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

In [2]:
```
df = pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
df.head(3)
```

Out[2]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	1	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1
1	49	0	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2
2	37	1	Travel_Rarely	1373	Research & Development	2	2	Other	1	4

3 rows × 35 columns

In [3]:
```
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   int64 
 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(27), object(8)
memory usage: 402.1+ KB
```

In [4]:
```
df.duplicated().sum()
```

Out[4]:
```
0
```

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
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
```

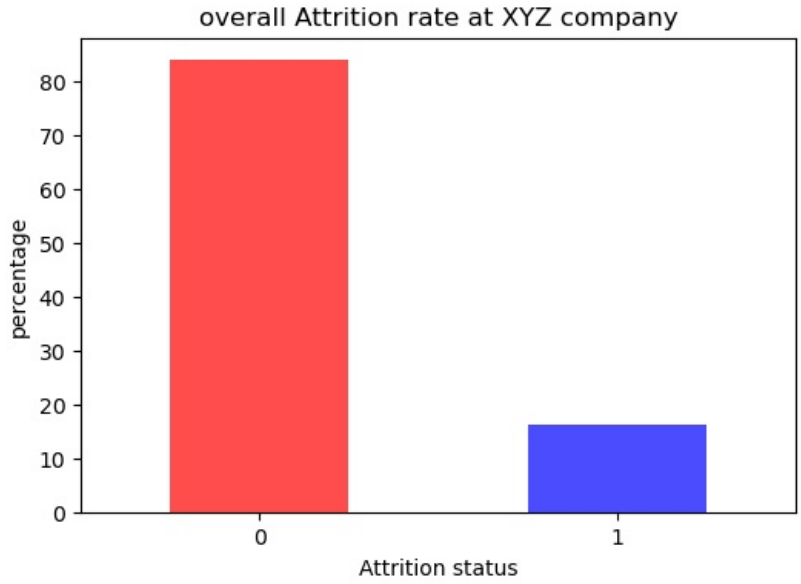
In [6]:
```
#describe
df.describe().T
```

Out[6]:

	count	mean	std	min	25%	50%	75%	max
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0	43.00	60.0
Attrition	1470.0	0.161224	0.367863	0.0	0.00	0.0	0.00	1.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0	4.00	4.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	2.0	3.00	5.0
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4919.0	8379.00	19999.0
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14235.5	20461.50	26999.0
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00	2.0	4.00	9.0
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00	14.0	18.00	25.0
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00	3.0	3.00	4.0
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00	3.0	4.00	4.0
StandardHours	1470.0	80.000000	0.000000	80.0	80.00	80.0	80.00	80.0
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00	1.0	1.00	3.0
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00	10.0	15.00	40.0
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00	3.0	3.00	6.0
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00	3.0	3.00	4.0
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00	5.0	9.00	40.0
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00	3.0	7.00	18.0
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00	1.0	3.00	15.0
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00	3.0	7.00	17.0

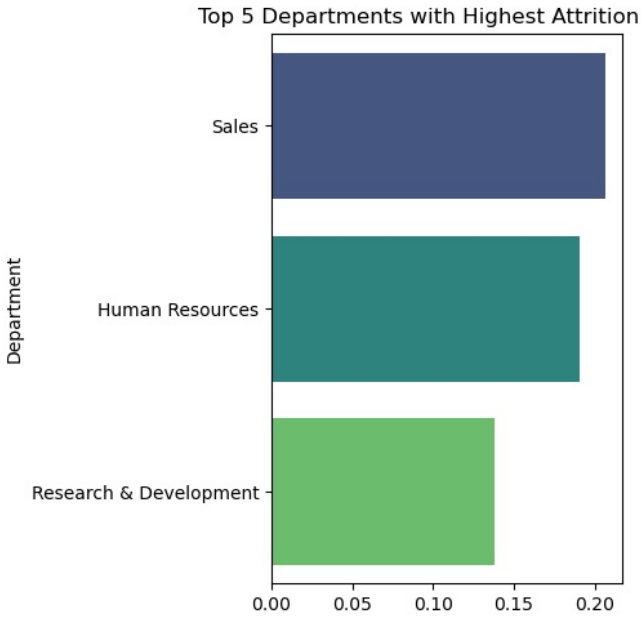
1.OVERALL ATTRITION RATE OF XYZ COMPANY

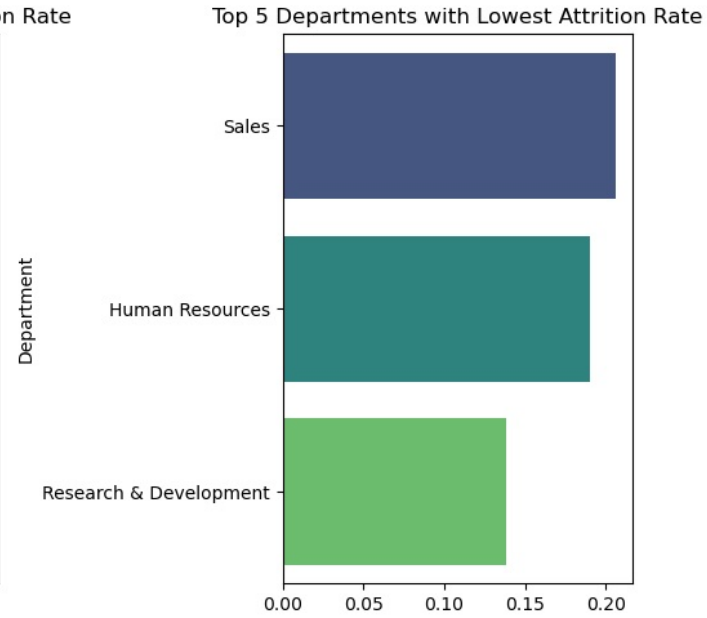
In [7]:
```
#calculate the overall attrition rate

attrition_rate = df['Attrition'].value_counts(normalize=True)*100

#plotting the attrition rate
plt.figure(figsize=(6,4))
attrition_rate.plot(kind='bar',color=['red','blue'],alpha=0.7)
plt.title('overall Attrition rate at XYZ company')
plt.xlabel('Attrition status')
plt.ylabel('percentage')
plt.xticks(rotation=0)
plt.show()
```



2.TOP 5 DEPARTMENT WITH HIGHEST AND LOWEST ATTRITION:

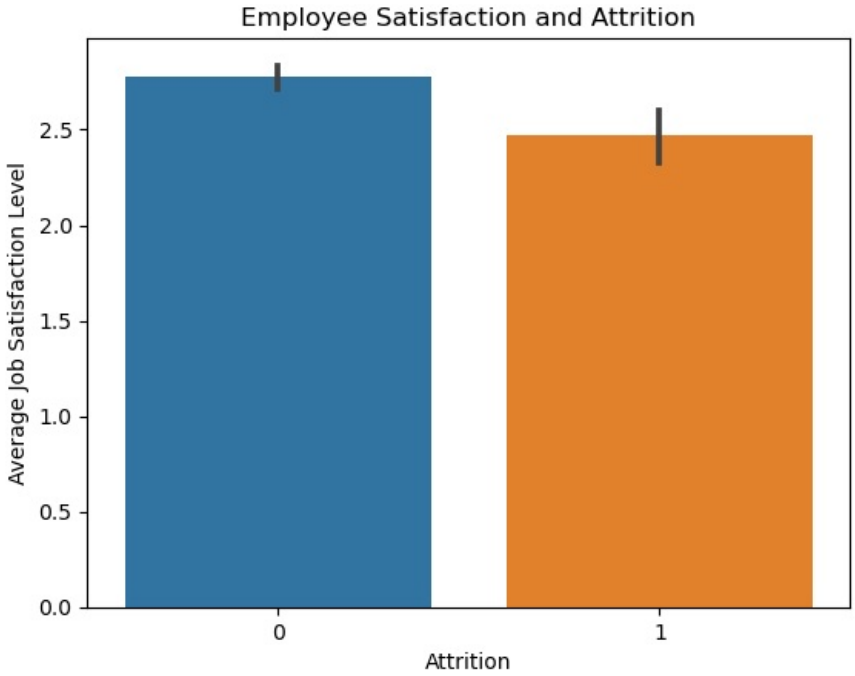
In [8]:
```
#calculate attrition rate of each departments;
attrition_rate= df.groupby('Department')['Attrition'].mean().sort_values(ascending=False)

#Extract top 5 and bottom 5 departments;
top_5_departments = attrition_rate.head(5)
bottom_5_departments = attrition_rate.tail(5)

#visulization;
plt.figure(figsize=(10,5))

#Top 5 departments with highest attrtion rate;
plt.subplot(1,2,1)
sns.barplot(x=top_5_departments.values,y=top_5_departments.index,palette='viridis')
plt.title('Top 5 Departments with Highest Attrition Rate')

#Top 5 Depatments with LOwest Attrtion Rate;
plt.subplot(1,2,2)
sns.barplot(x=bottom_5_departments.values,y=bottom_5_departments.index,palette='viridis')
plt.title('Top 5 Departments with Lowest Attrition Rate')

plt.tight_layout()
plt.show()
```





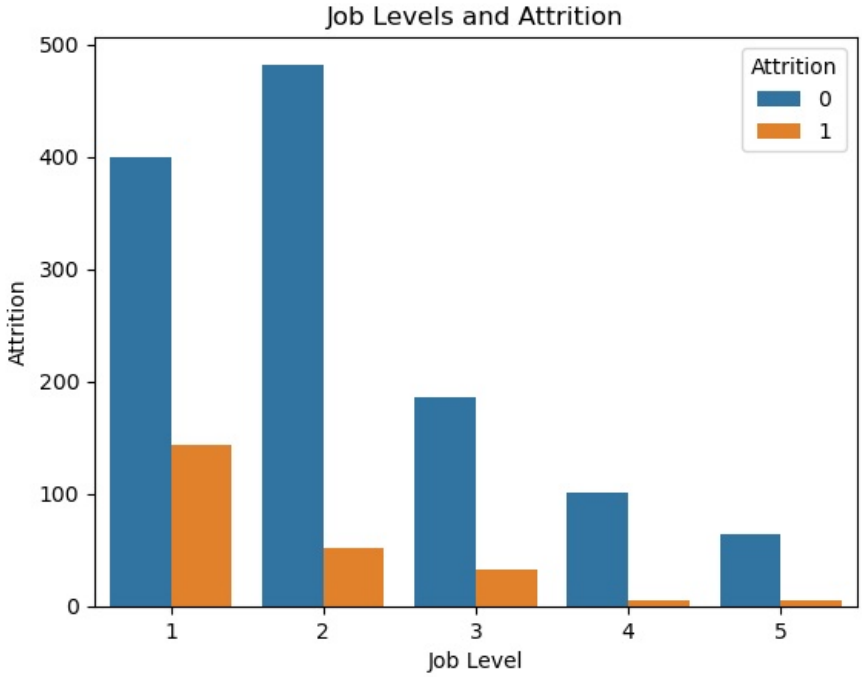
3.EMPLOYEE SATISFACTION AND ATTRITION :

In [9]:
```
#calculate relationship between employee satisfaction and attrition;
sns.barplot(x=df['Attrition'], y=df['JobSatisfaction'])
plt.title('Employee Satisfaction and Attrition')
plt.xlabel('Attrition')
plt.ylabel('Average Job Satisfaction Level')
plt.show()
```



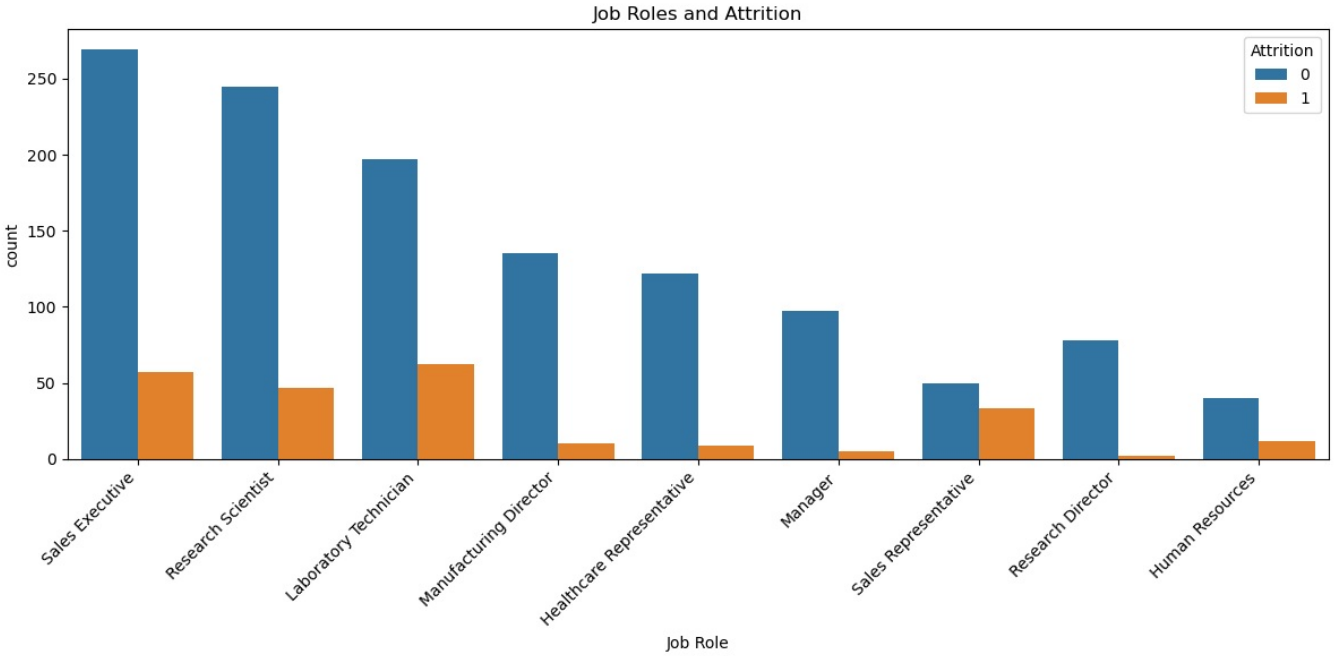
4.JOB LEVEL AND ATTRITION;

In [15]:
```
#Analyse relationship between job level and attrition;
sns.countplot(x='JobLevel', hue='Attrition' ,data=df)
plt.title('Job Levels and Attrition')
plt.xlabel('Job Level')
plt.ylabel('Attrition')
plt.show()
```



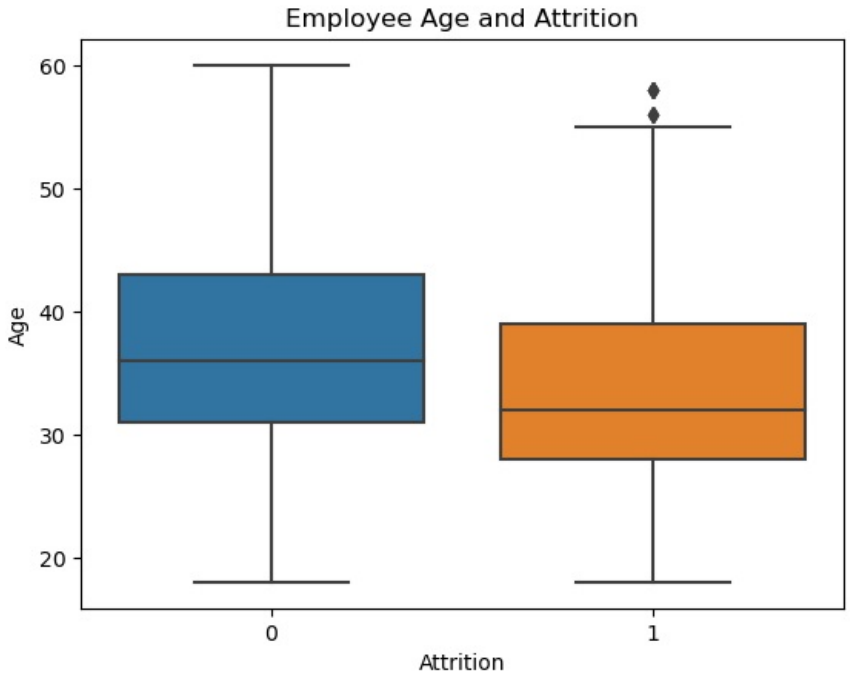
5.JOB ROLES AND ATTRITION;

In [25]:
```
#Analyse the relationship between job roles and attrition;
plt.figure(figsize=(12,6))
sns.countplot(x='JobRole', hue='Attrition', data=df)
plt.title('Job Roles and Attrition')
plt.xlabel('Job Role')
plt.ylabel('count')
plt.xticks(rotation=45 ,ha='right')
plt.tight_layout()
plt.show()
```



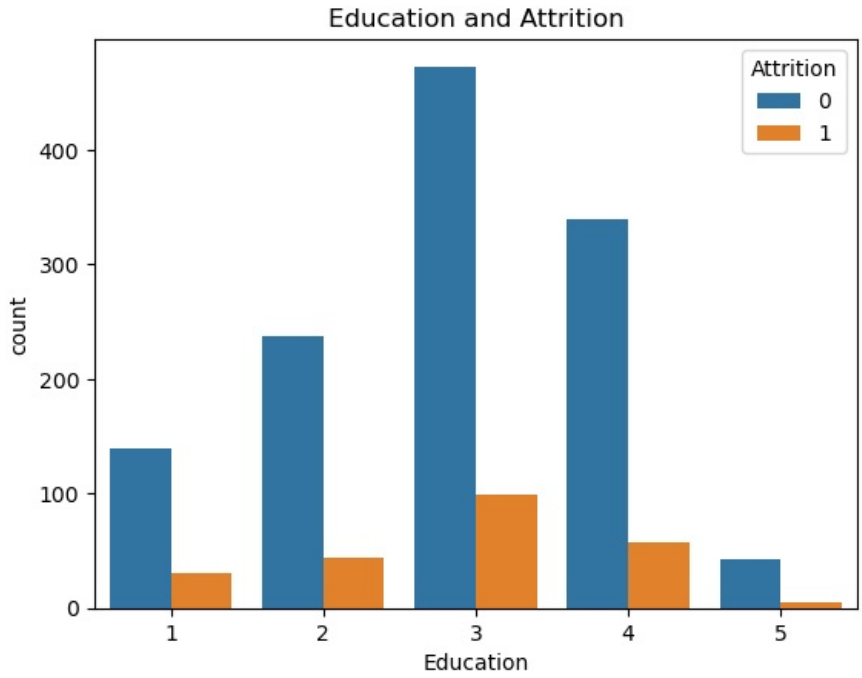
6.EMPLOYEE AGE AND ATTRITION;

In [27]:
```
#Analyse relationship between employee age and attrition;
sns.boxplot(x='Attrition',y='Age', data=df)
plt.xlabel('Attrition')
plt.ylabel('Age')
plt.title('Employee Age and Attrition')
plt.show()
```



7.EDUCATION AND ATTRITION;

In [29]:
```
#Analyse relationship between education and attrition;
sns.countplot(x='Education', hue='Attrition', data=df)
plt.title('Education and Attrition')
plt.xlabel('Education')
plt.ylabel('count')
plt.show()
```

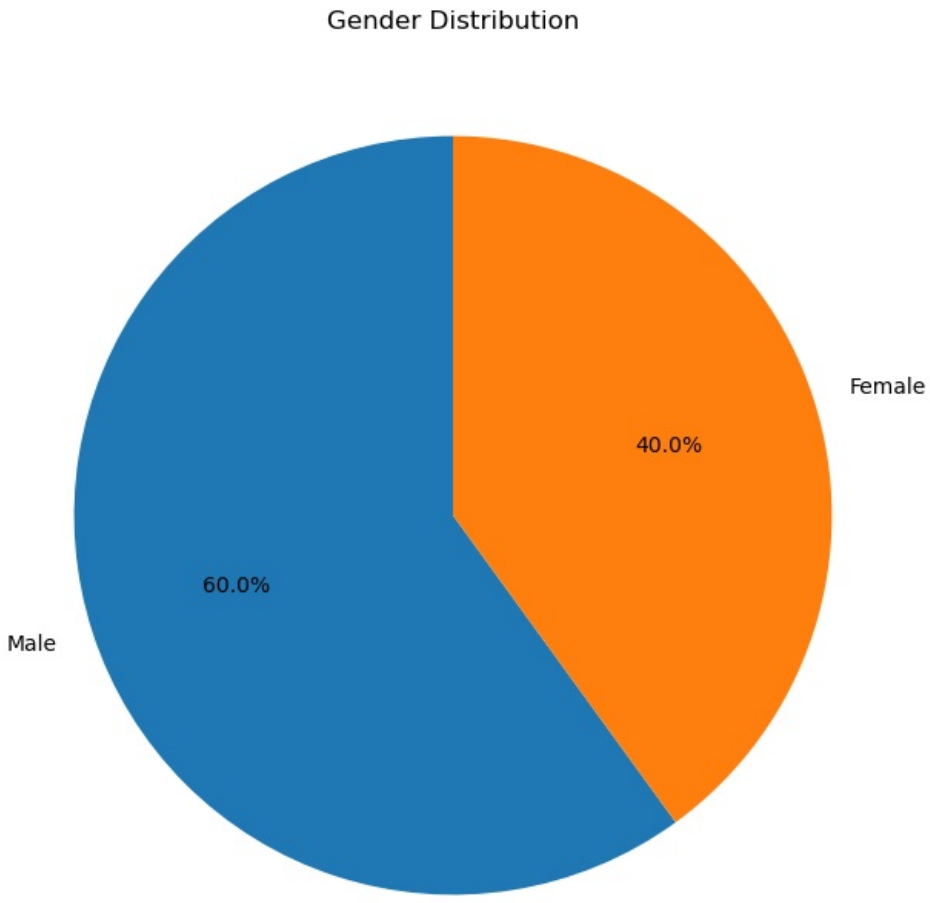


8.WORK EXPERIENCE AND ATTRITION;#violin plot for work experience

In [31]:
```
#Analyse the relationship between work experience and attrition;
plt.figure(figsize=(10,6))
sns.violinplot(x='Attrition', y='TotalWorkingYears', data=df)
plt.xlabel('Attrition')
plt.ylabel('TotalWorkingYears')
plt.title('Work Experience and Attrition')
plt.show()
```

9.GENDER AND ATTRITION;

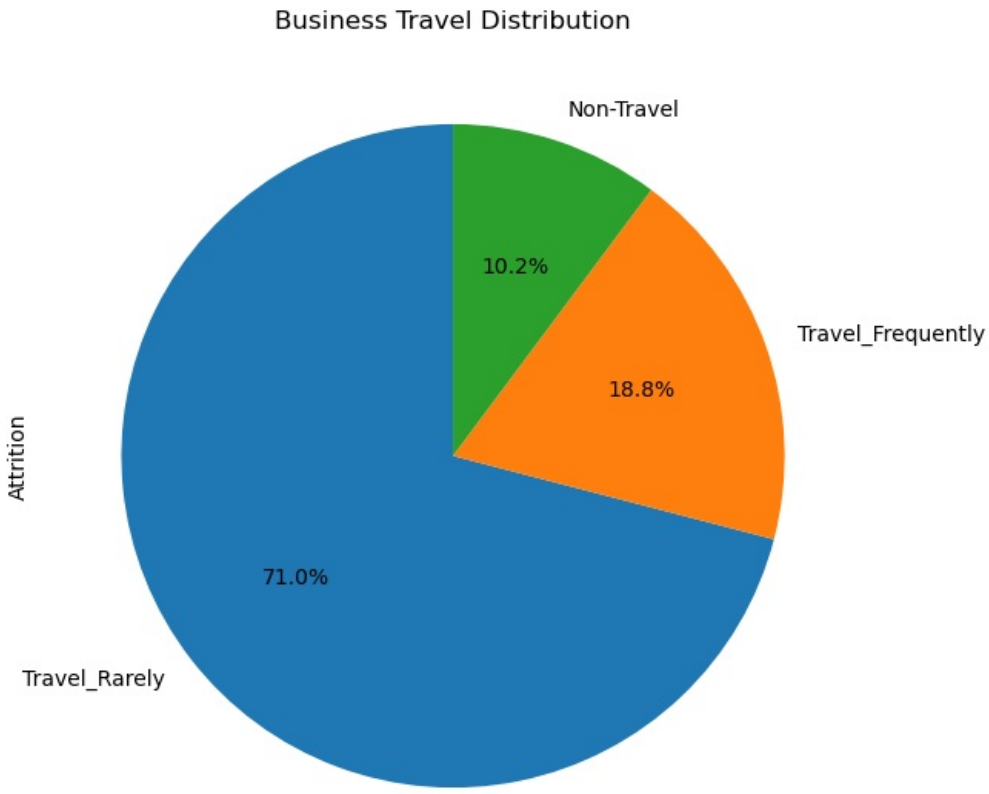
In [33]:
```
#Pie chart for gender distribution;
gender_distribution= df['Gender'].value_counts()
plt.figure(figsize=(8,8))
plt.pie(gender_distribution, labels=gender_distribution.index, autopct='%1.1f%%', startangle=90)
plt.title('Gender Distribution')
plt.show()
```



10.BUSINESS TRAVEL AND ATTRITION;

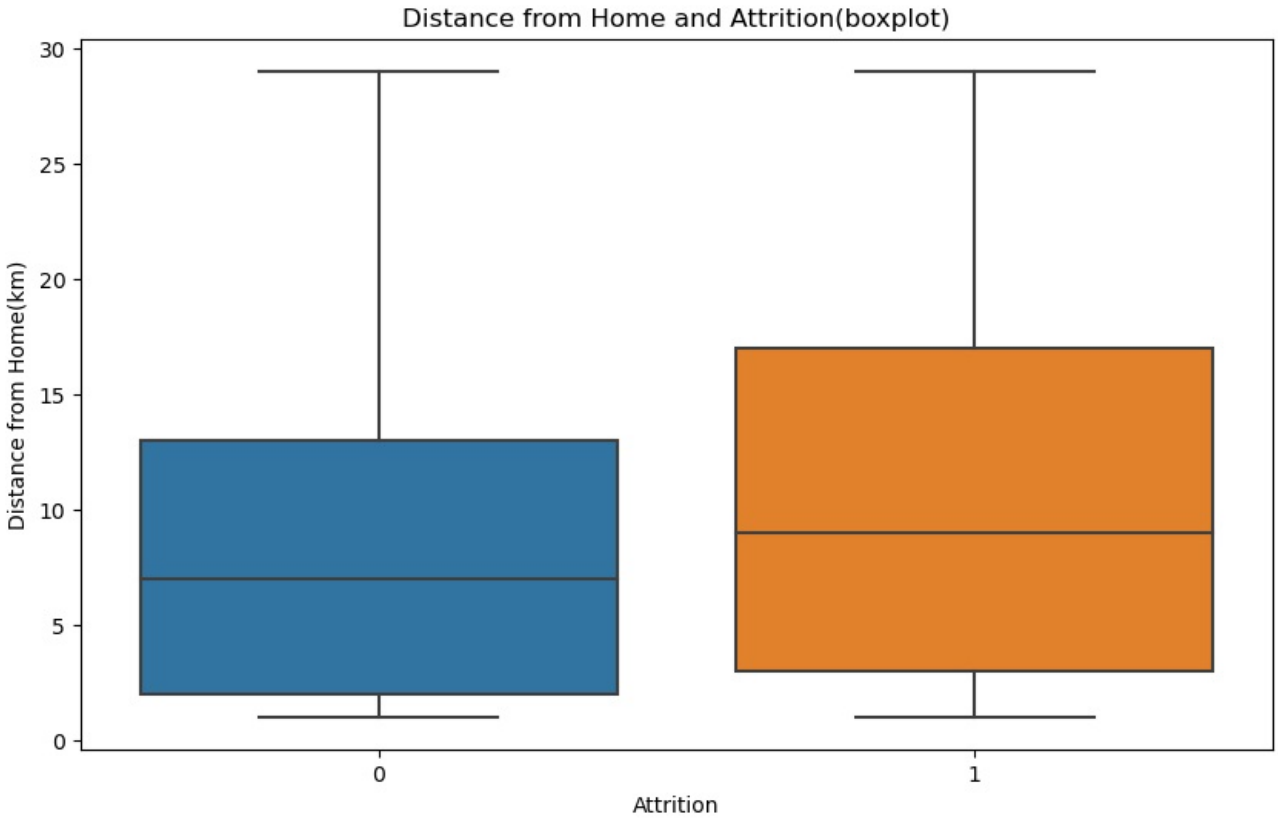
10.BUSINESS TRAVEL AND ATTRITION;

In [34]:
```
#Pie chart for business travel and attrition;
travel_distribution = df['BusinessTravel'].value_counts()
plt.figure(figsize=(7,7))
plt.pie(travel_distribution, labels=travel_distribution.index, autopct='%1.1f%%',startangle=90)
plt.ylabel('Attrition')
plt.title('Business Travel Distribution')
plt.show()
```



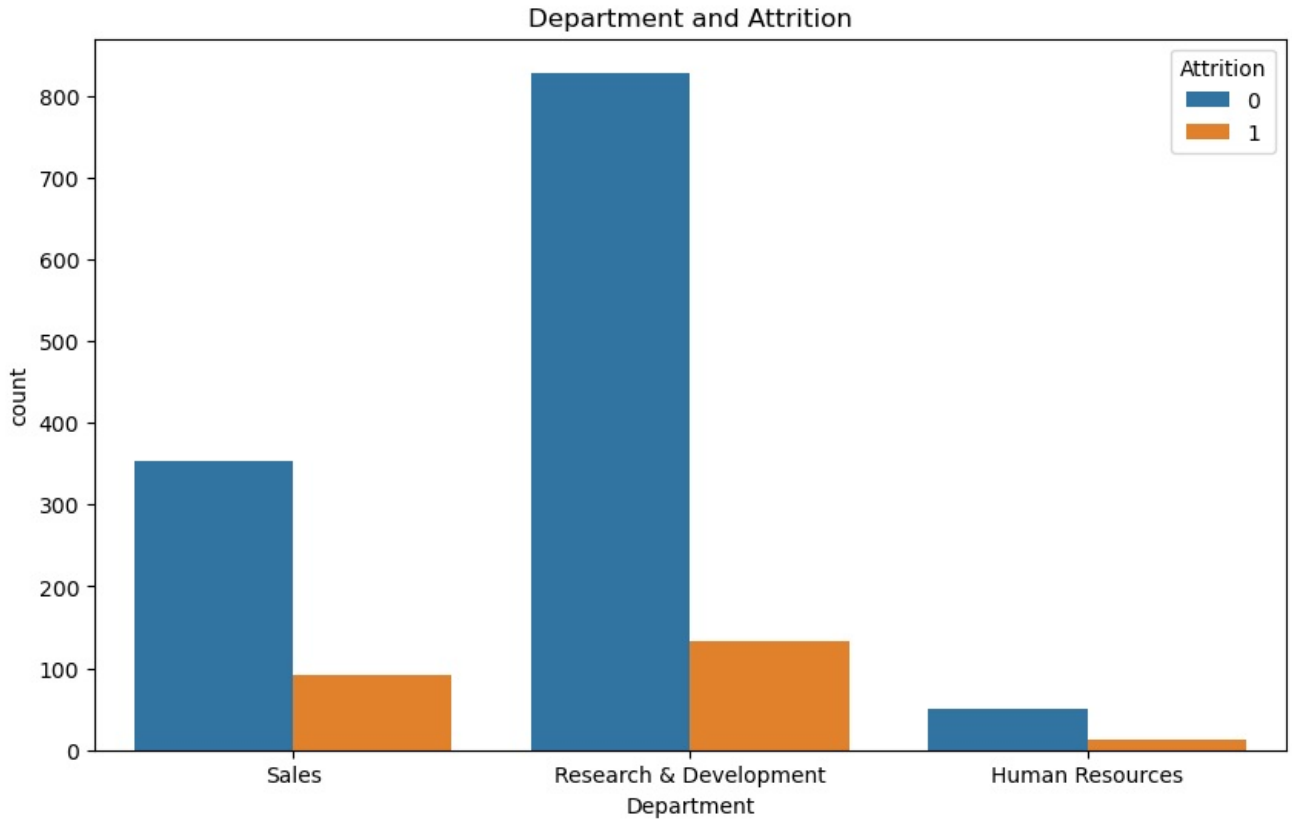
11.DISTANCE FROM HOME AND ATTRITION;

In [35]:
```
#Boxplot for distance from home and attrition;
plt.figure(figsize=(10,6))
sns.boxplot(x='Attrition', y='DistanceFromHome', data=df)
plt.title('Distance from Home and Attrition(boxplot)')
plt.xlabel('Attrition')
plt.ylabel('Distance from Home(km)')
plt.show()
```



12.DEPARTMENT AND ATTRITION

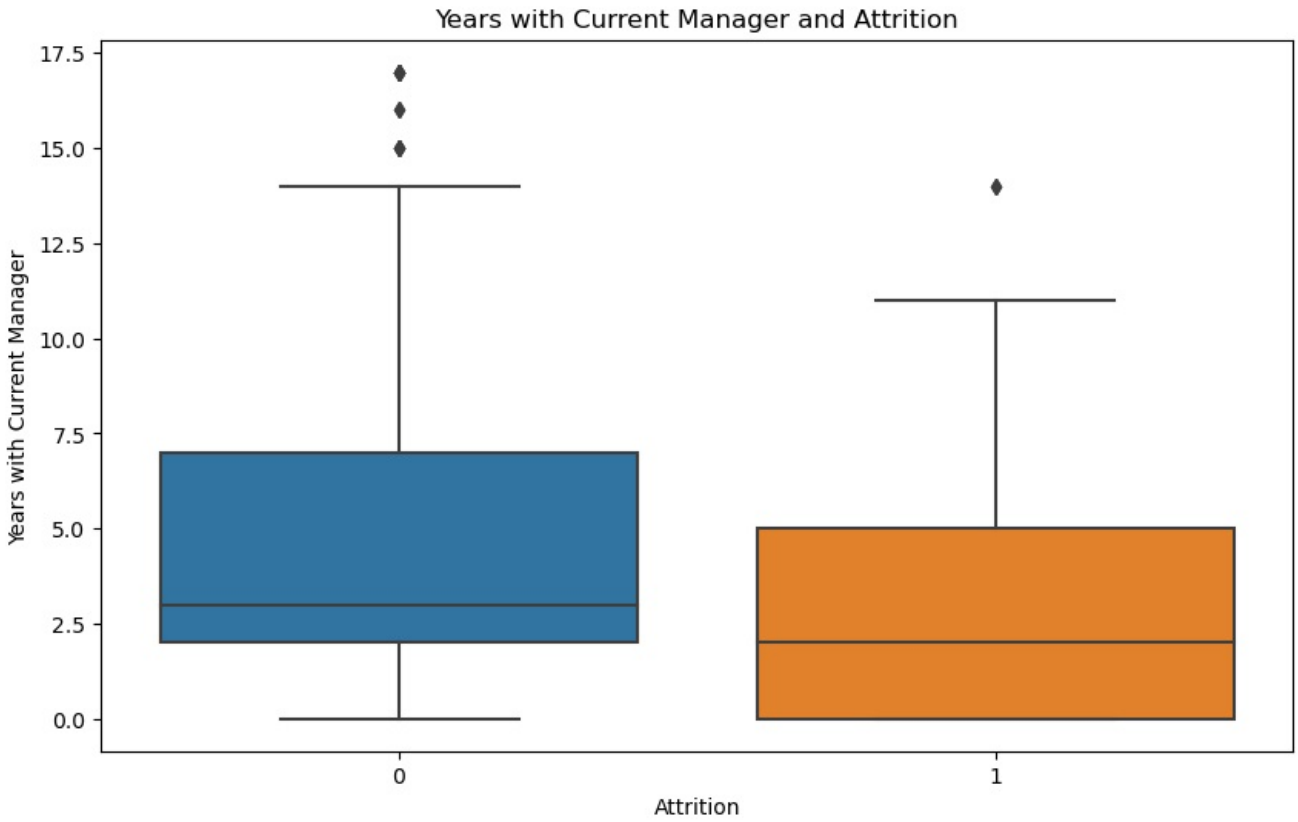
In [36]:
```
#Bar chart for department and attrition;
plt.figure(figsize=(10,6))
sns.countplot(x='Department',hue='Attrition', data=df)
plt.xlabel('Department')
plt.ylabel('count')
plt.title('Department and Attrition')
plt.show()
```



13.YEARS WITH CURRENT MANAGER AND ATTRITION;

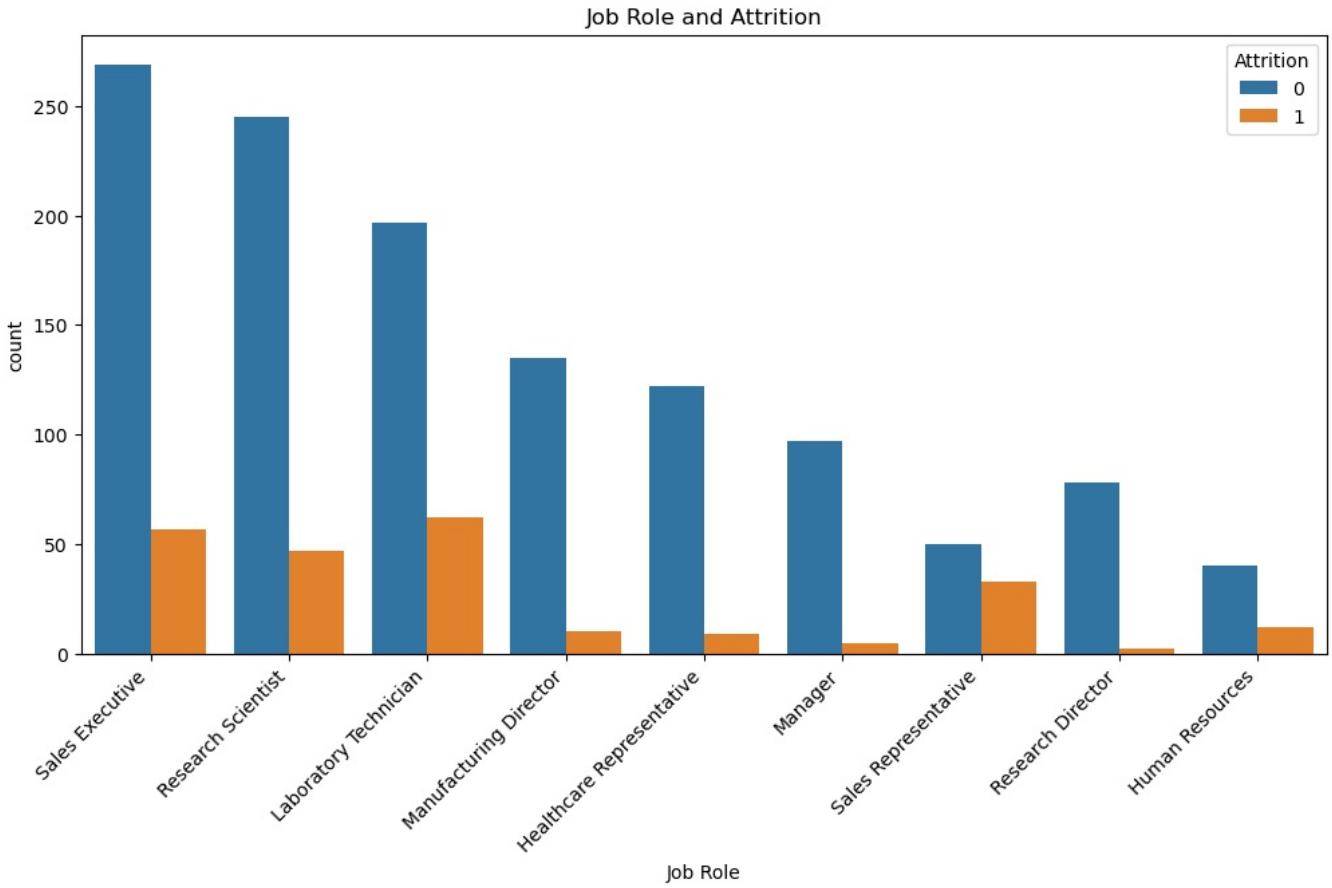
In [37]:
```
#Boxplot for year with current manager and attrition;
plt.figure(figsize=(10,6))
sns.boxplot(x='Attrition',y='YearsWithCurrManager', data=df)
```

```
plt.xlabel('Attrition')
plt.ylabel('Years with Current Manager')
plt.title('Years with Current Manager and Attrition')
plt.show()
```



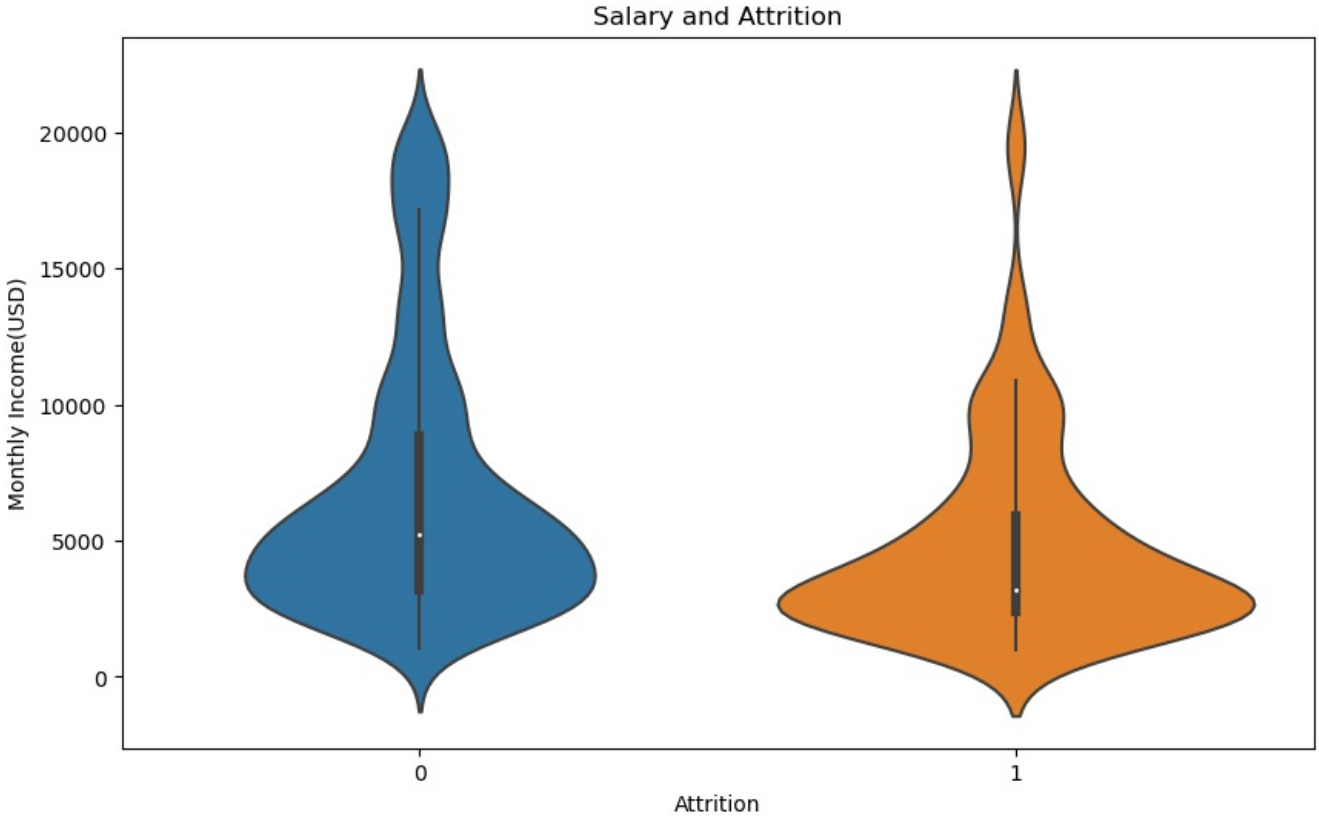
14.JOB ROLE AND ATTRITION

In [38]:
```
#barchart for job role and attrition;
plt.figure(figsize=(12,6))
sns.countplot(x='JobRole',hue='Attrition',data=df)
plt.xlabel('Job Role')
plt.ylabel('count')
plt.title('Job Role and Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```



15.SALARY AND ATTRITION;

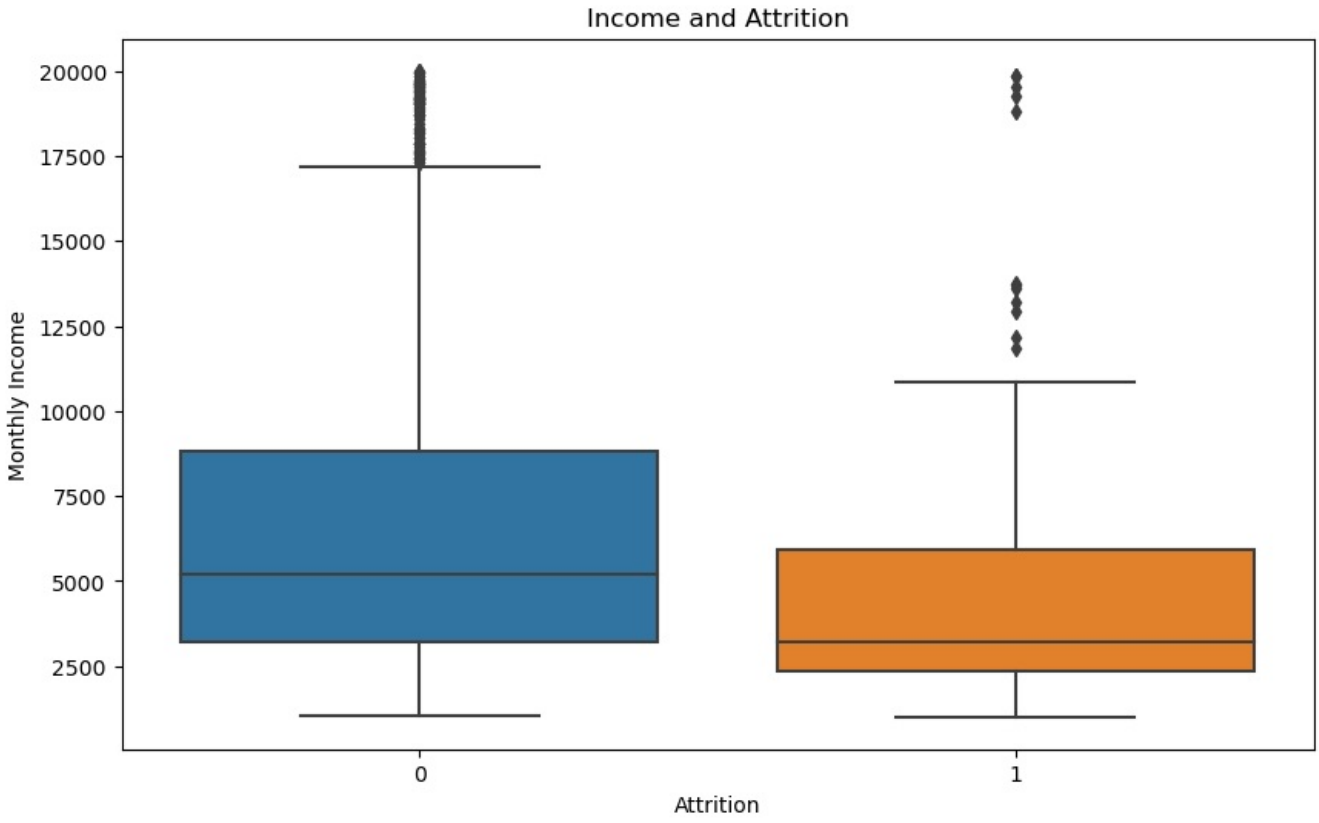
In [39]:
```
#Violinplot for salary and attrition:
plt.figure(figsize=(10,6))
sns.violinplot(x='Attrition', y='MonthlyIncome', data=df)
plt.xlabel('Attrition')
plt.ylabel('Monthly Income(USD)')
plt.title('Salary and Attrition')
plt.show()
```



16.INCOME AND ATTRITION:

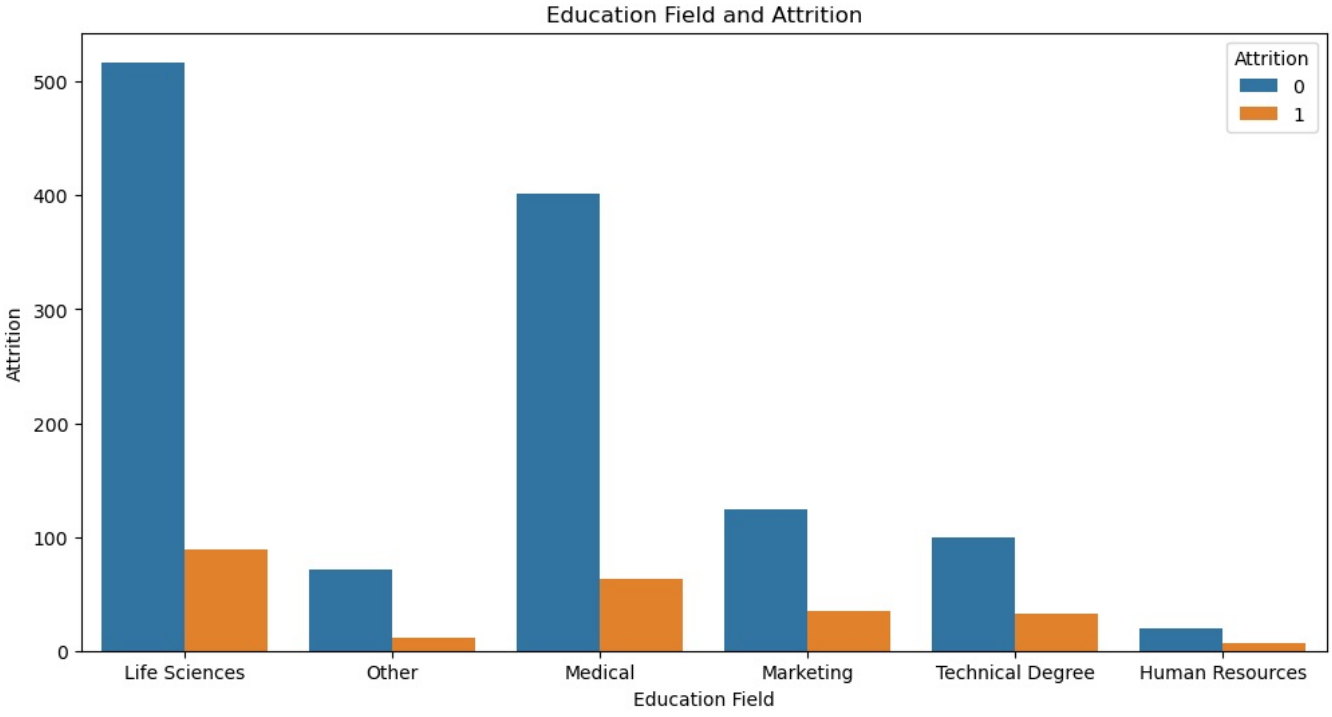
In [41]:
```
#Boxplot for income and attrition:
```

```
plt.figure(figsize=(10,6))
sns.boxplot(x='Attrition',y='MonthlyIncome', data=df)
plt.xlabel('Attrition')
plt.ylabel('Monthly Income')
plt.title('Income and Attrition')
plt.show()
```



17.EDUCATION FIELD AND ATTRITION:

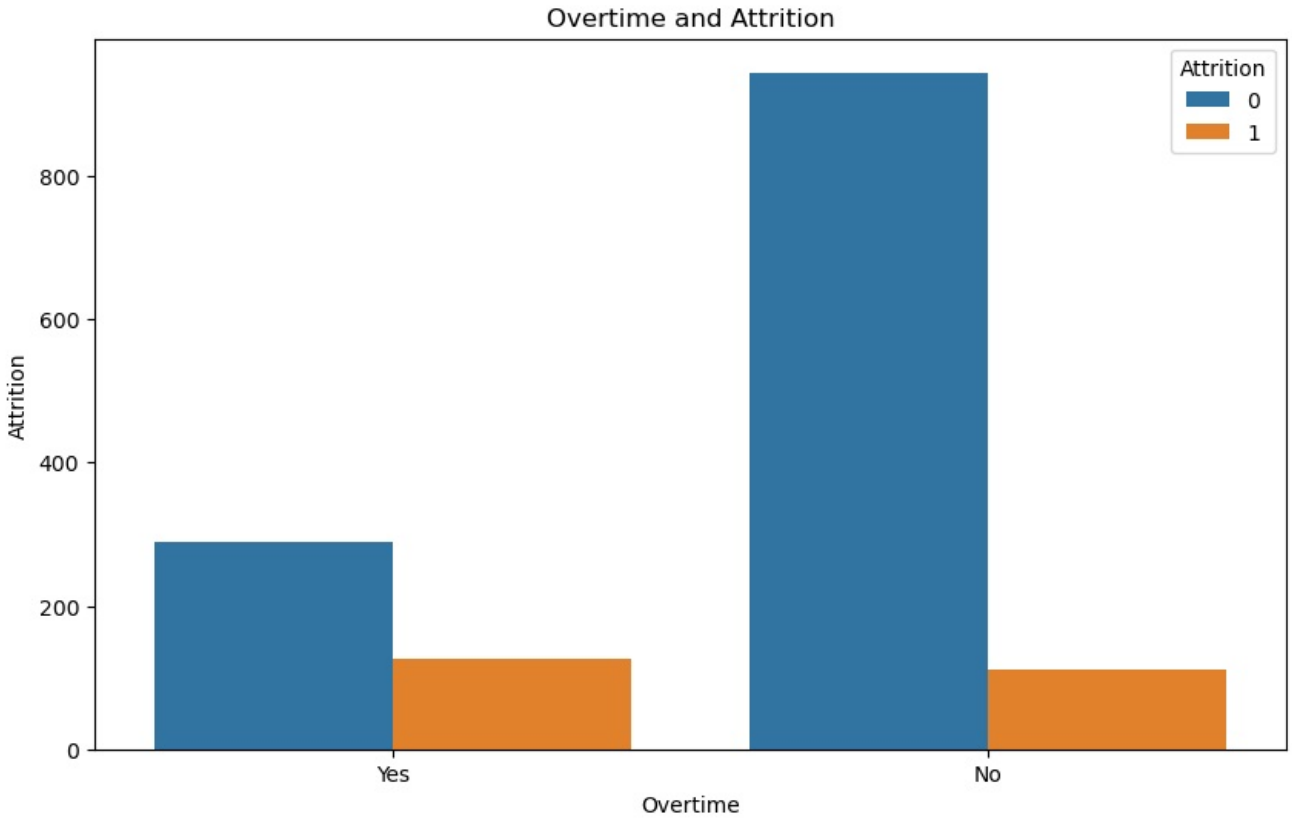
In [42]:
```
#Bar chart for education field and attrition:
plt.figure(figsize=(12,6))
sns.countplot(x='EducationField',hue='Attrition', data=df)
plt.xlabel('Education Field')
plt.ylabel('Attrition')
plt.title('Education Field and Attrition')
plt.show()
```



18.OVERTIME AND ATTRITION:

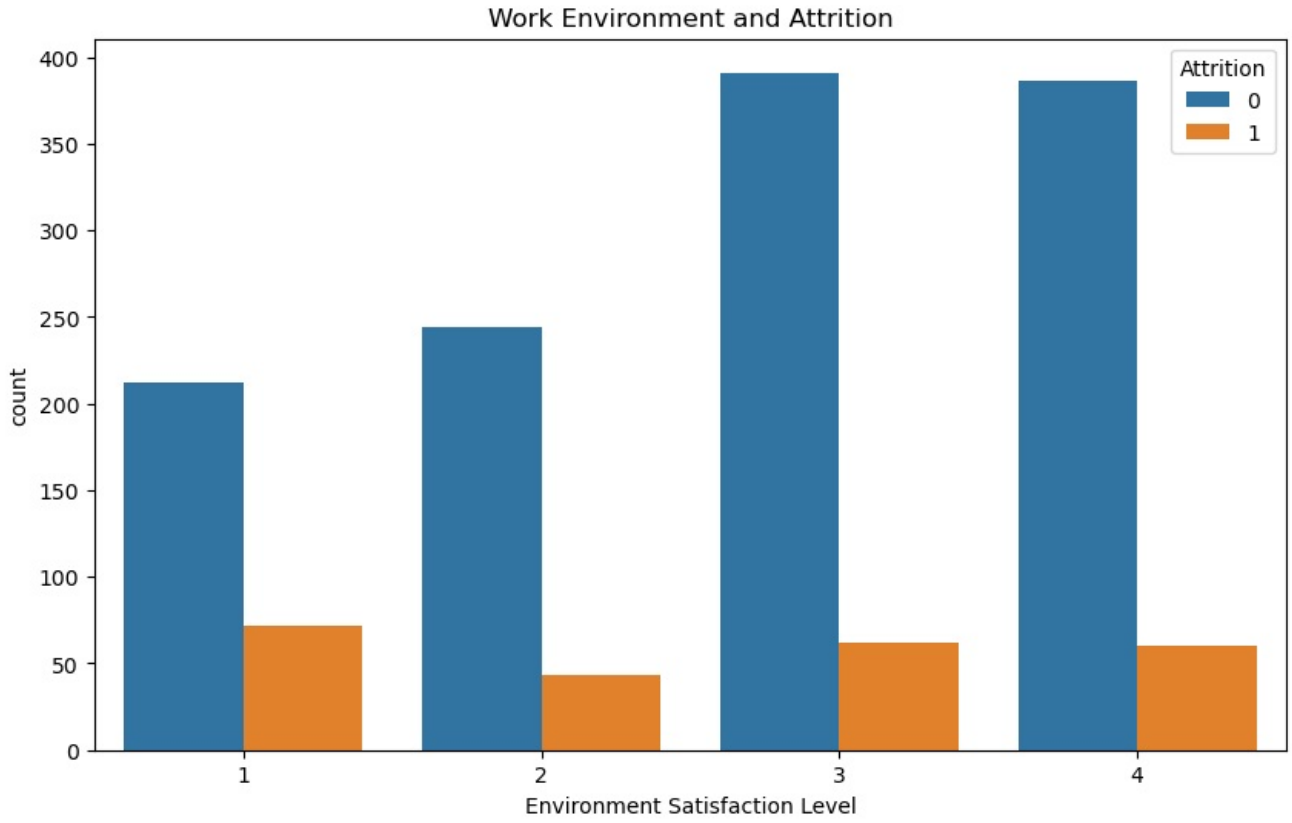
In [45]:
```
#Stacked bar chart for overtime and attrition:
```

```
plt.figure(figsize=(10,6))
sns.countplot(x='OverTime',hue='Attrition', data=df)
plt.xlabel('Overtime')
plt.ylabel('Attrition')
plt.title('Overtime and Attrition')
plt.show()
```



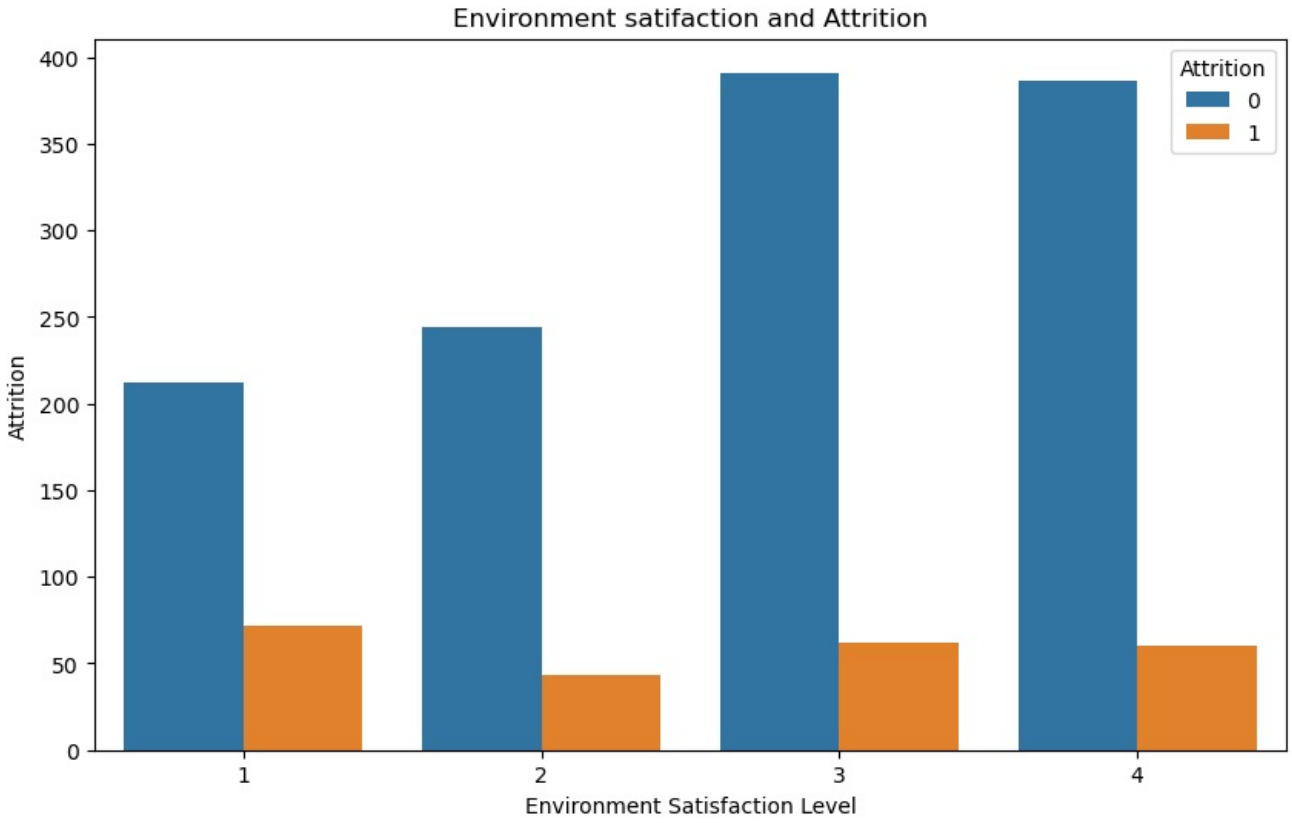
19.WORK EXPERIENCE AND ATTRITION:

In [54]:
```
#Bar chart for work experience and attrition:
plt.figure(figsize=(10,6))
sns.countplot(x='EnvironmentSatisfaction' ,hue='Attrition', data=df)
plt.xlabel('Environment Satisfaction Level')
plt.ylabel('count')
plt.title('Work Environment and Attrition')
plt.show()
```



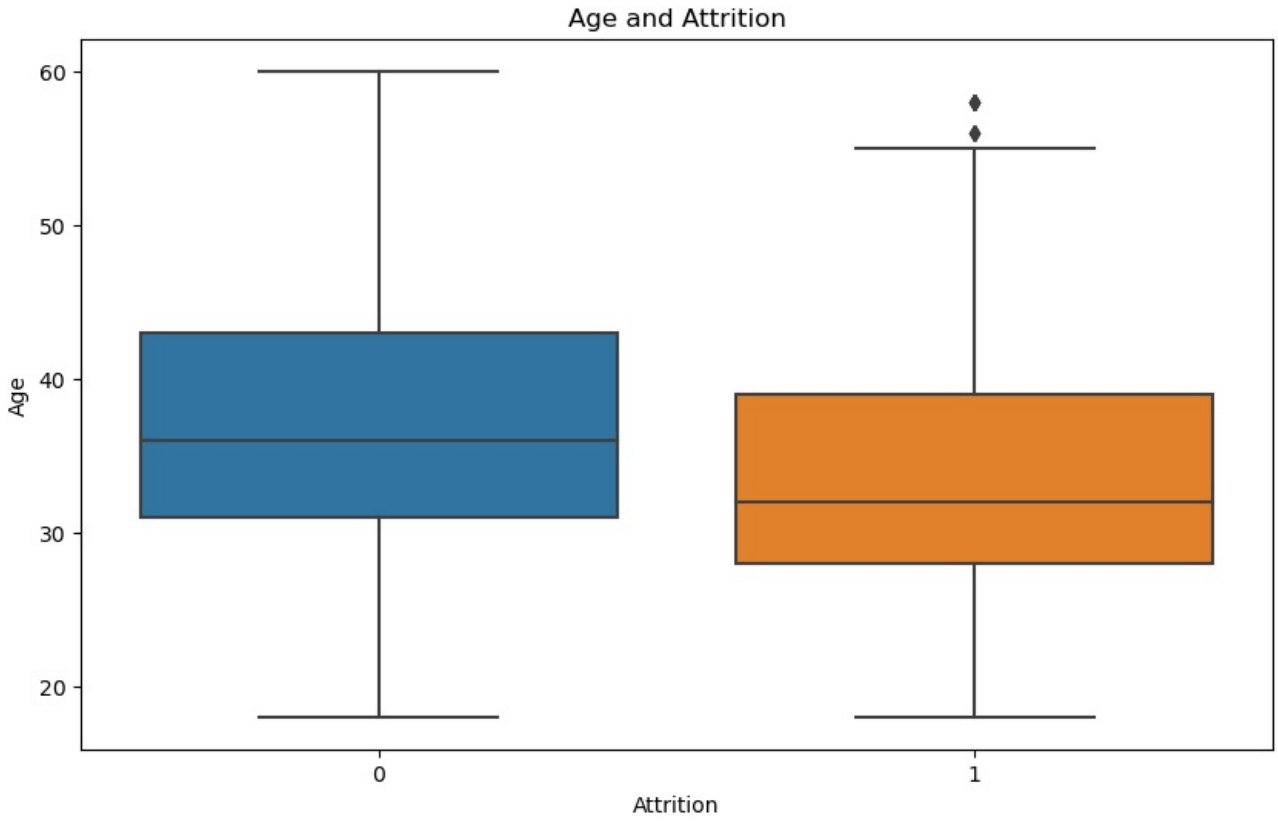
20.ENVIRONMENT SATISFACTION AND ATTRITION:

In [55]:
```
#Bar chart for environment satisfaction and attrition:
plt.figure(figsize=(10,6))
sns.countplot(x='EnvironmentSatisfaction',hue='Attrition', data=df)
plt.xlabel('Environment Satisfaction Level')
plt.ylabel('Attrition')
plt.title('Environment satifaction and Attrition')
plt.show()
```



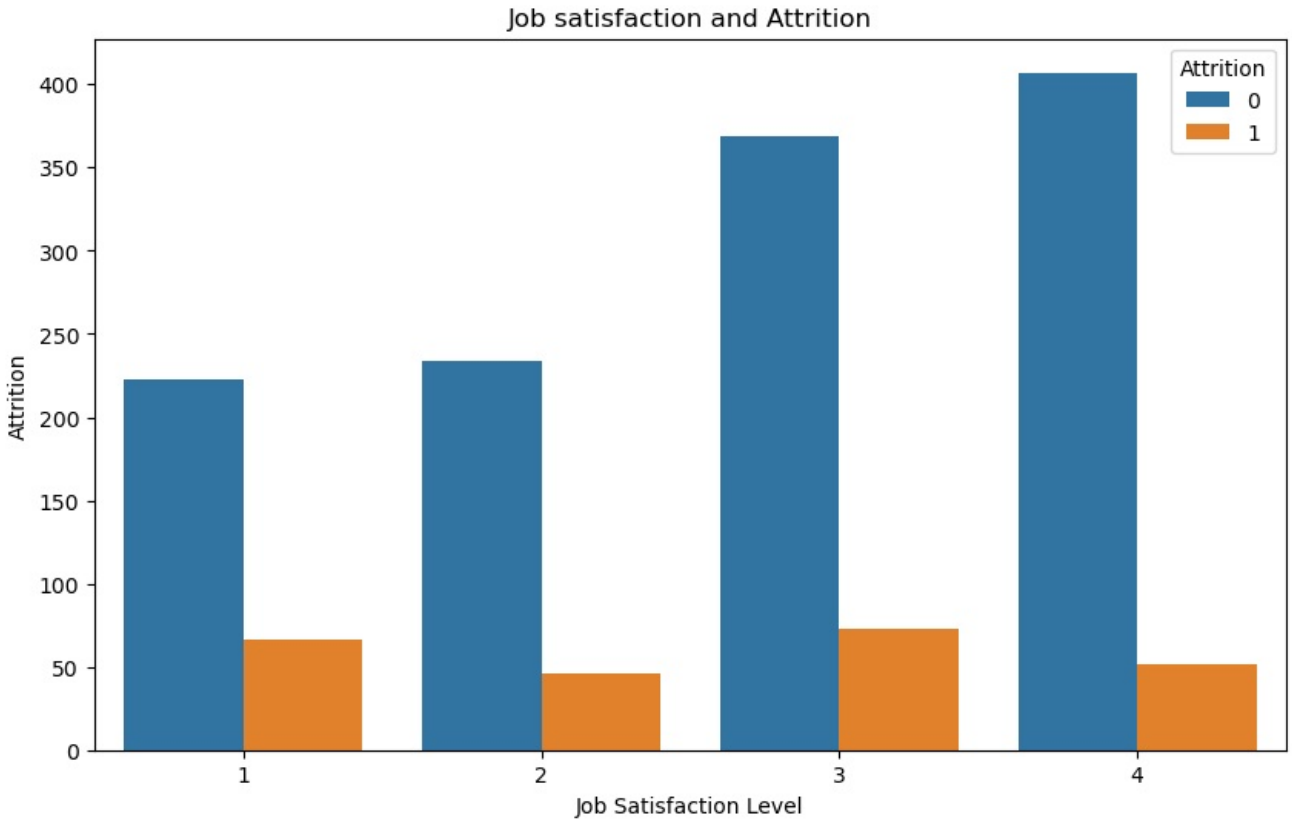
21.AGE AND ATTRITION:

In [57]:
```
#Box plot for age and attrition:
plt.figure(figsize=(10,6))
sns.boxplot(x='Attrition',y='Age', data=df)
plt.xlabel('Attrition')
plt.ylabel('Age')
plt.title('Age and Attrition')
plt.show()
```



22.SELF- JOB SATISFACTION AND ATTRITION:

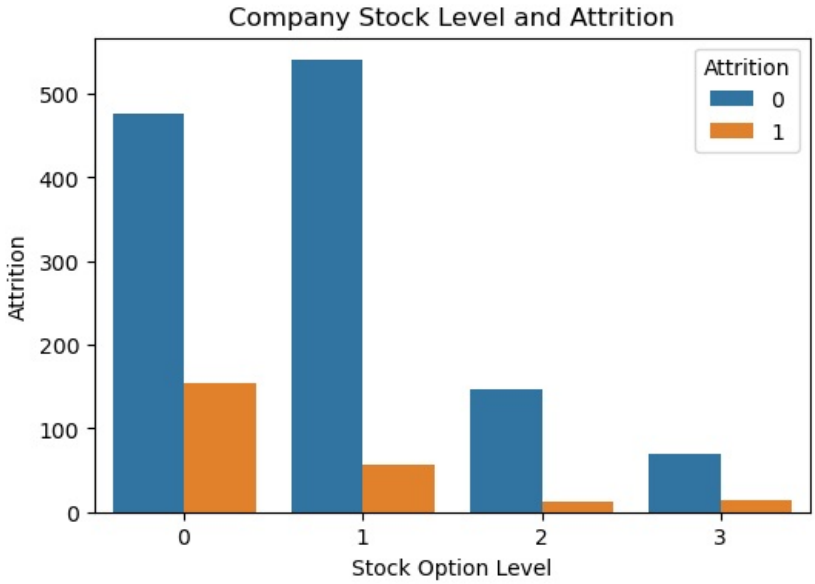
In [59]:
```
#Group bar chart for self-job satisfaction and attrition:
plt.figure(figsize=(10,6))
sns.countplot(x='JobSatisfaction',hue='Attrition', data=df)
plt.xlabel('Job Satisfaction Level')
plt.ylabel('Attrition')
plt.title('Job satisfaction and Attrition')
plt.show()
```



23.COMPANY STOCK AND ATTRITION:

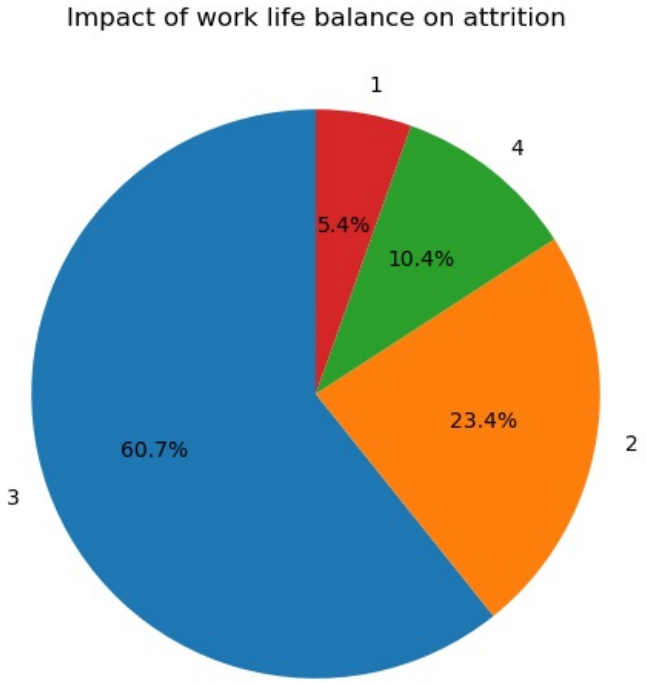
In [65]:
```
#Bar chart for company stock and attrition:
plt.figure(figsize=(6,4))
sns.countplot(x='StockOptionLevel',hue='Attrition', data=df)
```

```
plt.xlabel('Stock Option Level')
plt.ylabel('Attrition')
plt.title('Company Stock Level and Attrition')
plt.show()
```



24.WORK LIFE BALANCE AND ATTRITION:

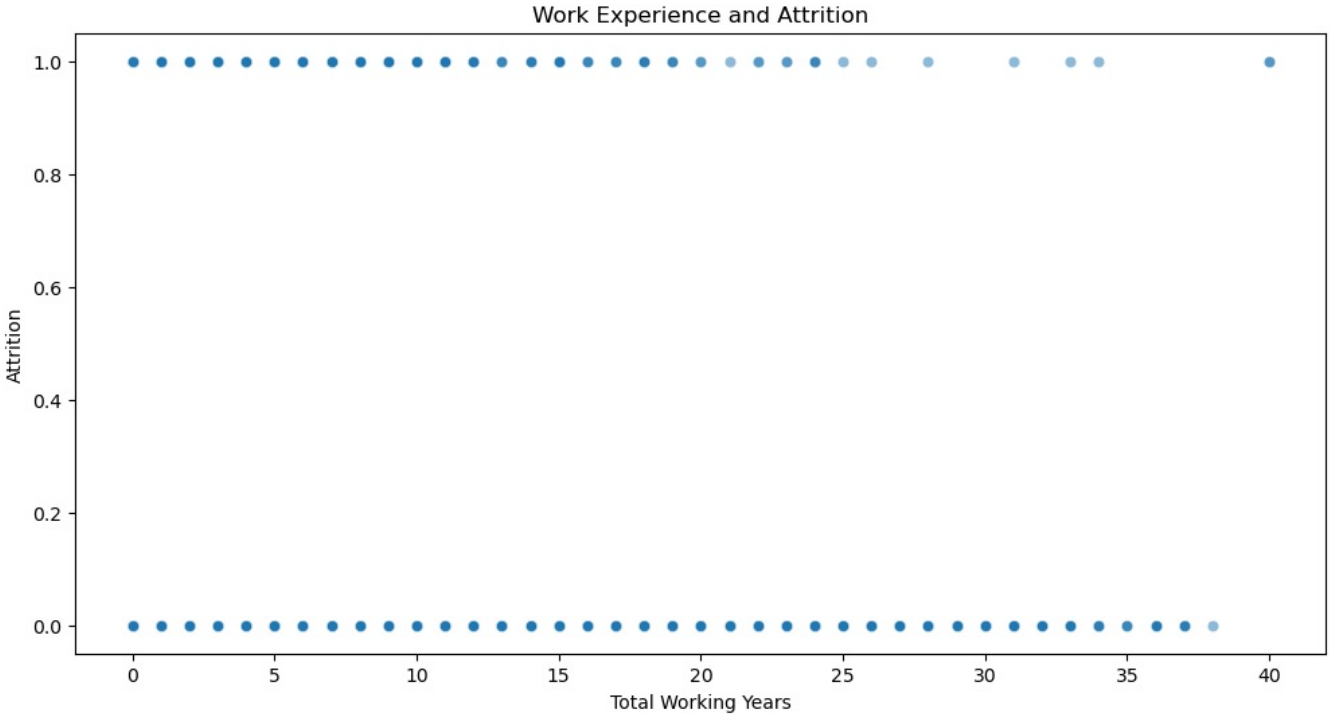
In [69]:
```
#Pie chart for worklife balance and attrition:
plt.figure(figsize=(6,6))
work_life_balance_distribution= df['WorkLifeBalance'].value_counts()
plt.pie(work_life_balance_distribution,labels=work_life_balance_distribution.index, autopct='%1.1f%%', startang
plt.title('Impact of work life balance on attrition')
plt.show()
```



25.WORK EXPERIENCE AND ATTRITION:

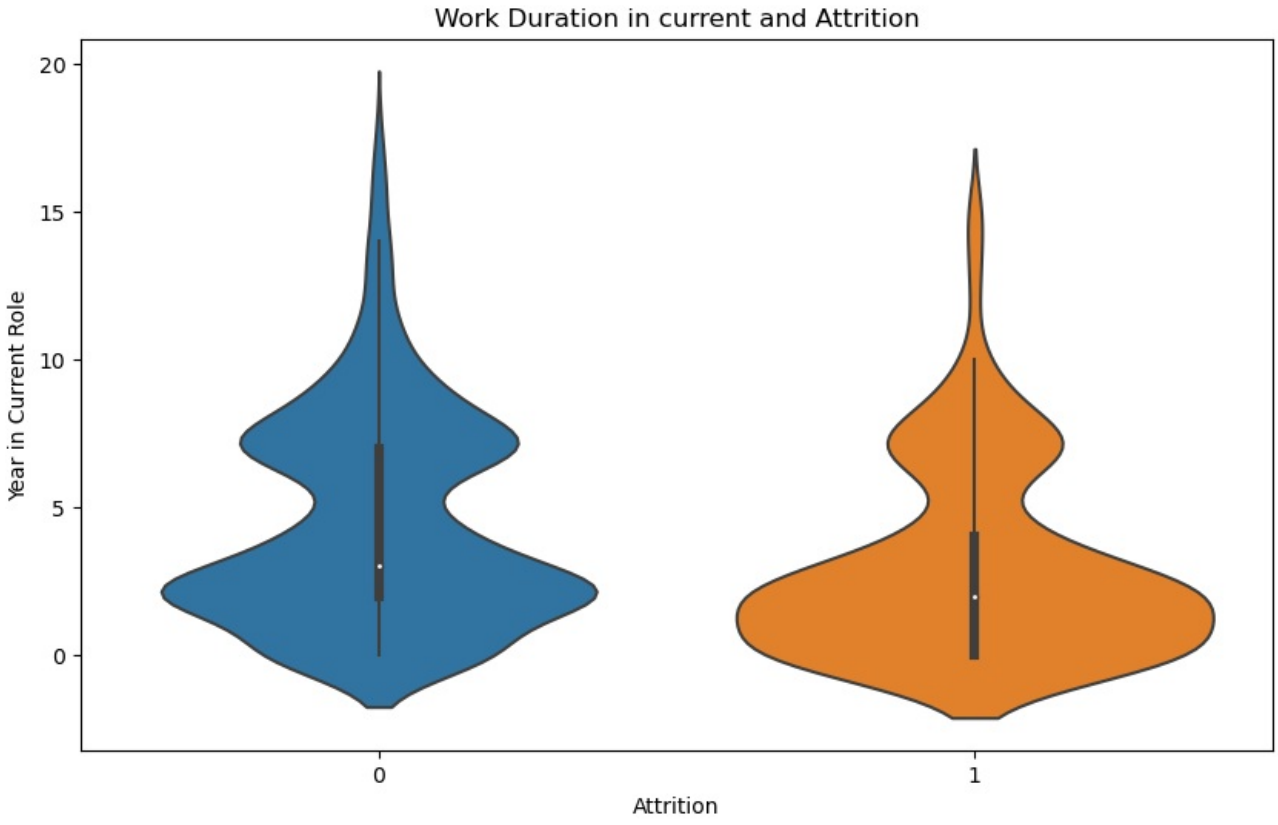
#scatter plot for work experience and attrition:

In [72]:
```
plt.figure(figsize=(12,6))
sns.scatterplot(x='TotalWorkingYears',y='Attrition', data=df,alpha=0.5)
plt.xlabel('Total Working Years')
plt.ylabel('Attrition')
plt.title('Work Experience and Attrition')
plt.show()
```



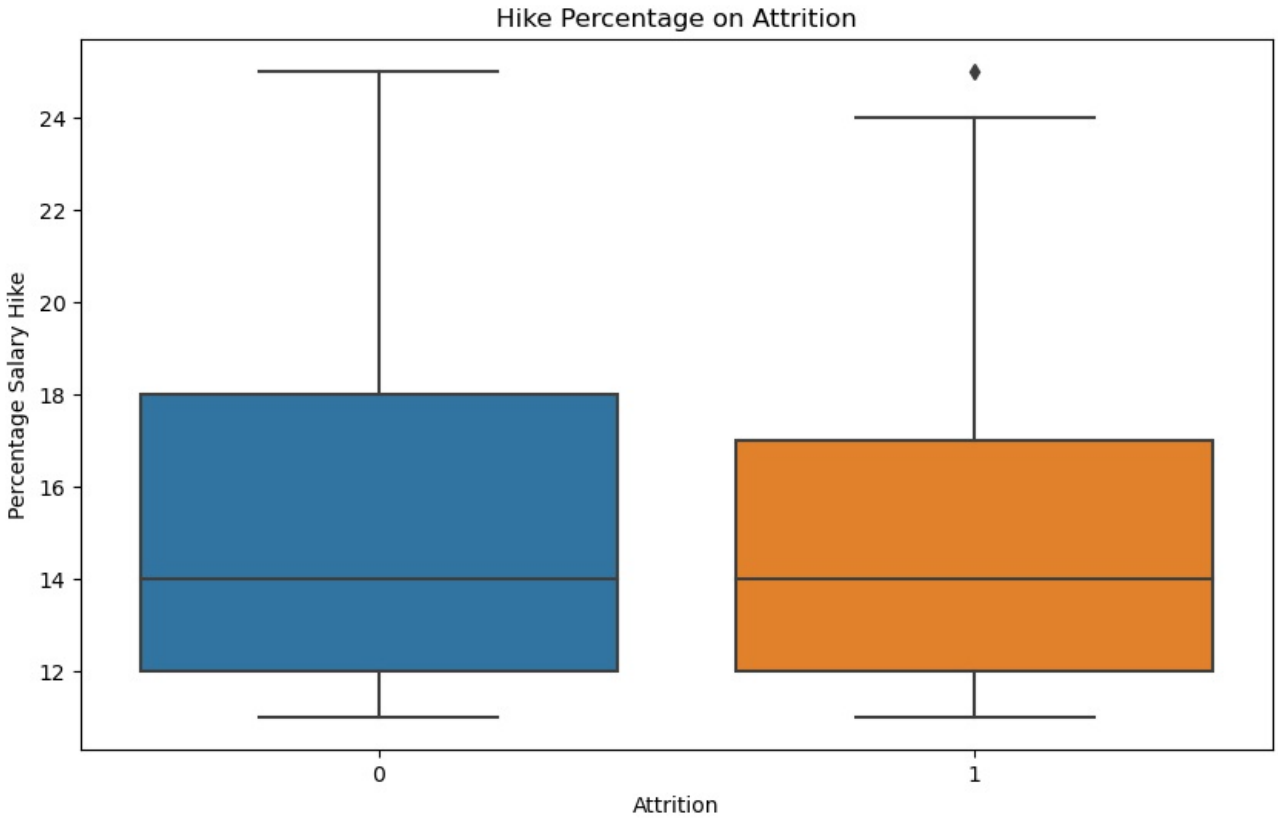
26.WORK DURATION IN CURRENT ROLE AND ATTRITION:

In [76]:
```
#Violin plot for work duration in current and attrition:
plt.figure(figsize=(10,6))
sns.violinplot(x='Attrition',y='YearsInCurrentRole', data=df)
plt.xlabel('Attrition')
plt.ylabel('Year in Current Role')
plt.title('Work Duration in current and Attrition')
plt.show()
```



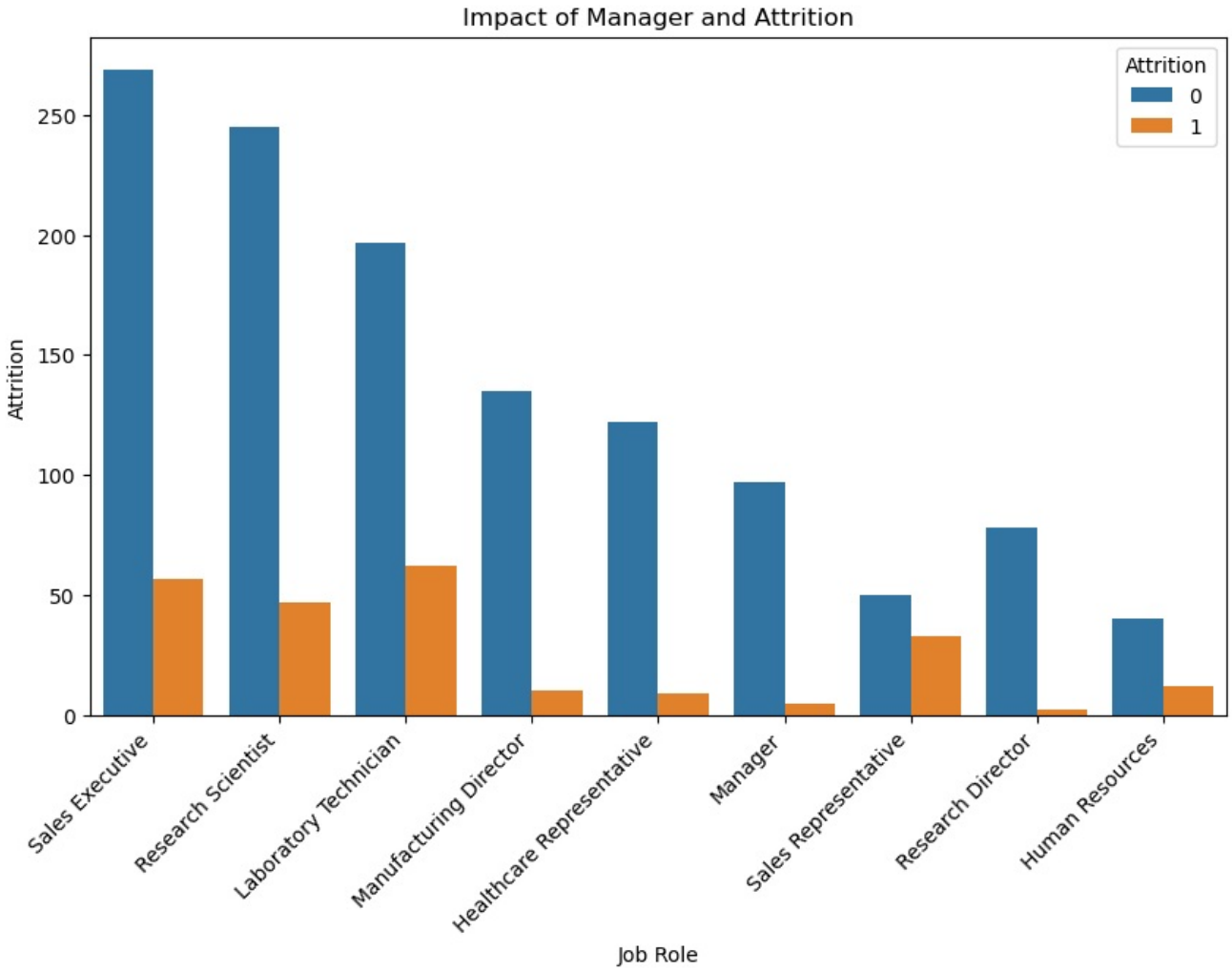
27.HIKE PERCENTAGE AND ATTRITION:

In [78]:
```
#Box plot for hike percentage and attrition:
plt.figure(figsize=(10,6))
sns.boxplot(x='Attrition',y='PercentSalaryHike', data=df)
plt.xlabel('Attrition')
plt.ylabel('Percentage Salary Hike')
plt.title('Hike Percentage on Attrition')
plt.show()
```



28.MANAGER AND ATTRITION:

In [80]:
```
#Bar chart for manager and attrition:
plt.figure(figsize=(10,6))
sns.countplot(x='JobRole',hue='Attrition', data=df)
plt.xlabel('Job Role')
plt.ylabel('Attrition')
plt.title('Impact of Manager and Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```



In [ ]:

In [ ]:

In [ ]: