Project 1 HR Analytics (MeriSkill)

October 20, 2023

1 HR Analytics Project

1.1 Project Description

This Jupyter Notebook contains the analysis and findings of the HR Analytics project. The project's main objective is to analyze employee attrition within the organization and identify factors that contribute to attrition. We explore various aspects such as employee satisfaction, career progression, work-life balance, and more to gain insights into attrition patterns.

The analysis includes data cleaning, data visualization, and recommendations for HR strategies based on the findings.

LinkedIn: Chandramani Kumar Chandravnsi

GitHub: ckrazaditya

1.2 Import Libraries

```
[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

Load the Data

```
[2]: df = pd.read_csv("C:/Users/Lenovo/Desktop/Meri skill Project/
     ↪drive-download-20231003T163556Z-001/Project 3 - HR Analytics/Data P3␣
     ↪MeriSKILL/HR-Employee-Attrition.csv")
```

1.3 Data Exploration

```
[3]: df.head()
```

```
[3]:    Age Attrition     BusinessTravel  DailyRate              Department  \
     0   41       Yes      Travel_Rarely       1102                   Sales
     1   49        No  Travel_Frequently        279  Research & Development
     2   37       Yes      Travel_Rarely       1373  Research & Development
     3   33        No  Travel_Frequently       1392  Research & Development
     4   27        No      Travel_Rarely        591  Research & Development
```

```
      DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
    0                1          2  Life Sciences              1               1
    1                8          1  Life Sciences              1               2
    2                2          2          Other              1               4
    3                3          4  Life Sciences              1               5
    4                2          1        Medical              1               7

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

       TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
    0                  8                      0               1               6
    1                 10                      3               3              10
    2                  7                      3               3               0
    3                  8                      3               3               8
    4                  6                      3               3               2

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

    [5 rows x 35 columns]
```

[4]: `df.tail()`

[4]:
```
          Age Attrition       BusinessTravel  DailyRate              Department  \
    1465   36        No  Travel_Frequently        884  Research & Development
    1466   39        No      Travel_Rarely        613  Research & Development
    1467   27        No      Travel_Rarely        155  Research & Development
    1468   49        No  Travel_Frequently       1023                   Sales
    1469   34        No      Travel_Rarely        628  Research & Development

          DistanceFromHome  Education EducationField  EmployeeCount  \
    1465                23          2        Medical              1
    1466                 6          1        Medical              1
    1467                 4          3  Life Sciences              1
    1468                 2          3        Medical              1
    1469                 8          3        Medical              1

          EmployeeNumber  …  RelationshipSatisfaction StandardHours  \
```

```
1465           2061  …                        3          80
1466           2062  …                        1          80
1467           2064  …                        2          80
1468           2065  …                        4          80
1469           2068  …                        1          80


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


      WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
1465                3               5                   2
1466                3               7                   7
1467                3               6                   2
1468                2               9                   6
1469                4               4                   3


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

[5 rows x 35 columns]
```

[5]: `df.describe()`

```
[5]:                Age      DailyRate  DistanceFromHome    Education  EmployeeCount  \
     count  1470.000000  1470.000000       1470.000000  1470.000000         1470.0
     mean     36.923810   802.485714          9.192517     2.912925            1.0
     std       9.135373   403.509100          8.106864     1.024165            0.0
     min      18.000000   102.000000          1.000000     1.000000            1.0
     25%      30.000000   465.000000          2.000000     2.000000            1.0
     50%      36.000000   802.000000          7.000000     3.000000            1.0
     75%      43.000000  1157.000000         14.000000     4.000000            1.0
     max      60.000000  1499.000000         29.000000     5.000000            1.0


            EmployeeNumber  EnvironmentSatisfaction    HourlyRate  JobInvolvement  \
     count     1470.000000              1470.000000   1470.000000     1470.000000
     mean      1024.865306                 2.721769     65.891156        2.729932
     std        602.024335                 1.093082     20.329428        0.711561
     min          1.000000                 1.000000     30.000000        1.000000
     25%        491.250000                 2.000000     48.000000        2.000000
```

```
50%      1020.500000              3.000000   66.000000              3.000000
75%      1555.750000              4.000000   83.750000              3.000000
max      2068.000000              4.000000  100.000000              4.000000

           JobLevel  …  RelationshipSatisfaction  StandardHours  \
count  1470.000000  …               1470.000000         1470.0
mean      2.063946  …                  2.712245           80.0
std       1.106940  …                  1.081209            0.0
min       1.000000  …                  1.000000           80.0
25%       1.000000  …                  2.000000           80.0
50%       2.000000  …                  3.000000           80.0
75%       3.000000  …                  4.000000           80.0
max       5.000000  …                  4.000000           80.0

       StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
count       1470.000000        1470.000000            1470.000000
mean           0.793878          11.279592               2.799320
std            0.852077           7.780782               1.289271
min            0.000000           0.000000               0.000000
25%            0.000000           6.000000               2.000000
50%            1.000000          10.000000               3.000000
75%            1.000000          15.000000               3.000000
max            3.000000          40.000000               6.000000

       WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
count      1470.000000     1470.000000         1470.000000
mean          2.761224        7.008163            4.229252
std           0.706476        6.126525            3.623137
min           1.000000        0.000000            0.000000
25%           2.000000        3.000000            2.000000
50%           3.000000        5.000000            3.000000
75%           3.000000        9.000000            7.000000
max           4.000000       40.000000           18.000000

       YearsSinceLastPromotion  YearsWithCurrManager
count              1470.000000           1470.000000
mean                  2.187755              4.123129
std                   3.222430              3.568136
min                   0.000000              0.000000
25%                   0.000000              2.000000
50%                   1.000000              3.000000
75%                   3.000000              7.000000
max                  15.000000             17.000000

[8 rows x 26 columns]
```

```python
[6]: 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
```

1.4 Data Cleaning and Preprocessing

```
[7]: df.dropna(inplace=True)
```

```
[8]: df.isnull().sum()
```

```
[8]: 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
```

```
[9]: df = df.drop(["EmployeeCount", "Over18", "StandardHours"], axis=1)
```

```
[10]: df = df.rename(columns={"Attrition": "Attrition", "JobSatisfaction":␣
      ↪"Job_Satisfaction",})
```

```
[11]: df = df.dropna()
```

```
[12]: df.head()
```

```
[12]:    Age Attrition     BusinessTravel  DailyRate              Department  \
      0   41       Yes      Travel_Rarely       1102                   Sales
      1   49        No  Travel_Frequently        279  Research & Development
      2   37       Yes      Travel_Rarely       1373  Research & Development
      3   33        No  Travel_Frequently       1392  Research & Development
      4   27        No      Travel_Rarely        591  Research & Development

         DistanceFromHome  Education EducationField  EmployeeNumber  \
      0                 1          2  Life Sciences               1
      1                 8          1  Life Sciences               2
      2                 2          2          Other               4
      3                 3          4  Life Sciences               5
      4                 2          1        Medical               7

         EnvironmentSatisfaction  … PerformanceRating  RelationshipSatisfaction  \
      0                        2  …                 3                         1
      1                        3  …                 4                         4
      2                        4  …                 3                         2
      3                        4  …                 3                         3
      4                        1  …                 3                         4

         StockOptionLevel  TotalWorkingYears TrainingTimesLastYear  WorkLifeBalance  \
      0                 0                  8                     0                1
      1                 1                 10                     3                3
      2                 0                  7                     3                3
      3                 0                  8                     3                3
      4                 1                  6                     3                3

         YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
      0               6                   4                        0
      1              10                   7                        1
      2               0                   0                        0
      3               8                   7                        3
      4               2                   2                        2

         YearsWithCurrManager
      0                     5
      1                     7
      2                     0
      3                     0
      4                     2

      [5 rows x 32 columns]
```

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

```
[14]: categorical_vars = ["BusinessTravel", "Department", "Gender", "JobRole",
      ↪"MaritalStatus"]
      for var in categorical_vars:
          plt.figure(figsize=(8, 5))
          sns.countplot(data=df, x=var, hue="Attrition")
          plt.title(f"Distribution of {var} vs. Attrition")
          plt.xticks(rotation=45)
          plt.show()
```
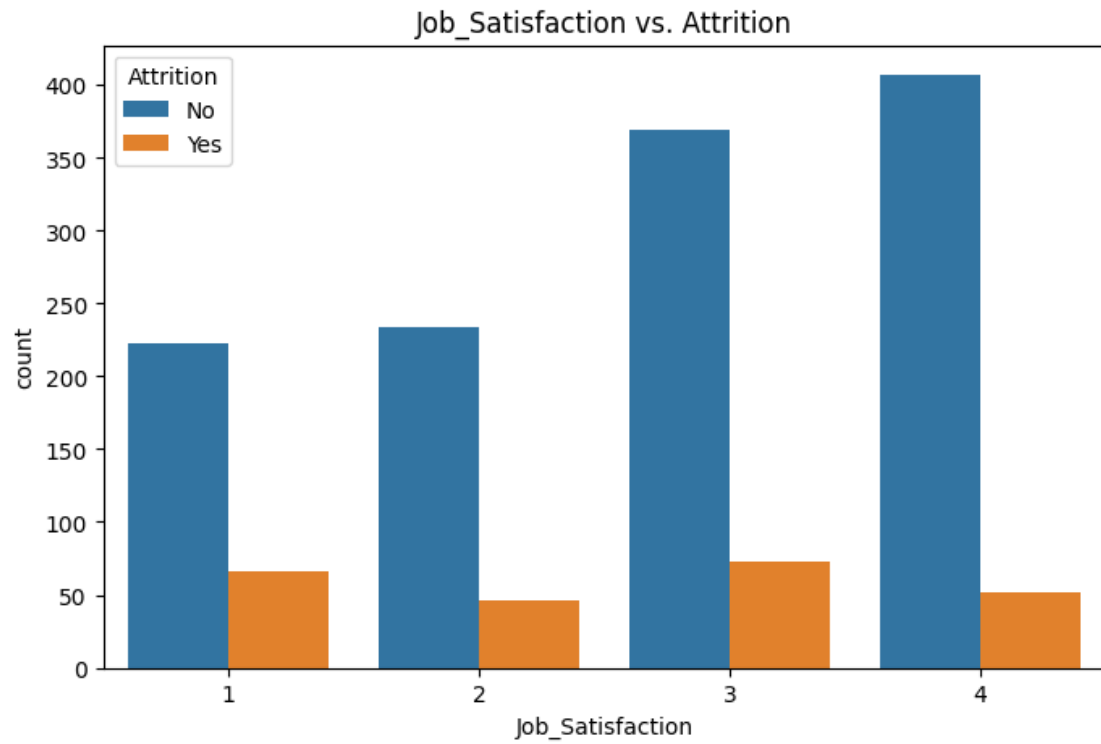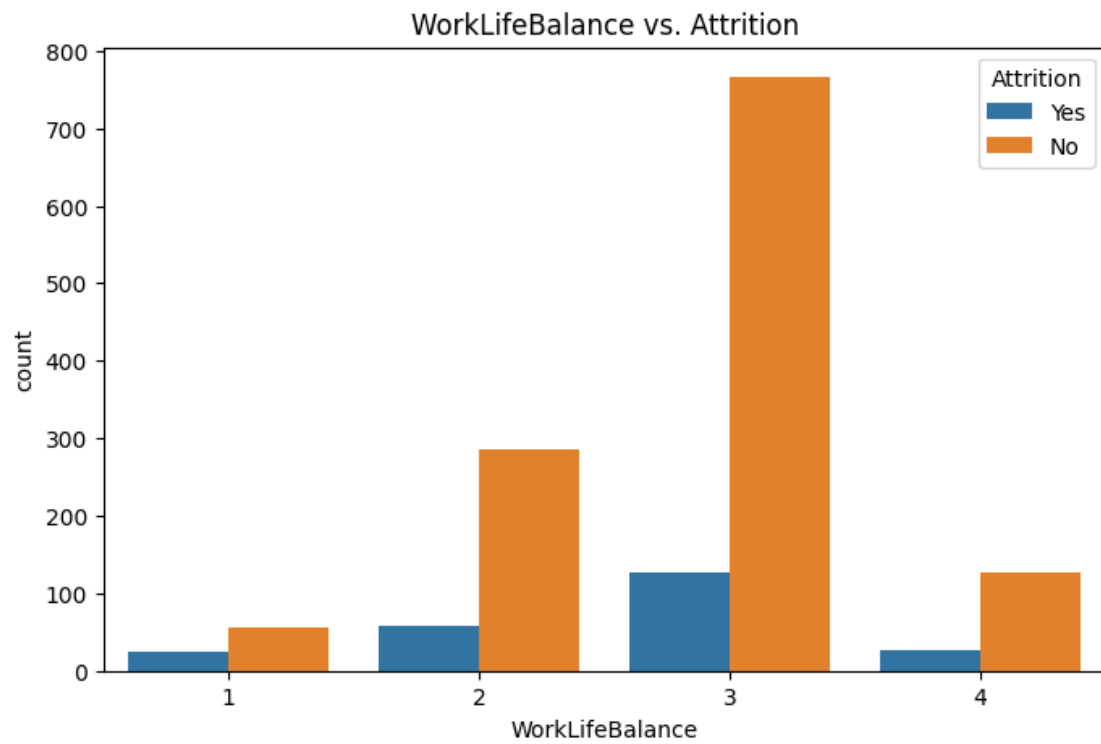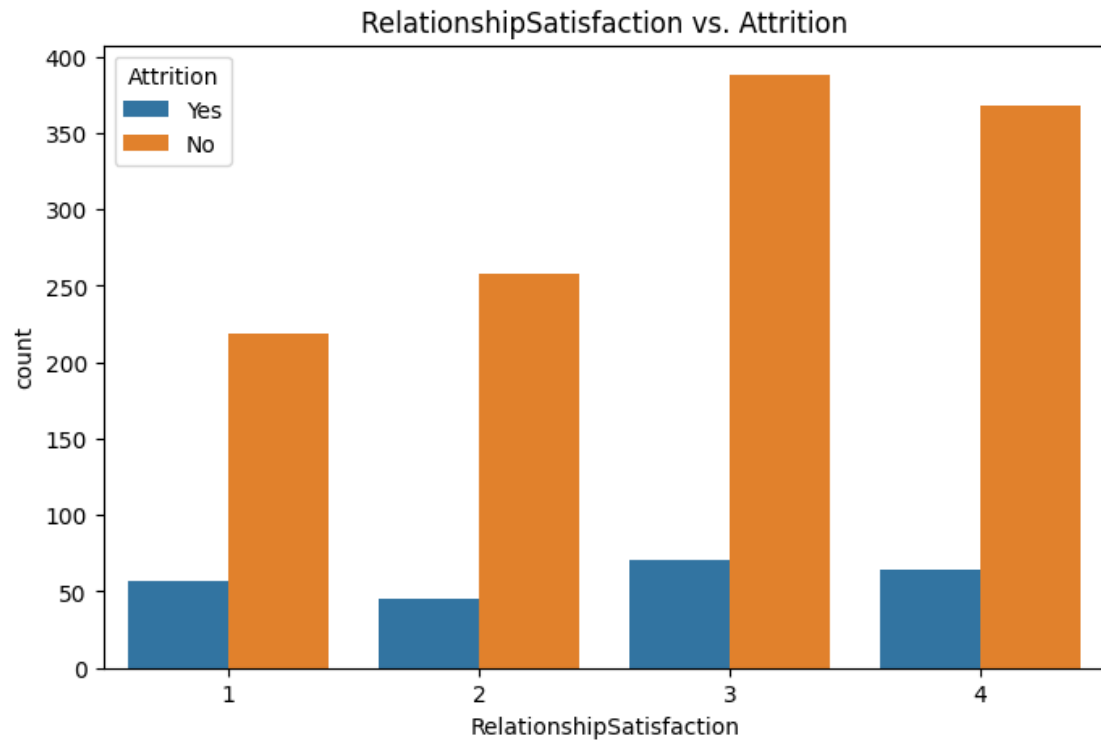

Distribution of BusinessTravel vs. Attrition

Distribution of Department vs. Attrition

Distribution of Gender vs. Attrition

Distribution of JobRole vs. Attrition

Distribution of MaritalStatus vs. Attrition
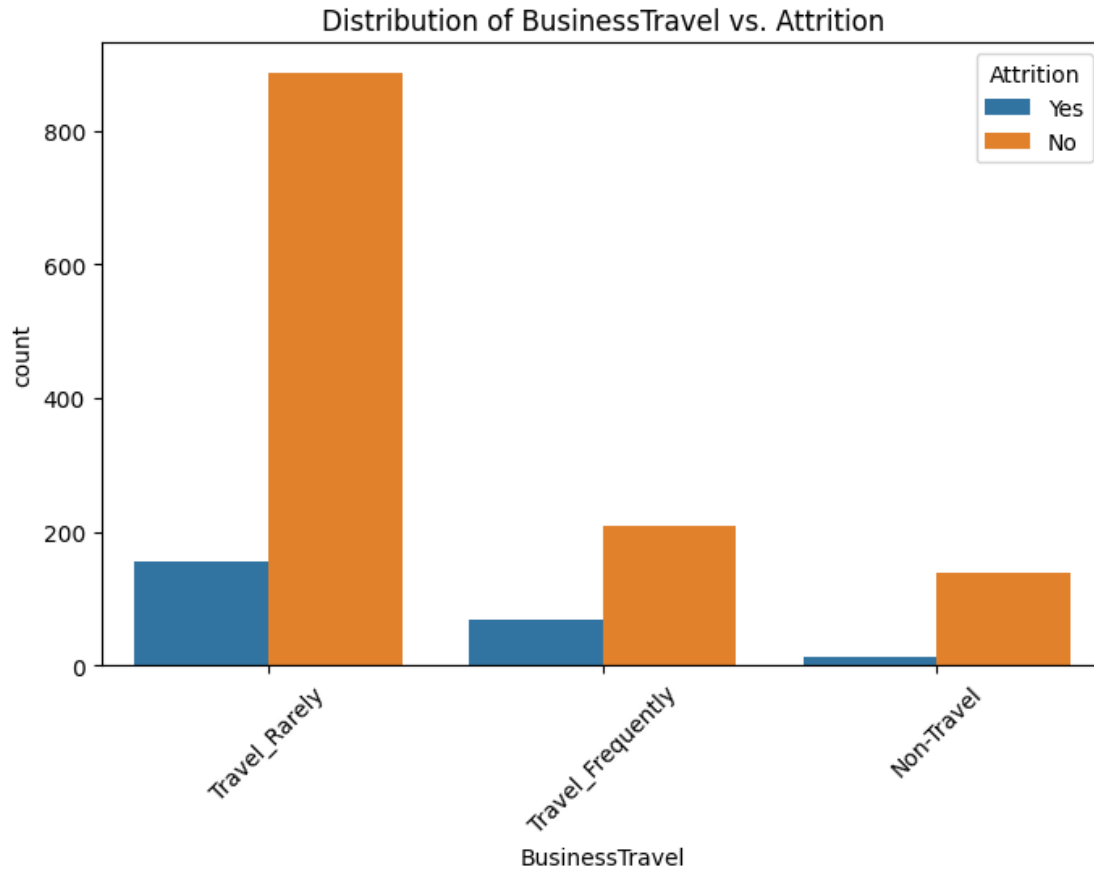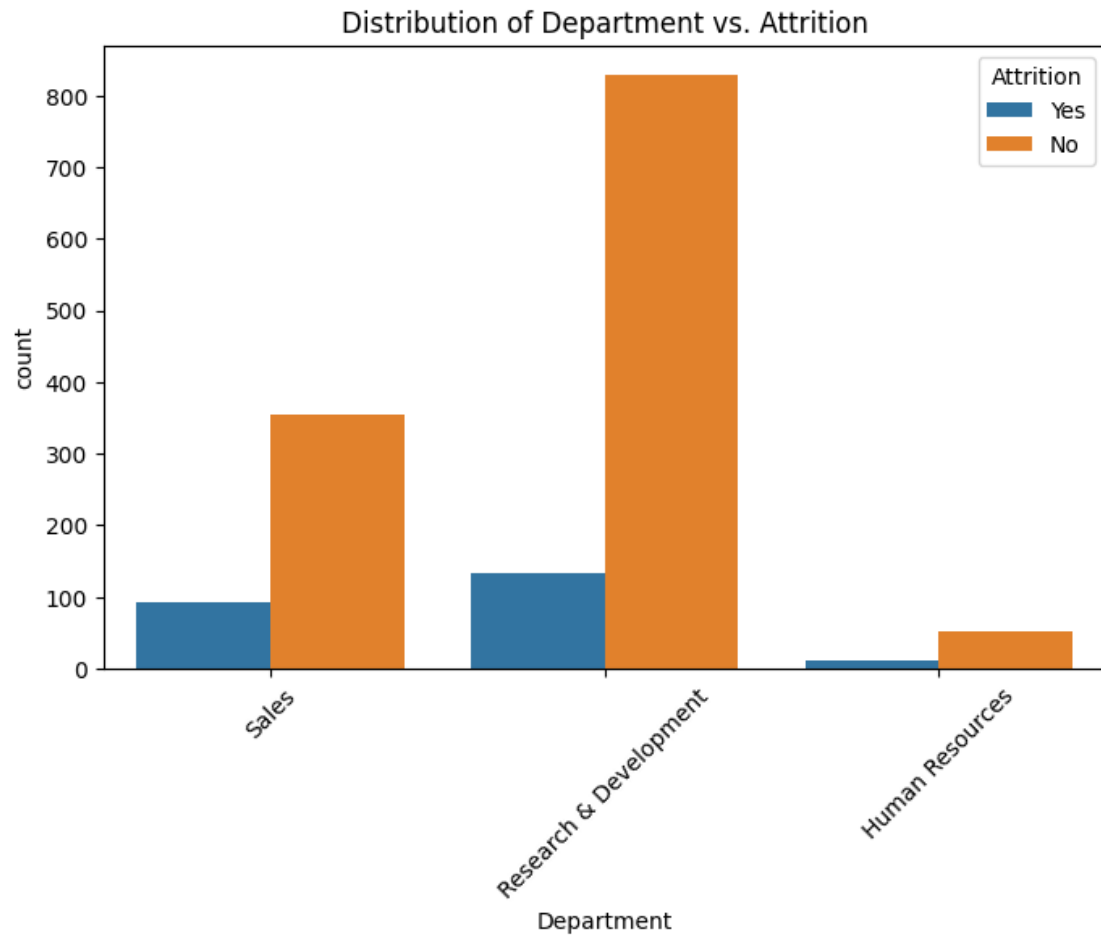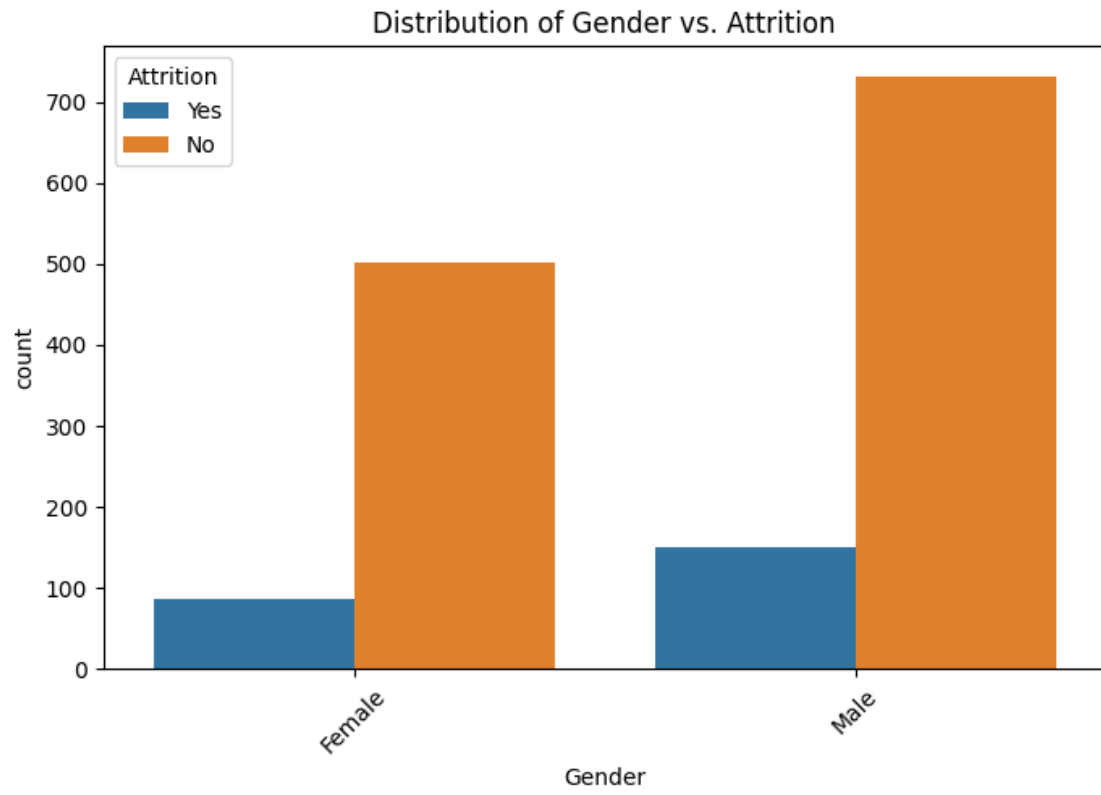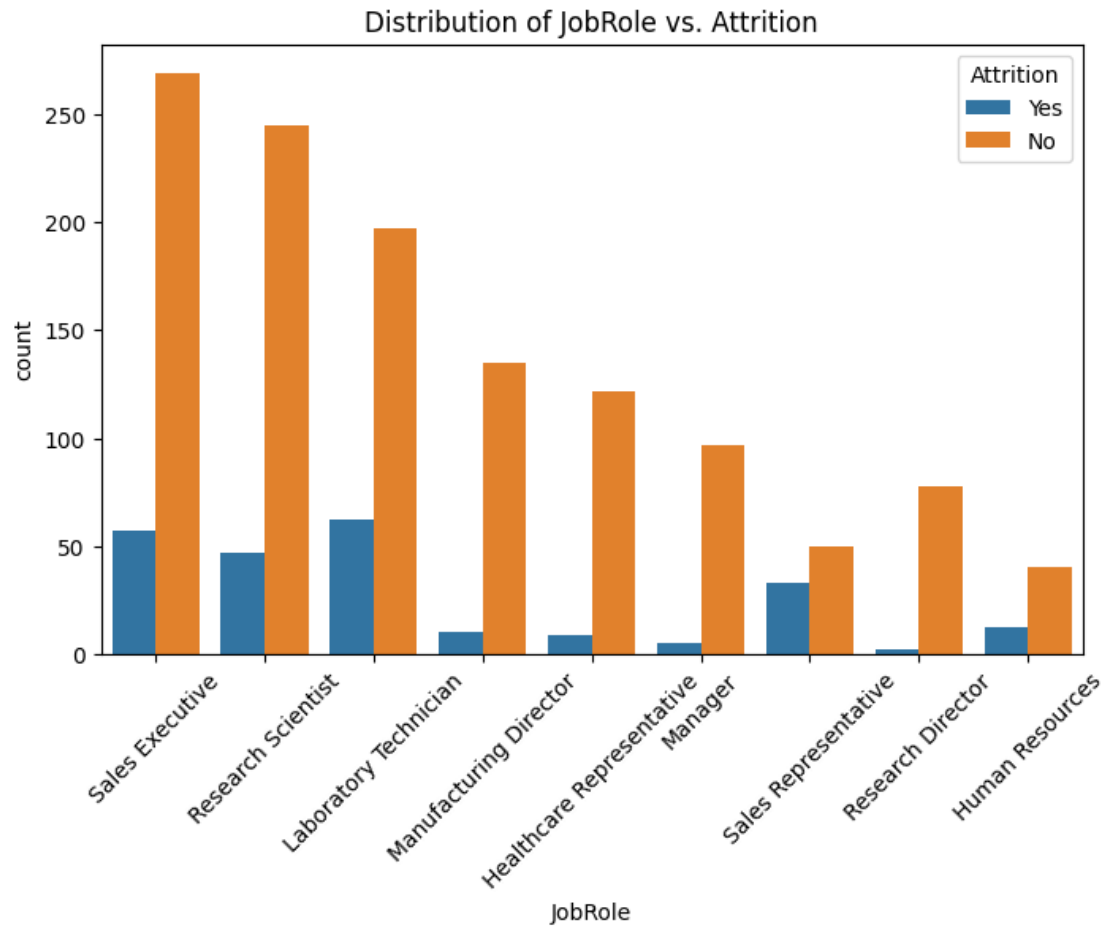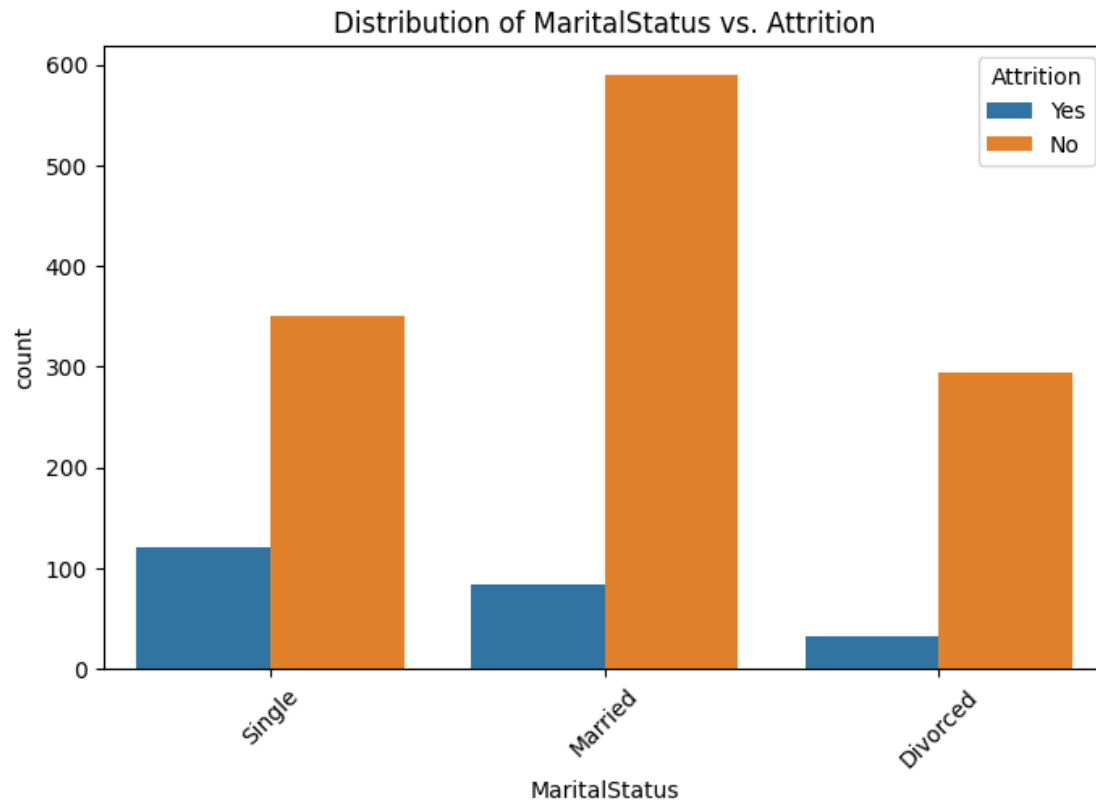

```
[15]:  # Explore the distribution of numeric variables
       numeric_vars = ["Age", "MonthlyIncome", "YearsAtCompany", "YearsInCurrentRole"]
       for var in numeric_vars:
           plt.figure(figsize=(8, 5))
           sns.boxplot(data=df, y=var, x="Attrition")
           plt.title(f"Distribution of {var} vs. Attrition")
           plt.show()
```

Distribution of Age vs. Attrition

Distribution of MonthlyIncome vs. Attrition

Distribution of YearsAtCompany vs. Attrition

Distribution of YearsInCurrentRole vs. Attrition
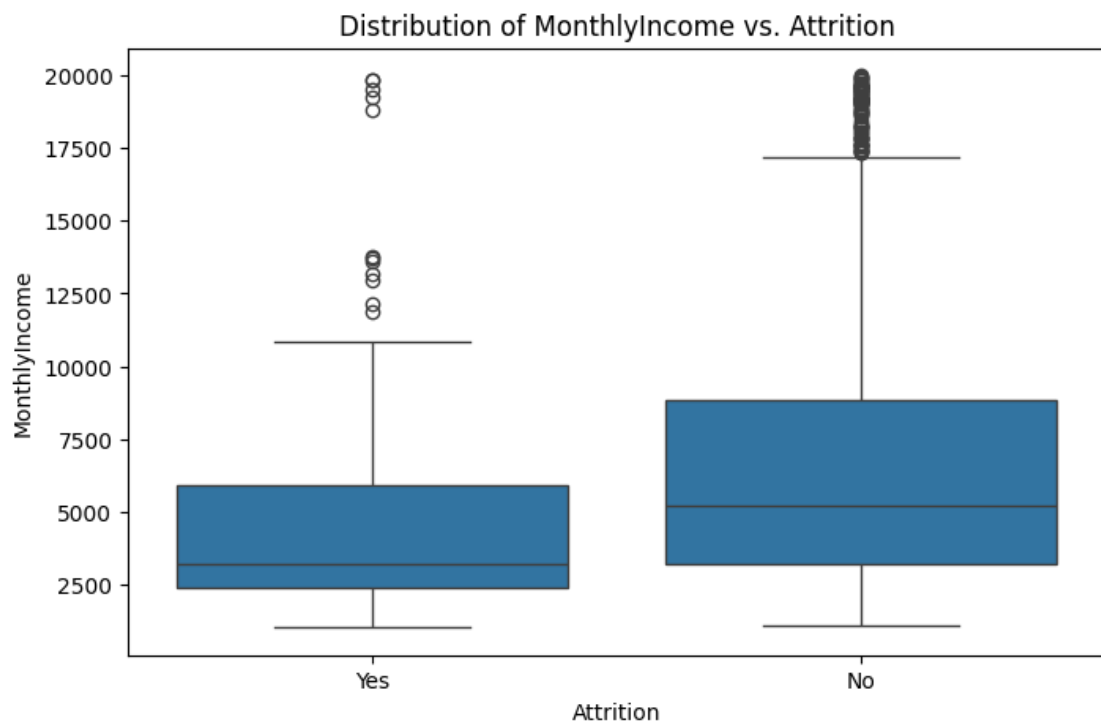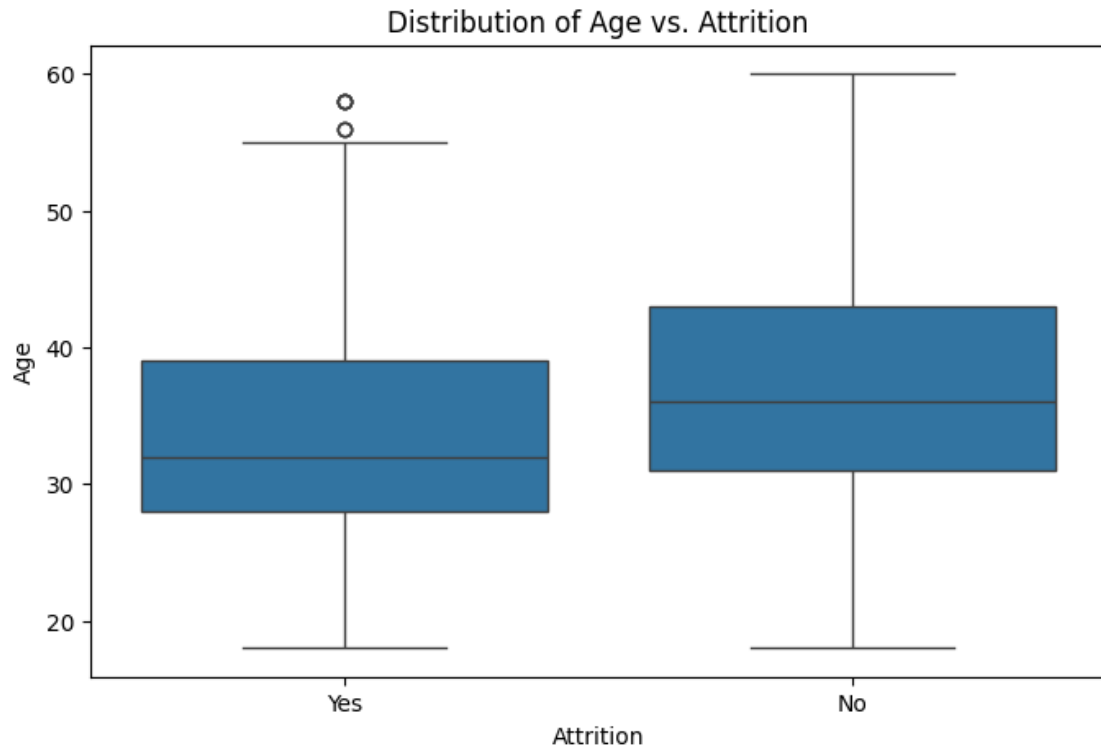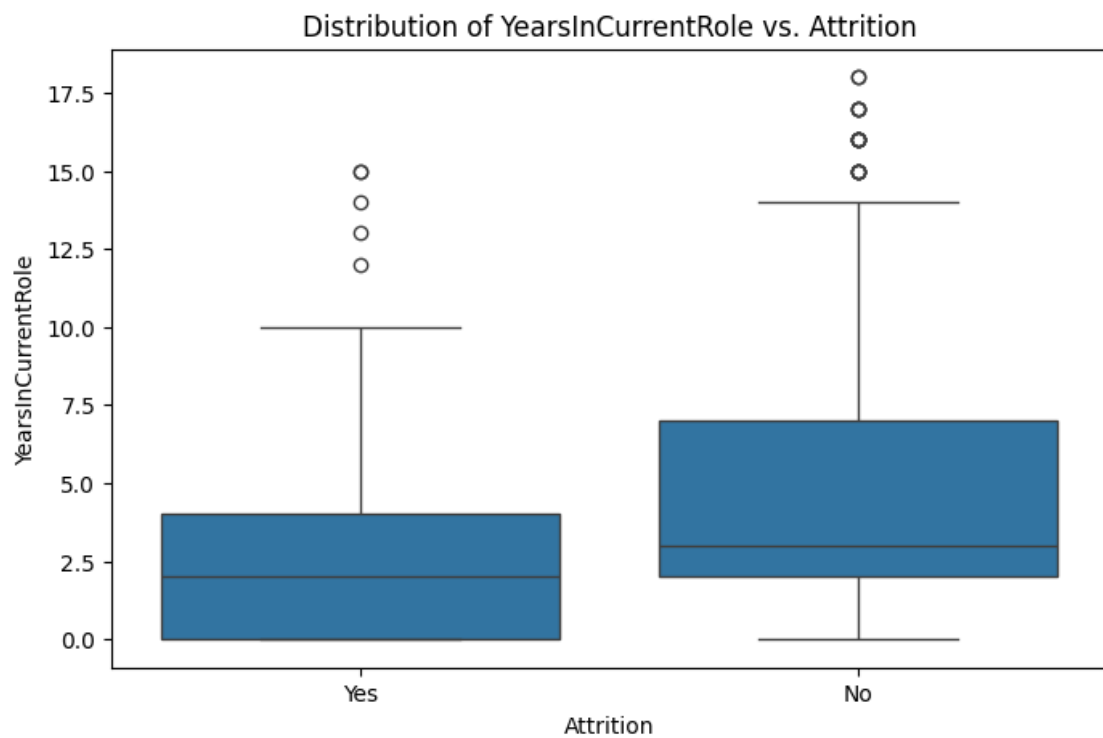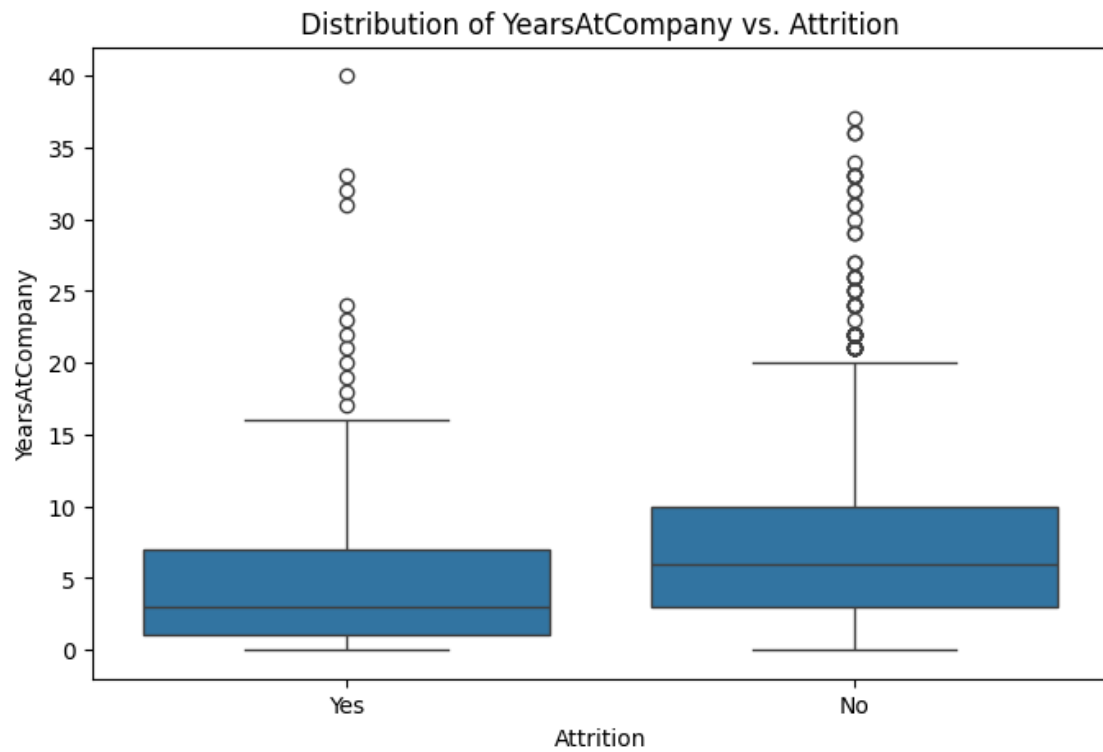
```
[16]: # Calculate and visualize the correlation between numeric variables
      correlation_matrix = df[numeric_vars].corr()
      sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
      plt.title("Correlation Matrix")
      plt.show()
```
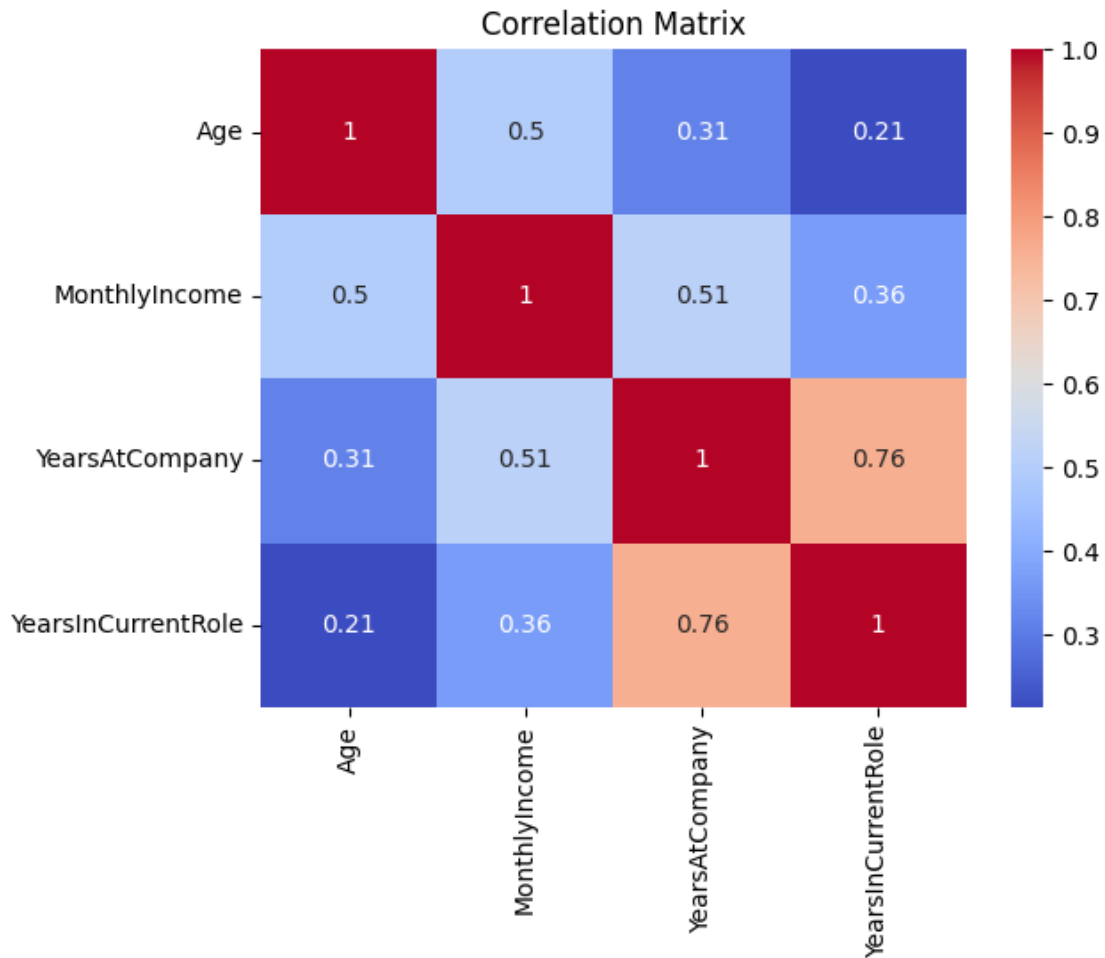

1.5 Analyze Attrition

```
[17]: # Understand the distribution of Attrition
      plt.figure(figsize=(6, 4))
      sns.countplot(data=df, x="Attrition")
      plt.title("Attrition Distribution")
      plt.show()
```

Attrition Distribution
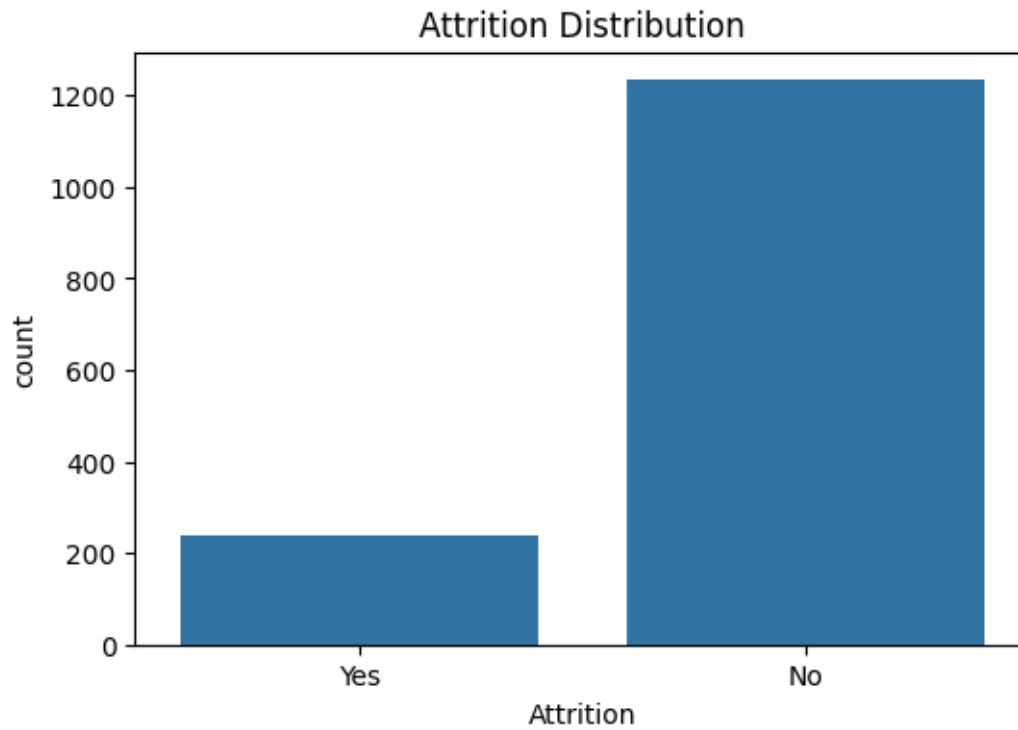

```
[18]:  # Explore factors related to attrition
       attrition_factors = ["MaritalStatus", "JobRole", "Education"]
       for var in attrition_factors:
           plt.figure(figsize=(8, 5))
           sns.countplot(data=df, x=var, hue="Attrition")
           plt.title(f"{var} vs. Attrition")
           plt.xticks(rotation=45)
           plt.show()
```

MaritalStatus vs. Attrition

JobRole vs. Attrition
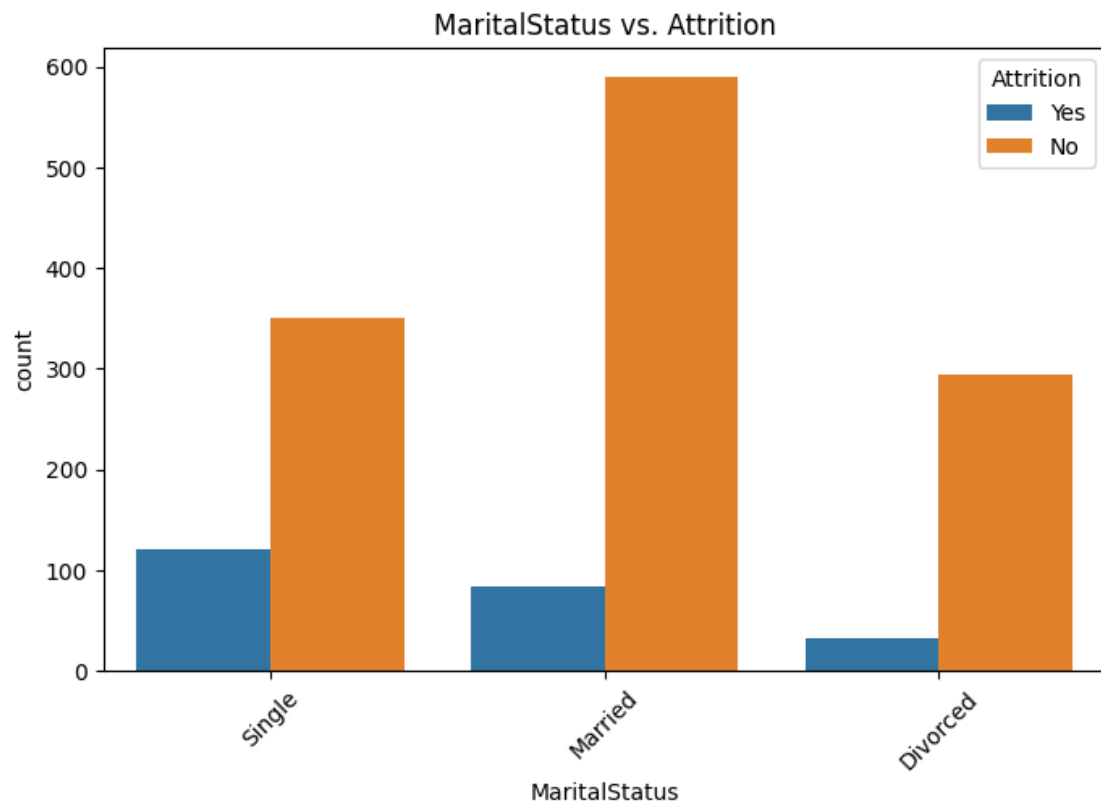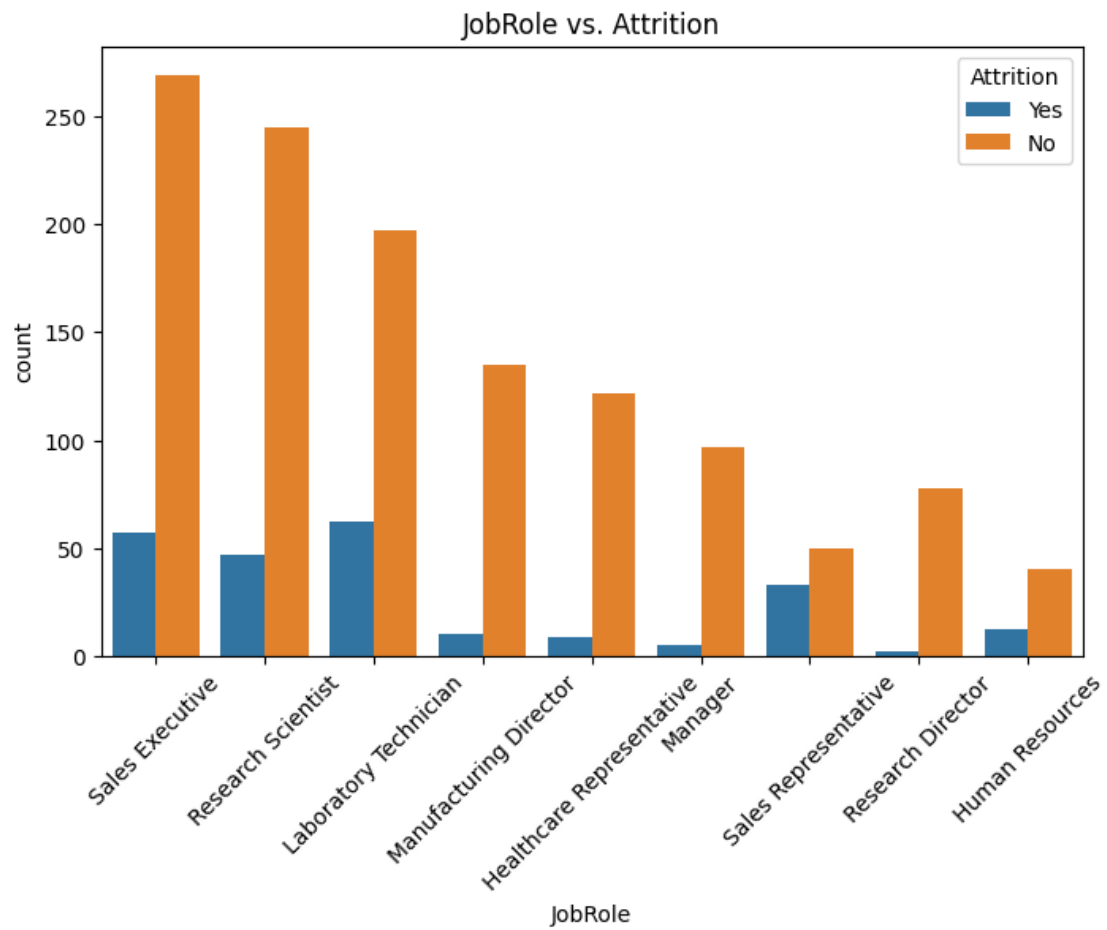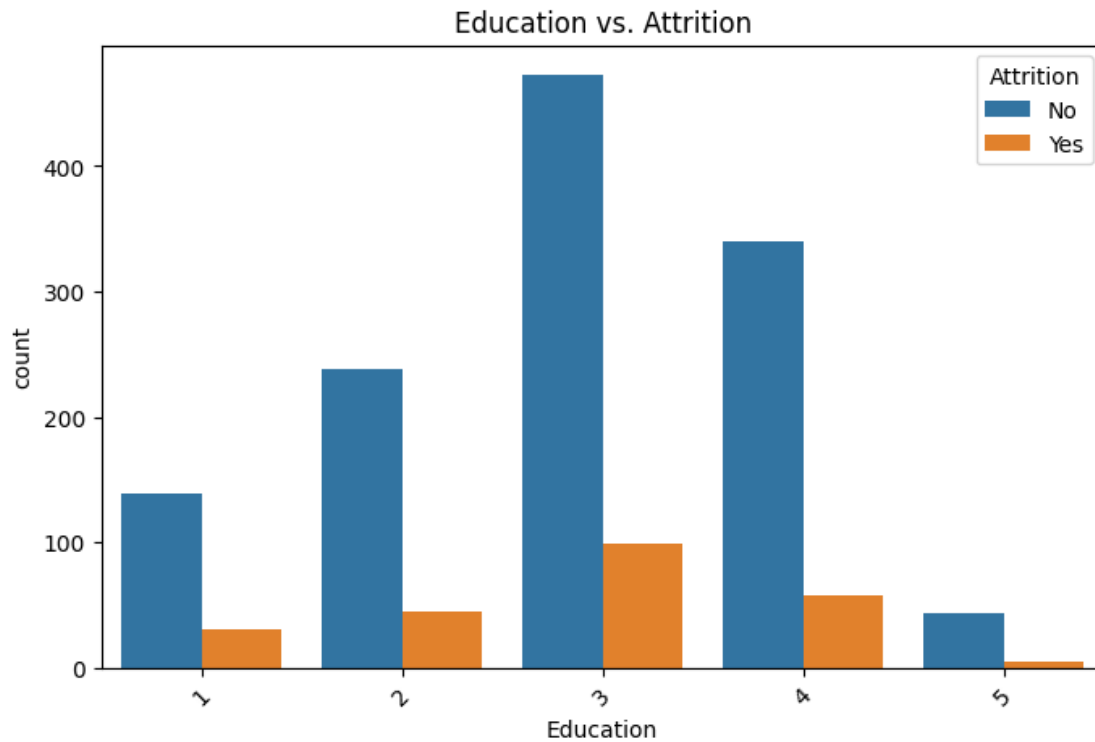
1.6 Employee Satisfaction and Engagement

```
[19]: # Analyze employee satisfaction and engagement
      satisfaction_vars = ["Job_Satisfaction", "EnvironmentSatisfaction",␣
       ↪"RelationshipSatisfaction", "WorkLifeBalance"]
      for var in satisfaction_vars:
          plt.figure(figsize=(8, 5))
          sns.countplot(data=df, x=var, hue="Attrition")
          plt.title(f"{var} vs. Attrition")
          plt.show()
```
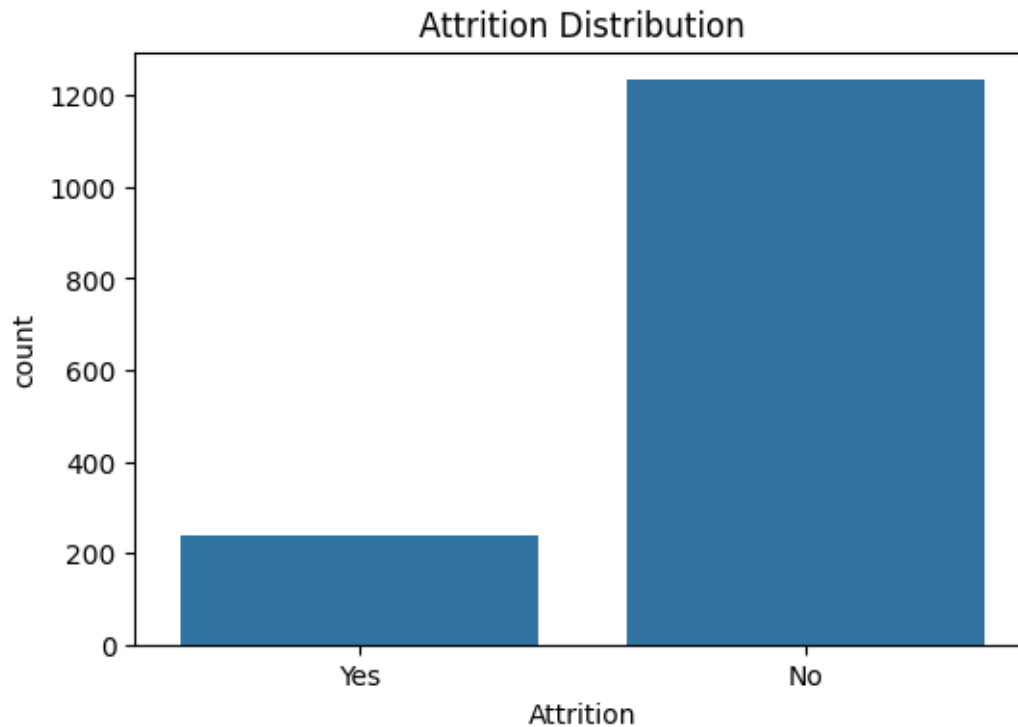
Job_Satisfaction vs. Attrition

EnvironmentSatisfaction vs. Attrition

RelationshipSatisfaction vs. Attrition

WorkLifeBalance vs. Attrition
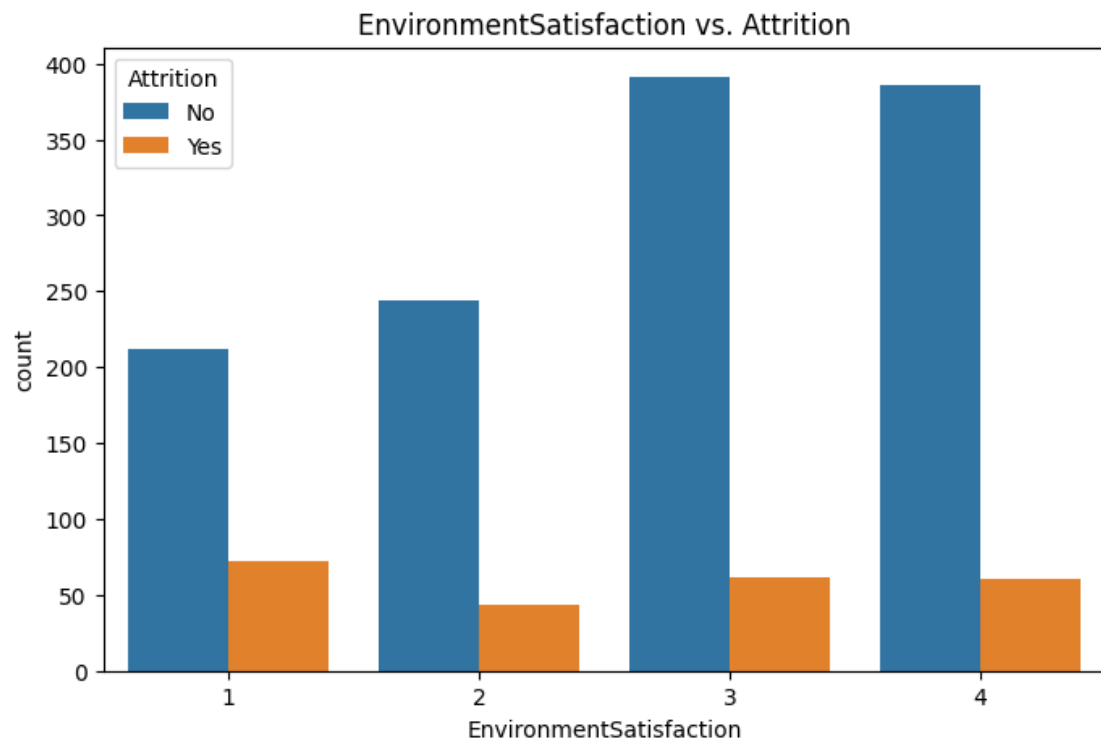
[20]:
```python
# Investigate the relationship between JobInvolvement and Attrition
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x="JobInvolvement", hue="Attrition")
plt.title("JobInvolvement vs. Attrition")
plt.show()
```
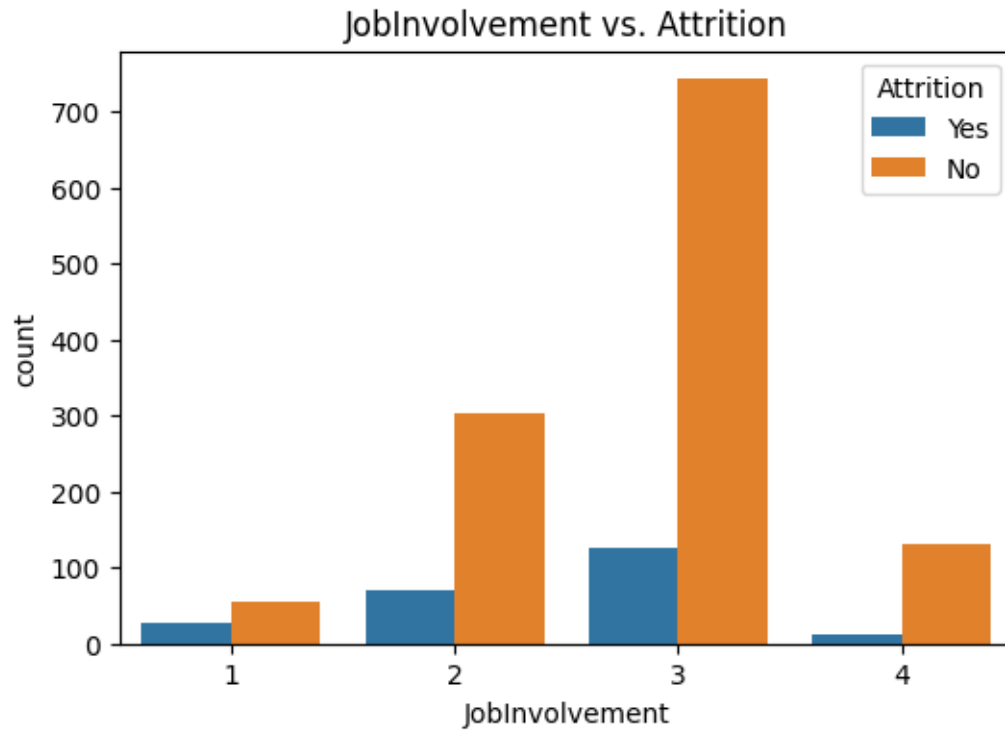

1.7 Career Progression

[21]:
```python
# Analyze employee career progression
plt.figure(figsize=(8, 5))
sns.boxplot(data=df, y="JobLevel", x="YearsInCurrentRole")
plt.title("Job Level vs. Years in Current Role")
plt.show()
```
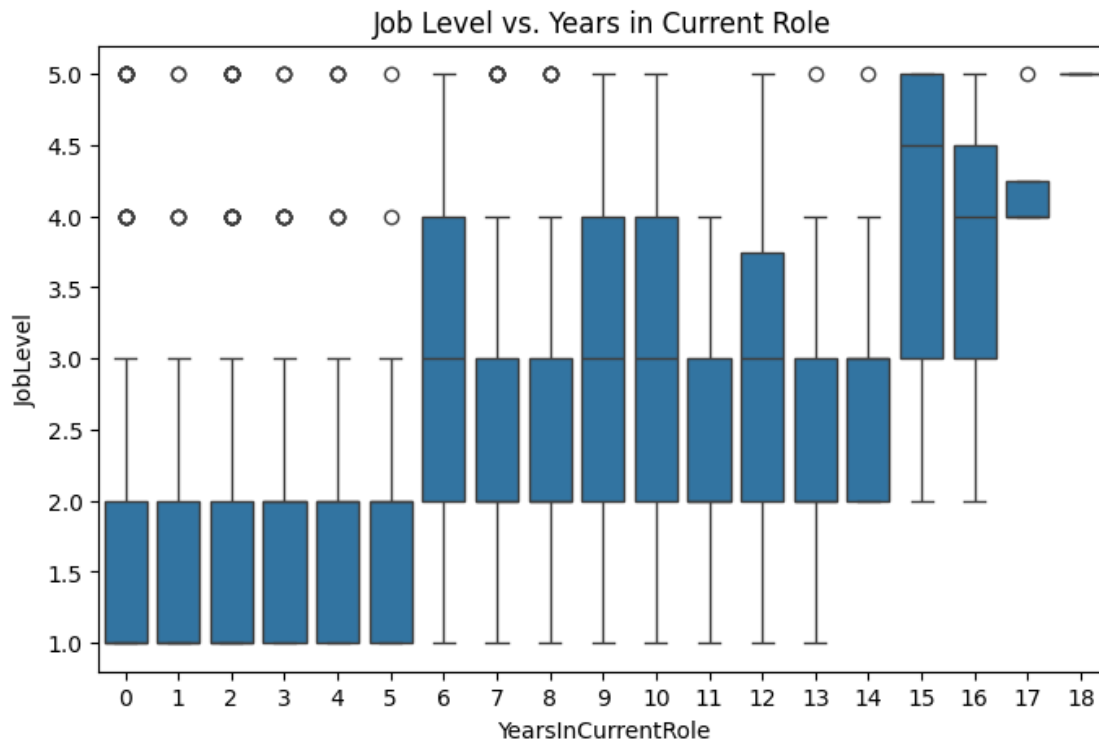
Job Level vs. Years in Current Role

```
# Investigate the relationship between promotions and attrition
plt.figure(figsize=(6, 4))
sns.countplot(data=df, x="YearsSinceLastPromotion", hue="Attrition")
plt.title("Years Since Last Promotion vs. Attrition")
plt.xticks(rotation=45)
plt.show()
```

Years Since Last Promotion vs. Attrition
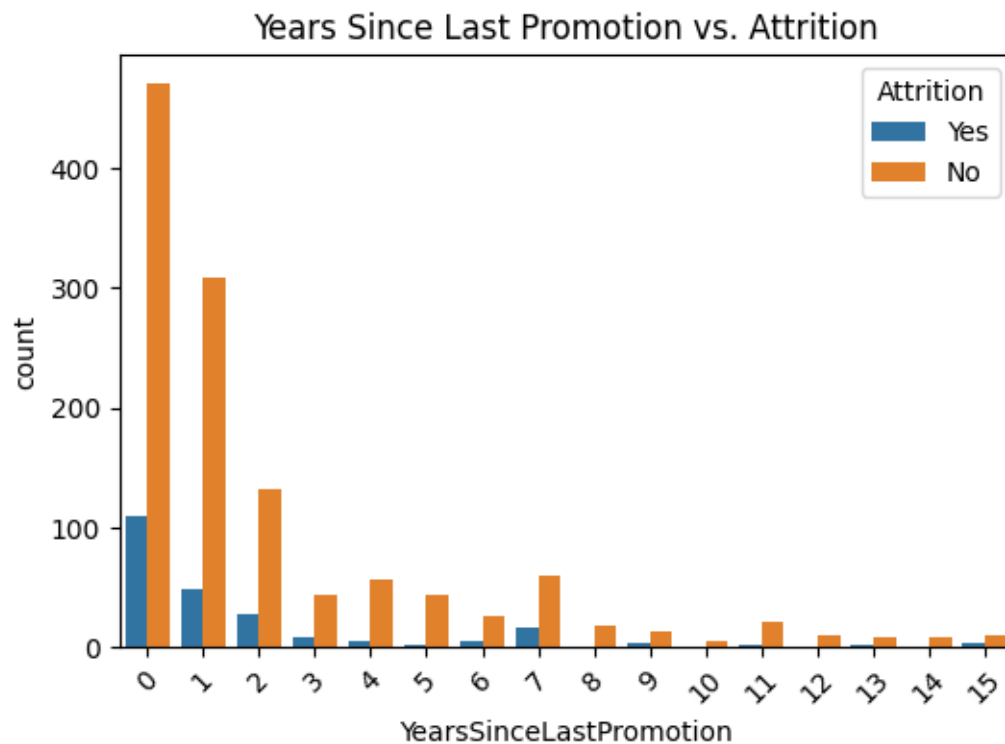
1.8 Work-Life Balance and Overtime

```
[23]: # Analyze the impact of work-life balance and overtime on attrition
      plt.figure(figsize=(6, 4))
      sns.countplot(data=df, x="OverTime", hue="Attrition")
      plt.title("OverTime vs. Attrition")
      plt.show()
```

OverTime vs. Attrition

```
[24]: plt.figure(figsize=(8, 5))
      sns.countplot(data=df, x="WorkLifeBalance", hue="Attrition")
      plt.title("WorkLifeBalance vs. Attrition")
      plt.show()
```
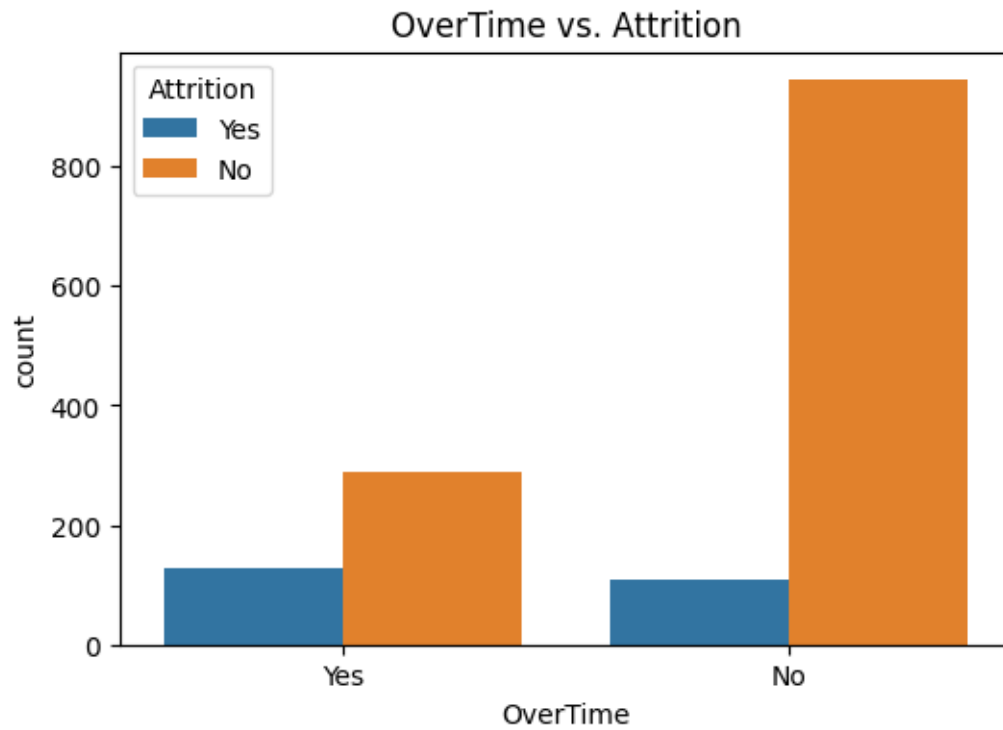
WorkLifeBalance vs. Attrition
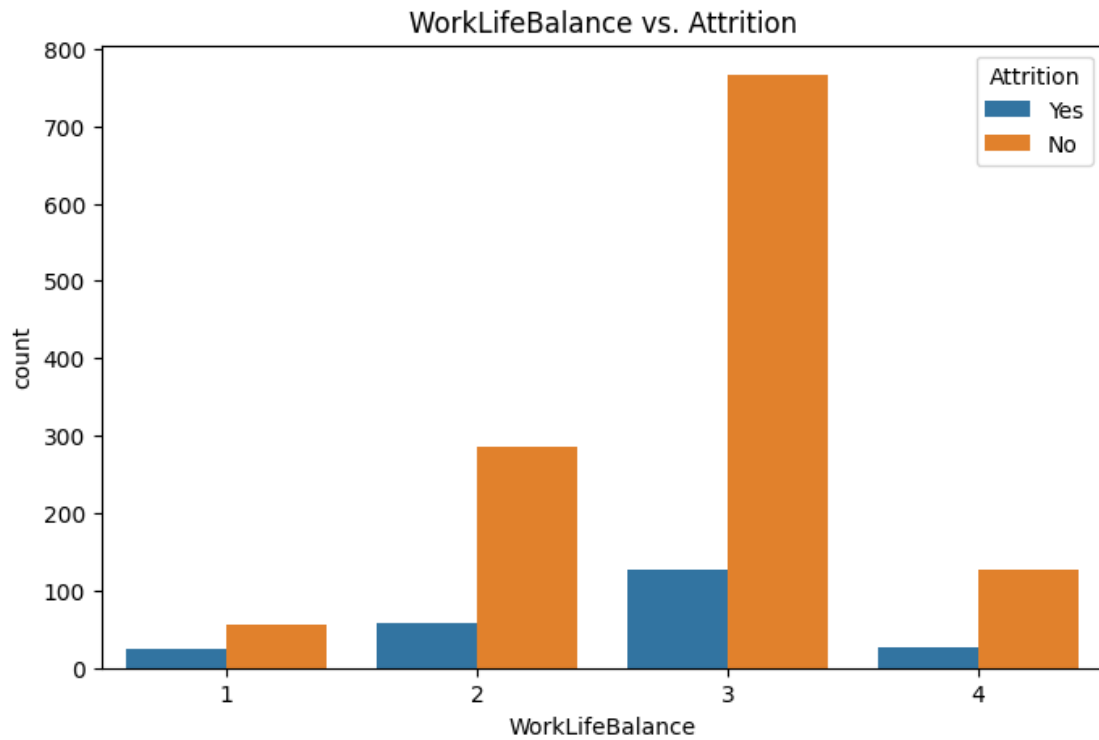
1.9 Analyze Attrition

```
[25]:  # Understand the distribution of Attrition
       plt.figure(figsize=(6, 4))
       sns.countplot(data=df, x="Attrition")
       plt.title("Attrition Distribution")
       plt.show()
```
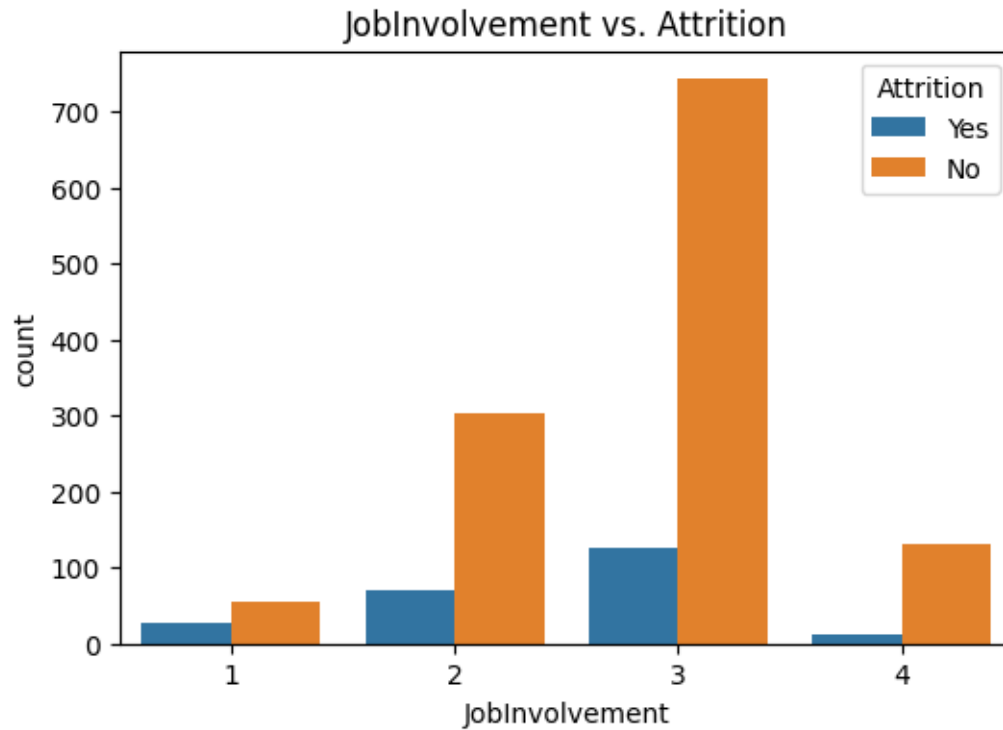
Attrition Distribution

2 Explore factors related to attrition

```
# Analyze employee satisfaction and engagement
satisfaction_vars = ["Job_Satisfaction", "EnvironmentSatisfaction",
 ↪"RelationshipSatisfaction", "WorkLifeBalance"]
for var in satisfaction_vars:
    plt.figure(figsize=(8, 5))
    sns.countplot(data=df, x=var, hue="Attrition")
    plt.title(f"{var} vs. Attrition")
    plt.show()
```

Job_Satisfaction vs. Attrition

EnvironmentSatisfaction vs. Attrition

RelationshipSatisfaction vs. Attrition

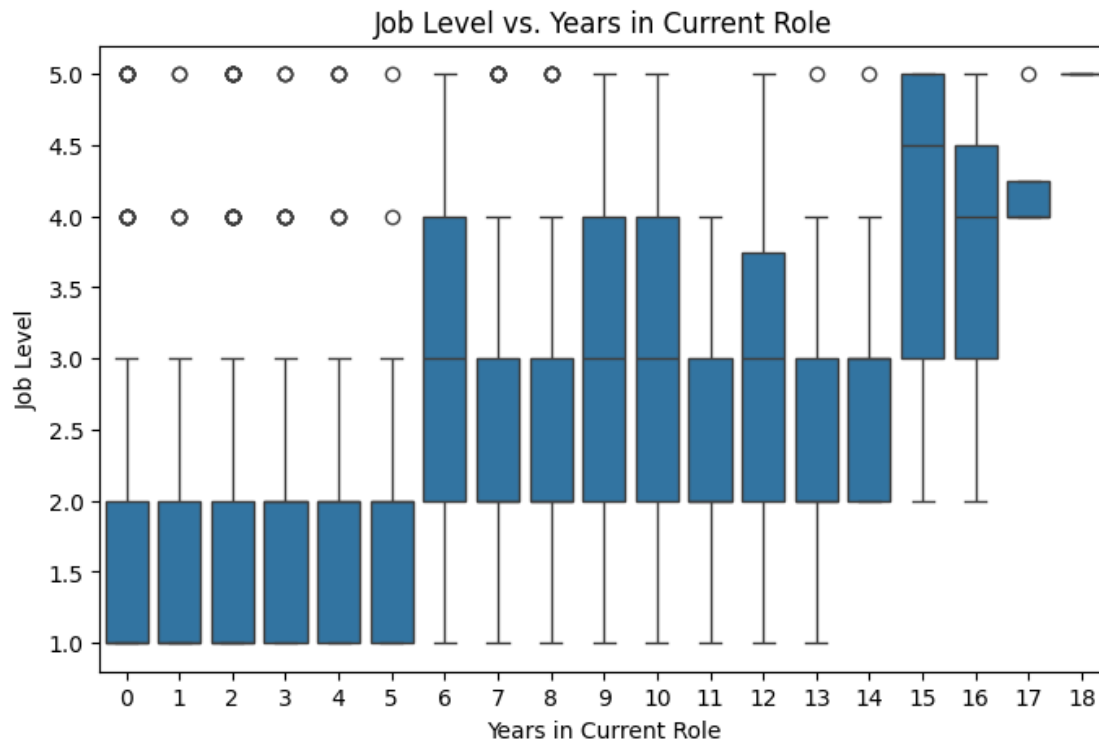WorkLifeBalance vs. Attrition
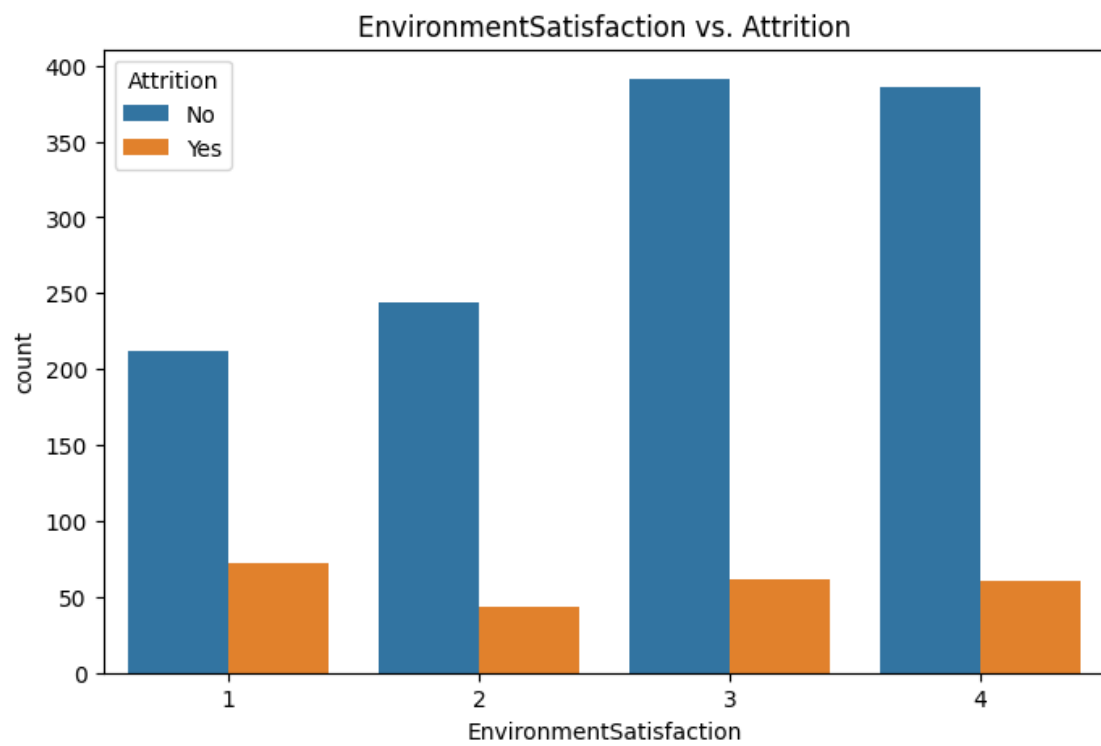
```
[27]: # Investigate the relationship between JobInvolvement and Attrition
      plt.figure(figsize=(6, 4))
      sns.countplot(data=df, x="JobInvolvement", hue="Attrition")
      plt.title("JobInvolvement vs. Attrition")
      plt.show()
```


2.1 Career Progression

```
[28]: # Analyze employee career progression
      plt.figure(figsize=(8, 5))
      sns.boxplot(data=df, y="JobLevel", x="YearsInCurrentRole")
      plt.title("Job Level vs. Years in Current Role")
      plt.xlabel("Years in Current Role")
      plt.ylabel("Job Level")
      plt.show()
```

Job Level vs. Years in Current Role

```
# Investigate the relationship between promotions and attrition
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x="YearsSinceLastPromotion", hue="Attrition")
plt.title("Years Since Last Promotion vs. Attrition")
plt.xlabel("Years Since Last Promotion")
plt.xticks(rotation=45)
plt.show()
```
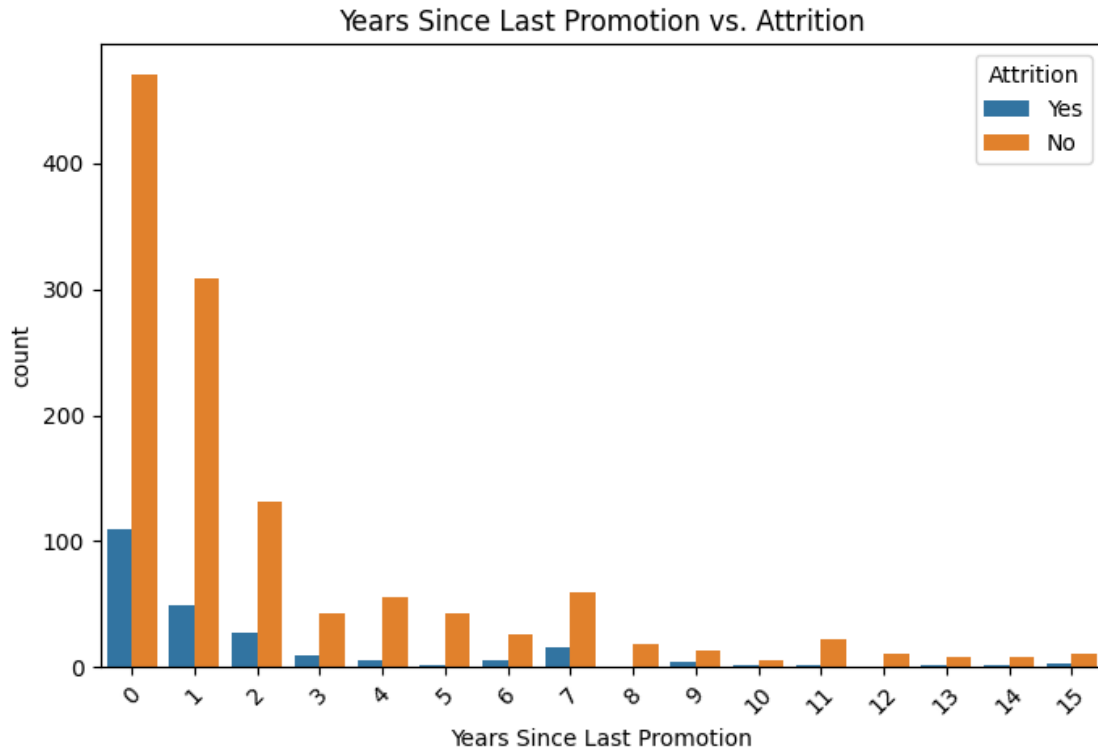
2.2 Work-Life Balance and Overtime

```
[30]: # Analyze the impact of work-life balance on attrition
      plt.figure(figsize=(8, 5))
      sns.countplot(data=df, x="OverTime", hue="Attrition")
      plt.title("OverTime vs. Attrition")
      plt.xlabel("Overtime")
      plt.show()
```

OverTime vs. Attrition
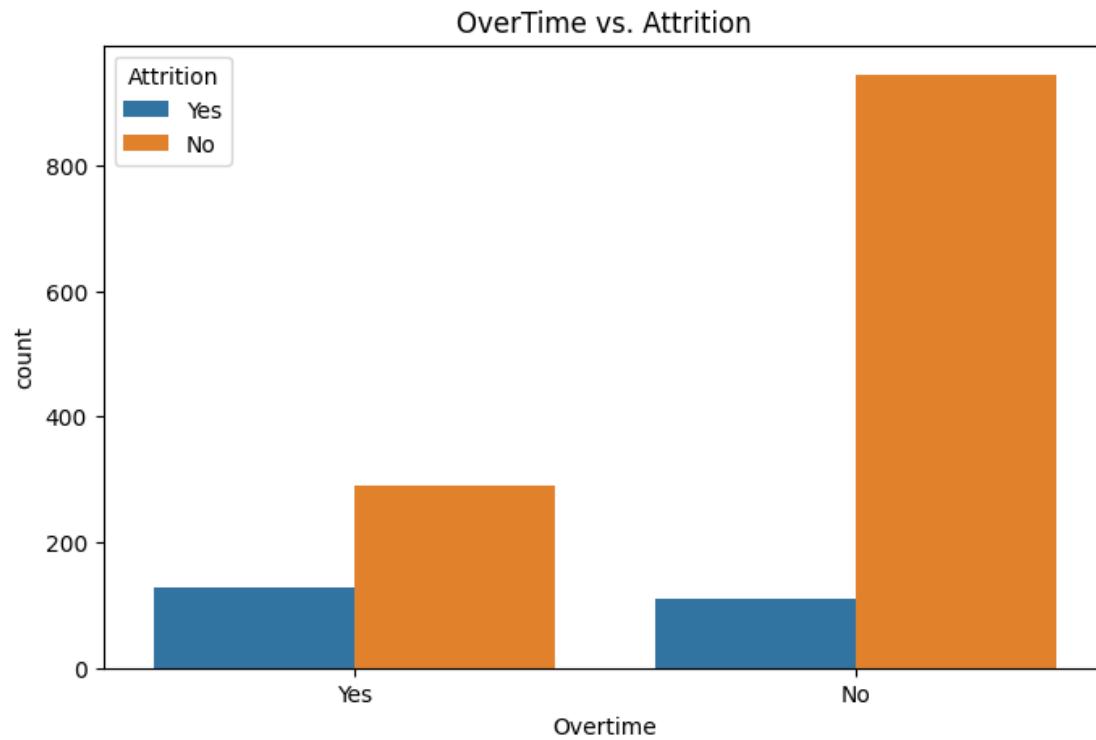

```
[31]:  # Investigate the relationship between work-life balance and attrition
       plt.figure(figsize=(8, 5))
       sns.countplot(data=df, x="WorkLifeBalance", hue="Attrition")
       plt.title("WorkLifeBalance vs. Attrition")
       plt.xlabel("Work-Life Balance")
       plt.show()
```
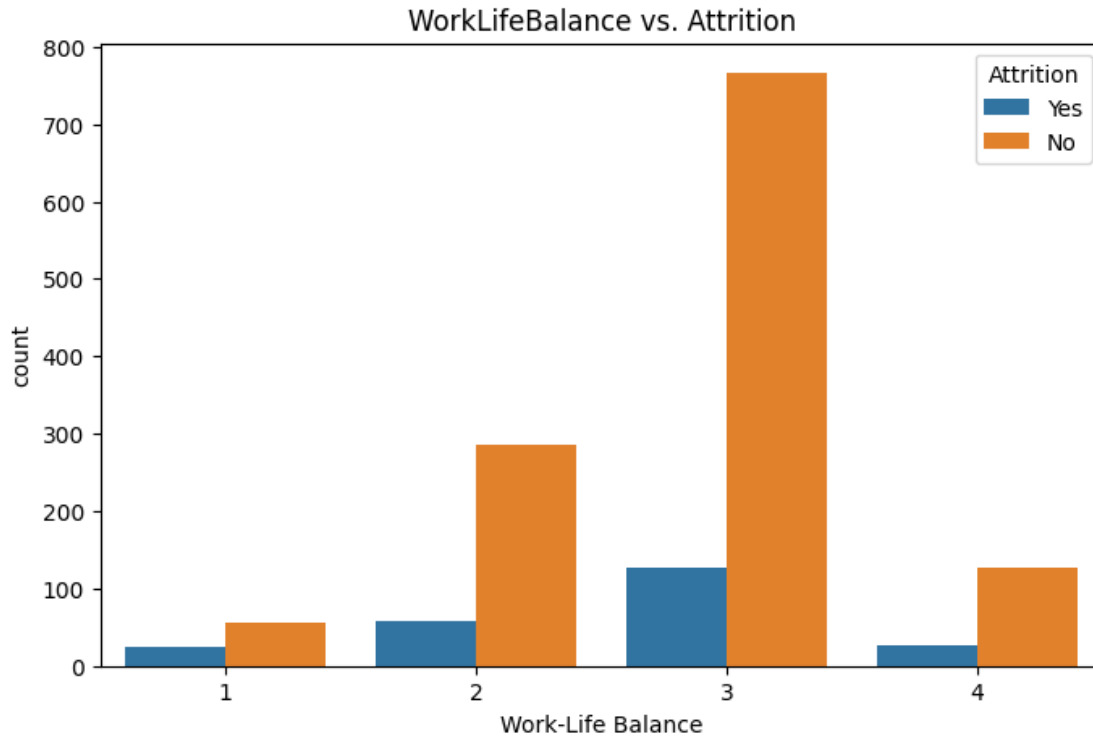
2.3 Conclusion and Recommendations

```python
[32]: # Summarize your findings
      print("Summary of Findings:")
      print("- Attrition Distribution:")
      attrition_counts = df['Attrition'].value_counts()
      print(attrition_counts)
```

```
Summary of Findings:
- Attrition Distribution:
Attrition
No      1233
Yes      237
Name: count, dtype: int64
```

```python
[33]: # Provide recommendations
      print("\nRecommendations:")
      print("- Consider improving work-life balance to reduce attrition among␣
        ↪employees.")
      print("- Monitor the impact of overtime work on attrition and take necessary␣
        ↪actions to manage workload.")
      print("- Focus on career development opportunities, such as promotions and␣
        ↪skill development, to enhance job satisfaction.")
```

```
print("- Conduct exit interviews with departing employees to gather more␣
  ↪insights into attrition reasons.")
```

Recommendations:
- Consider improving work-life balance to reduce attrition among employees.
- Monitor the impact of overtime work on attrition and take necessary actions to
manage workload.
- Focus on career development opportunities, such as promotions and skill
development, to enhance job satisfaction.
- Conduct exit interviews with departing employees to gather more insights into
attrition reasons.

[34]:
```python
# Overall Conclusion
print("\nOverall Conclusion:")
print("Based on the analysis, we have identified several factors that are␣
  ↪related to attrition within the organization. It's important for the company␣
  ↪to address these factors in order to improve employee retention and␣
  ↪satisfaction. By implementing the recommended actions, the company can work␣
  ↪towards reducing attrition and creating a more positive work environment.")
```

Overall Conclusion:
Based on the analysis, we have identified several factors that are related to
attrition within the organization. It's important for the company to address
these factors in order to improve employee retention and satisfaction. By
implementing the recommended actions, the company can work towards reducing
attrition and creating a more positive work environment.

[ ]: