
        WorkLifeBalance            0
        YearsAtCompany             0
        YearsInCurrentRole         0
        YearsSinceLastPromotion    0
        YearsWithCurrManager       0
        dtype: int64
```

```python
In [10]:   print(df.dtypes)
```

```
Age                         int64
Attrition                   object
BusinessTravel              object
DailyRate                   int64
Department                  object
DistanceFromHome            int64
Education                   int64
EducationField              object
EmployeeCount               int64
EmployeeNumber              int64
EnvironmentSatisfaction     int64
Gender                      object
HourlyRate                  int64
JobInvolvement              int64
JobLevel                    int64
JobRole                     object
JobSatisfaction             int64
MaritalStatus               object
MonthlyIncome               int64
MonthlyRate                 int64
NumCompaniesWorked          int64
Over18                      object
OverTime                    object
PercentSalaryHike           int64
PerformanceRating           int64
RelationshipSatisfaction    int64
StandardHours               int64
StockOptionLevel            int64
TotalWorkingYears           int64
TrainingTimesLastYear       int64
WorkLifeBalance             int64
YearsAtCompany              int64
YearsInCurrentRole          int64
YearsSinceLastPromotion     int64
YearsWithCurrManager        int64
dtype: object
```

```python
In [11]:   print(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
None
```

```
In [6]: #question1 What is the distribution of attrition in the dataset?
        print("question1 What is the distribution of attrition in the dataset?")
        print("i:")
        attrition_distribution = df['Attrition'].value_counts()
        print(attrition_distribution)

        print("ii overall attrition rate")
        total_employees = 1233 + 237
        attrition_rate = 237 / total_employees
        print(f"Overall Attrition Rate: {attrition_rate:.2%}")

        print("ii retention rate")
        retention_rate = 1233 / total_employees
        print(f"Overall Retention Rate: {retention_rate:.2%}")

        print("iii proportions")
        prop_no = 1233 / total_employees
        prop_yes = 237 / total_employees
        print(f"Proportion of employees who stayed (No Attrition): {prop_no:.2%}")
        print(f"Proportion of employees who left (Yes Attrition): {prop_yes:.2%}")


        sns.countplot(x='Attrition', data=df)
        plt.title('Distribution of Attrition')
        plt.xlabel('Attrition Status')
        plt.ylabel('Number of Employees')
        plt.show()
```

```
question1 What is the distribution of attrition in the dataset?
i:
Attrition
No      1233
Yes      237
Name: count, dtype: int64
ii overall attrition rate
Overall Attrition Rate: 16.12%
ii retention rate
Overall Retention Rate: 83.88%
iii proportions
Proportion of employees who stayed (No Attrition): 83.88%
Proportion of employees who left (Yes Attrition): 16.12%
```


Distribution of Attrition
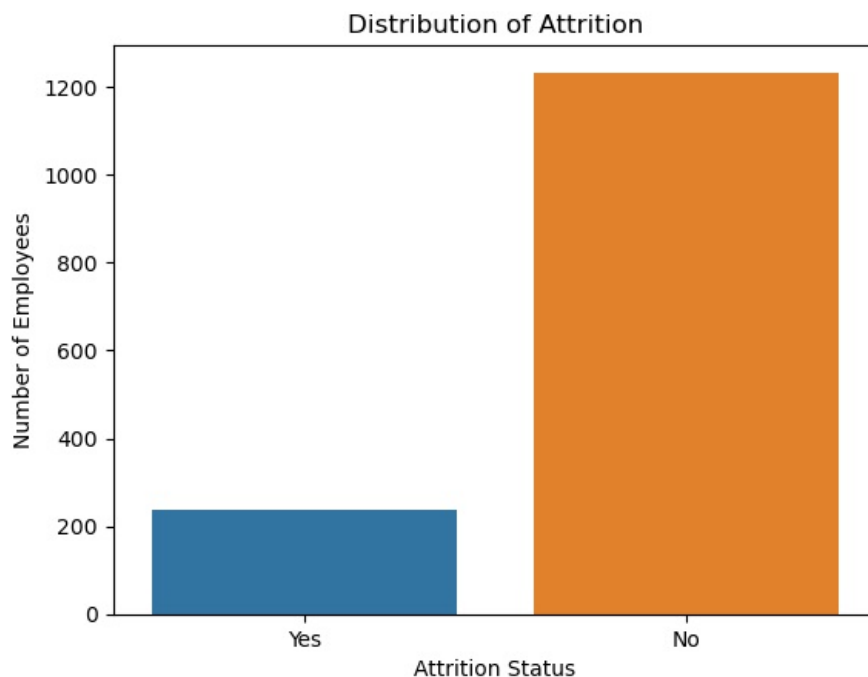
```
In [71]: #question2 How does age affect attrition?
         print("insights:")
         age_attrition = df.groupby('Attrition')['Age'].describe()
         print(age_attrition)

         # Data insights for age and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the relationship between age and attrition.

         ### Summary Statistics for Age by Attrition Status

         #### No Attrition (Employees who stayed)
```

```
  - **Count**: 1233 employees did not leave the company.
  - **Mean Age**: The average age of employees who stayed is approximately 37.56 years.
  - **Standard Deviation**: The age variability among employees who stayed is about 8.89 years.
  - **Minimum Age**: The youngest employee who stayed is 18 years old.
  - **25th Percentile (Q1)**: 25% of the employees who stayed are younger than 31 years.
  - **Median (50th Percentile)**: The median age of employees who stayed is 36 years.
  - **75th Percentile (Q3)**: 75% of the employees who stayed are younger than 43 years.
  - **Maximum Age**: The oldest employee who stayed is 60 years old.

#### Yes Attrition (Employees who left)
  - **Count**: 237 employees left the company.
  - **Mean Age**: The average age of employees who left is approximately 33.61 years.
  - **Standard Deviation**: The age variability among employees who left is about 9.69 years.
  - **Minimum Age**: The youngest employee who left is 18 years old.
  - **25th Percentile (Q1)**: 25% of the employees who left are younger than 28 years.
  - **Median (50th Percentile)**: The median age of employees who left is 32 years.
  - **75th Percentile (Q3)**: 75% of the employees who left are younger than 39 years.
  - **Maximum Age**: The oldest employee who left is 58 years old.

### Insights

1. **Age Distribution**:
   - Employees who left are generally younger (mean age of 33.61) compared to those who stayed (mean age of 37.5
   - The age range of employees who stayed (18 to 60 years) is slightly broader compared to those who left (18 '

2. **Central Tendency**:
   - The median age of employees who stayed (36 years) is higher than that of employees who left (32 years). Th

3. **Variability**:
   - The standard deviation of ages for employees who left (9.69) is higher than for those who stayed (8.89), in

4. **Percentiles**:
   - 25% of employees who left are younger than 28 years, whereas 25% of employees who stayed are younger than :
   - 75% of employees who left are younger than 39 years, whereas 75% of employees who stayed are younger than ·

5. **Age Distribution Range**:
   - Both groups have the same minimum age (18 years), but the maximum age for employees who stayed is slightly

### Conclusion
These insights indicate that younger employees are more likely to leave the company, whereas older employees te
"""

print(insights)


sns.histplot(data=df, x='Age', hue='Attrition', multiple='stack')
plt.title('Age Distribution by Attrition Status')
plt.show()
```

```
insights:
          count     mean       std   min   25%   50%   75%    max
Attrition
No        1233.0  37.561233  8.88836  18.0  31.0  36.0  43.0  60.0
Yes        237.0  33.607595  9.68935  18.0  28.0  32.0  39.0  58.0
```

Let's break down the insights from the provided data for the relationship between age and attrition.

Summary Statistics for Age by Attrition Status

No Attrition (Employees who stayed)
- **Count**: 1233 employees did not leave the company.
- **Mean Age**: The average age of employees who stayed is approximately 37.56 years.
- **Standard Deviation**: The age variability among employees who stayed is about 8.89 years.
- **Minimum Age**: The youngest employee who stayed is 18 years old.
- **25th Percentile (Q1)**: 25% of the employees who stayed are younger than 31 years.
- **Median (50th Percentile)**: The median age of employees who stayed is 36 years.
- **75th Percentile (Q3)**: 75% of the employees who stayed are younger than 43 years.
- **Maximum Age**: The oldest employee who stayed is 60 years old.

Yes Attrition (Employees who left)
- **Count**: 237 employees left the company.
- **Mean Age**: The average age of employees who left is approximately 33.61 years.
- **Standard Deviation**: The age variability among employees who left is about 9.69 years.
- **Minimum Age**: The youngest employee who left is 18 years old.
- **25th Percentile (Q1)**: 25% of the employees who left are younger than 28 years.
- **Median (50th Percentile)**: The median age of employees who left is 32 years.
- **75th Percentile (Q3)**: 75% of the employees who left are younger than 39 years.
- **Maximum Age**: The oldest employee who left is 58 years old.
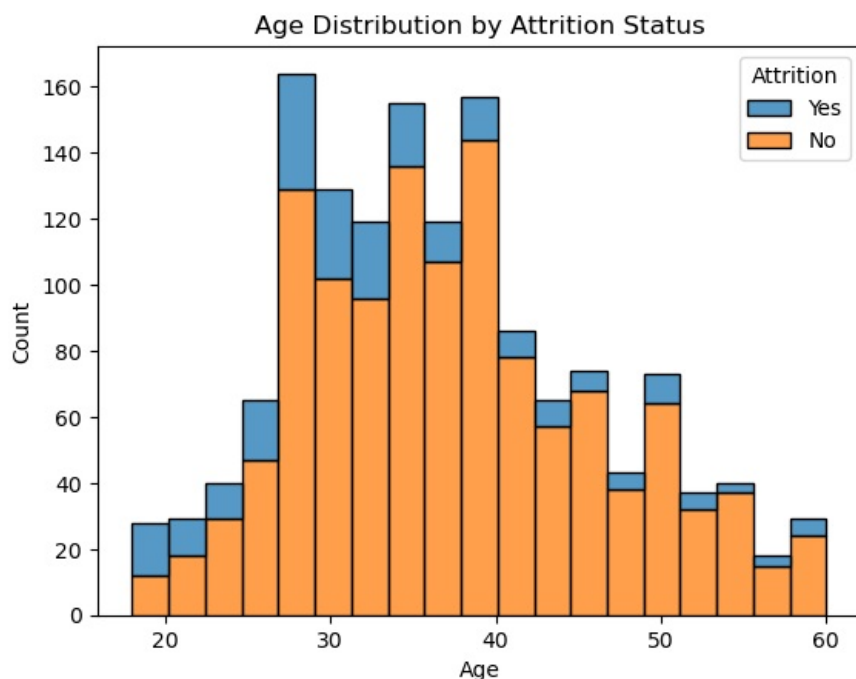
Insights

1. **Age Distribution**:
   - Employees who left are generally younger (mean age of 33.61) compared to those who stayed (mean age of 37.56).
   - The age range of employees who stayed (18 to 60 years) is slightly broader compared to those who left (18 to 58 years).

2. **Central Tendency**:
   - The median age of employees who stayed (36 years) is higher than that of employees who left (32 years). This suggests that older employees are more likely to stay with the company.

3. **Variability**:
   - The standard deviation of ages for employees who left (9.69) is higher than for those who stayed (8.89), indicating greater age variability among the former group.

4. **Percentiles**:
   - 25% of employees who left are younger than 28 years, whereas 25% of employees who stayed are younger than 31 years.
   - 75% of employees who left are younger than 39 years, whereas 75% of employees who stayed are younger than 43 years.

5. **Age Distribution Range**:
   - Both groups have the same minimum age (18 years), but the maximum age for employees who stayed is slightly higher (60 years) compared to those who left (58 years).

Conclusion
These insights indicate that younger employees are more likely to leave the company, whereas older employees tend to stay longer. The higher age variability among employees who left might suggest that attrition is influenced by other factors that interact with age, such as career stage, opportunities for growth, and work-life balance preferences. Understanding these patterns can help the company tailor its retention strategies to different age groups, potentially addressing specific needs or concerns that younger employees might have.

C:\Users\Admin\anaconda3\Lib\site-packages\seaborn_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):

Age Distribution by Attrition Status

```
In [72]: #question3 Is there a relationship between monthly income and attrition?
         ("insights:")
         income_attrition = df.groupby('Attrition')['MonthlyIncome'].describe()
         print(income_attrition)

         # Data insights for monthly income and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the relationship between monthly income and attrition.

         ### Summary Statistics for Monthly Income by Attrition Status

         #### No Attrition (Employees who stayed)
         - **Count**: 1233 employees did not leave the company.
         - **Mean Monthly Income**: The average monthly income of employees who stayed is approximately $6832.74.
         - **Standard Deviation**: The variability in monthly income among employees who stayed is about $4818.21.
         - **Minimum Monthly Income**: The lowest monthly income among employees who stayed is $1051.
         - **25th Percentile (Q1)**: 25% of the employees who stayed have a monthly income less than $3211.
         - **Median (50th Percentile)**: The median monthly income of employees who stayed is $5204.
         - **75th Percentile (Q3)**: 75% of the employees who stayed have a monthly income less than $8834.
         - **Maximum Monthly Income**: The highest monthly income among employees who stayed is $19999.

         #### Yes Attrition (Employees who left)
         - **Count**: 237 employees left the company.
         - **Mean Monthly Income**: The average monthly income of employees who left is approximately $4787.09.
         - **Standard Deviation**: The variability in monthly income among employees who left is about $3640.21.
         - **Minimum Monthly Income**: The lowest monthly income among employees who left is $1009.
         - **25th Percentile (Q1)**: 25% of the employees who left have a monthly income less than $2373.
         - **Median (50th Percentile)**: The median monthly income of employees who left is $3202.
         - **75th Percentile (Q3)**: 75% of the employees who left have a monthly income less than $5916.
         - **Maximum Monthly Income**: The highest monthly income among employees who left is $19859.

         ### Insights

         1. **Monthly Income Distribution**:
            - Employees who left have a lower average monthly income ($4787.09) compared to those who stayed ($6832.74).
            - The variability in monthly income is higher among employees who stayed ($4818.21) compared to those who le

         2. **Central Tendency**:
            - The median monthly income of employees who left ($3202) is lower than that of employees who stayed ($5204)

         3. **Percentiles**:
            - 25% of employees who left have a monthly income less than $2373, whereas 25% of employees who stayed have a
            - 75% of employees who left have a monthly income less than $5916, whereas 75% of employees who stayed have a

         ### Conclusion
         These insights suggest that employees with higher monthly incomes are more likely to stay with the company, whi
         """

         print(insights)


         sns.boxplot(data=df, x='Attrition', y='MonthlyIncome')
         plt.title('Monthly Income vs Attrition')
         plt.show()
```

```
           count       mean          std      min      25%      50%      75%  \
Attrition
No         1233.0  6832.739659  4818.208001  1051.0  3211.0  5204.0  8834.0
Yes         237.0  4787.092827  3640.210367  1009.0  2373.0  3202.0  5916.0

               max
Attrition
No         19999.0
Yes        19859.0
```

Let's break down the insights from the provided data for the relationship between monthly income and attrition.

Summary Statistics for Monthly Income by Attrition Status

No Attrition (Employees who stayed)
- **Count**: 1233 employees did not leave the company.
- **Mean Monthly Income**: The average monthly income of employees who stayed is approximately $6832.74.
- **Standard Deviation**: The variability in monthly income among employees who stayed is about $4818.21.
- **Minimum Monthly Income**: The lowest monthly income among employees who stayed is $1051.
- **25th Percentile (Q1)**: 25% of the employees who stayed have a monthly income less than $3211.
- **Median (50th Percentile)**: The median monthly income of employees who stayed is $5204.
- **75th Percentile (Q3)**: 75% of the employees who stayed have a monthly income less than $8834.
- **Maximum Monthly Income**: The highest monthly income among employees who stayed is $19999.

Yes Attrition (Employees who left)
- **Count**: 237 employees left the company.
- **Mean Monthly Income**: The average monthly income of employees who left is approximately $4787.09.
- **Standard Deviation**: The variability in monthly income among employees who left is about $3640.21.
- **Minimum Monthly Income**: The lowest monthly income among employees who left is $1009.
- **25th Percentile (Q1)**: 25% of the employees who left have a monthly income less than $2373.
- **Median (50th Percentile)**: The median monthly income of employees who left is $3202.
- **75th Percentile (Q3)**: 75% of the employees who left have a monthly income less than $5916.
- **Maximum Monthly Income**: The highest monthly income among employees who left is $19859.

Insights

1. **Monthly Income Distribution**:
   - Employees who left have a lower average monthly income ($4787.09) compared to those who stayed ($6832.74).
   - The variability in monthly income is higher among employees who stayed ($4818.21) compared to those who left ($3640.21).

2. **Central Tendency**:
   - The median monthly income of employees who left ($3202) is lower than that of employees who stayed ($5204), indicating that employees with higher incomes are more likely to stay.

3. **Percentiles**:
   - 25% of employees who left have a monthly income less than $2373, whereas 25% of employees who stayed have a monthly income less than $3211.
   - 75% of employees who left have a monthly income less than $5916, whereas 75% of employees who stayed have a monthly income less than $8834.

Conclusion
These insights suggest that employees with higher monthly incomes are more likely to stay with the company, while those with lower incomes are more likely to leave. However, other factors such as job satisfaction, work-life balance, and career advancement opportunities may also influence attrition rates.

Monthly Income vs Attrition
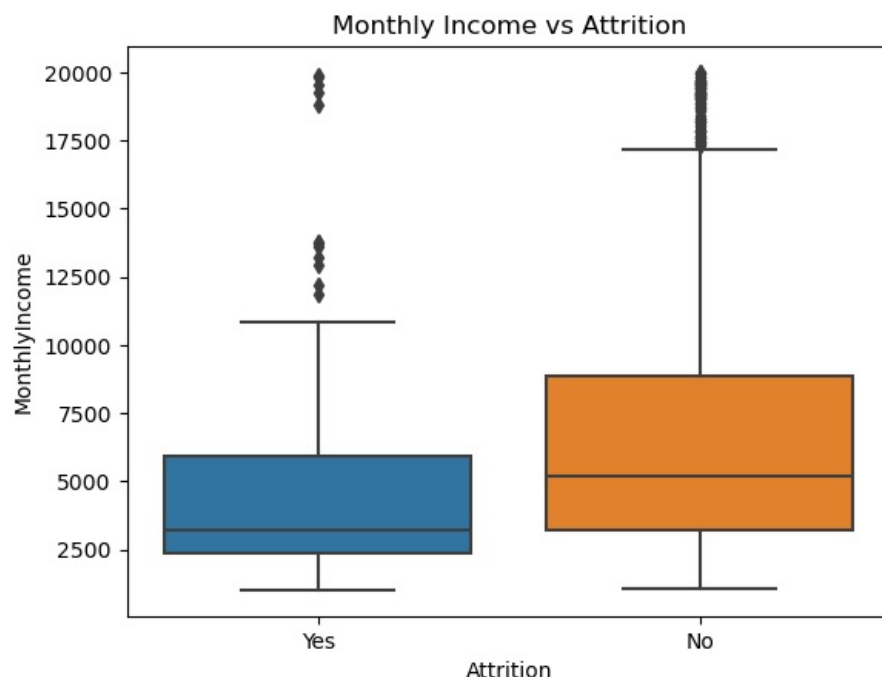
```
In [73]: #question5 What is the distribution of job satisfaction levels among employees?
         job_satisfaction_distribution = df['JobSatisfaction'].value_counts()
         print(job_satisfaction_distribution)

         # Data insights for job satisfaction and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the relationship between job satisfaction and attrition

         ### Summary Statistics for Job Satisfaction by Attrition Status

         - **Job Satisfaction Level 4**: 459 employees
         - **Job Satisfaction Level 3**: 442 employees
         - **Job Satisfaction Level 1**: 289 employees
         - **Job Satisfaction Level 2**: 280 employees

         ### Insights

         1. **Job Satisfaction Distribution**:
            - The majority of employees in the dataset have job satisfaction levels rated as 4 and 3, with 459 and 442 em
            - There are fewer employees with job satisfaction levels rated as 1 and 2, with 289 and 280 employees respec

         2. **Attrition Across Job Satisfaction Levels**:
            - We need additional data or information to analyze the attrition rates across different job satisfaction lev

         ### Conclusion
         This summary provides the distribution of employees across different job satisfaction levels. However, to under
         """

         print(insights)

         sns.countplot(data=df, x='JobSatisfaction')
         plt.title('Distribution of Job Satisfaction Levels')
         plt.show()
```

```
JobSatisfaction
4    459
3    442
1    289
2    280
Name: count, dtype: int64

Let's break down the insights from the provided data for the relationship between job satisfaction and attrition
.

### Summary Statistics for Job Satisfaction by Attrition Status

- **Job Satisfaction Level 4**: 459 employees
- **Job Satisfaction Level 3**: 442 employees
- **Job Satisfaction Level 1**: 289 employees
- **Job Satisfaction Level 2**: 280 employees

### Insights

1. **Job Satisfaction Distribution**:
   - The majority of employees in the dataset have job satisfaction levels rated as 4 and 3, with 459 and 442 em
ployees respectively.
   - There are fewer employees with job satisfaction levels rated as 1 and 2, with 289 and 280 employees respect
ively.

2. **Attrition Across Job Satisfaction Levels**:
   - We need additional data or information to analyze the attrition rates across different job satisfaction lev
els. This summary provides the count of employees for each job satisfaction level but does not directly indicate
the attrition status of these employees.

### Conclusion
This summary provides the distribution of employees across different job satisfaction levels. However, to unders
tand the relationship between job satisfaction and attrition, we need additional information on attrition rates
within each job satisfaction level. Analyzing attrition rates within each job satisfaction level can provide ins
ights into whether job satisfaction influences employee retention.
```

Distribution of Job Satisfaction Levels
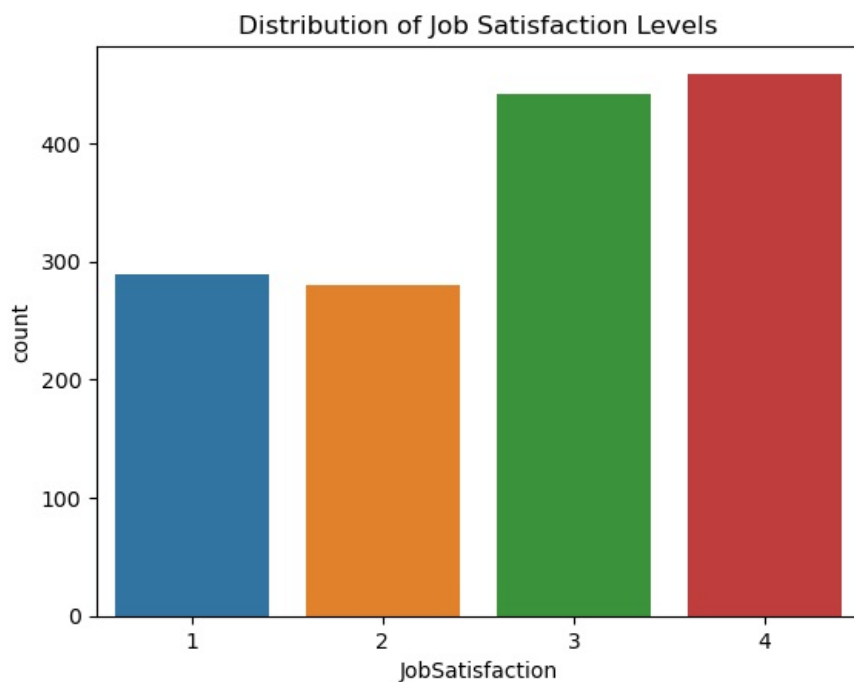
In [75]:
```python
#question6 Is there a significant difference in education levels between employees who left and those who stayed
education_attrition = df.groupby('Attrition')['Education'].value_counts()
print(education_attrition)

# Data insights for education level and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between education level and attrition

### Summary Statistics for Education Level by Attrition Status

#### No Attrition (Employees who stayed)
- **Education Level 3**: 473 employees
- **Education Level 4**: 340 employees
- **Education Level 2**: 238 employees
- **Education Level 1**: 139 employees
- **Education Level 5**: 43 employees

#### Yes Attrition (Employees who left)
- **Education Level 3**: 99 employees
- **Education Level 4**: 58 employees
- **Education Level 2**: 44 employees
- **Education Level 1**: 31 employees
- **Education Level 5**: 5 employees

### Insights

1. **Distribution of Education Levels**:
   - Education levels 3 and 4 are the most common among employees who both stayed and left the company.
   - Education levels 1, 2, and 5 have lower representation among employees who both stayed and left.

2. **Attrition Across Education Levels**:
   - The majority of employees across all education levels stayed in the company.
   - Education level 3 has the highest count of employees who both stayed and left.
   - Education level 5 has the lowest count of employees in both categories, indicating that employees with higl

### Conclusion
This summary provides insights into the distribution of employees across different education levels and their at
"""

print(insights)
```

```
sns.countplot(data=df, x='Education', hue='Attrition')
plt.title('Education Level by Attrition Status')
plt.show()
```

```
Attrition  Education
No         3            473
           4            340
           2            238
           1            139
           5             43
Yes        3             99
           4             58
           2             44
           1             31
           5              5
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between education level and attrition.

Summary Statistics for Education Level by Attrition Status

No Attrition (Employees who stayed)
- **Education Level 3**: 473 employees
- **Education Level 4**: 340 employees
- **Education Level 2**: 238 employees
- **Education Level 1**: 139 employees
- **Education Level 5**: 43 employees

Yes Attrition (Employees who left)
- **Education Level 3**: 99 employees
- **Education Level 4**: 58 employees
- **Education Level 2**: 44 employees
- **Education Level 1**: 31 employees
- **Education Level 5**: 5 employees

Insights

1. **Distribution of Education Levels**:
    - Education levels 3 and 4 are the most common among employees who both stayed and left the company.
    - Education levels 1, 2, and 5 have lower representation among employees who both stayed and left.

2. **Attrition Across Education Levels**:
    - The majority of employees across all education levels stayed in the company.
    - Education level 3 has the highest count of employees who both stayed and left.
    - Education level 5 has the lowest count of employees in both categories, indicating that employees with higher education levels are less likely to leave.

Conclusion
This summary provides insights into the distribution of employees across different education levels and their attrition status. While education level 3 is the most common among employees who stayed and left, further analysis is needed to understand the impact of education level on attrition rates. Factors such as career advancement opportunities, job satisfaction, and work-life balance may also influence attrition rates across different education levels.

Education Level by Attrition Status
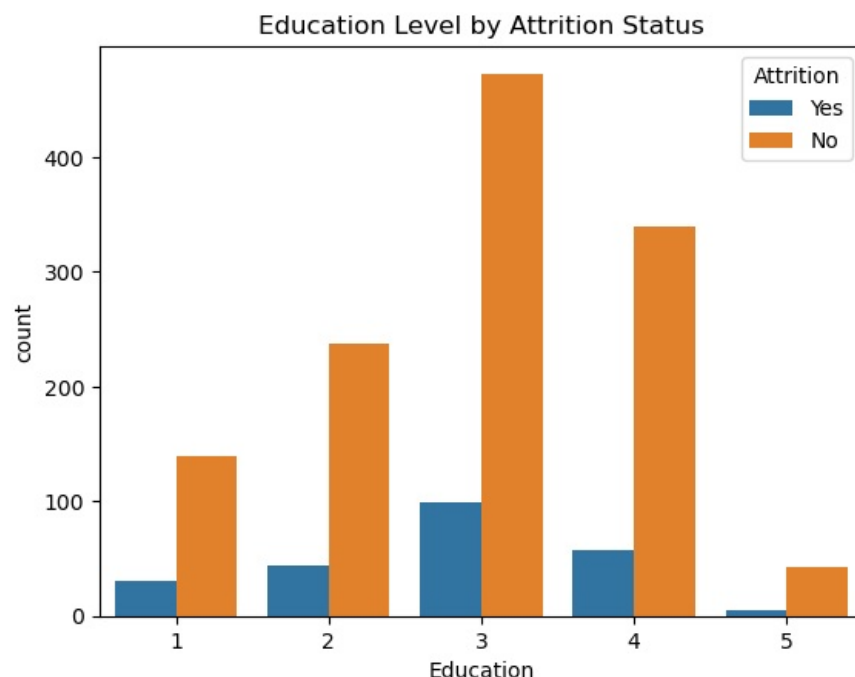
In [76]: ```
#question7 How does the number of companies worked affect attrition?
num_companies_attrition = df.groupby('Attrition')['NumCompaniesWorked'].value_counts()
```

```python
print(num_companies_attrition)

# Data insights for the number of companies worked and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between the number of companies worked

### Summary Statistics for Number of Companies Worked by Attrition Status

#### No Attrition (Employees who stayed)
- **1 Company**: 423 employees
- **0 Companies**: 174 employees
- **3 Companies**: 143 employees
- **2 Companies**: 130 employees
- **4 Companies**: 122 employees
- **7 Companies**: 57 employees
- **6 Companies**: 54 employees
- **5 Companies**: 47 employees
- **8 Companies**: 43 employees
- **9 Companies**: 40 employees

#### Yes Attrition (Employees who left)
- **1 Company**: 98 employees
- **0 Companies**: 23 employees
- **4 Companies**: 17 employees
- **7 Companies**: 17 employees
- **2 Companies**: 16 employees
- **3 Companies**: 16 employees
- **5 Companies**: 16 employees
- **6 Companies**: 16 employees
- **9 Companies**: 12 employees
- **8 Companies**: 6 employees

### Insights

1. **Distribution of Companies Worked**:
   - The highest number of employees who stayed have worked at only 1 company (423 employees), followed by those
   - Similarly, the highest number of employees who left have also worked at only 1 company (98 employees).

2. **Attrition Rates Across Number of Companies Worked**:
   - Employees who have worked at 1 company represent the highest count in both categories, suggesting that a s:
   - Employees who have worked at more than one company show lower counts for both staying and leaving, with a

3. **Comparative Analysis**:
   - The ratio of employees who stayed to those who left is highest for those who worked at 0 or 1 company, ind:
   - Attrition seems to be more evenly distributed among employees who have worked at multiple companies (2-9),

### Conclusion
These insights suggest that employees who have only worked at 1 or no other company are more likely to stay with
"""

print(insights)


sns.countplot(data=df, x='NumCompaniesWorked', hue='Attrition')
plt.title('Number of Companies Worked vs Attrition')
plt.show()
```

```
Attrition  NumCompaniesWorked
No         1                   423
           0                   174
           3                   143
           2                   130
           4                   122
           7                    57
           6                    54
           5                    47
           8                    43
           9                    40
Yes        1                    98
           0                    23
           4                    17
           7                    17
           2                    16
           3                    16
           5                    16
           6                    16
           9                    12
           8                     6
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between the number of companies worked and attrition.

Summary Statistics for Number of Companies Worked by Attrition Status

No Attrition (Employees who stayed)
- **1 Company**: 423 employees
- **0 Companies**: 174 employees
- **3 Companies**: 143 employees
- **2 Companies**: 130 employees
- **4 Companies**: 122 employees
- **7 Companies**: 57 employees
- **6 Companies**: 54 employees
- **5 Companies**: 47 employees
- **8 Companies**: 43 employees
- **9 Companies**: 40 employees

Yes Attrition (Employees who left)
- **1 Company**: 98 employees
- **0 Companies**: 23 employees
- **4 Companies**: 17 employees
- **7 Companies**: 17 employees
- **2 Companies**: 16 employees
- **3 Companies**: 16 employees
- **5 Companies**: 16 employees
- **6 Companies**: 16 employees
- **9 Companies**: 12 employees
- **8 Companies**: 6 employees

Insights

1. **Distribution of Companies Worked**:
   - The highest number of employees who stayed have worked at only 1 company (423 employees), followed by those who have worked at 0 companies (174 employees).
   - Similarly, the highest number of employees who left have also worked at only 1 company (98 employees).

2. **Attrition Rates Across Number of Companies Worked**:
   - Employees who have worked at 1 company represent the highest count in both categories, suggesting that a significant number of employees either stay or leave after working at only one company.
   - Employees who have worked at more than one company show lower counts for both staying and leaving, with a generally decreasing trend as the number of companies worked increases.

3. **Comparative Analysis**:
   - The ratio of employees who stayed to those who left is highest for those who worked at 0 or 1 company, indicating higher retention in this group.
   - Attrition seems to be more evenly distributed among employees who have worked at multiple companies (2-9), but the absolute numbers are much lower.

Conclusion
These insights suggest that employees who have only worked at 1 or no other company are more likely to stay with the current company. As the number of companies worked increases, the likelihood of staying decreases slightly, though the total number of employees in these categories is lower. This trend could imply that employees with more job changes are less committed to staying long-term or that they have higher expectations for career growth and opportunities.

Number of Companies Worked vs Attrition

In [77]:
```python
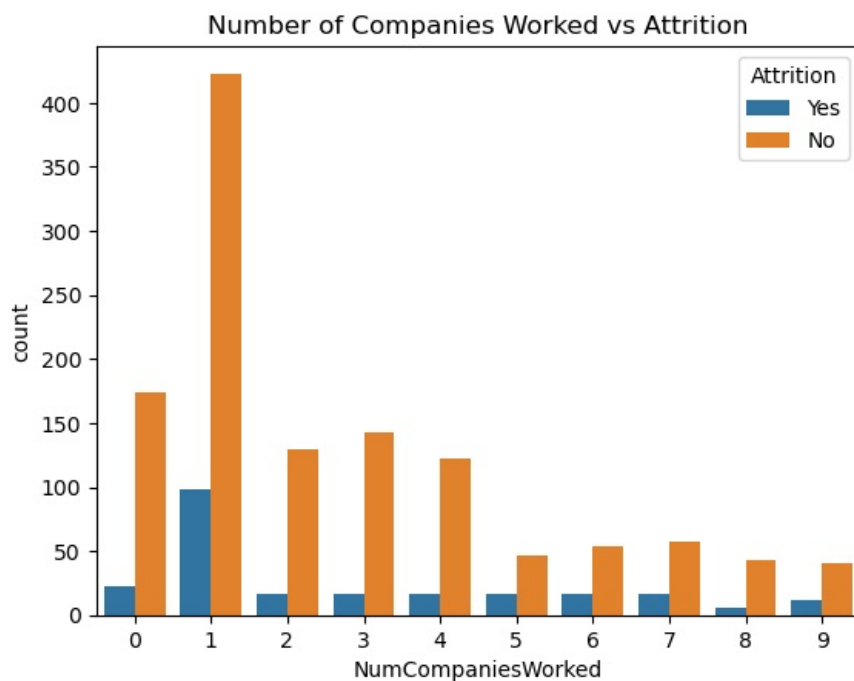#question8 What is the correlation between total working years and attrition?
total_working_years_attrition = df.groupby('Attrition')['TotalWorkingYears'].describe()
print(total_working_years_attrition)
# Data insights for total working years and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between total working years and attri

### Summary Statistics for Total Working Years by Attrition Status

#### No Attrition (Employees who stayed)
- **Count**: 1233 employees did not leave the company.
- **Mean Total Working Years**: The average total working years of employees who stayed is approximately 11.86 y
- **Standard Deviation**: The variability in total working years among employees who stayed is about 7.76 years
- **Minimum Total Working Years**: The minimum total working years among employees who stayed is 0 years.
- **25th Percentile (Q1)**: 25% of the employees who stayed have total working years less than 6 years.
- **Median (50th Percentile)**: The median total working years of employees who stayed is 10 years.
- **75th Percentile (Q3)**: 75% of the employees who stayed have total working years less than 16 years.
- **Maximum Total Working Years**: The maximum total working years among employees who stayed is 38 years.

#### Yes Attrition (Employees who left)
- **Count**: 237 employees left the company.
- **Mean Total Working Years**: The average total working years of employees who left is approximately 8.24 yea
- **Standard Deviation**: The variability in total working years among employees who left is about 7.17 years.
- **Minimum Total Working Years**: The minimum total working years among employees who left is 0 years.
- **25th Percentile (Q1)**: 25% of the employees who left have total working years less than 3 years.
- **Median (50th Percentile)**: The median total working years of employees who left is 7 years.
- **75th Percentile (Q3)**: 75% of the employees who left have total working years less than 10 years.
- **Maximum Total Working Years**: The maximum total working years among employees who left is 40 years.

### Insights

1. **Total Working Years Distribution**:
   - Employees who stayed have a higher average total working years (11.86 years) compared to those who left (8
   - The median total working years for employees who stayed (10 years) is also higher than that for employees w

2. **Variability**:
   - The standard deviation of total working years for employees who stayed (7.76 years) is slightly higher than

3. **Percentiles**:
   - 25% of employees who stayed have total working years less than 6 years, whereas 25% of employees who left l
   - 75% of employees who stayed have total working years less than 16 years, whereas 75% of employees who left

4. **Working Years Range**:
   - Both groups have a minimum of 0 years, but the maximum total working years for employees who left (40 years

### Conclusion
These insights indicate that employees with more total working years are more likely to stay with the company, w
"""

print(insights)


sns.boxplot(data=df, x='Attrition', y='TotalWorkingYears')
plt.title('Total Working Years vs Attrition')
```

```
plt.show()
```

	count	mean	std	min	25%	50%	75%	max
Attrition								
No	1233.0	11.862936	7.760719	0.0	6.0	10.0	16.0	38.0
Yes	237.0	8.244726	7.169204	0.0	3.0	7.0	10.0	40.0

Let's break down the insights from the provided data for the relationship between total working years and attrition.

Summary Statistics for Total Working Years by Attrition Status

No Attrition (Employees who stayed)
- **Count**: 1233 employees did not leave the company.
- **Mean Total Working Years**: The average total working years of employees who stayed is approximately 11.86 years.
- **Standard Deviation**: The variability in total working years among employees who stayed is about 7.76 years.
- **Minimum Total Working Years**: The minimum total working years among employees who stayed is 0 years.
- **25th Percentile (Q1)**: 25% of the employees who stayed have total working years less than 6 years.
- **Median (50th Percentile)**: The median total working years of employees who stayed is 10 years.
- **75th Percentile (Q3)**: 75% of the employees who stayed have total working years less than 16 years.
- **Maximum Total Working Years**: The maximum total working years among employees who stayed is 38 years.

Yes Attrition (Employees who left)
- **Count**: 237 employees left the company.
- **Mean Total Working Years**: The average total working years of employees who left is approximately 8.24 years.
- **Standard Deviation**: The variability in total working years among employees who left is about 7.17 years.
- **Minimum Total Working Years**: The minimum total working years among employees who left is 0 years.
- **25th Percentile (Q1)**: 25% of the employees who left have total working years less than 3 years.
- **Median (50th Percentile)**: The median total working years of employees who left is 7 years.
- **75th Percentile (Q3)**: 75% of the employees who left have total working years less than 10 years.
- **Maximum Total Working Years**: The maximum total working years among employees who left is 40 years.

Insights

1. **Total Working Years Distribution**:
   - Employees who stayed have a higher average total working years (11.86 years) compared to those who left (8.24 years).
   - The median total working years for employees who stayed (10 years) is also higher than that for employees who left (7 years).

2. **Variability**:
   - The standard deviation of total working years for employees who stayed (7.76 years) is slightly higher than for those who left (7.17 years), indicating a wider spread of working years among the former group.

3. **Percentiles**:
   - 25% of employees who stayed have total working years less than 6 years, whereas 25% of employees who left have total working years less than 3 years.
   - 75% of employees who stayed have total working years less than 16 years, whereas 75% of employees who left have total working years less than 10 years.

4. **Working Years Range**:
   - Both groups have a minimum of 0 years, but the maximum total working years for employees who left (40 years) is slightly higher than for those who stayed (38 years).

Conclusion
These insights indicate that employees with more total working years are more likely to stay with the company, whereas those with fewer total working years are more likely to leave. The data shows a correlation where higher total working years are associated with lower attrition rates. This trend could suggest that more experienced employees are more stable and less likely to leave, possibly due to greater job satisfaction, better compensation, or stronger company loyalty developed over time.

Total Working Years vs Attrition

```python
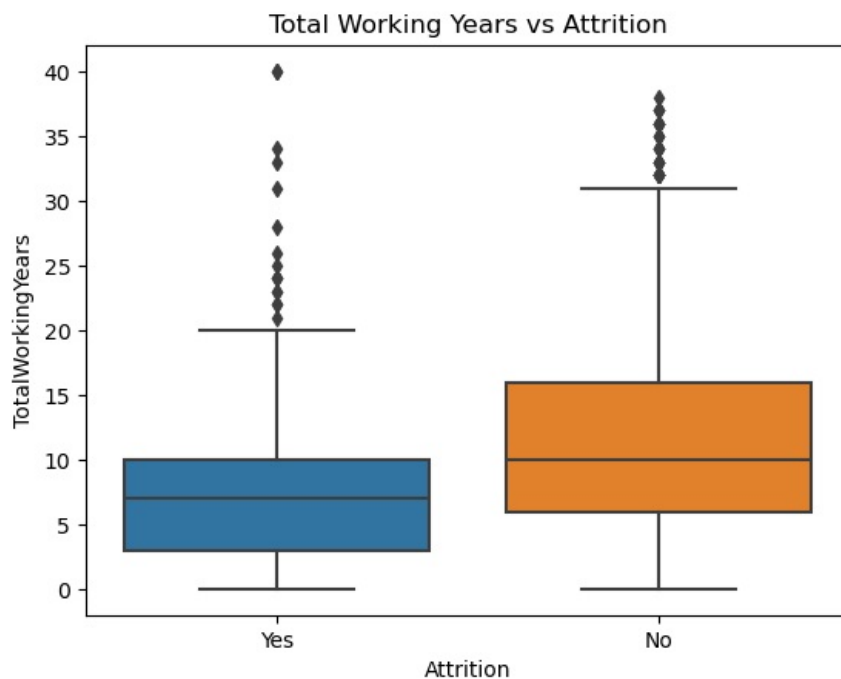#question9 what is the gender distribution in the dataset?
gender_distribution = df['Gender'].value_counts()
print(gender_distribution)
# Data insights for gender distribution in the dataset

# Insights
insights = """
Let's break down the insights from the provided data for the gender distribution in the dataset.

### Gender Distribution

- **Male**: 882 employees
- **Female**: 588 employees

### Insights

1. **Overall Distribution**:
   - The dataset contains 882 male employees and 588 female employees.
   - This indicates that approximately 60% of the employees are male and 40% are female.

2. **Representation**:
   - The gender distribution shows a higher number of male employees compared to female employees.
   - The ratio of male to female employees is roughly 1.5:1.

### Conclusion
The dataset exhibits a gender distribution where males are the majority, making up about 60% of the total employ
"""

print(insights)


sns.countplot(data=df, x='Gender')
plt.title('Gender Distribution')
plt.show()
```

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

Let's break down the insights from the provided data for the gender distribution in the dataset.

Gender Distribution

- **Male**: 882 employees
- **Female**: 588 employees

Insights

1. **Overall Distribution**:
   - The dataset contains 882 male employees and 588 female employees.
   - This indicates that approximately 60% of the employees are male and 40% are female.

2. **Representation**:
   - The gender distribution shows a higher number of male employees compared to female employees.
   - The ratio of male to female employees is roughly 1.5:1.

Conclusion
The dataset exhibits a gender distribution where males are the majority, making up about 60% of the total employee count. Understanding the gender distribution is crucial for analyzing various aspects of the workforce, such as diversity, inclusion, and potential disparities in attrition rates, job satisfaction, and other metrics. This insight can help the company address any gender-related issues and work towards a more balanced and equitable work environment.

Gender Distribution
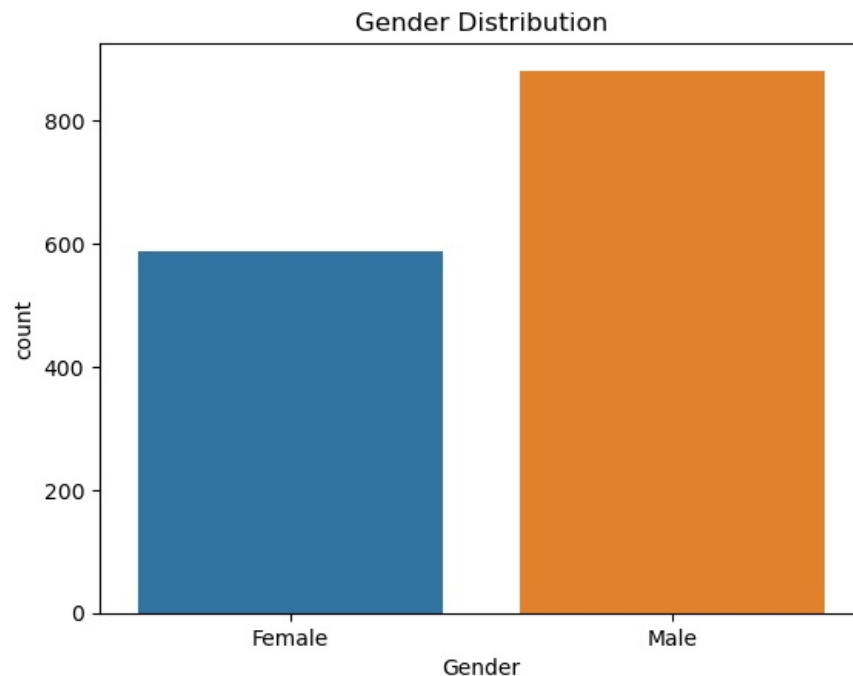
```python
#question 10  Does gender affect attrition?
gender_attrition = df.groupby('Gender')['Attrition'].value_counts()
print(gender_attrition)

# Data insights for gender and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between gender and attrition.

### Gender and Attrition

#### Female
- **No Attrition**: 501 employees
- **Yes Attrition**: 87 employees

#### Male
- **No Attrition**: 732 employees
- **Yes Attrition**: 150 employees

### Insights

1. **Attrition Rates by Gender**:
   - **Female Employees**:
     - 501 female employees stayed with the company.
     - 87 female employees left the company.
     - Attrition rate for females = (87 / (501 + 87)) * 100 ≈ 14.8%
```

```
       - **Male Employees**:
         - 732 male employees stayed with the company.
         - 150 male employees left the company.
         - Attrition rate for males = (150 / (732 + 150)) * 100 ≈ 17%

     2. **Comparison of Attrition Rates**:
        - The attrition rate for female employees is approximately 14.8%.
        - The attrition rate for male employees is approximately 17%.
        - This suggests that male employees have a slightly higher attrition rate compared to female employees.

     3. **Overall Gender Distribution in Attrition**:
        - Of the total 237 employees who left the company, 87 were female and 150 were male.
        - Female employees make up approximately 36.7% of the attrition cases.
        - Male employees make up approximately 63.3% of the attrition cases.

     ### Conclusion
     The analysis indicates that gender does have an effect on attrition, with male employees exhibiting a slightly l
     """

     print(insights)


     sns.countplot(data=df, x='Gender', hue='Attrition')
     plt.title('Attrition by Gender')
     plt.show()
```

```
Gender  Attrition
Female  No              501
        Yes              87
Male    No              732
        Yes             150
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between gender and attrition.

Gender and Attrition

Female
- **No Attrition**: 501 employees
- **Yes Attrition**: 87 employees

Male
- **No Attrition**: 732 employees
- **Yes Attrition**: 150 employees

Insights

1. **Attrition Rates by Gender**:
   - **Female Employees**:
     - 501 female employees stayed with the company.
     - 87 female employees left the company.
     - Attrition rate for females = (87 / (501 + 87)) * 100 ≈ 14.8%
   - **Male Employees**:
     - 732 male employees stayed with the company.
     - 150 male employees left the company.
     - Attrition rate for males = (150 / (732 + 150)) * 100 ≈ 17%

2. **Comparison of Attrition Rates**:
   - The attrition rate for female employees is approximately 14.8%.
   - The attrition rate for male employees is approximately 17%.
   - This suggests that male employees have a slightly higher attrition rate compared to female employees.

3. **Overall Gender Distribution in Attrition**:
   - Of the total 237 employees who left the company, 87 were female and 150 were male.
   - Female employees make up approximately 36.7% of the attrition cases.
   - Male employees make up approximately 63.3% of the attrition cases.

Conclusion
The analysis indicates that gender does have an effect on attrition, with male employees exhibiting a slightly higher attrition rate (17%) compared to female employees (14.8%). This could suggest that male employees may be more likely to leave the company than female employees. Understanding these patterns can help the company tailor its retention strategies to address the specific needs and concerns of both male and female employees, potentially reducing overall attrition rates.

Attrition by Gender
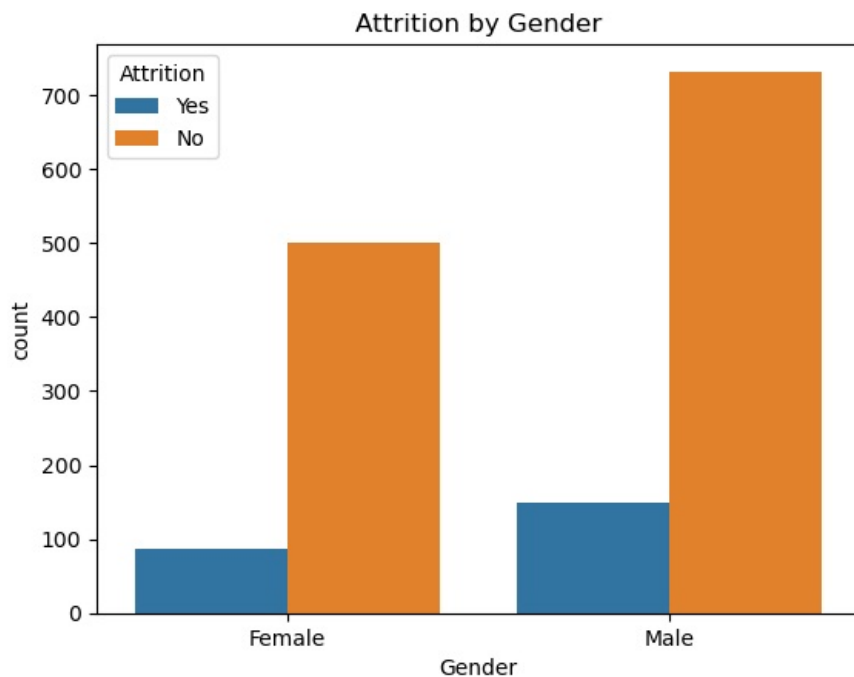

```
In [80]: #question 11What is the relationship between marital status and attrition?
         marital_status_attrition = df.groupby('Attrition')['MaritalStatus'].value_counts()
         print(marital_status_attrition)

         # Data insights for marital status and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the relationship between marital status and attrition.

         ### Marital Status and Attrition

         #### No Attrition (Employees who stayed)
         - **Married**: 589 employees
         - **Single**: 350 employees
         - **Divorced**: 294 employees

         #### Yes Attrition (Employees who left)
         - **Single**: 120 employees
         - **Married**: 84 employees
         - **Divorced**: 33 employees

         ### Insights

         1. **Attrition Rates by Marital Status**:
            - **Married Employees**:
              - 589 married employees stayed with the company.
              - 84 married employees left the company.
              - Attrition rate for married employees = (84 / (589 + 84)) * 100 ≈ 12.5%
            - **Single Employees**:
              - 350 single employees stayed with the company.
              - 120 single employees left the company.
              - Attrition rate for single employees = (120 / (350 + 120)) * 100 ≈ 25.5%
            - **Divorced Employees**:
              - 294 divorced employees stayed with the company.
              - 33 divorced employees left the company.
              - Attrition rate for divorced employees = (33 / (294 + 33)) * 100 ≈ 10.1%

         2. **Comparison of Attrition Rates**:
            - The attrition rate for married employees is approximately 12.5%.
            - The attrition rate for single employees is significantly higher at approximately 25.5%.
            - The attrition rate for divorced employees is the lowest at approximately 10.1%.

         3. **Overall Marital Status Distribution in Attrition**:
            - Of the total 237 employees who left the company:
```

```
Attrition  MaritalStatus
No         Married          589
           Single           350
           Divorced         294
Yes        Single           120
           Married           84
           Divorced          33
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between marital status and attrition.

Marital Status and Attrition

No Attrition (Employees who stayed)
- **Married**: 589 employees
- **Single**: 350 employees
- **Divorced**: 294 employees

Yes Attrition (Employees who left)
- **Single**: 120 employees
- **Married**: 84 employees
- **Divorced**: 33 employees

Insights

1. **Attrition Rates by Marital Status**:
   - **Married Employees**:
     - 589 married employees stayed with the company.
     - 84 married employees left the company.
     - Attrition rate for married employees = (84 / (589 + 84)) * 100 ≈ 12.5%
   - **Single Employees**:
     - 350 single employees stayed with the company.
     - 120 single employees left the company.
     - Attrition rate for single employees = (120 / (350 + 120)) * 100 ≈ 25.5%
   - **Divorced Employees**:
     - 294 divorced employees stayed with the company.
     - 33 divorced employees left the company.
     - Attrition rate for divorced employees = (33 / (294 + 33)) * 100 ≈ 10.1%

2. **Comparison of Attrition Rates**:
   - The attrition rate for married employees is approximately 12.5%.
   - The attrition rate for single employees is significantly higher at approximately 25.5%.
   - The attrition rate for divorced employees is the lowest at approximately 10.1%.

3. **Overall Marital Status Distribution in Attrition**:
   - Of the total 237 employees who left the company:
     - 120 were single, making up approximately 50.6% of the attrition cases.
     - 84 were married, making up approximately 35.4% of the attrition cases.
     - 33 were divorced, making up approximately 13.9% of the attrition cases.

Conclusion
The analysis indicates a significant relationship between marital status and attrition. Single employees have the highest attrition rate at 25.5%, suggesting they are more likely to leave the company compared to their married (12.5%) and divorced (10.1%) counterparts. This trend may reflect different priorities or life circumstances associated with marital status that influence the decision to stay or leave a job. Understanding these patterns can help the company develop targeted retention strategies that address the specific needs and concerns of employees based on their marital status, potentially improving overall employee retention.

Marital Status vs Attrition

```
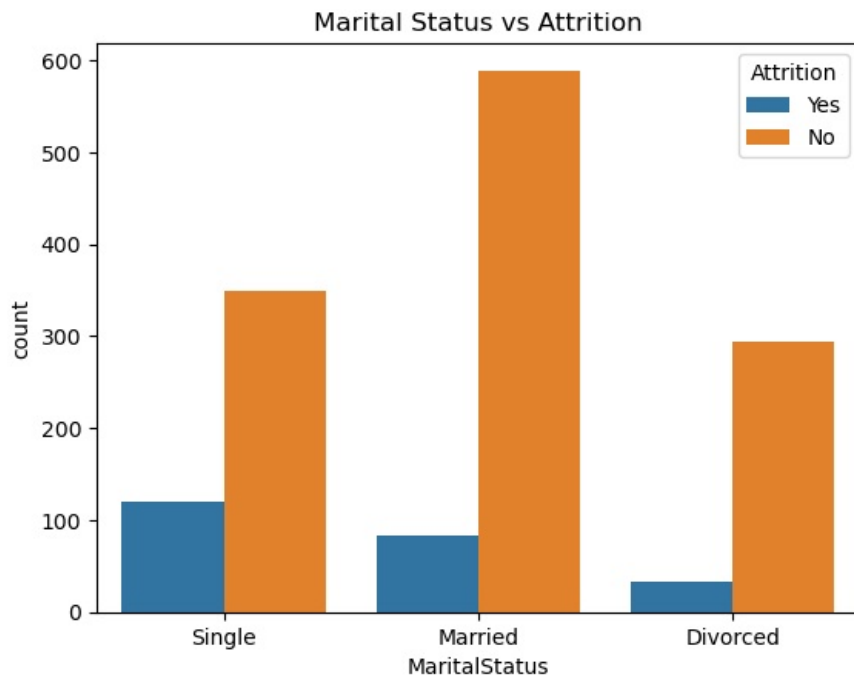In [81]:  #question12 How does overtime work relate to attrition?
          overtime_attrition = df.groupby('Attrition')['OverTime'].value_counts()
          print(overtime_attrition)

          # Data insights for overtime work and attrition

          # Insights
          insights = """
          Let's break down the insights from the provided data for the relationship between overtime work and attrition.

          ### Overtime Work and Attrition

          #### No Attrition (Employees who stayed)
          - **No Overtime**: 944 employees
          - **Yes Overtime**: 289 employees

          #### Yes Attrition (Employees who left)
          - **Yes Overtime**: 127 employees
          - **No Overtime**: 110 employees

          ### Insights

          1. **Attrition Rates by Overtime Work**:
             - **Employees with No Overtime**:
               - 944 employees with no overtime stayed with the company.
               - 110 employees with no overtime left the company.
               - Attrition rate for employees with no overtime = (110 / (944 + 110)) * 100 ≈ 10.4%
             - **Employees with Overtime**:
               - 289 employees with overtime stayed with the company.
               - 127 employees with overtime left the company.
               - Attrition rate for employees with overtime = (127 / (289 + 127)) * 100 ≈ 30.5%

          2. **Comparison of Attrition Rates**:
             - The attrition rate for employees with no overtime is approximately 10.4%.
             - The attrition rate for employees with overtime is significantly higher at approximately 30.5%.
             - This indicates that employees who work overtime are much more likely to leave the company compared to thos

          3. **Overall Overtime Work Distribution in Attrition**:
             - Of the total 237 employees who left the company:
               - 127 worked overtime, making up approximately 53.6% of the attrition cases.
               - 110 did not work overtime, making up approximately 46.4% of the attrition cases.

          ### Conclusion
          The analysis indicates a significant relationship between overtime work and attrition. Employees who work overti
```

```
"""

print(insights)


sns.countplot(data=df, x='OverTime', hue='Attrition')
plt.title('OverTime vs Attrition')
plt.show()
```

```
Attrition  OverTime
No         No          944
           Yes         289
Yes        Yes         127
           No          110
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between overtime work and attrition.

Overtime Work and Attrition

No Attrition (Employees who stayed)
- **No Overtime**: 944 employees
- **Yes Overtime**: 289 employees

Yes Attrition (Employees who left)
- **Yes Overtime**: 127 employees
- **No Overtime**: 110 employees

Insights

1. **Attrition Rates by Overtime Work**:
   - **Employees with No Overtime**:
     - 944 employees with no overtime stayed with the company.
     - 110 employees with no overtime left the company.
     - Attrition rate for employees with no overtime = (110 / (944 + 110)) * 100 ≈ 10.4%
   - **Employees with Overtime**:
     - 289 employees with overtime stayed with the company.
     - 127 employees with overtime left the company.
     - Attrition rate for employees with overtime = (127 / (289 + 127)) * 100 ≈ 30.5%

2. **Comparison of Attrition Rates**:
   - The attrition rate for employees with no overtime is approximately 10.4%.
   - The attrition rate for employees with overtime is significantly higher at approximately 30.5%.
   - This indicates that employees who work overtime are much more likely to leave the company compared to those who do not work overtime.

3. **Overall Overtime Work Distribution in Attrition**:
   - Of the total 237 employees who left the company:
     - 127 worked overtime, making up approximately 53.6% of the attrition cases.
     - 110 did not work overtime, making up approximately 46.4% of the attrition cases.

Conclusion
The analysis indicates a significant relationship between overtime work and attrition. Employees who work overtime have a much higher attrition rate (30.5%) compared to those who do not work overtime (10.4%). This suggests that overtime work is a strong predictor of employee attrition, possibly due to factors such as burnout, work-life imbalance, or dissatisfaction with working conditions. Understanding these patterns can help the company develop strategies to manage overtime work more effectively, improve employee satisfaction, and reduce overall attrition rates.

OverTime vs Attrition

```python
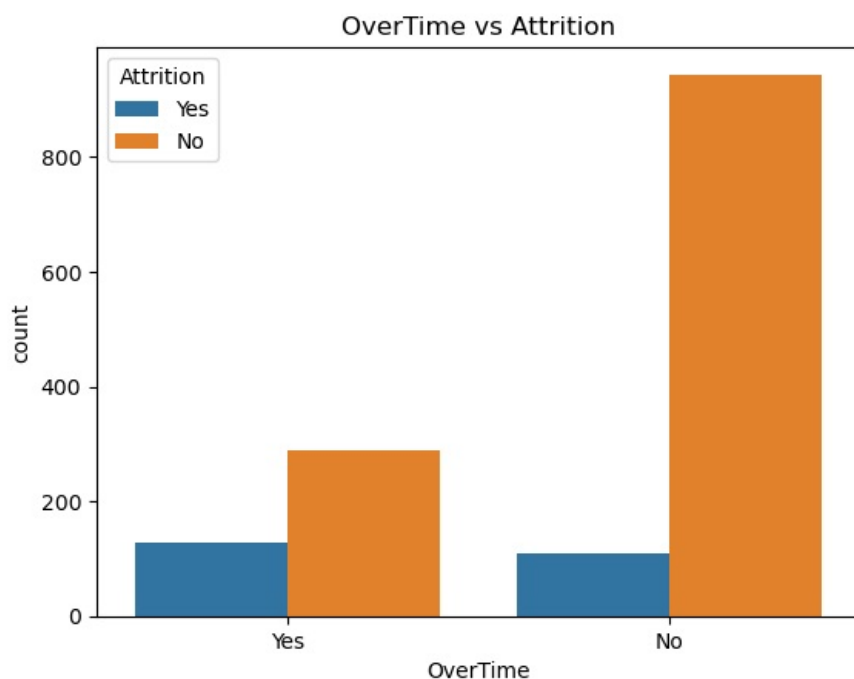#question13 Is there a difference in attrition rates across different job roles?
job_role_attrition = df.groupby('Attrition')['JobRole'].value_counts()
print(job_role_attrition)
# Data insights for job roles and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between job roles and attrition.

### Job Roles and Attrition

#### No Attrition (Employees who stayed)
- **Sales Executive**: 269 employees
- **Research Scientist**: 245 employees
- **Laboratory Technician**: 197 employees
- **Manufacturing Director**: 135 employees
- **Healthcare Representative**: 122 employees
- **Manager**: 97 employees
- **Research Director**: 78 employees
- **Sales Representative**: 50 employees
- **Human Resources**: 40 employees

#### Yes Attrition (Employees who left)
- **Laboratory Technician**: 62 employees
- **Sales Executive**: 57 employees
- **Research Scientist**: 47 employees
- **Sales Representative**: 33 employees
- **Human Resources**: 12 employees
- **Manufacturing Director**: 10 employees
- **Healthcare Representative**: 9 employees
```

```
    - **Manager**: 5 employees
    - **Research Director**: 2 employees

### Insights

1. **Attrition Rates by Job Role**:
    - **Sales Executive**:
      - Stayed: 269 employees
      - Left: 57 employees
      - Attrition rate = (57 / (269 + 57)) * 100 ≈ 17.5%
    - **Research Scientist**:
      - Stayed: 245 employees
      - Left: 47 employees
      - Attrition rate = (47 / (245 + 47)) * 100 ≈ 16.1%
    - **Laboratory Technician**:
      - Stayed: 197 employees
      - Left: 62 employees
      - Attrition rate = (62 / (197 + 62)) * 100 ≈ 23.9%
    - **Manufacturing Director**:
      - Stayed: 135 employees
      - Left: 10 employees
      - Attrition rate = (10 / (135 + 10)) * 100 ≈ 6.9%
    - **Healthcare Representative**:
      - Stayed: 122 employees
      - Left: 9 employees
      - Attrition rate = (9 / (122 + 9)) * 100 ≈ 6.9%
    - **Manager**:
      - Stayed: 97 employees
      - Left: 5 employees
      - Attrition rate = (5 / (97 + 5)) * 100 ≈ 4.9%
    - **Research Director**:
      - Stayed: 78 employees
      - Left: 2 employees
      - Attrition rate = (2 / (78 + 2)) * 100 ≈ 2.5%
    - **Sales Representative**:
      - Stayed: 50 employees
      - Left: 33 employees
      - Attrition rate = (33 / (50 + 33)) * 100 ≈ 39.8%
    - **Human Resources**:
      - Stayed: 40 employees
      - Left: 12 employees
      - Attrition rate = (12 / (40 + 12)) * 100 ≈ 23.1%

2. **Comparison of Attrition Rates**:
    - The highest attrition rate is among Sales Representatives at approximately 39.8%.
    - The lowest attrition rate is among Research Directors at approximately 2.5%.
    - Other job roles with relatively high attrition rates include Laboratory Technicians (23.9%) and Human Resou
    - Job roles with relatively low attrition rates include Managers (4.9%), Healthcare Representatives (6.9%), a

3. **Overall Job Role Distribution in Attrition**:
    - Sales Representatives and Laboratory Technicians are notable for their high attrition rates, suggesting th

### Conclusion
The analysis indicates significant differences in attrition rates across different job roles. Job roles such as
"""

print(insights)


plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='JobRole', hue='Attrition')
plt.title('Job Role vs Attrition')
plt.xticks(rotation=45)
plt.show()
```

```
Attrition  JobRole
No         Sales Executive            269
           Research Scientist         245
           Laboratory Technician      197
           Manufacturing Director     135
           Healthcare Representative  122
           Manager                     97
           Research Director           78
           Sales Representative        50
           Human Resources             40
Yes        Laboratory Technician       62
           Sales Executive             57
           Research Scientist          47
           Sales Representative        33
           Human Resources             12
           Manufacturing Director      10
           Healthcare Representative    9
           Manager                      5
           Research Director            2
```

Name: count, dtype: int64

Let's break down the insights from the provided data for the relationship between job roles and attrition.

Job Roles and Attrition

No Attrition (Employees who stayed)
- **Sales Executive**: 269 employees
- **Research Scientist**: 245 employees
- **Laboratory Technician**: 197 employees
- **Manufacturing Director**: 135 employees
- **Healthcare Representative**: 122 employees
- **Manager**: 97 employees
- **Research Director**: 78 employees
- **Sales Representative**: 50 employees
- **Human Resources**: 40 employees

Yes Attrition (Employees who left)
- **Laboratory Technician**: 62 employees
- **Sales Executive**: 57 employees
- **Research Scientist**: 47 employees
- **Sales Representative**: 33 employees
- **Human Resources**: 12 employees
- **Manufacturing Director**: 10 employees
- **Healthcare Representative**: 9 employees
- **Manager**: 5 employees
- **Research Director**: 2 employees

Insights

1. **Attrition Rates by Job Role**:
   - **Sales Executive**:
     - Stayed: 269 employees
     - Left: 57 employees
     - Attrition rate = (57 / (269 + 57)) * 100 ≈ 17.5%
   - **Research Scientist**:
     - Stayed: 245 employees
     - Left: 47 employees
     - Attrition rate = (47 / (245 + 47)) * 100 ≈ 16.1%
   - **Laboratory Technician**:
     - Stayed: 197 employees
     - Left: 62 employees
     - Attrition rate = (62 / (197 + 62)) * 100 ≈ 23.9%
   - **Manufacturing Director**:
     - Stayed: 135 employees
     - Left: 10 employees
     - Attrition rate = (10 / (135 + 10)) * 100 ≈ 6.9%
   - **Healthcare Representative**:
     - Stayed: 122 employees
     - Left: 9 employees
     - Attrition rate = (9 / (122 + 9)) * 100 ≈ 6.9%
   - **Manager**:
     - Stayed: 97 employees
     - Left: 5 employees
     - Attrition rate = (5 / (97 + 5)) * 100 ≈ 4.9%
   - **Research Director**:
     - Stayed: 78 employees
     - Left: 2 employees
     - Attrition rate = (2 / (78 + 2)) * 100 ≈ 2.5%
   - **Sales Representative**:
     - Stayed: 50 employees
     - Left: 33 employees
     - Attrition rate = (33 / (50 + 33)) * 100 ≈ 39.8%
   - **Human Resources**:
     - Stayed: 40 employees
     - Left: 12 employees
     - Attrition rate = (12 / (40 + 12)) * 100 ≈ 23.1%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among Sales Representatives at approximately 39.8%.
   - The lowest attrition rate is among Research Directors at approximately 2.5%.
   - Other job roles with relatively high attrition rates include Laboratory Technicians (23.9%) and Human Resources (23.1%).
   - Job roles with relatively low attrition rates include Managers (4.9%), Healthcare Representatives (6.9%), and Manufacturing Directors (6.9%).

3. **Overall Job Role Distribution in Attrition**:
   - Sales Representatives and Laboratory Technicians are notable for their high attrition rates, suggesting these roles might face more challenges or have less job satisfaction.

Conclusion
The analysis indicates significant differences in attrition rates across different job roles. Job roles such as Sales Representatives and Laboratory Technicians have notably higher attrition rates, suggesting that employees

in these positions are more likely to leave the company. In contrast, roles like Research Directors and Managers have much lower attrition rates, indicating higher job stability. Understanding these patterns can help the comp any identify specific job roles that require targeted retention strategies to reduce attrition rates and improve employee satisfaction.

In [83]:
```python
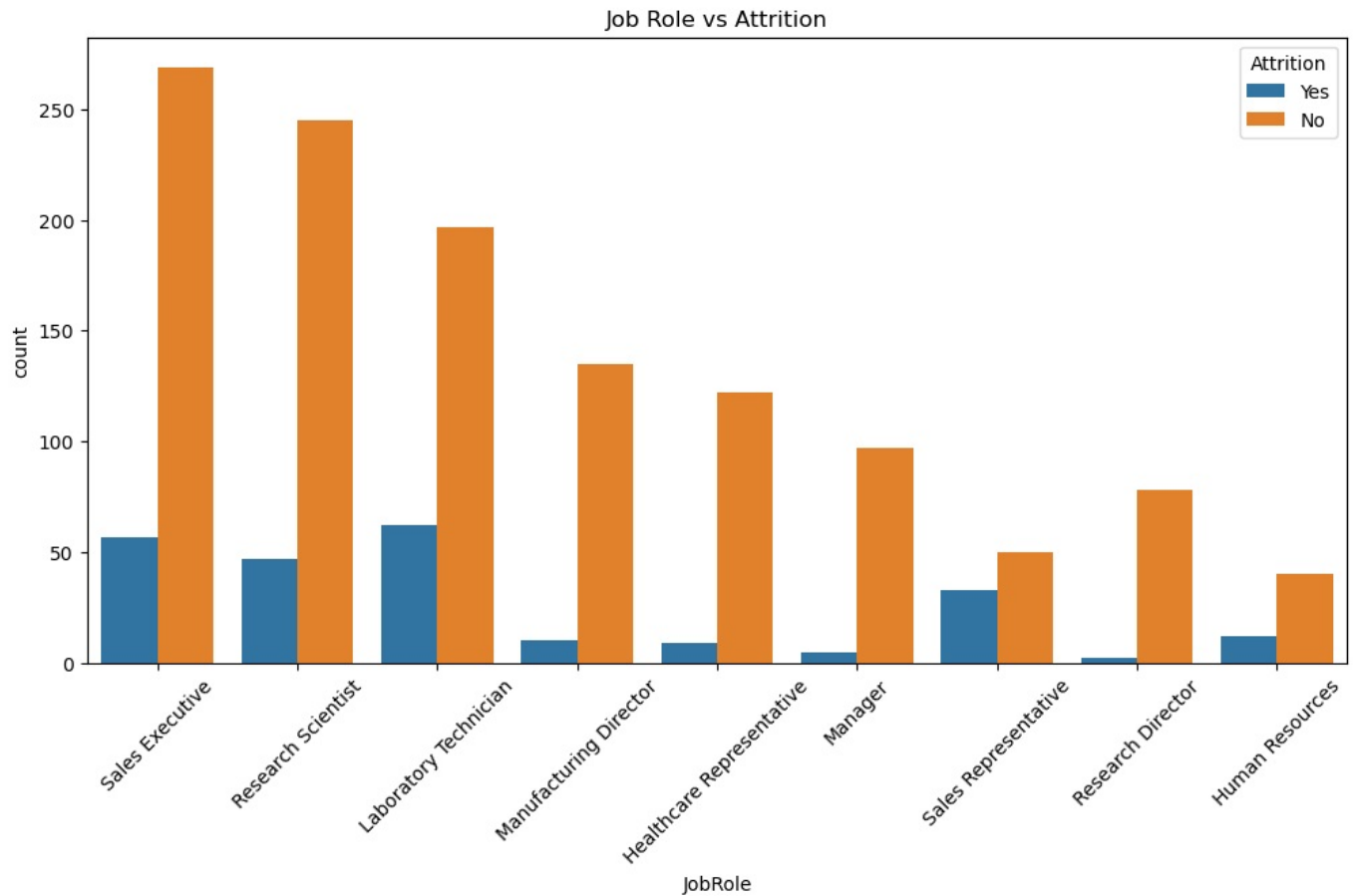#question14 How does the environment satisfaction score vary between employees who left and stayed?
environment_satisfaction_attrition = df.groupby('Attrition')['EnvironmentSatisfaction'].value_counts()
print(environment_satisfaction_attrition)
# Data insights for environment satisfaction score and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between environment satisfaction score

### Environment Satisfaction and Attrition

#### No Attrition (Employees who stayed)
- **Satisfaction Score 3**: 391 employees
- **Satisfaction Score 4**: 386 employees
- **Satisfaction Score 2**: 244 employees
- **Satisfaction Score 1**: 212 employees

#### Yes Attrition (Employees who left)
- **Satisfaction Score 1**: 72 employees
- **Satisfaction Score 3**: 62 employees
- **Satisfaction Score 4**: 60 employees
- **Satisfaction Score 2**: 43 employees

### Insights

1. **Attrition Rates by Environment Satisfaction Score**:
   - **Satisfaction Score 1**:
     - Stayed: 212 employees
     - Left: 72 employees
     - Attrition rate = (72 / (212 + 72)) * 100 ≈ 25.4%
   - **Satisfaction Score 2**:
     - Stayed: 244 employees
     - Left: 43 employees
     - Attrition rate = (43 / (244 + 43)) * 100 ≈ 15.0%
   - **Satisfaction Score 3**:
     - Stayed: 391 employees
     - Left: 62 employees
     - Attrition rate = (62 / (391 + 62)) * 100 ≈ 13.7%
   - **Satisfaction Score 4**:
     - Stayed: 386 employees
     - Left: 60 employees
     - Attrition rate = (60 / (386 + 60)) * 100 ≈ 13.4%
```

```
2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest satisfaction score (Score 1) at approximately
   - Employees with satisfaction scores of 2, 3, and 4 have lower attrition rates, with the lowest rates among
   - This suggests that higher environment satisfaction is associated with lower attrition rates.

3. **Overall Distribution in Attrition**:
   - Employees with the lowest satisfaction score (1) make up a significant portion of the attrition cases, ind:
   - Higher satisfaction scores (3 and 4) are associated with lower attrition rates, indicating that employees

### Conclusion
The analysis indicates a clear relationship between environment satisfaction scores and attrition rates. Employe
"""

print(insights)


sns.countplot(data=df, x='EnvironmentSatisfaction', hue='Attrition')
plt.title('Environment Satisfaction vs Attrition')
plt.show()
```

```
Attrition  EnvironmentSatisfaction
No         3                        391
           4                        386
           2                        244
           1                        212
Yes        1                         72
           3                         62
           4                         60
           2                         43
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between environment satisfaction score and attrition.

Environment Satisfaction and Attrition

No Attrition (Employees who stayed)
- **Satisfaction Score 3**: 391 employees
- **Satisfaction Score 4**: 386 employees
- **Satisfaction Score 2**: 244 employees
- **Satisfaction Score 1**: 212 employees

Yes Attrition (Employees who left)
- **Satisfaction Score 1**: 72 employees
- **Satisfaction Score 3**: 62 employees
- **Satisfaction Score 4**: 60 employees
- **Satisfaction Score 2**: 43 employees

Insights

1. **Attrition Rates by Environment Satisfaction Score**:
   - **Satisfaction Score 1**:
     - Stayed: 212 employees
     - Left: 72 employees
     - Attrition rate = (72 / (212 + 72)) * 100 ≈ 25.4%
   - **Satisfaction Score 2**:
     - Stayed: 244 employees
     - Left: 43 employees
     - Attrition rate = (43 / (244 + 43)) * 100 ≈ 15.0%
   - **Satisfaction Score 3**:
     - Stayed: 391 employees
     - Left: 62 employees
     - Attrition rate = (62 / (391 + 62)) * 100 ≈ 13.7%
   - **Satisfaction Score 4**:
     - Stayed: 386 employees
     - Left: 60 employees
     - Attrition rate = (60 / (386 + 60)) * 100 ≈ 13.4%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest satisfaction score (Score 1) at approximately 25.4%.
   - Employees with satisfaction scores of 2, 3, and 4 have lower attrition rates, with the lowest rates among those with scores of 3 (13.7%) and 4 (13.4%).
   - This suggests that higher environment satisfaction is associated with lower attrition rates.

3. **Overall Distribution in Attrition**:
   - Employees with the lowest satisfaction score (1) make up a significant portion of the attrition cases, indicating dissatisfaction with the work environment as a potential reason for leaving.
   - Higher satisfaction scores (3 and 4) are associated with lower attrition rates, indicating that employees with better perceived work environments are more likely to stay.

Conclusion
The analysis indicates a clear relationship between environment satisfaction scores and attrition rates. Employees with lower environment satisfaction scores are more likely to leave the company, with the highest attrition rate observed among those with a satisfaction score of 1 (25.4%). In contrast, employees with higher satisfaction scores (3 and 4) have much lower attrition rates (13.7% and 13.4%, respectively). These insights suggest that improving the work environment and increasing employee satisfaction can be effective strategies for reducing attrition rates and retaining talent within the company.

Environment Satisfaction vs Attrition

In [84]:
```python
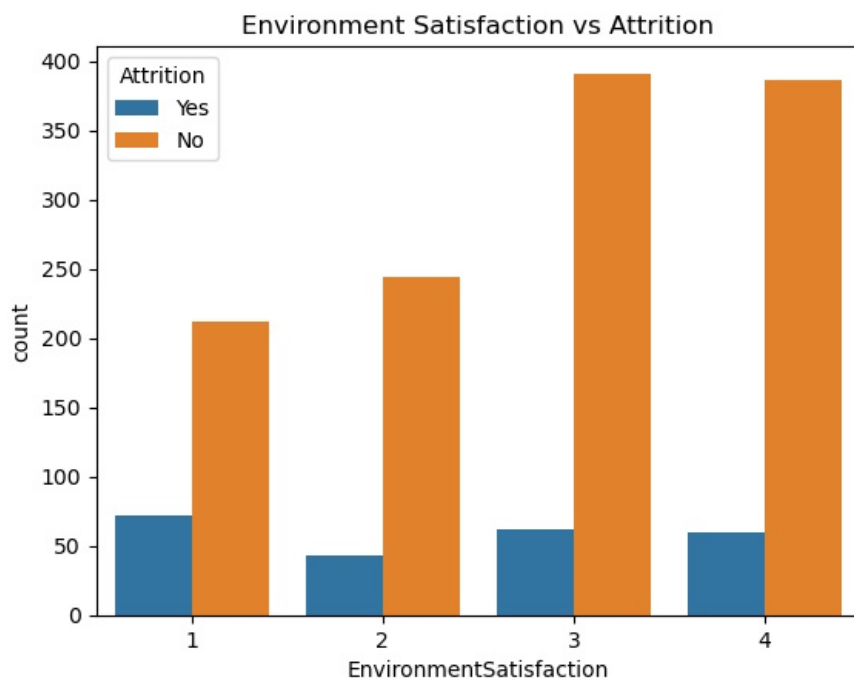#question15 What is the distribution of employees' performance ratings?
performance_rating_distribution = df['PerformanceRating'].value_counts()
print(performance_rating_distribution)

# Data insights for the distribution of employees' performance ratings

# Insights
insights = """
Let's break down the insights from the provided data for the distribution of employees' performance ratings.

### Distribution of Employees' Performance Ratings

#### Performance Rating Counts
- **Rating 3**: 1244 employees
- **Rating 4**: 226 employees

### Insights

1. **Performance Rating Distribution**:
   - The majority of employees (1244 out of 1470) have a performance rating of 3.
   - A smaller portion of employees (226 out of 1470) have a performance rating of 4.

2. **Percentage Distribution**:
   - **Rating 3**:
     - Number of employees: 1244
     - Percentage of total employees = (1244 / (1244 + 226)) * 100 ≈ 84.6%
   - **Rating 4**:
     - Number of employees: 226
     - Percentage of total employees = (226 / (1244 + 226)) * 100 ≈ 15.4%

3. **Interpretation**:
   - A performance rating of 3 is the most common rating among employees, indicating that the majority of the wo
   - A performance rating of 4, while less common, still represents a significant portion of the workforce, ind

### Conclusion
The distribution of performance ratings shows that the vast majority of employees have a performance rating of
"""

print(insights)


sns.countplot(data=df, x='PerformanceRating')
plt.title('Distribution of Performance Ratings')
plt.show()
```

```
PerformanceRating
3    1244
4     226
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the distribution of employees' performance ratings.

Distribution of Employees' Performance Ratings

Performance Rating Counts
- **Rating 3**: 1244 employees
- **Rating 4**: 226 employees

Insights

1. **Performance Rating Distribution**:
   - The majority of employees (1244 out of 1470) have a performance rating of 3.
   - A smaller portion of employees (226 out of 1470) have a performance rating of 4.

2. **Percentage Distribution**:
   - **Rating 3**:
     - Number of employees: 1244
     - Percentage of total employees = (1244 / (1244 + 226)) * 100 ≈ 84.6%
   - **Rating 4**:
     - Number of employees: 226
     - Percentage of total employees = (226 / (1244 + 226)) * 100 ≈ 15.4%

3. **Interpretation**:
   - A performance rating of 3 is the most common rating among employees, indicating that the majority of the workforce is meeting expectations.
   - A performance rating of 4, while less common, still represents a significant portion of the workforce, indicating that there is a noteworthy segment of employees who are exceeding expectations.

Conclusion
The distribution of performance ratings shows that the vast majority of employees have a performance rating of 3 (approximately 84.6%), suggesting that most employees are performing at an expected level. Meanwhile, around 15.4% of employees have a performance rating of 4, indicating a smaller group of high performers. These insights can help the company understand the overall performance levels of its workforce and identify areas where performance improvements or recognitions are needed.

Distribution of Performance Ratings

In [85]:
```python
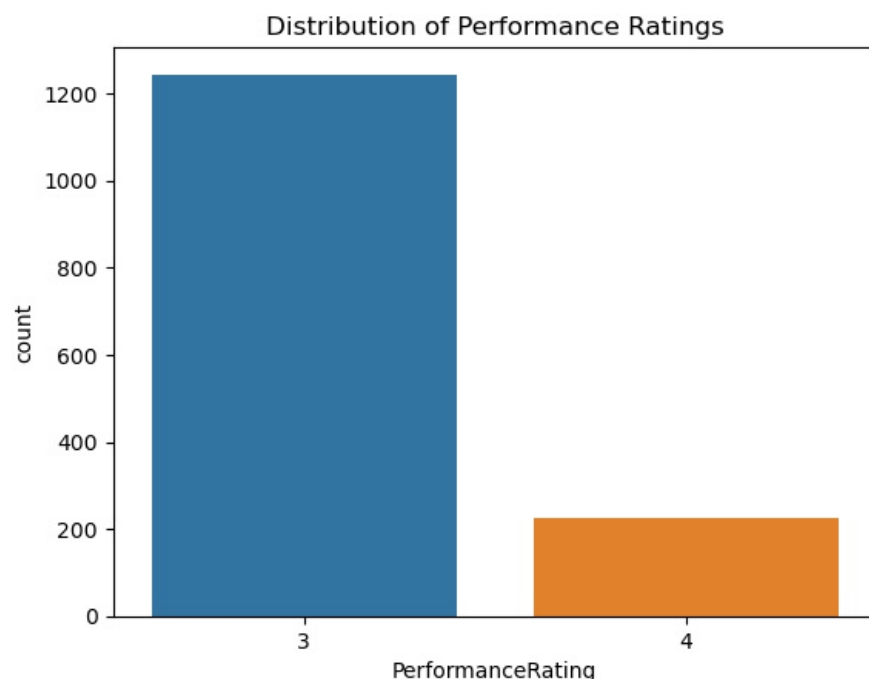#question16 Does the number of training times last year affect attrition?
training_times_attrition = df.groupby('Attrition')['TrainingTimesLastYear'].value_counts()
print(training_times_attrition)

# Data insights for training times last year and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between the number of training times

### Training Times Last Year and Attrition

#### No Attrition (Employees who stayed)
- **2 Trainings**: 449 employees
- **3 Trainings**: 422 employees
- **5 Trainings**: 105 employees
```

```
   - **4 Trainings**: 97 employees
   - **1 Training**: 62 employees
   - **6 Trainings**: 59 employees
   - **0 Trainings**: 39 employees

#### Yes Attrition (Employees who left)
   - **2 Trainings**: 98 employees
   - **3 Trainings**: 69 employees
   - **4 Trainings**: 26 employees
   - **0 Trainings**: 15 employees
   - **5 Trainings**: 14 employees
   - **1 Training**: 9 employees
   - **6 Trainings**: 6 employees

### Insights

1. **Attrition Rates by Training Times Last Year**:
   - **2 Trainings**:
     - Stayed: 449 employees
     - Left: 98 employees
     - Attrition rate = (98 / (449 + 98)) * 100 ≈ 17.9%
   - **3 Trainings**:
     - Stayed: 422 employees
     - Left: 69 employees
     - Attrition rate = (69 / (422 + 69)) * 100 ≈ 14.1%
   - **4 Trainings**:
     - Stayed: 97 employees
     - Left: 26 employees
     - Attrition rate = (26 / (97 + 26)) * 100 ≈ 21.1%
   - **0 Trainings**:
     - Stayed: 39 employees
     - Left: 15 employees
     - Attrition rate = (15 / (39 + 15)) * 100 ≈ 27.8%
   - **5 Trainings**:
     - Stayed: 105 employees
     - Left: 14 employees
     - Attrition rate = (14 / (105 + 14)) * 100 ≈ 11.8%
   - **1 Training**:
     - Stayed: 62 employees
     - Left: 9 employees
     - Attrition rate = (9 / (62 + 9)) * 100 ≈ 12.7%
   - **6 Trainings**:
     - Stayed: 59 employees
     - Left: 6 employees
     - Attrition rate = (6 / (59 + 6)) * 100 ≈ 9.2%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with 0 trainings (approximately 27.8%).
   - Employees with 4 trainings also have a relatively high attrition rate (21.1%).
   - The lowest attrition rate is among employees with 6 trainings (approximately 9.2%).
   - Employees with 5 trainings and 1 training have relatively low attrition rates (11.8% and 12.7%, respective

3. **Overall Distribution in Attrition**:
   - Employees who received 2 and 3 trainings last year make up a significant portion of both the stayed and le
   - Employees with no training have the highest attrition rate, suggesting a lack of training might contribute

### Conclusion
The analysis indicates a clear relationship between the number of training times last year and attrition rates.
"""

print(insights)


sns.countplot(data=df, x='TrainingTimesLastYear', hue='Attrition')
plt.title('Training Times Last Year vs Attrition')
plt.show()
```
```
Attrition  TrainingTimesLastYear
No         2                        449
           3                        422
           5                        105
           4                         97
           1                         62
           6                         59
           0                         39
Yes        2                         98
           3                         69
           4                         26
           0                         15
           5                         14
           1                          9
           6                          6
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between the number of training times last year and attrition.

Training Times Last Year and Attrition

No Attrition (Employees who stayed)
- **2 Trainings**: 449 employees
- **3 Trainings**: 422 employees
- **5 Trainings**: 105 employees
- **4 Trainings**: 97 employees
- **1 Training**: 62 employees
- **6 Trainings**: 59 employees
- **0 Trainings**: 39 employees

Yes Attrition (Employees who left)
- **2 Trainings**: 98 employees
- **3 Trainings**: 69 employees
- **4 Trainings**: 26 employees
- **0 Trainings**: 15 employees
- **5 Trainings**: 14 employees
- **1 Training**: 9 employees
- **6 Trainings**: 6 employees

Insights

1. **Attrition Rates by Training Times Last Year**:
   - **2 Trainings**:
     - Stayed: 449 employees
     - Left: 98 employees
     - Attrition rate = (98 / (449 + 98)) * 100 ≈ 17.9%
   - **3 Trainings**:
     - Stayed: 422 employees
     - Left: 69 employees
     - Attrition rate = (69 / (422 + 69)) * 100 ≈ 14.1%
   - **4 Trainings**:
     - Stayed: 97 employees
     - Left: 26 employees
     - Attrition rate = (26 / (97 + 26)) * 100 ≈ 21.1%
   - **0 Trainings**:
     - Stayed: 39 employees
     - Left: 15 employees
     - Attrition rate = (15 / (39 + 15)) * 100 ≈ 27.8%
   - **5 Trainings**:
     - Stayed: 105 employees
     - Left: 14 employees
     - Attrition rate = (14 / (105 + 14)) * 100 ≈ 11.8%
   - **1 Training**:
     - Stayed: 62 employees
     - Left: 9 employees
     - Attrition rate = (9 / (62 + 9)) * 100 ≈ 12.7%
   - **6 Trainings**:
     - Stayed: 59 employees
     - Left: 6 employees
     - Attrition rate = (6 / (59 + 6)) * 100 ≈ 9.2%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with 0 trainings (approximately 27.8%).
   - Employees with 4 trainings also have a relatively high attrition rate (21.1%).
   - The lowest attrition rate is among employees with 6 trainings (approximately 9.2%).
   - Employees with 5 trainings and 1 training have relatively low attrition rates (11.8% and 12.7%, respectively).

3. **Overall Distribution in Attrition**:
   - Employees who received 2 and 3 trainings last year make up a significant portion of both the stayed and left groups, but their attrition rates are moderate.
   - Employees with no training have the highest attrition rate, suggesting a lack of training might contribute to higher attrition.

Conclusion
The analysis indicates a clear relationship between the number of training times last year and attrition rates. Employees who did not receive any training have the highest attrition rate (27.8%), while those who received 6 trainings have the lowest attrition rate (9.2%). This suggests that providing more training opportunities may help reduce attrition rates. Companies should consider increasing training sessions and ensuring all employees have access to training to enhance their skills and job satisfaction, thereby reducing the likelihood of them leaving the company.

Training Times Last Year vs Attrition

In [86]:
```python
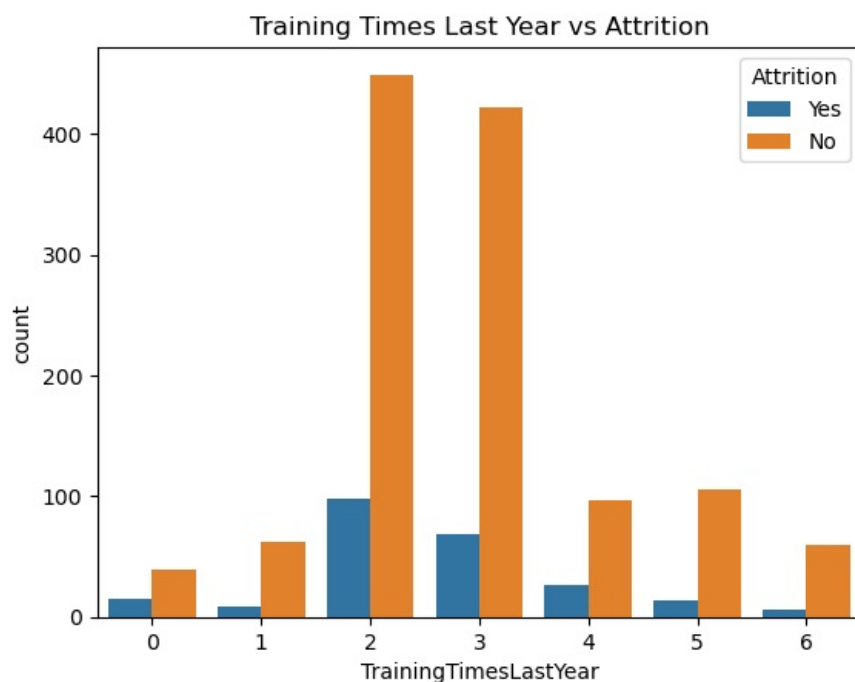#question17 How does work-life balance relate to attrition?
work_life_balance_attrition = df.groupby('Attrition')['WorkLifeBalance'].value_counts()
print(work_life_balance_attrition)
# Data insights for work-life balance and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the relationship between work-life balance and attriti

### Work-Life Balance and Attrition

#### No Attrition (Employees who stayed)
- **Work-Life Balance 3**: 766 employees
- **Work-Life Balance 2**: 286 employees
- **Work-Life Balance 4**: 126 employees
- **Work-Life Balance 1**: 55 employees

#### Yes Attrition (Employees who left)
- **Work-Life Balance 3**: 127 employees
- **Work-Life Balance 2**: 58 employees
- **Work-Life Balance 4**: 27 employees
- **Work-Life Balance 1**: 25 employees

### Insights

1. **Attrition Rates by Work-Life Balance**:
   - **Work-Life Balance 3**:
     - Stayed: 766 employees
     - Left: 127 employees
     - Attrition rate = (127 / (766 + 127)) * 100 ≈ 14.2%
   - **Work-Life Balance 2**:
     - Stayed: 286 employees
     - Left: 58 employees
     - Attrition rate = (58 / (286 + 58)) * 100 ≈ 16.8%
   - **Work-Life Balance 4**:
     - Stayed: 126 employees
     - Left: 27 employees
     - Attrition rate = (27 / (126 + 27)) * 100 ≈ 17.6%
   - **Work-Life Balance 1**:
     - Stayed: 55 employees
     - Left: 25 employees
     - Attrition rate = (25 / (55 + 25)) * 100 ≈ 31.2%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest work-life balance rating (Balance 1) at appro
   - Employees with work-life balance ratings of 2, 3, and 4 also have relatively high attrition rates, ranging

3. **Overall Distribution in Attrition**:
```

```
Attrition  WorkLifeBalance
No         3                766
           2                286
           4                126
           1                 55
Yes        3                127
           2                 58
           4                 27
           1                 25
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the relationship between work-life balance and attrition.

Work-Life Balance and Attrition

No Attrition (Employees who stayed)
- **Work-Life Balance 3**: 766 employees
- **Work-Life Balance 2**: 286 employees
- **Work-Life Balance 4**: 126 employees
- **Work-Life Balance 1**: 55 employees

Yes Attrition (Employees who left)
- **Work-Life Balance 3**: 127 employees
- **Work-Life Balance 2**: 58 employees
- **Work-Life Balance 4**: 27 employees
- **Work-Life Balance 1**: 25 employees

Insights

1. **Attrition Rates by Work-Life Balance**:
   - **Work-Life Balance 3**:
     - Stayed: 766 employees
     - Left: 127 employees
     - Attrition rate = (127 / (766 + 127)) * 100 ≈ 14.2%
   - **Work-Life Balance 2**:
     - Stayed: 286 employees
     - Left: 58 employees
     - Attrition rate = (58 / (286 + 58)) * 100 ≈ 16.8%
   - **Work-Life Balance 4**:
     - Stayed: 126 employees
     - Left: 27 employees
     - Attrition rate = (27 / (126 + 27)) * 100 ≈ 17.6%
   - **Work-Life Balance 1**:
     - Stayed: 55 employees
     - Left: 25 employees
     - Attrition rate = (25 / (55 + 25)) * 100 ≈ 31.2%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest work-life balance rating (Balance 1) at approximately 31.2%.
   - Employees with work-life balance ratings of 2, 3, and 4 also have relatively high attrition rates, ranging from 14.2% to 17.6%.

3. **Overall Distribution in Attrition**:
   - Employees with moderate work-life balance ratings (3 and 2) make up a significant portion of both the stayed and left groups, with attrition rates ranging from 14.2% to 16.8%.
   - Employees with the highest work-life balance rating (4) have a lower attrition rate compared to those with lower ratings.

Conclusion
The analysis indicates a relationship between work-life balance ratings and attrition rates. Employees with lower work-life balance ratings (1, 2) have higher attrition rates, with the highest attrition rate observed among those with the lowest rating (31.2%). Even employees with moderate work-life balance ratings (3, 4) experience relatively high attrition rates, suggesting that work-life balance is an important factor influencing employee retention. Companies should prioritize initiatives to improve work-life balance for employees, which may include flexible work arrangements, wellness programs, and supportive management practices, to reduce attrition and retain talent.

WorkLifeBalance vs Attrition

```python
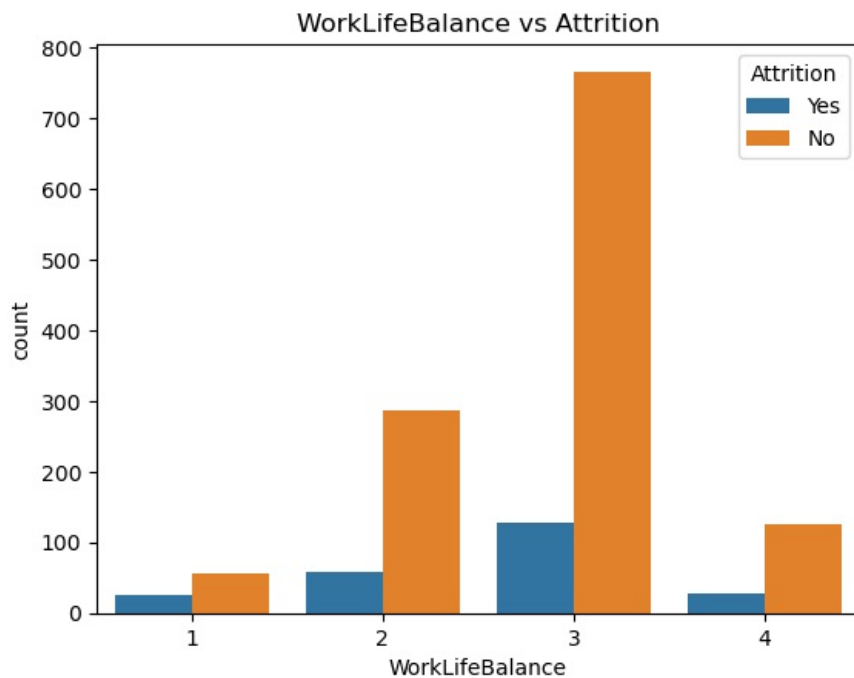In [87]: #question18 Is there a correlation between years at company and attrition?
         years_at_company_attrition = df.groupby('Attrition')['YearsAtCompany'].describe()
         print(years_at_company_attrition)
         # Data insights for years at company and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the correlation between years at the company and attrit

         ### Years at Company and Attrition

         #### No Attrition (Employees who stayed)
         - **Mean Years at Company**: 7.37 years
         - **Standard Deviation**: 6.10 years
         - **Minimum Years at Company**: 0 years
         - **25th Percentile (Q1)**: 3 years
         - **Median (50th Percentile)**: 6 years
         - **75th Percentile (Q3)**: 10 years
         - **Maximum Years at Company**: 37 years

         #### Yes Attrition (Employees who left)
         - **Mean Years at Company**: 5.13 years
         - **Standard Deviation**: 5.95 years
         - **Minimum Years at Company**: 0 years
         - **25th Percentile (Q1)**: 1 year
         - **Median (50th Percentile)**: 3 years
         - **75th Percentile (Q3)**: 7 years
         - **Maximum Years at Company**: 40 years

         ### Insights

         1. **Mean Years at Company**:
            - Employees who stayed have, on average, worked at the company for approximately 7.37 years, whereas employee
            - This suggests that employees who stayed tend to have longer tenures at the company compared to those who l

         2. **Spread of Years at Company**:
            - Both groups have a wide range of years at the company, from 0 to 37 years for employees who stayed, and fr
            - However, the spread of years at the company is slightly lower for employees who left, as indicated by a lo

         3. **Percentiles**:
            - Employees who left generally have shorter years at the company across all percentiles compared to those wh

         ### Conclusion
         The analysis suggests that there is a correlation between years at the company and attrition. Employees who sta
         """

         print(insights)


         sns.boxplot(data=df, x='Attrition', y='YearsAtCompany')
         plt.title('Years at Company vs Attrition')
         plt.show()
```

```
               count      mean       std  min  25%  50%   75%    max
Attrition
No            1233.0  7.369019  6.096298  0.0  3.0  6.0  10.0  37.0
Yes            237.0  5.130802  5.949984  0.0  1.0  3.0   7.0  40.0
```

Let's break down the insights from the provided data for the correlation between years at the company and attrition.

Years at Company and Attrition

No Attrition (Employees who stayed)
- **Mean Years at Company**: 7.37 years
- **Standard Deviation**: 6.10 years
- **Minimum Years at Company**: 0 years
- **25th Percentile (Q1)**: 3 years
- **Median (50th Percentile)**: 6 years
- **75th Percentile (Q3)**: 10 years
- **Maximum Years at Company**: 37 years

Yes Attrition (Employees who left)
- **Mean Years at Company**: 5.13 years
- **Standard Deviation**: 5.95 years
- **Minimum Years at Company**: 0 years
- **25th Percentile (Q1)**: 1 year
- **Median (50th Percentile)**: 3 years
- **75th Percentile (Q3)**: 7 years
- **Maximum Years at Company**: 40 years

Insights

1. **Mean Years at Company**:
   - Employees who stayed have, on average, worked at the company for approximately 7.37 years, whereas employees who left have worked for approximately 5.13 years on average.
   - This suggests that employees who stayed tend to have longer tenures at the company compared to those who left.

2. **Spread of Years at Company**:
   - Both groups have a wide range of years at the company, from 0 to 37 years for employees who stayed, and from 0 to 40 years for employees who left.
   - However, the spread of years at the company is slightly lower for employees who left, as indicated by a lower standard deviation (5.95 years) compared to those who stayed (6.10 years).

3. **Percentiles**:
   - Employees who left generally have shorter years at the company across all percentiles compared to those who stayed. For example, the median years at the company for employees who left is 3 years, whereas it's 6 years for employees who stayed.

Conclusion
The analysis suggests that there is a correlation between years at the company and attrition. Employees who stayed tend to have longer tenures at the company, with a higher average number of years compared to those who left. Additionally, employees who left generally have shorter years at the company across all percentiles, indicating that they are more likely to leave earlier in their careers. Understanding this correlation can help companies identify potential risk factors for attrition and develop strategies to improve employee retention, such as enhancing career development opportunities and fostering a positive work environment.

Years at Company vs Attrition

```
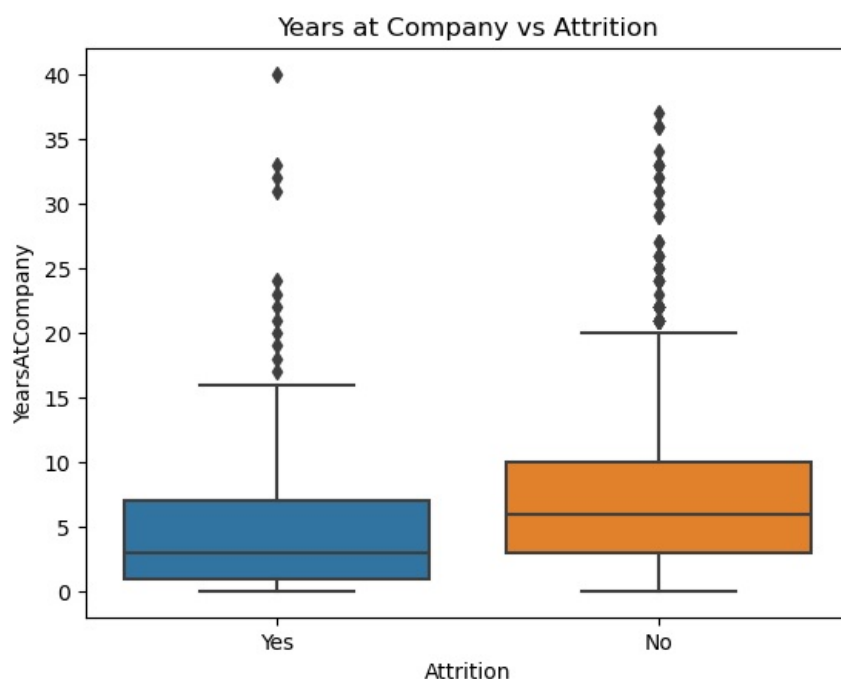In [88]:  #question19 How does job involvement score correlate with attrition?
```

```python
job_involvement_attrition = df.groupby('Attrition')['JobInvolvement'].value_counts()
print(job_involvement_attrition)

# Data insights for job involvement score and attrition

# Insights
insights = """
Let's break down the insights from the provided data for the correlation between job involvement score and attri

### Job Involvement Score and Attrition

#### No Attrition (Employees who stayed)
- **Job Involvement Score 3**: 743 employees
- **Job Involvement Score 2**: 304 employees
- **Job Involvement Score 4**: 131 employees
- **Job Involvement Score 1**: 55 employees

#### Yes Attrition (Employees who left)
- **Job Involvement Score 3**: 125 employees
- **Job Involvement Score 2**: 71 employees
- **Job Involvement Score 1**: 28 employees
- **Job Involvement Score 4**: 13 employees

### Insights

1. **Attrition Rates by Job Involvement Score**:
   - **Job Involvement Score 3**:
     - Stayed: 743 employees
     - Left: 125 employees
     - Attrition rate = (125 / (743 + 125)) * 100 ≈ 14.4%
   - **Job Involvement Score 2**:
     - Stayed: 304 employees
     - Left: 71 employees
     - Attrition rate = (71 / (304 + 71)) * 100 ≈ 18.9%
   - **Job Involvement Score 1**:
     - Stayed: 55 employees
     - Left: 28 employees
     - Attrition rate = (28 / (55 + 28)) * 100 ≈ 33.7%
   - **Job Involvement Score 4**:
     - Stayed: 131 employees
     - Left: 13 employees
     - Attrition rate = (13 / (131 + 13)) * 100 ≈ 9.0%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest job involvement score (Score 1) at approxima
   - Employees with job involvement scores of 2 and 3 also have relatively high attrition rates, ranging from 1
   - Employees with the highest job involvement score (4) have the lowest attrition rate at approximately 9.0%.

### Conclusion
The analysis suggests a correlation between job involvement score and attrition rates. Employees with lower job
"""

print(insights)


sns.countplot(data=df, x='JobInvolvement', hue='Attrition')
plt.title('Job Involvement vs Attrition')
plt.show()
```

```
Attrition  JobInvolvement
No         3                 743
           2                 304
           4                 131
           1                  55
Yes        3                 125
           2                  71
           1                  28
           4                  13
Name: count, dtype: int64
```

Let's break down the insights from the provided data for the correlation between job involvement score and attrition.

Job Involvement Score and Attrition

No Attrition (Employees who stayed)
- **Job Involvement Score 3**: 743 employees
- **Job Involvement Score 2**: 304 employees
- **Job Involvement Score 4**: 131 employees
- **Job Involvement Score 1**: 55 employees

Yes Attrition (Employees who left)
- **Job Involvement Score 3**: 125 employees
- **Job Involvement Score 2**: 71 employees
- **Job Involvement Score 1**: 28 employees
- **Job Involvement Score 4**: 13 employees

Insights

1. **Attrition Rates by Job Involvement Score**:
   - **Job Involvement Score 3**:
     - Stayed: 743 employees
     - Left: 125 employees
     - Attrition rate = (125 / (743 + 125)) * 100 ≈ 14.4%
   - **Job Involvement Score 2**:
     - Stayed: 304 employees
     - Left: 71 employees
     - Attrition rate = (71 / (304 + 71)) * 100 ≈ 18.9%
   - **Job Involvement Score 1**:
     - Stayed: 55 employees
     - Left: 28 employees
     - Attrition rate = (28 / (55 + 28)) * 100 ≈ 33.7%
   - **Job Involvement Score 4**:
     - Stayed: 131 employees
     - Left: 13 employees
     - Attrition rate = (13 / (131 + 13)) * 100 ≈ 9.0%

2. **Comparison of Attrition Rates**:
   - The highest attrition rate is among employees with the lowest job involvement score (Score 1) at approximately 33.7%.
   - Employees with job involvement scores of 2 and 3 also have relatively high attrition rates, ranging from 14.4% to 18.9%.
   - Employees with the highest job involvement score (4) have the lowest attrition rate at approximately 9.0%.

Conclusion
The analysis suggests a correlation between job involvement score and attrition rates. Employees with lower job involvement scores tend to have higher attrition rates, with the highest attrition rate observed among those with the lowest score (33.7%). Conversely, employees with higher job involvement scores have lower attrition rates. This indicates that employees who are more involved in their jobs are less likely to leave the company. Companies should focus on strategies to increase employee job involvement, such as providing opportunities for skill development, fostering a positive work environment, and promoting employee engagement initiatives, to reduce attrition rates and retain valuable talent.

Job Involvement vs Attrition


```
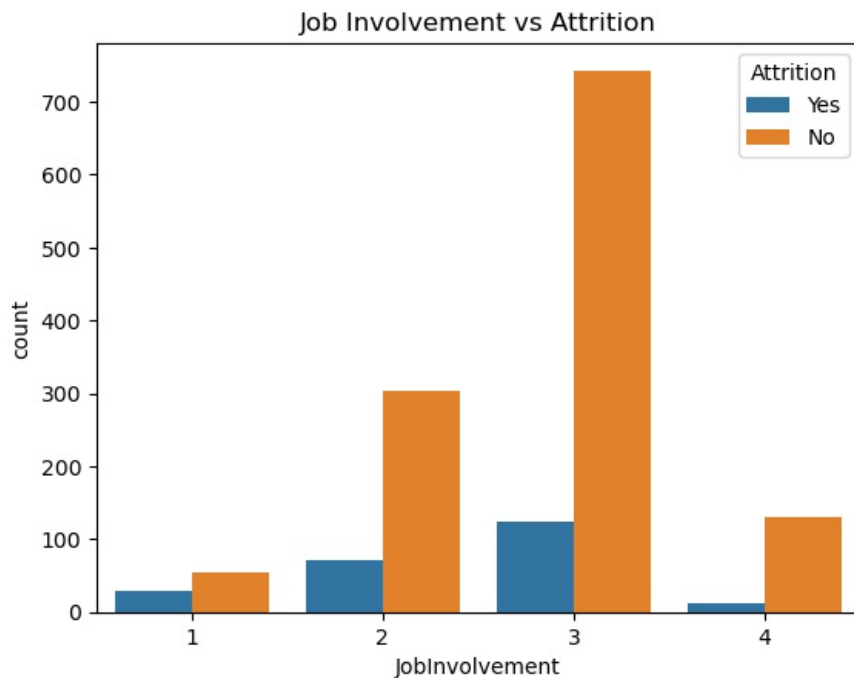In [89]: #question20 What is the relationship between years with current manager and attrition?
         years_with_manager_attrition = df.groupby('Attrition')['YearsWithCurrManager'].describe()
         print(years_with_manager_attrition)

         # Data insights for years with current manager and attrition

         # Insights
         insights = """
         Let's break down the insights from the provided data for the relationship between years with the current manage

         ### Years with Current Manager and Attrition

         #### No Attrition (Employees who stayed)
         - **Mean Years with Current Manager**: 4.37 years
         - **Standard Deviation**: 3.59 years
         - **Minimum Years with Current Manager**: 0 years
         - **25th Percentile (Q1)**: 2 years
         - **Median (50th Percentile)**: 3 years
         - **75th Percentile (Q3)**: 7 years
         - **Maximum Years with Current Manager**: 17 years

         #### Yes Attrition (Employees who left)
         - **Mean Years with Current Manager**: 2.85 years
         - **Standard Deviation**: 3.14 years
         - **Minimum Years with Current Manager**: 0 years
         - **25th Percentile (Q1)**: 0 years
         - **Median (50th Percentile)**: 2 years
         - **75th Percentile (Q3)**: 5 years
         - **Maximum Years with Current Manager**: 14 years

         ### Insights

         1. **Mean Years with Current Manager**:
            - Employees who stayed have, on average, worked with their current manager for approximately 4.37 years, whe
            - This suggests that employees who stayed tend to have longer tenures with their current manager compared to

         2. **Spread of Years with Current Manager**:
            - Both groups have a wide range of years with the current manager, from 0 to 17 years for employees who staye
            - However, the spread of years with the current manager is slightly lower for employees who left, as indicate

         3. **Percentiles**:
            - Employees who left generally have shorter years with the current manager across all percentiles compared to

         ### Conclusion
         The analysis suggests that there is a relationship between years with the current manager and attrition. Employe
         """

         print(insights)


         sns.boxplot(data=df, x='Attrition', y='YearsWithCurrManager')
         plt.title('Years with Current Manager vs Attrition')
         plt.show()
```

```
            count      mean       std  min  25%  50%  75%   max
Attrition
No         1233.0  4.367397  3.594116  0.0  2.0  3.0  7.0  17.0
Yes         237.0  2.852321  3.143349  0.0  0.0  2.0  5.0  14.0
```

Let's break down the insights from the provided data for the relationship between years with the current manager and attrition.

Years with Current Manager and Attrition

No Attrition (Employees who stayed)
- **Mean Years with Current Manager**: 4.37 years
- **Standard Deviation**: 3.59 years
- **Minimum Years with Current Manager**: 0 years
- **25th Percentile (Q1)**: 2 years
- **Median (50th Percentile)**: 3 years
- **75th Percentile (Q3)**: 7 years
- **Maximum Years with Current Manager**: 17 years

Yes Attrition (Employees who left)
- **Mean Years with Current Manager**: 2.85 years
- **Standard Deviation**: 3.14 years
- **Minimum Years with Current Manager**: 0 years
- **25th Percentile (Q1)**: 0 years
- **Median (50th Percentile)**: 2 years
- **75th Percentile (Q3)**: 5 years
- **Maximum Years with Current Manager**: 14 years

Insights

1. **Mean Years with Current Manager**:
   - Employees who stayed have, on average, worked with their current manager for approximately 4.37 years, whereas employees who left have worked for approximately 2.85 years on average.
   - This suggests that employees who stayed tend to have longer tenures with their current manager compared to those who left.

2. **Spread of Years with Current Manager**:
   - Both groups have a wide range of years with the current manager, from 0 to 17 years for employees who stayed, and from 0 to 14 years for employees who left.
   - However, the spread of years with the current manager is slightly lower for employees who left, as indicated by a lower standard deviation (3.14 years) compared to those who stayed (3.59 years).

3. **Percentiles**:
   - Employees who left generally have shorter years with the current manager across all percentiles compared to those who stayed. For example, the median years with the current manager for employees who left is 2 years, whereas it's 3 years for employees who stayed.

Conclusion
The analysis suggests that there is a relationship between years with the current manager and attrition. Employees who stayed tend to have longer tenures with their current manager, with a higher average number of years compared to those who left. Additionally, employees who left generally have shorter years with the current manager across all percentiles, indicating that they are more likely to leave earlier in their relationships with their managers. Understanding this relationship can help companies identify potential risk factors for attrition and develop strategies to improve employee-manager relationships, such as providing training for managers on effective leadership and communication skills, fostering a supportive work environment, and offering regular feedback and recognition to employees.