"Exploring Factors Contributing to Employee Attrition: An EDA on a Fictional IBM Dataset"

The dataset is a fictional dataset created by IBM data scientists to uncover the factors that lead to employee attrition. The dataset contains various factors such as employee age, job role, education level, monthly income, distance from home, and many more. The dataset contains a total of 1470 observations with 35 features.

The main objective of the dataset is to explore important questions related to employee attrition, such as breakdown of distance from home by job role and attrition, and comparison of average monthly income by education and attrition.

To achieve this objective, we conducted exploratory data analysis (EDA) on the dataset. We started by checking for missing values and found that there were no missing values in the dataset. We then performed descriptive statistics to understand the distribution of various features.

Next, we created various visualizations such as histograms, box plots, and scatter plots to visualize the relationship between different features and employee attrition. We found that job role, education level, monthly income, and distance from home were some of the most significant factors that contributed to employee attrition.

We also conducted a correlation analysis and found that there was a strong positive correlation between monthly income and job level, indicating that employees with higher job levels tended to have higher monthly incomes.

In conclusion, this dataset provides valuable insights into the factors that contribute to employee attrition. By analyzing and visualizing the data, we can gain a deeper understanding of the relationship between different factors and employee attrition, which can help organizations make informed decisions to reduce attrition rates and improve employee retention.

importing librarys

In [1]:
```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

import warnings
warnings.filterwarnings('ignore')
```

importing dataset

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

In [3]:
```
ibm
```

Out[3]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Ec
0	41	Yes	Travel_Rarely	1102	Sales	1	2	
1	49	No	Travel_Frequently	279	Research & Development	8	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	
4	27	No	Travel_Rarely	591	Research & Development	2	1	
...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	
1468	49	No	Travel_Frequently	1023	Sales	2	3	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	

1470 rows × 35 columns

Dataset info

In [4]:
```
ibm.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
```

Null values in dataset

In [5]:
```
ibm.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
```

Exploratory Data Analysis (EDA)

we have 8 object columns and remaining are int columns

'Over18', 'OverTime', 'JobRole', 'MaritalStatus', 'Gender', 'EducationField', 'Attrition', 'BusinessTravel', 'Department'

In [ ]:
```
```

In [6]:
```
ibm.duplicated().sum()
```

Out[6]:
```
0
```

In [7]:
```
ibm.nunique()
```

Out[7]:

Age	43
Attrition	2
BusinessTravel	3
DailyRate	886
Department	3
DistanceFromHome	29
Education	5
EducationField	6
EmployeeCount	1
EmployeeNumber	1470
EnvironmentSatisfaction	4
Gender	2
HourlyRate	71
JobInvolvement	4
JobLevel	5
JobRole	9
JobSatisfaction	4
MaritalStatus	3
MonthlyIncome	1349
MonthlyRate	1427
NumCompaniesWorked	10
Over18	1
OverTime	2
PercentSalaryHike	15
PerformanceRating	2
RelationshipSatisfaction	4
StandardHours	1
StockOptionLevel	4
TotalWorkingYears	40
TrainingTimesLastYear	7
WorkLifeBalance	4
YearsAtCompany	37
YearsInCurrentRole	19
YearsSinceLastPromotion	16
YearsWithCurrManager	18

dtype: int64

In [8]:
```
ibm
```

Out[8]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Ec
0	41	Yes	Travel_Rarely	1102	Sales	1	2	
1	49	No	Travel_Frequently	279	Research & Development	8	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	
4	27	No	Travel_Rarely	591	Research & Development	2	1	
...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	
1468	49	No	Travel_Frequently	1023	Sales	2	3	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	

1470 rows × 35 columns

In [9]:
```
ibm.iloc[:,10:-10]
```

Out[9]:

	EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	Job
0	2	Female	94	3	2	Sales Executive	
1	3	Male	61	2	2	Research Scientist	
2	4	Male	92	2	1	Laboratory Technician	
3	4	Female	56	3	1	Research Scientist	
4	1	Male	40	3	1	Laboratory Technician	
...	...	...	...	...	...	...	
1465	3	Male	41	4	2	Laboratory Technician	
1466	4	Male	42	2	3	Healthcare Representative	
1467	2	Male	87	4	2	Manufacturing Director	
1468	4	Male	63	2	2	Sales Executive	
1469	2	Male	82	4	2	Laboratory Technician	

1470 rows × 15 columns

-

EmployeeCount have employee count and its '1' for all , Over18 is about employer is above 18 or not, obviously every employe is over 18 years only , 'Y' for all , StandardHours is same for all '80'

so we are droping them and EmployeeNumber column also droping

In [10]:
```
print(list(ibm))
```

```
['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department', 'DistanceFromHom
e', 'Education', 'EducationField', 'EmployeeCount', 'EmployeeNumber', 'Environment
Satisfaction', 'Gender', 'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobRole', 'J
obSatisfaction', 'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWor
ked', 'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating', 'Relationshi
pSatisfaction', 'StandardHours', 'StockOptionLevel', 'TotalWorkingYears', 'Trainin
gTimesLastYear', 'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole', 'Years
SinceLastPromotion', 'YearsWithCurrManager']
```

In [11]:
```
ibm['Gender'].value_counts()
```

Out[11]:
```
Male      882
Female    588
Name: Gender, dtype: int64
```

In [12]:
```
ibm.drop(['EmployeeCount','Over18','StandardHours','EmployeeNumber'],axis=1,inplac
```

In [13]:
```
ibm
```

Out[13]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Ec
0	41	Yes	Travel_Rarely	1102	Sales	1	2	
1	49	No	Travel_Frequently	279	Research & Development	8	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	
4	27	No	Travel_Rarely	591	Research & Development	2	1	
...	...	...	...	...	...	...	...	
1465	36	No	Travel_Frequently	884	Research & Development	23	2	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	
1468	49	No	Travel_Frequently	1023	Sales	2	3	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	

1470 rows × 31 columns

In [14]:
```
column =['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department', 'Distanc
```

In [15]:
```
for col in column:
    if ibm[col].dtype == 'object':
        print(col)
```

```
Attrition
BusinessTravel
Department
EducationField
Gender
JobRole
MaritalStatus
OverTime
```

In [16]:
```
ibm = ibm.reindex(columns=['Attrition','BusinessTravel','Department','EducationFie
```

now dataset is frist replaced with string columns then intiger columns

In [17]:
```
ibm
```

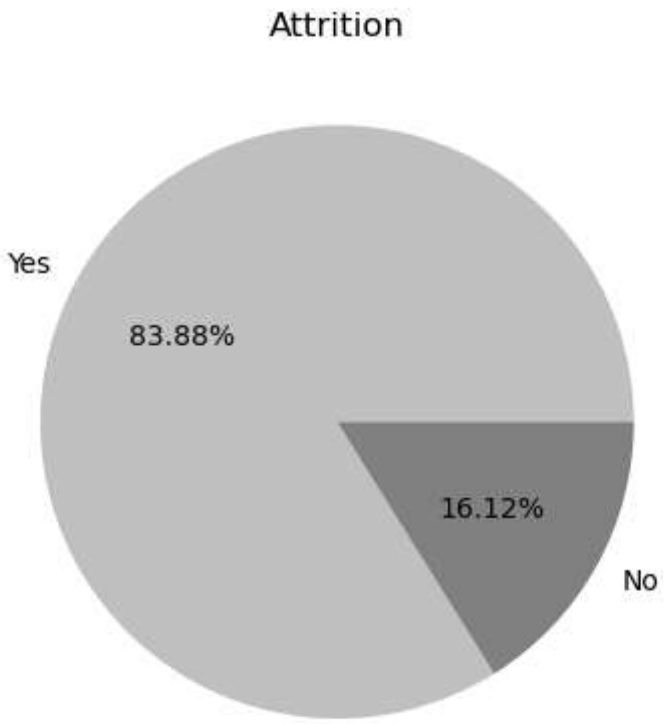
Out[17]:

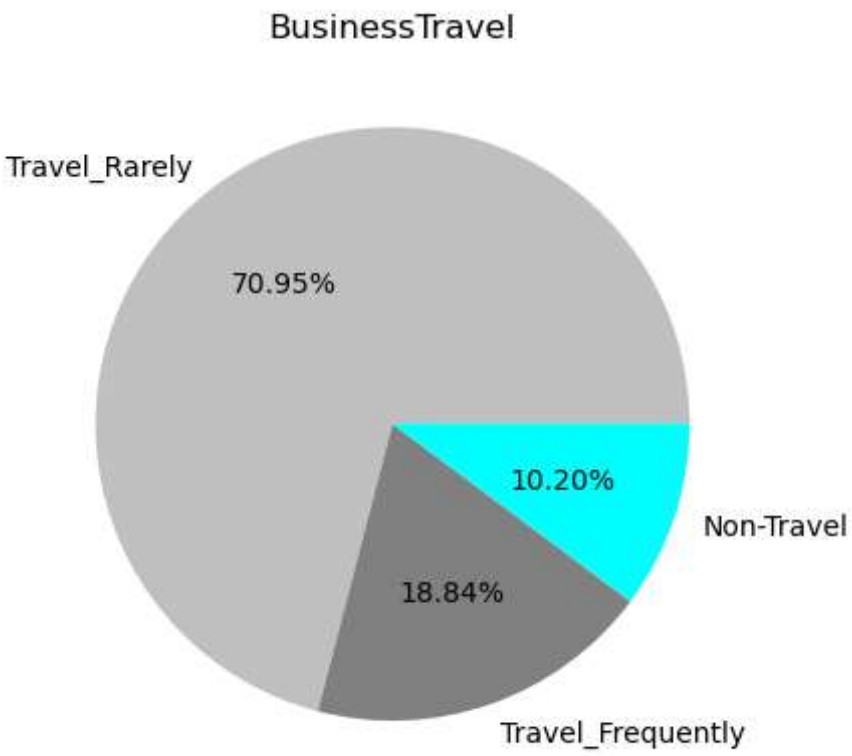
	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	MaritalStat
0	Yes	Travel_Rarely	Sales	Life Sciences	Female	Sales Executive	Sing
1	No	Travel_Frequently	Research & Development	Life Sciences	Male	Research Scientist	Marri
2	Yes	Travel_Rarely	Research & Development	Other	Male	Laboratory Technician	Sing
3	No	Travel_Frequently	Research & Development	Life Sciences	Female	Research Scientist	Marri
4	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri
...	...	...	...	...	...	...	
1465	No	Travel_Frequently	Research & Development	Medical	Male	Laboratory Technician	Marri
1466	No	Travel_Rarely	Research & Development	Medical	Male	Healthcare Representative	Marri
1467	No	Travel_Rarely	Research & Development	Life Sciences	Male	Manufacturing Director	Marri
1468	No	Travel_Frequently	Sales	Medical	Male	Sales Executive	Marri
1469	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri

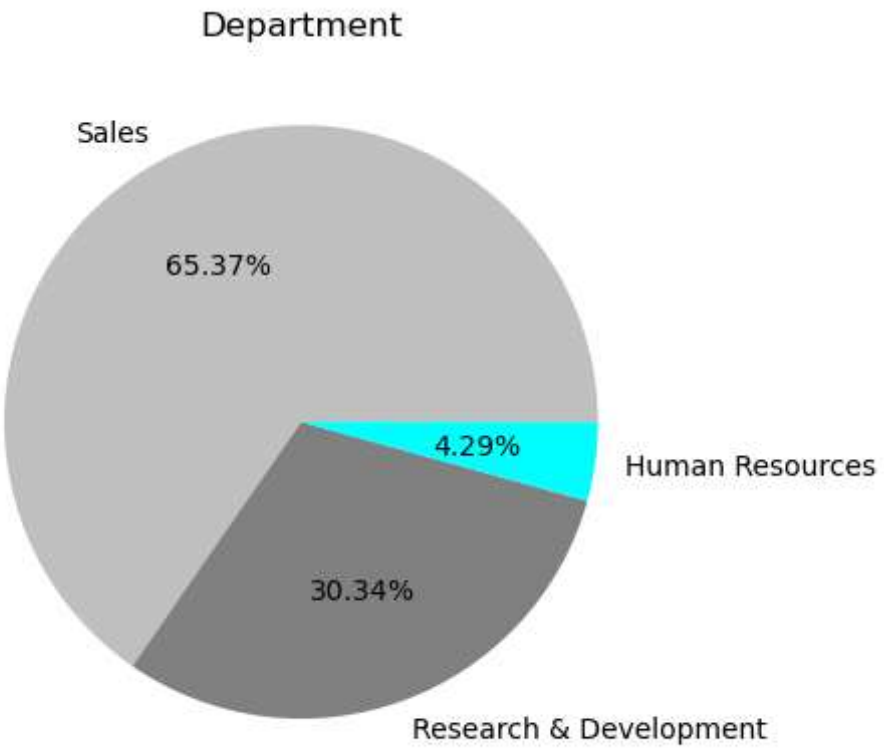
1470 rows × 31 columns

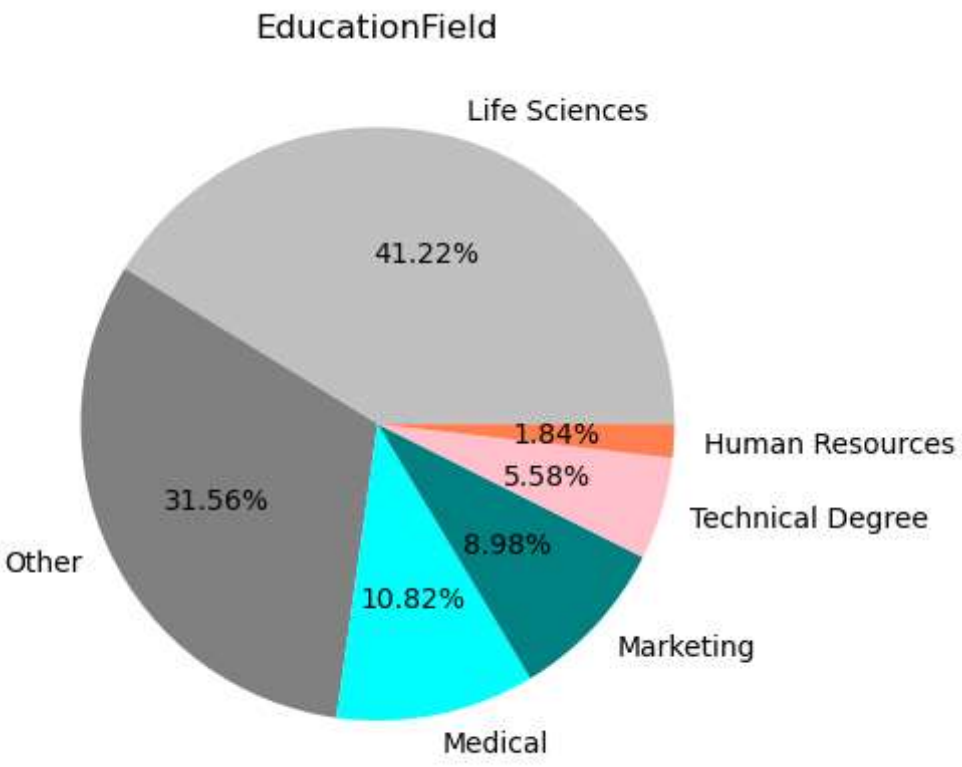
In [18]:
```
#colours=['magenta','red','white','gold','yellow','navy','maroon','indigo','purple
colours = ['silver','gray','cyan','teal','pink','coral','lavender','turquoise','ol
columns=['Attrition','BusinessTravel','Department','EducationField','Gender','JobR
```

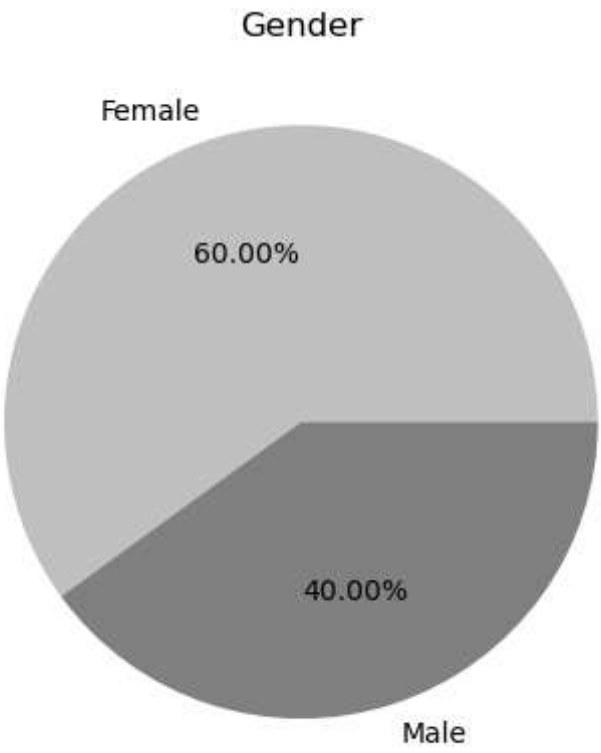
In [19]:
```
for col in columns:
    plt.pie(ibm[col].value_counts(),autopct='%0.2f%%',labels=ibm[col].unique(),col
    plt.title(col)
    plt.show()
```

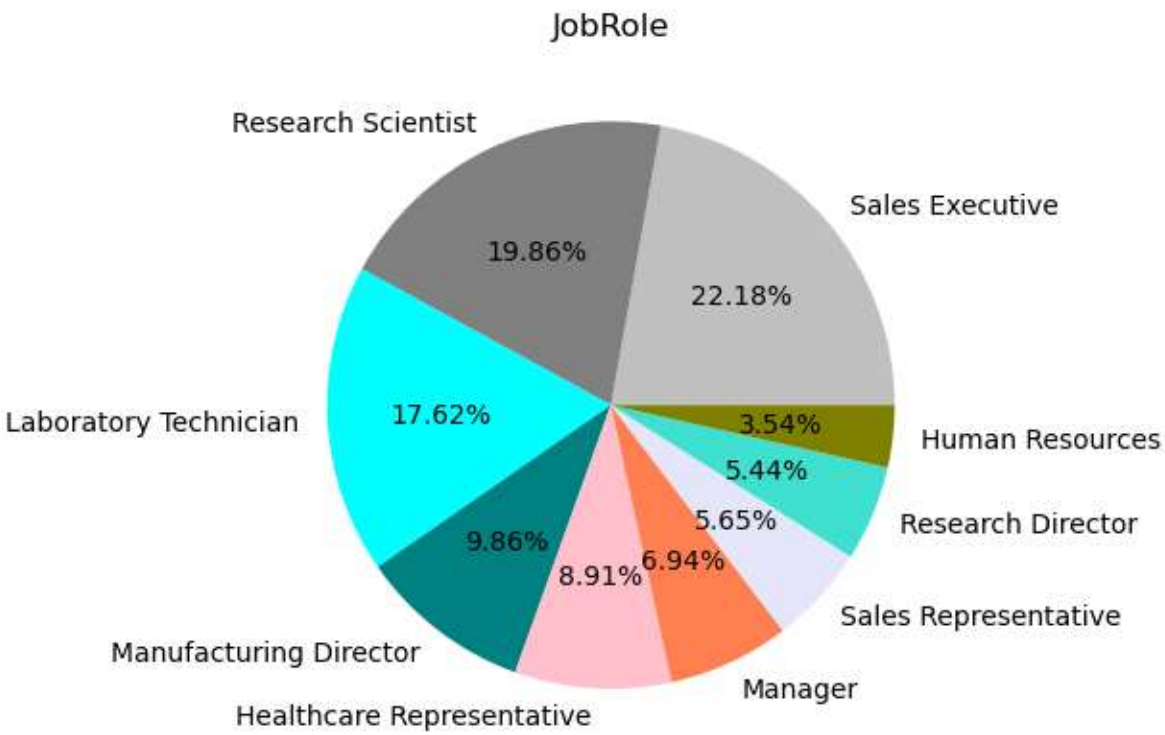

Attrition
Yes
83.88%
16.12%
No

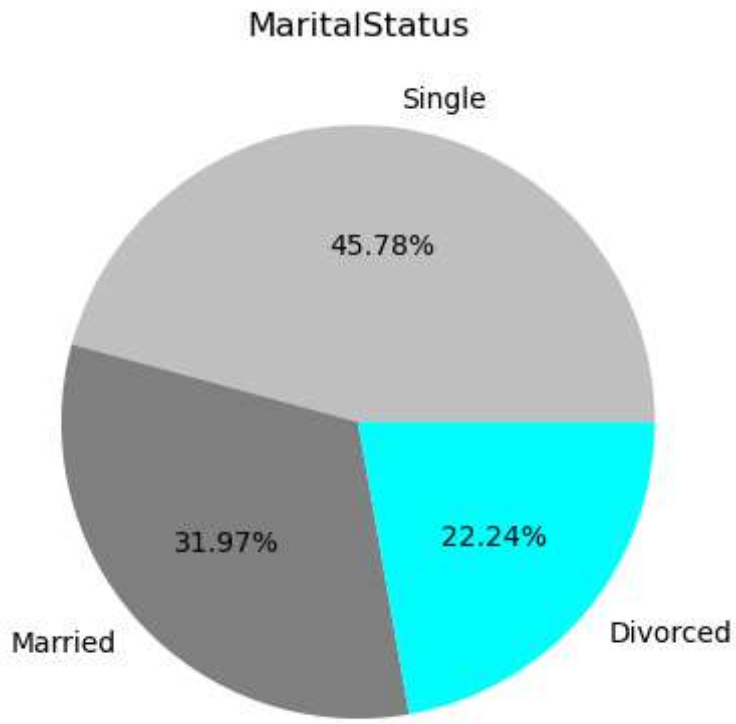

BusinessTravel
Travel_Rarely
70.95%
10.20%
Non-Travel
18.84%
Travel_Frequently

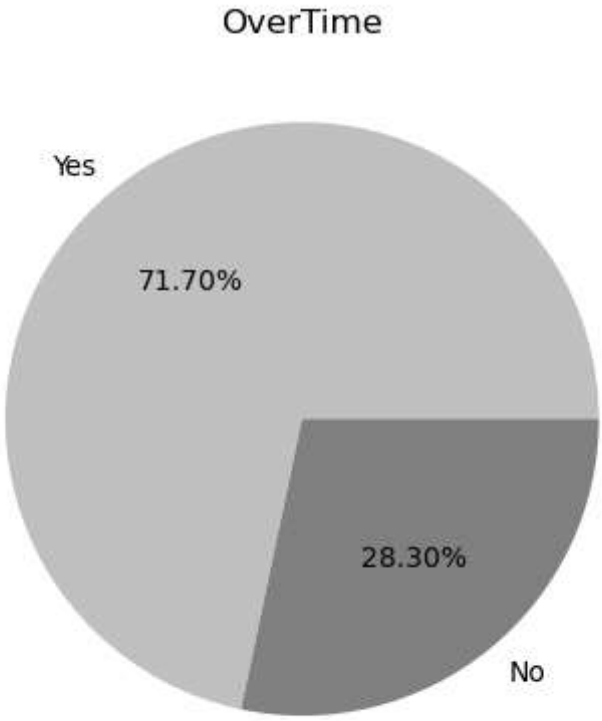

Department
Sales
65.37%
4.29%
Human Resources
30.34%
Research & Development

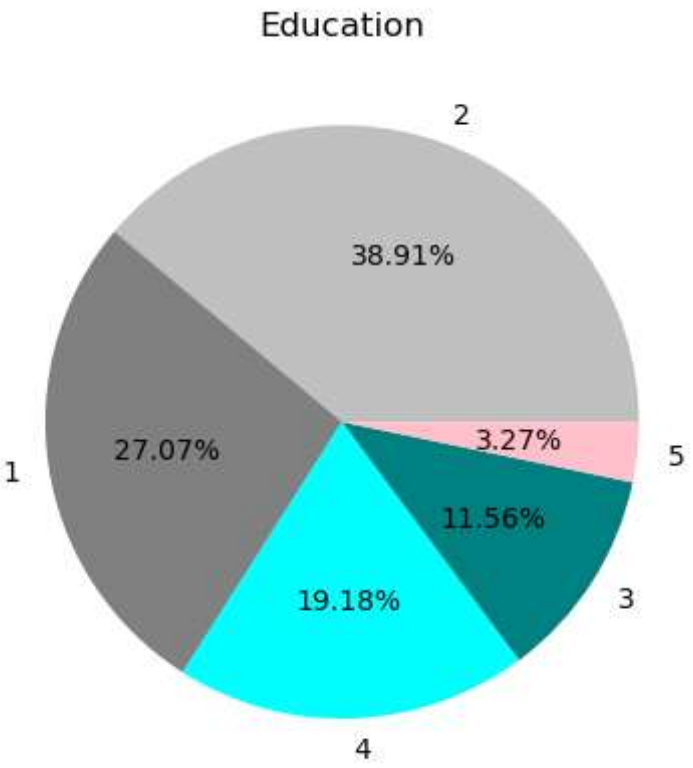

EducationField
Life Sciences
41.22%
1.84%
Human Resources
5.58%
Technical Degree
31.56%
Other
8.98%
10.82%
Marketing
Medical

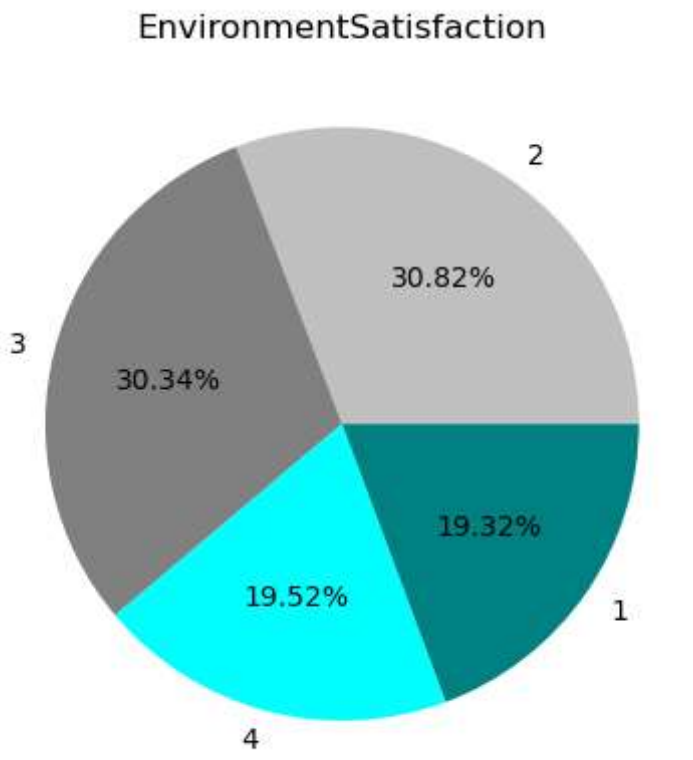

Gender
Female
60.00%
40.00%
Male

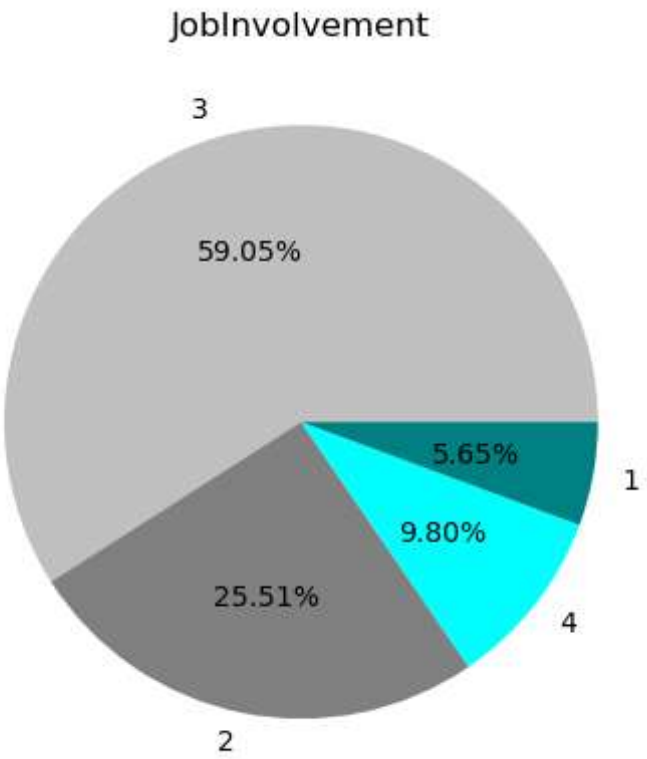

JobRole
Research Scientist
19.86%
Sales Executive
22.18%
Laboratory Technician
17.62%
3.54%
Human Resources
5.44%
Research Director
5.65%
9.86%
6.94%
8.91%
Sales Representative
Manufacturing Director
Manager
Healthcare Representative

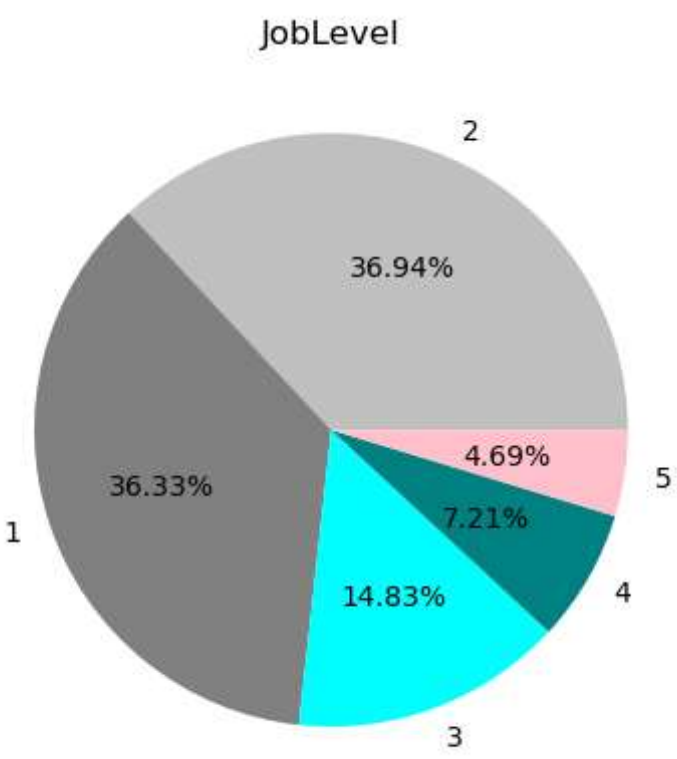

MaritalStatus
Single
45.78%
31.97%
22.24%
Married
Divorced

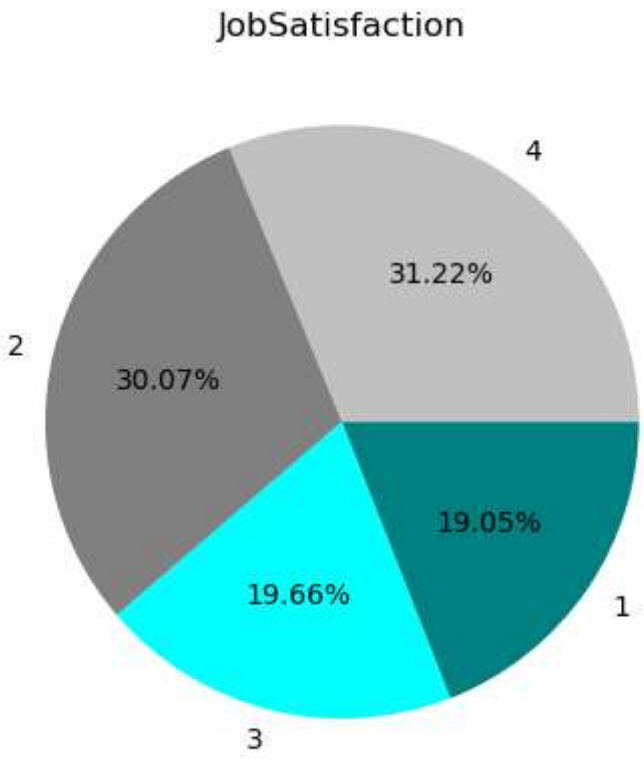

OverTime
Yes
71.70%
28.30%
No

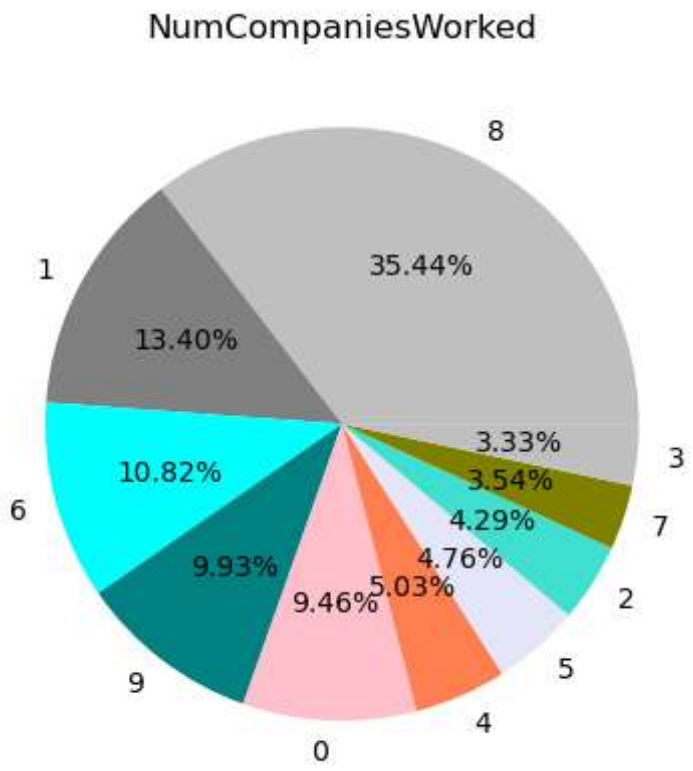

Education
2
38.91%
1
27.07%
3.27%
5
11.56%
3
19.18%
4

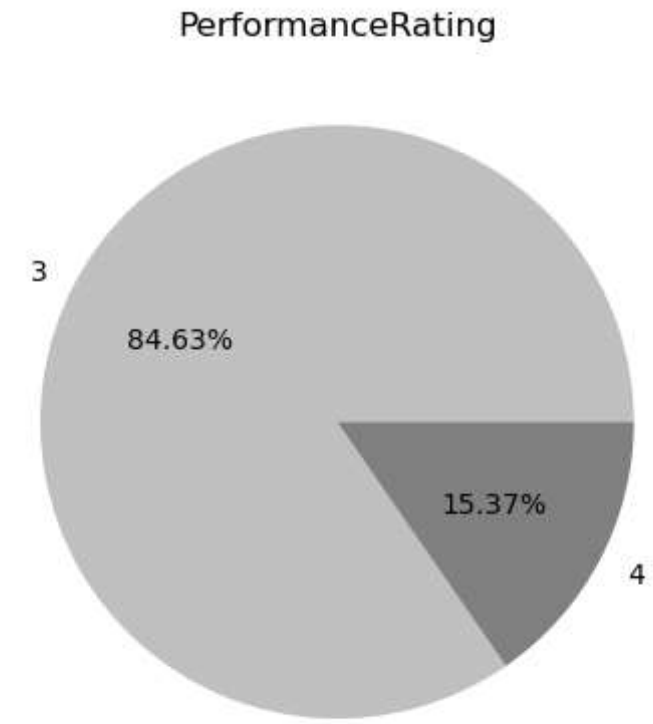

EnvironmentSatisfaction
2
30.82%
3
30.34%
19.32%
1
19.52%
4

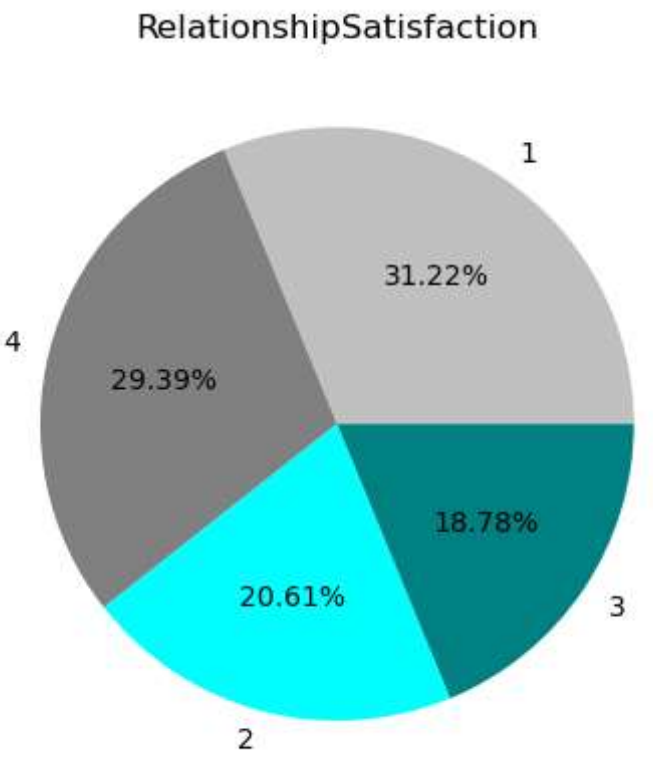

JobInvolvement
3
59.05%
5.65%
1
9.80%
25.51%
4
2

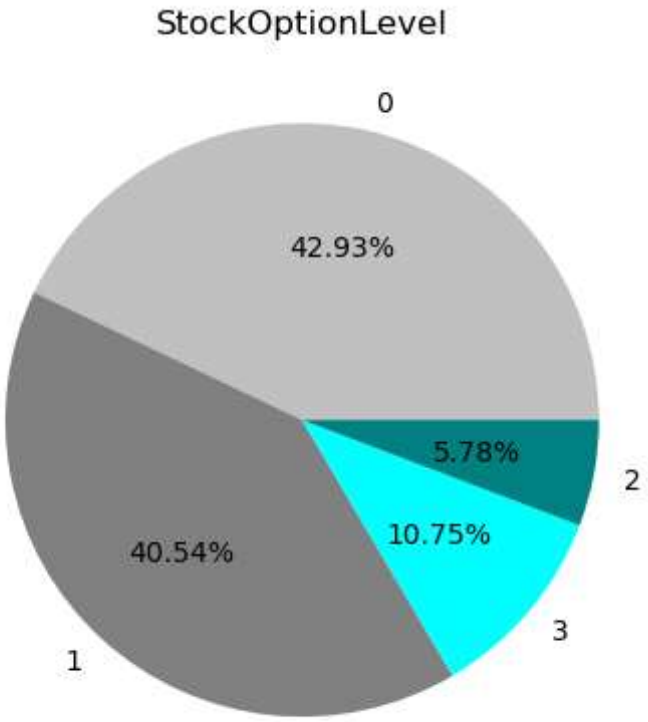

JobLevel
2
36.94%
4.69%
5
36.33%
1
7.21%
14.83%
4
3

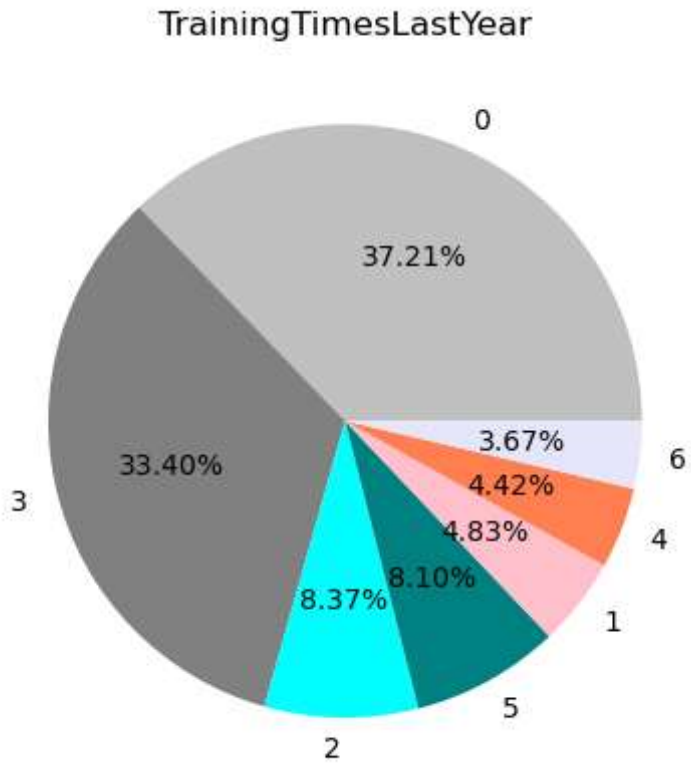

JobSatisfaction
4
31.22%
2
30.07%
19.05%
1
19.66%
3

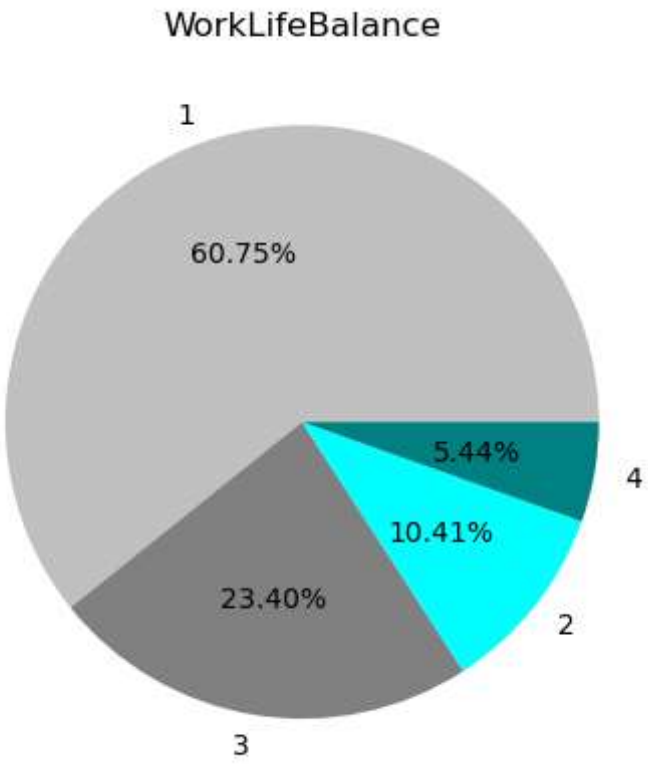

NumCompaniesWorked
8
35.44%
1
13.40%
3.33%
3
10.82%
3.54%
6
4.29%
7
4.76%
9.93%
5.03%
9.46%
2
5
9
4
0


PerformanceRating
3
84.63%
15.37%
4


RelationshipSatisfaction
1
31.22%
4
29.39%
18.78%
3
20.61%
2


StockOptionLevel
0
42.93%
5.78%
2
10.75%
40.54%
3
1


TrainingTimesLastYear
0
37.21%
3.67%
6
33.40%
3
4.42%
4.83%
4
8.10%
8.37%
1
5
2



Education 1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'

EnvironmentSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobInvolvement 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

JobSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

PerformanceRating 1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding'

RelationshipSatisfaction 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

WorkLifeBalance 1 'Bad' 2 'Good' 3 'Better' 4 'Best'

Inferences-

Attrition: Around 84% of the employees have not left the company (No), while the remaining 16% have (Yes). This suggests that employee retention is relatively good at this company.

Business travel: The majority of employees (nearly 71%) travel rarely for business, while about 19% travel frequently and 10% don't travel at all. This could indicate that the company has a relatively stable and consistent travel policy, with most employees not having to travel often.

Department: The largest department in the company is Sales, with over 65% of the employees working in this area. Research & Development comes in second, with about 30% of the employees, while Human Resources has the smallest number of employees, at just over 4%. This suggests that the company is focused on sales and innovation.

Education field: The most common education field among employees is Life Sciences, with over 41% of the employees having this background. Other fields such as Medical, Marketing, and Technical Degree also have a notable representation, while Human Resources has the smallest number of employees with this background. This could suggest that the company has a focus on science and technology.

Gender: The employee population is dominated by females, who make up 60% of the workforce, while males make up 40%. This suggests that the company may be doing well in terms of gender diversity and inclusion.

Job role: Sales Executive, Research Scientist, and Laboratory Technician are the most common job roles, with around 22%, 20%, and 18% of the employees respectively. Human Resources has the smallest number of employees in this area, at just over 3%. This suggests that the company places a strong emphasis on sales and research roles.

Work-life balance: The majority of employees (over 60%) have a work-life balance rating of 1, while about 23% have a rating of 3. A smaller proportion of employees have a rating of 2 or 4. This could suggest that the company may need to work on improving work-life balance for its employees.

Training time last year: Over 37% of employees did not receive any training last year, while over 33% received 3 hours of training. Other training hour categories, such as 1, 2, 4, 5, and 6 hours, had a smaller representation. This could indicate that the company may need to prioritize training and development for its employees.

Stock option level: The majority of employees (over 42%) have a stock option level of 0, while about 41% have a level of 1. A smaller proportion of employees have levels of 2 or 3. This could suggest that the company may need to revisit its stock option plan and make it more appealing to employees.

Relationship satisfaction: Almost 30% of employees have a relationship satisfaction rating of 4, while the other ratings (1, 2, and 3) have relatively similar representations. This suggests that the company may be doing well in terms of employee relationships and communication.

Performance rating: The majority of employees (over 84%) have a performance rating of 3, while the remaining 15% have a rating of 4. This could indicate that the company has a relatively consistent and effective performance evaluation system.

Number of Companies Worked: The majority of employees (35.44%) have worked at only one company, while around 10% of employees have worked at six or more companies. This suggests that the company may have a higher rate of employee turnover, which could impact productivity and team cohesion.

Job Satisfaction: About a third of employees (31.22%) report high job satisfaction (rating of 4), while 30.07% report moderate satisfaction (rating of 2). The remaining employees report either low or neutral satisfaction. This indicates that a significant proportion of employees may not be fully engaged or motivated at work.

Job Level: The majority of employees (36.94%) hold a job level of 2, while around 15% hold a job level of 3. This suggests that there may be limited opportunities for career advancement within the company, which could contribute to employee turnover.

Job Involvement: The majority of employees (59.05%) report high job involvement (rating of 3), while only a small percentage report low involvement (rating of 1). This indicates that employees are generally committed to their work and engaged in their roles.

Environment Satisfaction: About a third of employees (30.82%) report moderate environment satisfaction (rating of 2), while a similar percentage (30.34%) report high satisfaction (rating of 3). This suggests that employees generally have a positive view of their work environment, although there is room for improvement.

Over Time: The majority of employees (71.70%) report working overtime, while just under a third (28.30%) do not. This could suggest that the company has high workload demands or may struggle with efficient work processes.

Marital Status: The largest proportion of employees (45.78%) are single, followed by married (31.97%) and divorced (22.24%). This suggests that the company may have a younger workforce, and may want to consider implementing policies that support work-life balance for employees with family responsibilities.

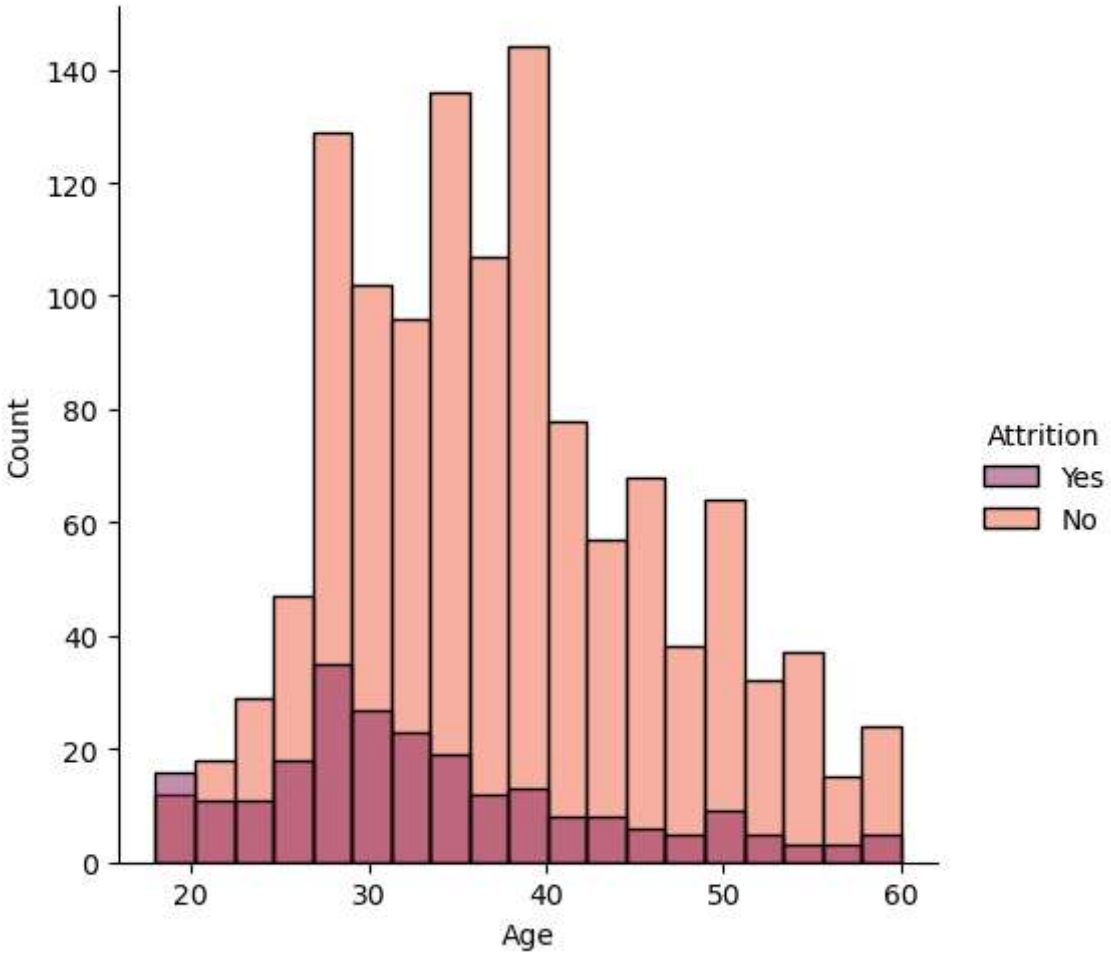
Education: The majority of employees (38.91%) have a background in Business, followed by Life Sciences (27.07%) and Medical (19.18%). This indicates that the company may have a diverse workforce with a range of educational backgrounds.

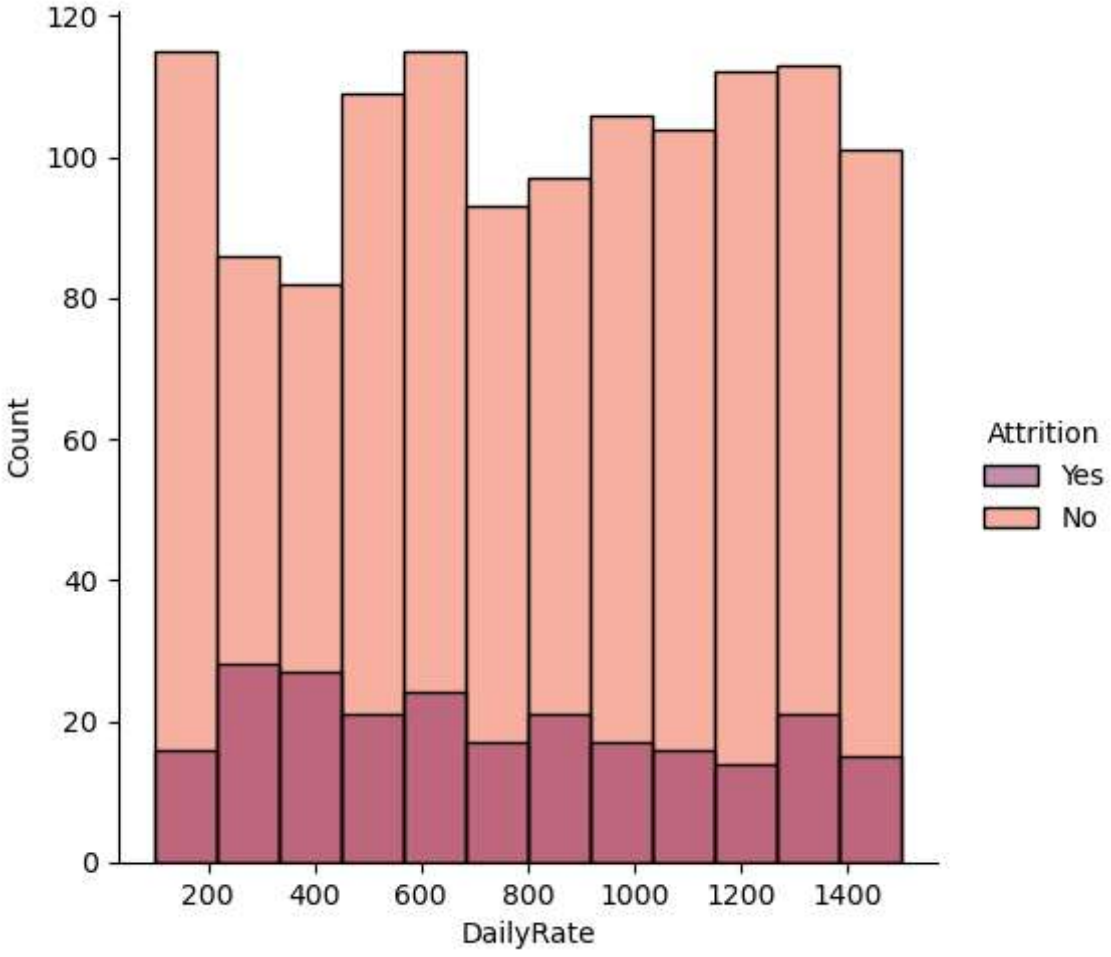
Overtime: A significant proportion of employees (71.70%) report working overtime, while the remaining employees (28.30%) do not. This suggests that the company may have a culture of working long hours, which could contribute to employee burnout and turnover.

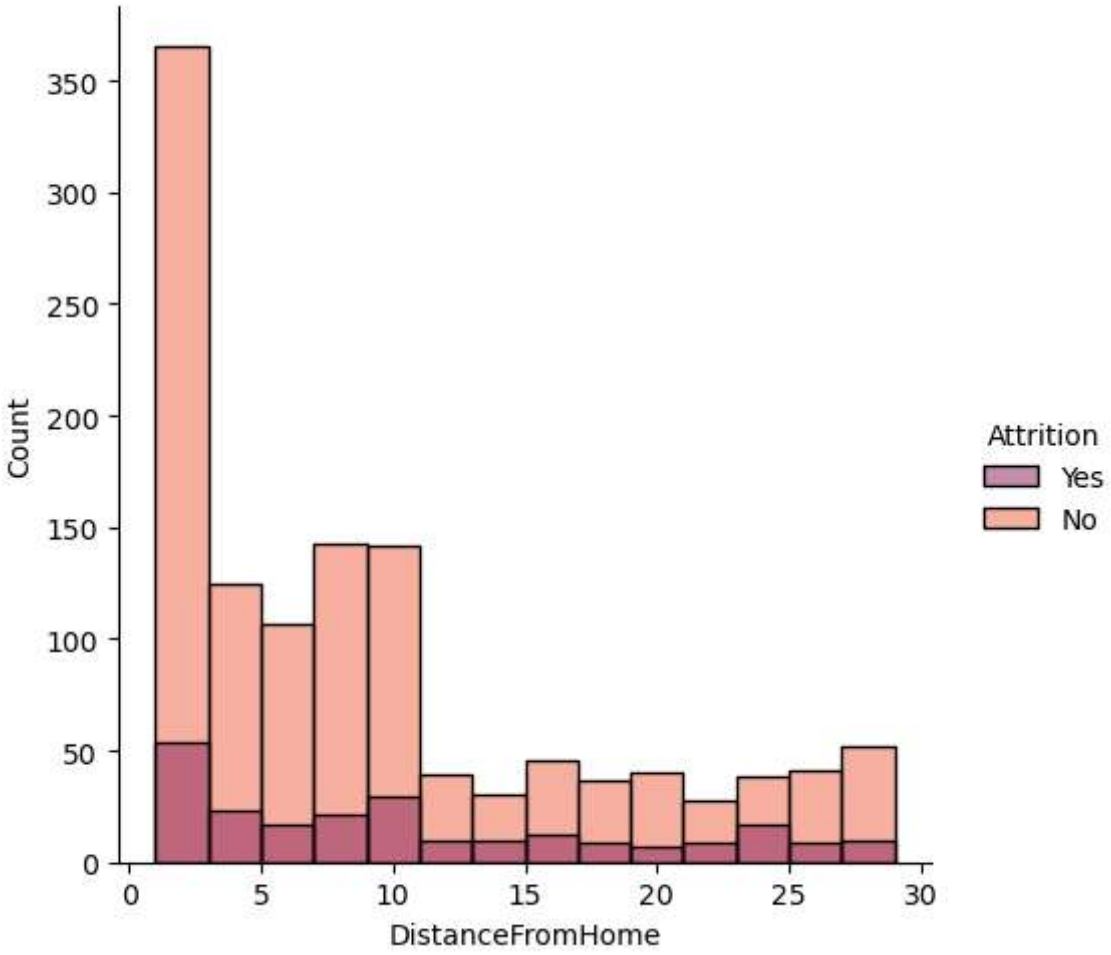
Overall, these inferences suggest that the company may have areas for improvement, particularly in terms of employee satisfaction and turnover. The company may benefit from implementing strategies to improve work-life balance, career advancement opportunities, and employee engagement. Additionally, the high rate of overtime reported by employees may be a cause for concern and may need to be addressed to avoid burnout and high turnover rates.

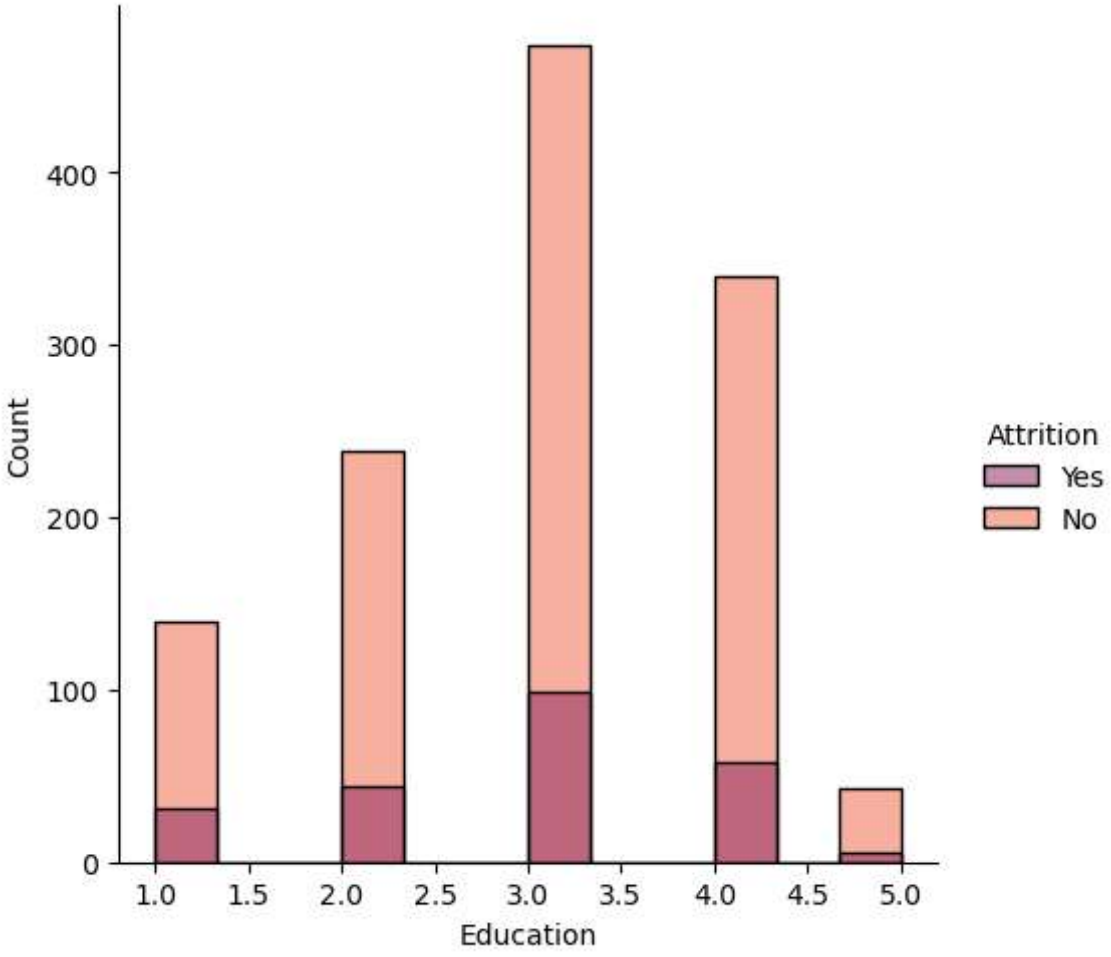
In [20]:
```
columns=['Age', 'DailyRate',  'DistanceFromHome', 'Education', 'EnvironmentSatisfa
```

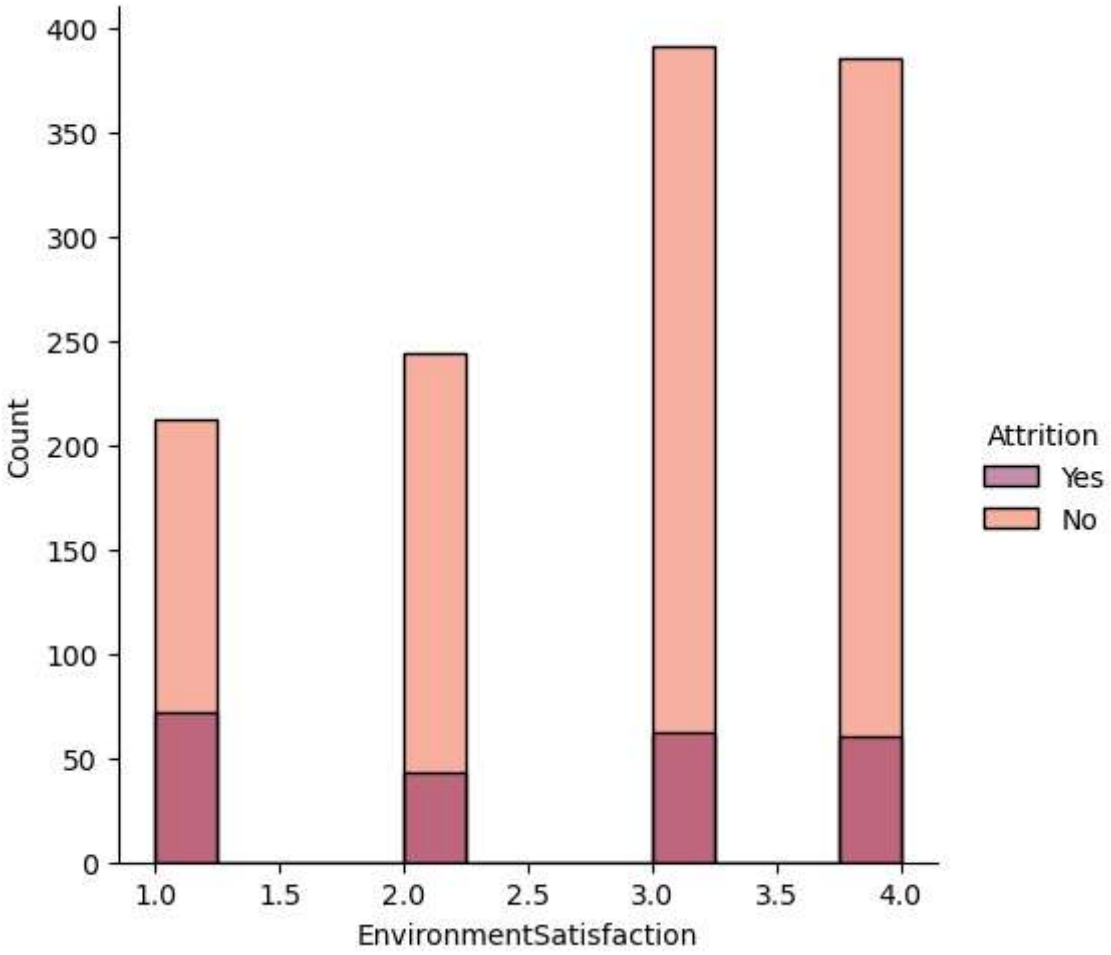
In [21]:
```
for x in columns:
    sns.displot(x=x,hue='Attrition',data=ibm,palette='rocket')
    plt.show()
```

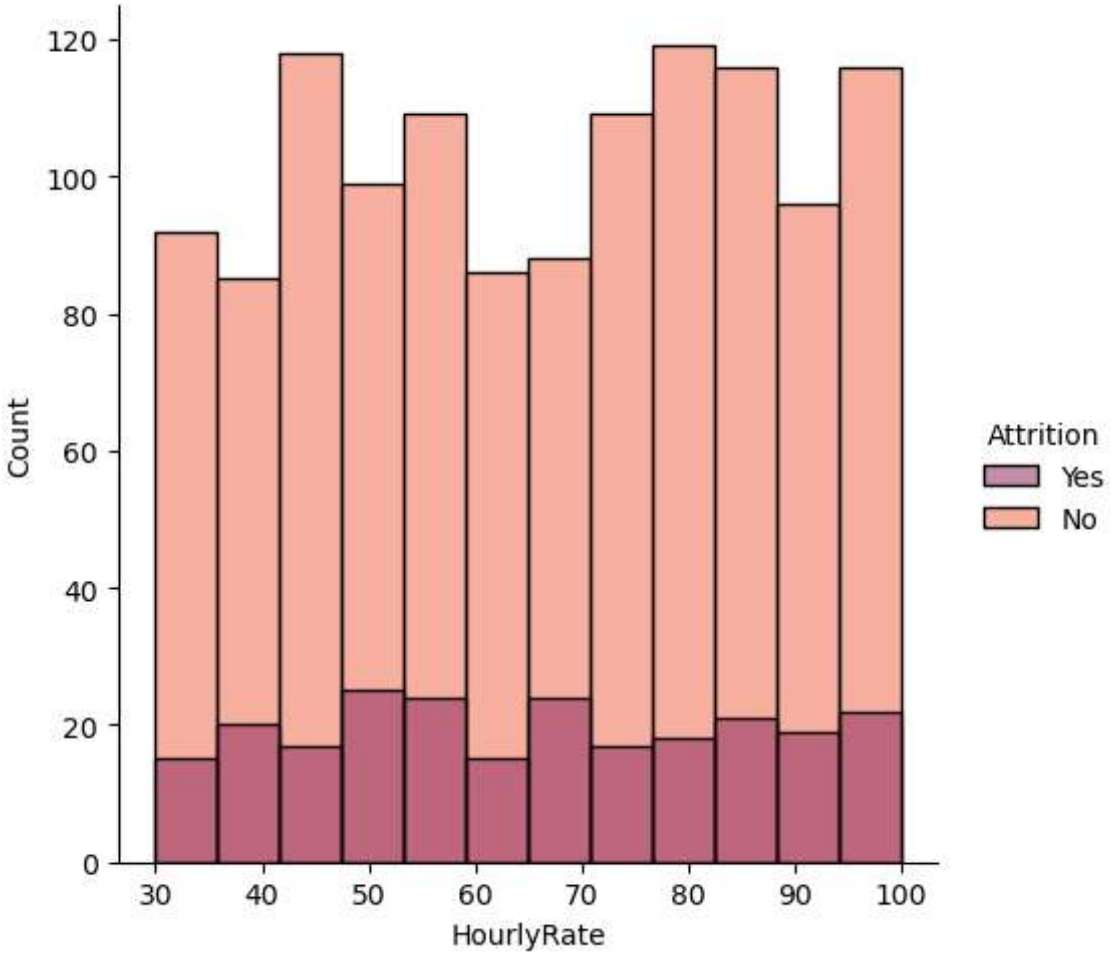

Count
Age
Attrition
Yes
No
0
20
40
60
80
100
120
140
20
30
40
50
60

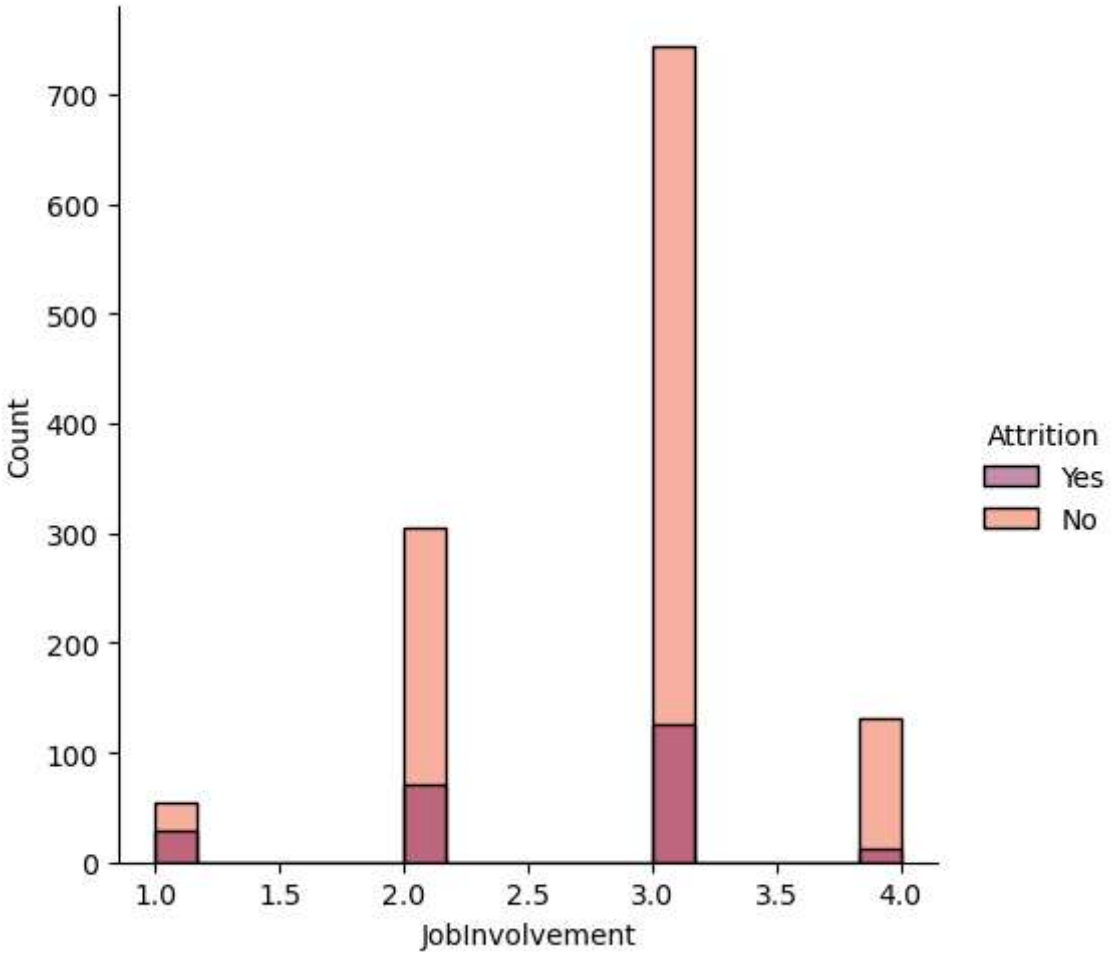

Count
DailyRate
Attrition
Yes
No
0
20
40
60
80
100
120
200
400
600
800
1000
1200
1400

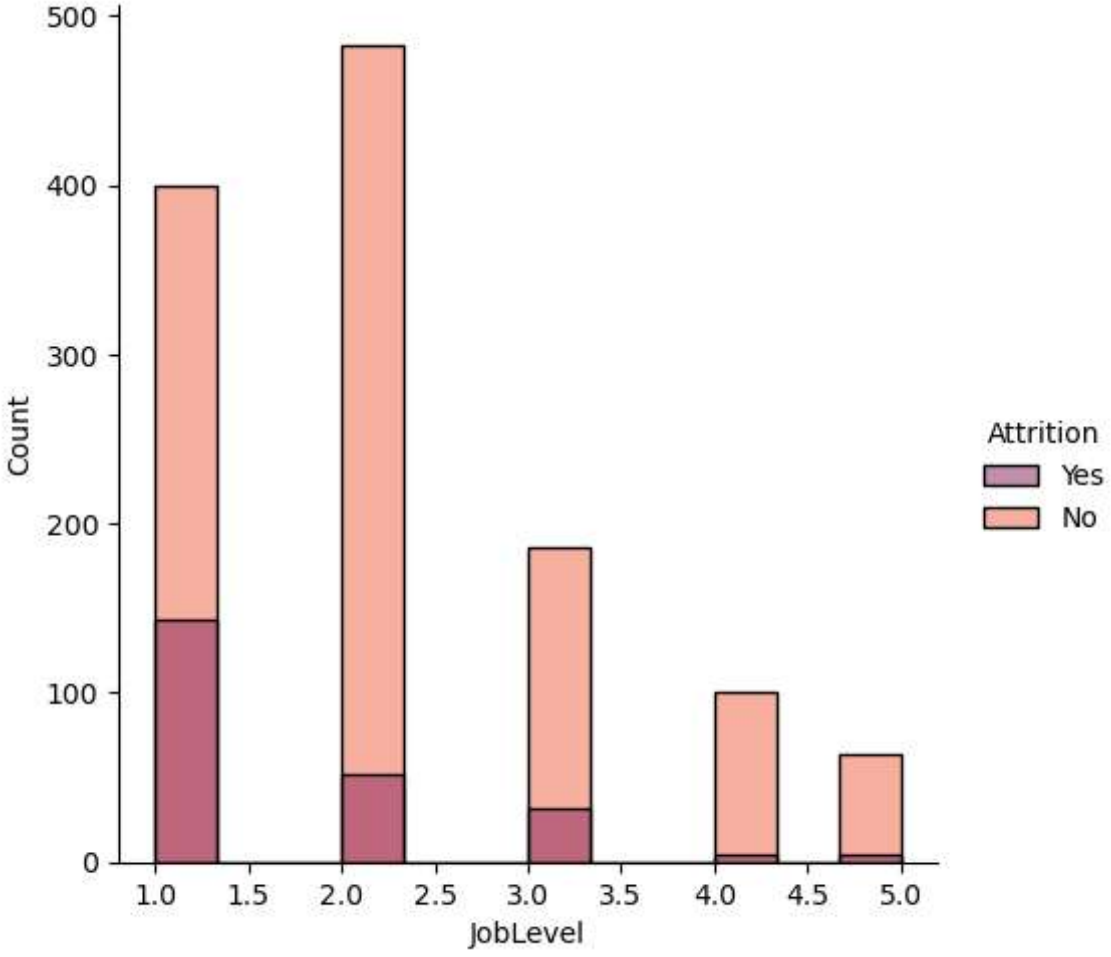

Count
DistanceFromHome
Attrition
Yes
No

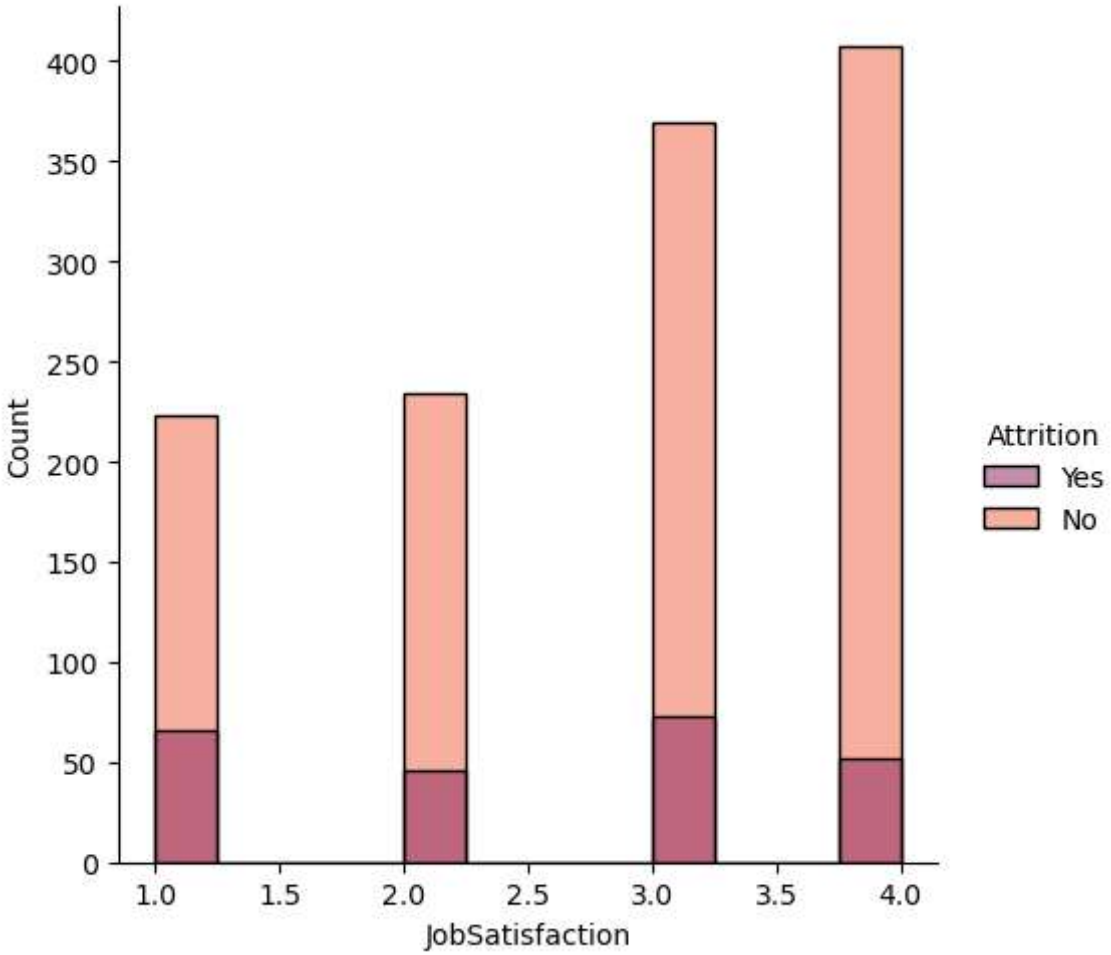

Count
Education
Attrition
Yes
No

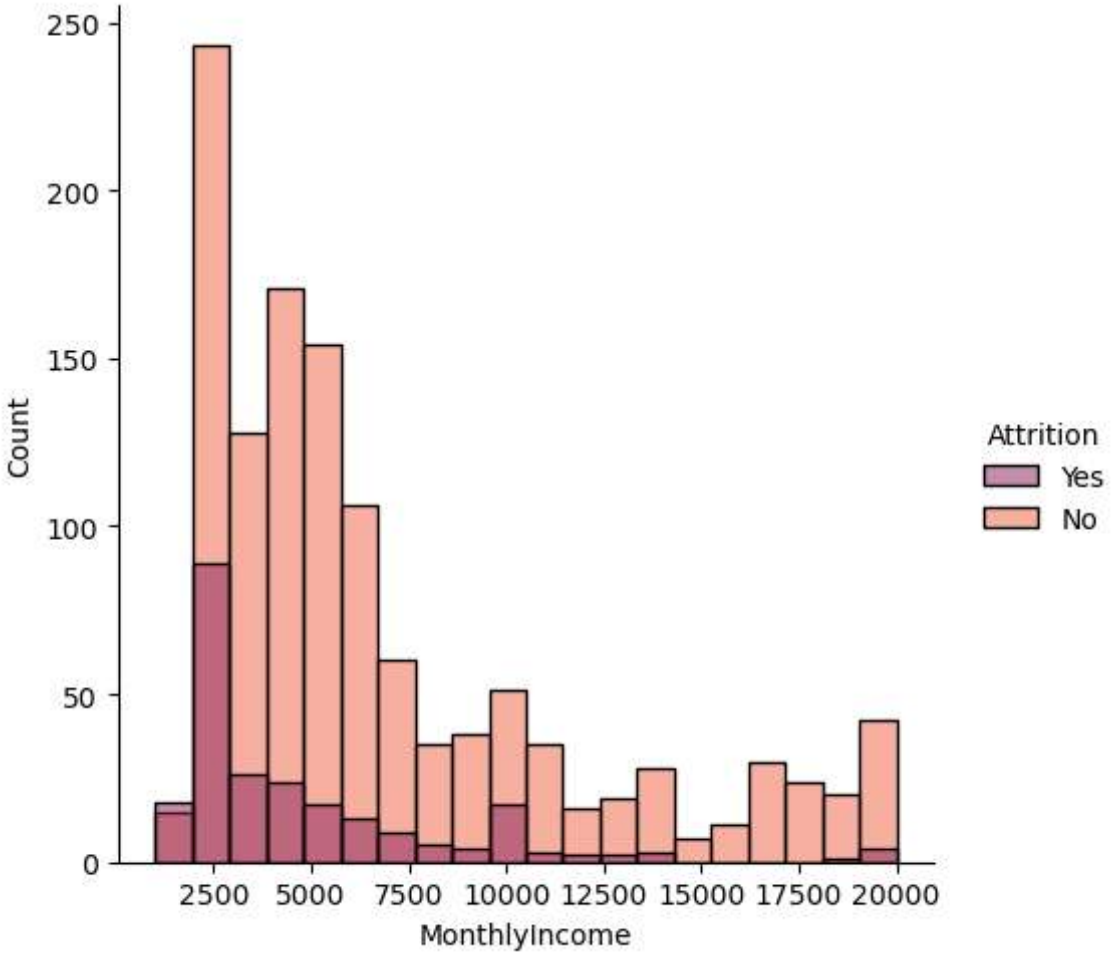

400
350
300
250
200
150
100
50
0
Count
1.0
1.5
2.0
2.5
3.0
3.5
4.0
EnvironmentSatisfaction
Attrition
Yes
No

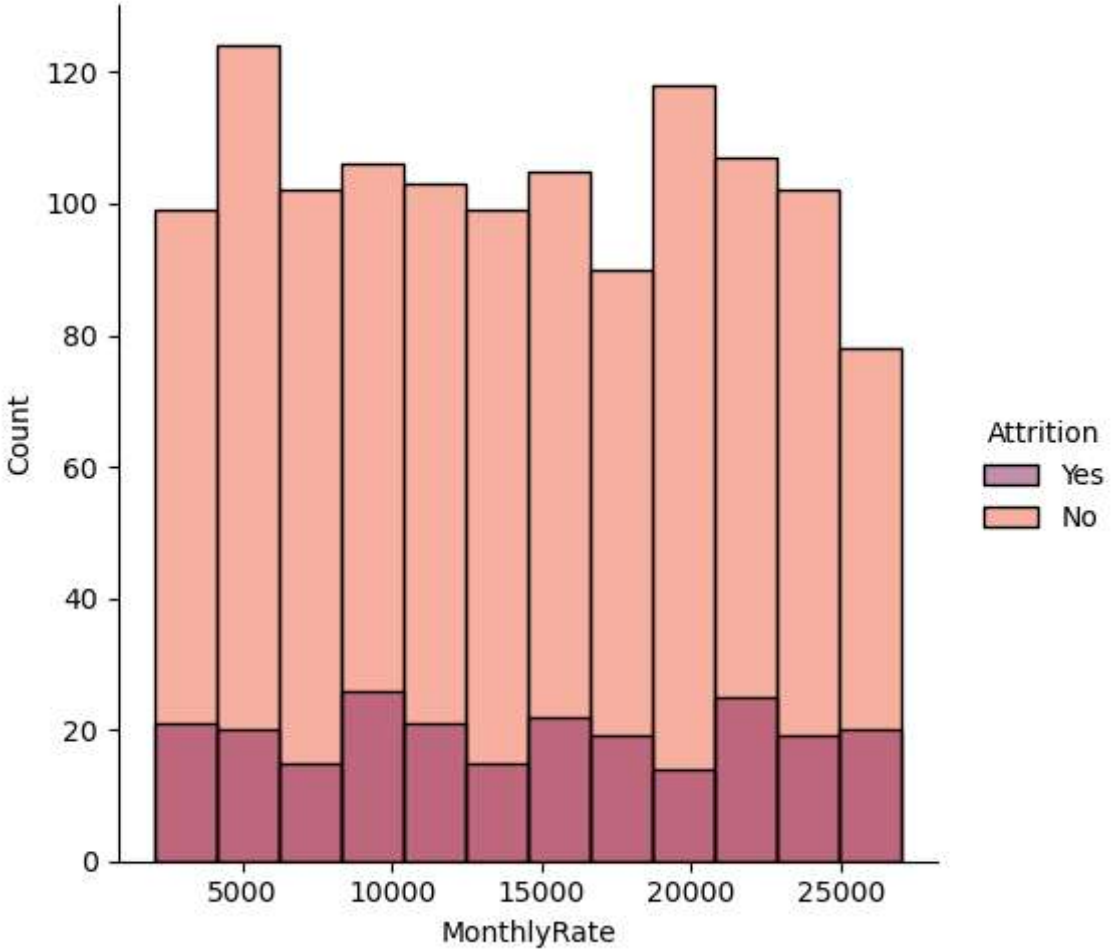

120
100
80
60
40
20
0
Count
30
40
50
60
70
80
90
100
HourlyRate
Attrition
Yes
No

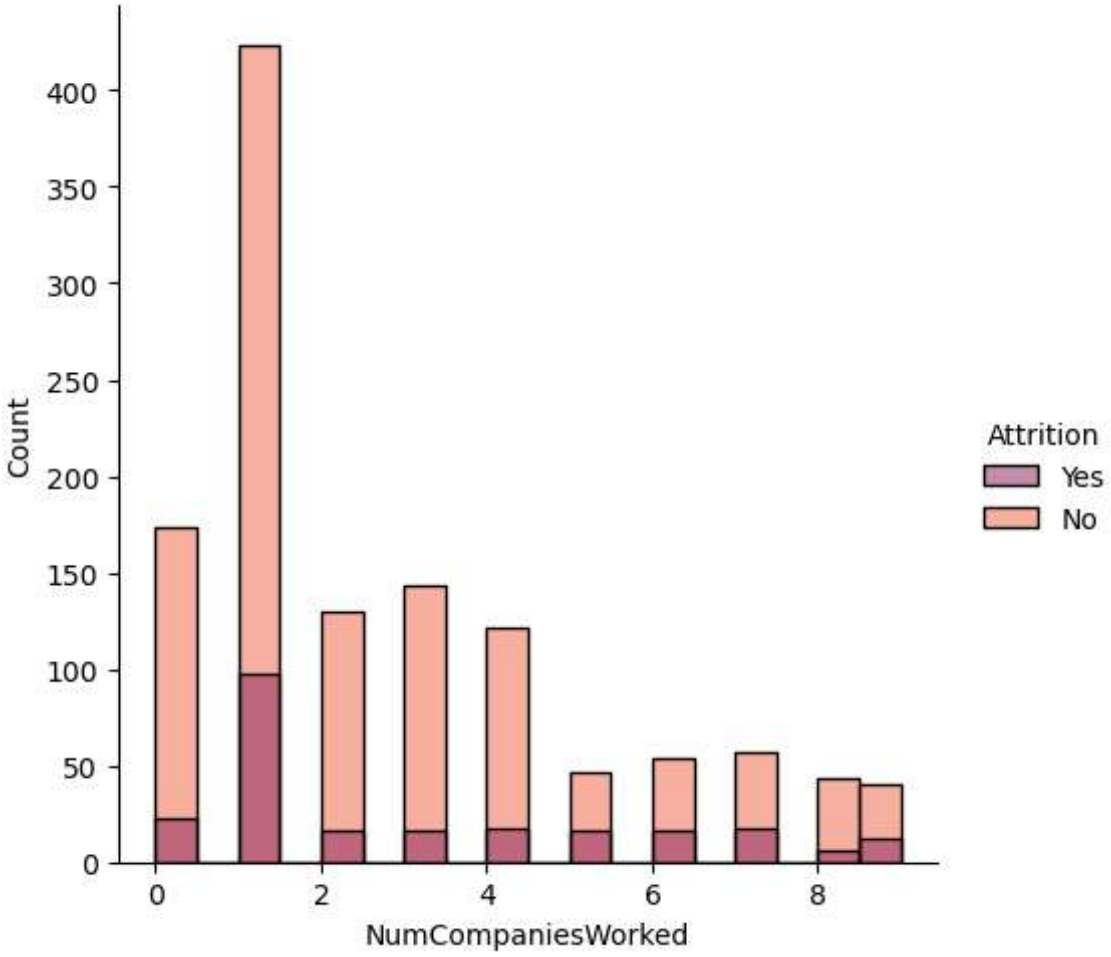

Count
JobInvolvement
Attrition
Yes
No

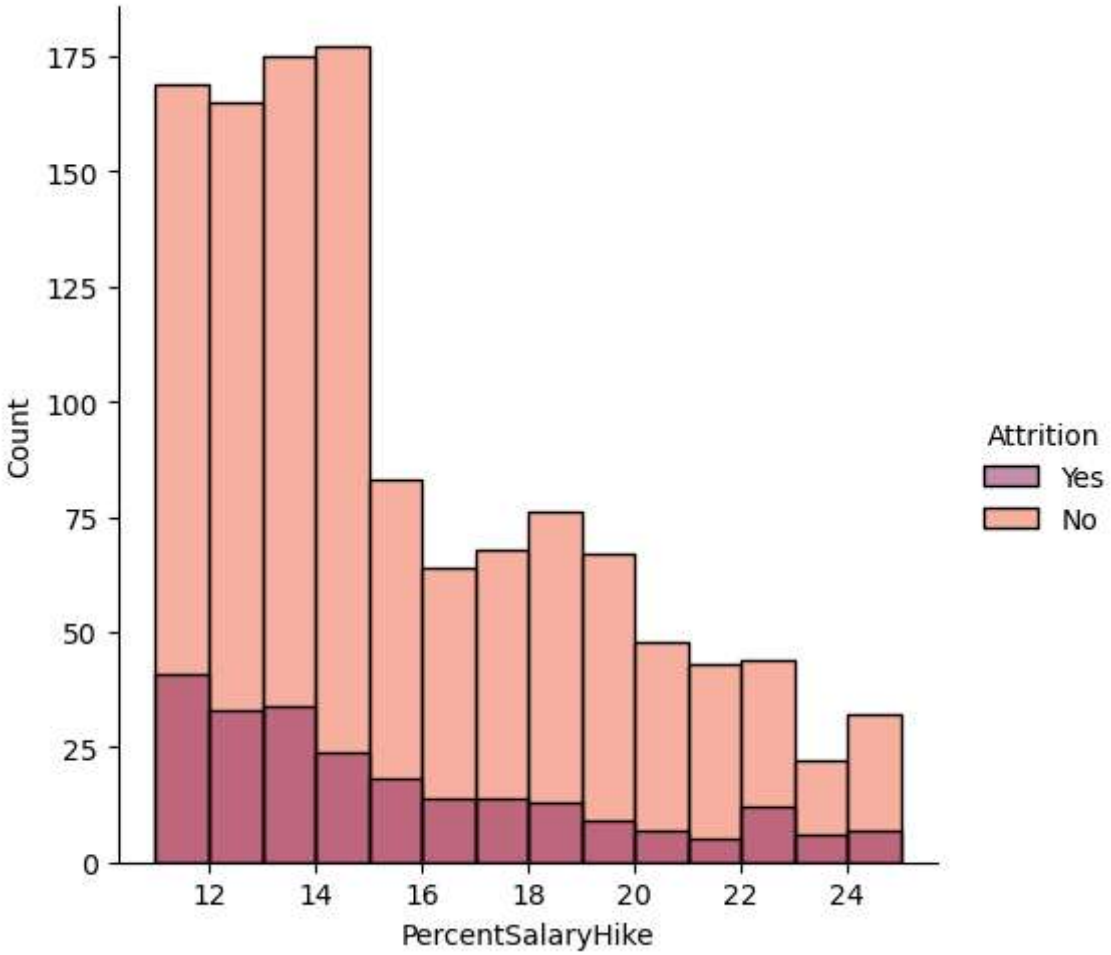

Count
JobLevel
Attrition
Yes
No

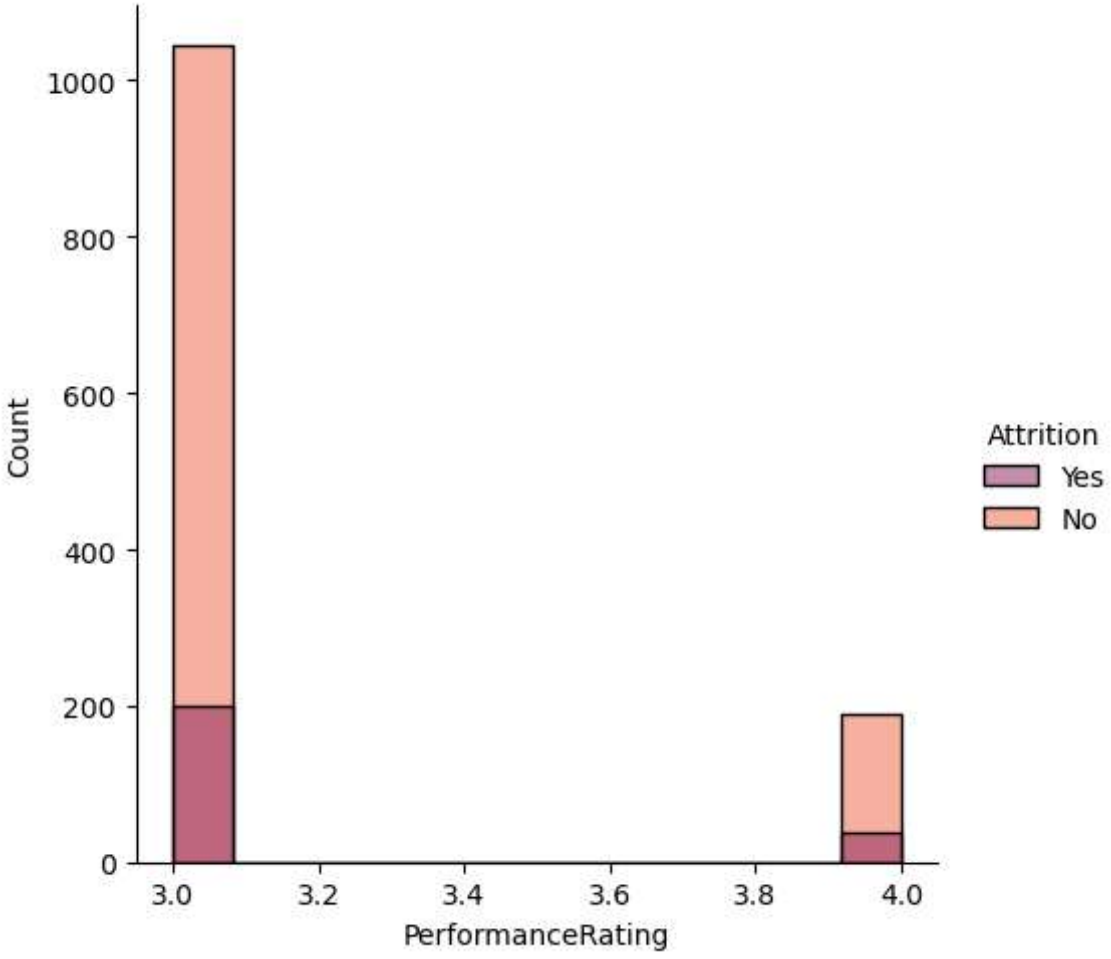

Count
JobSatisfaction
Attrition
Yes
No
0
50
100
150
200
250
300
350
400
1.0
1.5
2.0
2.5
3.0
3.5
4.0

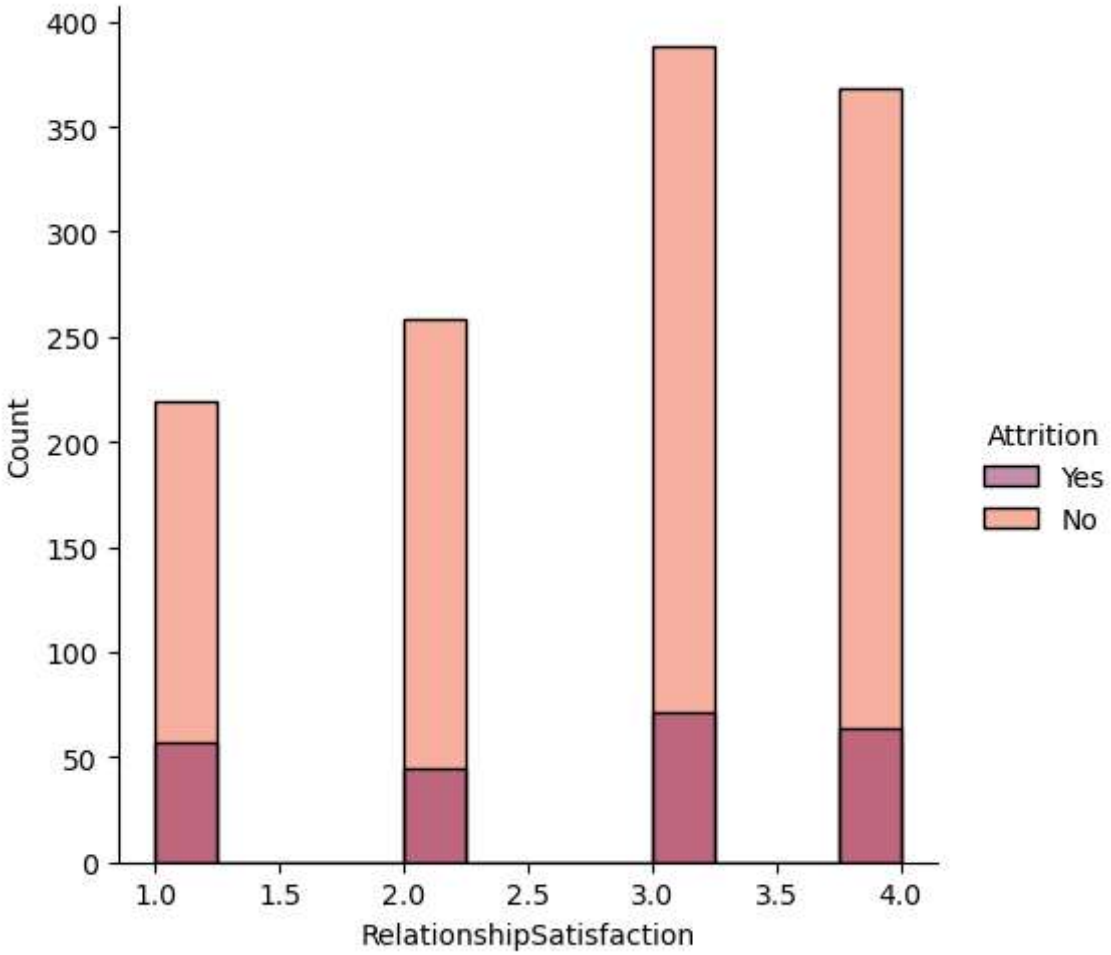

Count
MonthlyIncome
Attrition
Yes
No
0
50
100
150
200
250
2500
5000
7500
10000
12500
15000
17500
20000

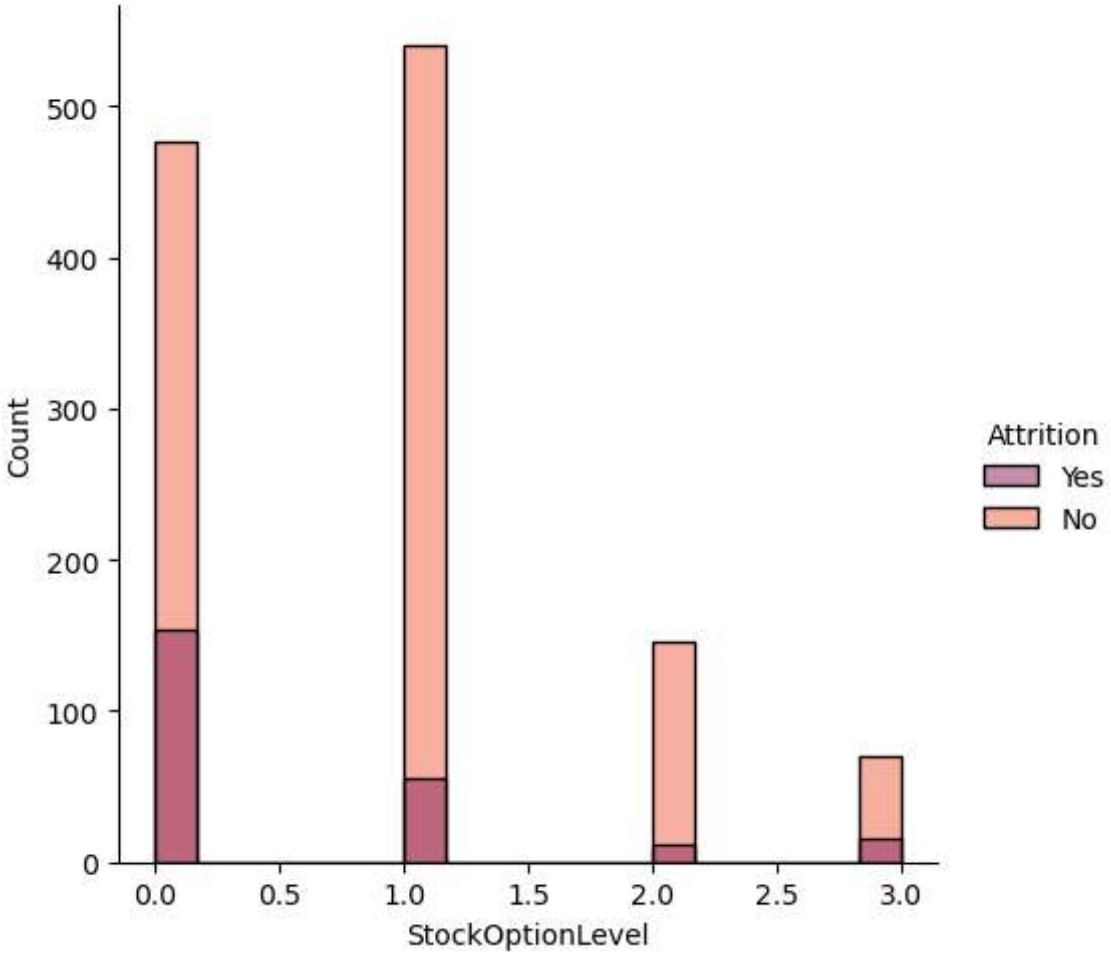

Count
MonthlyRate
Attrition
Yes
No
0
20
40
60
80
100
120
5000
10000
15000
20000
25000

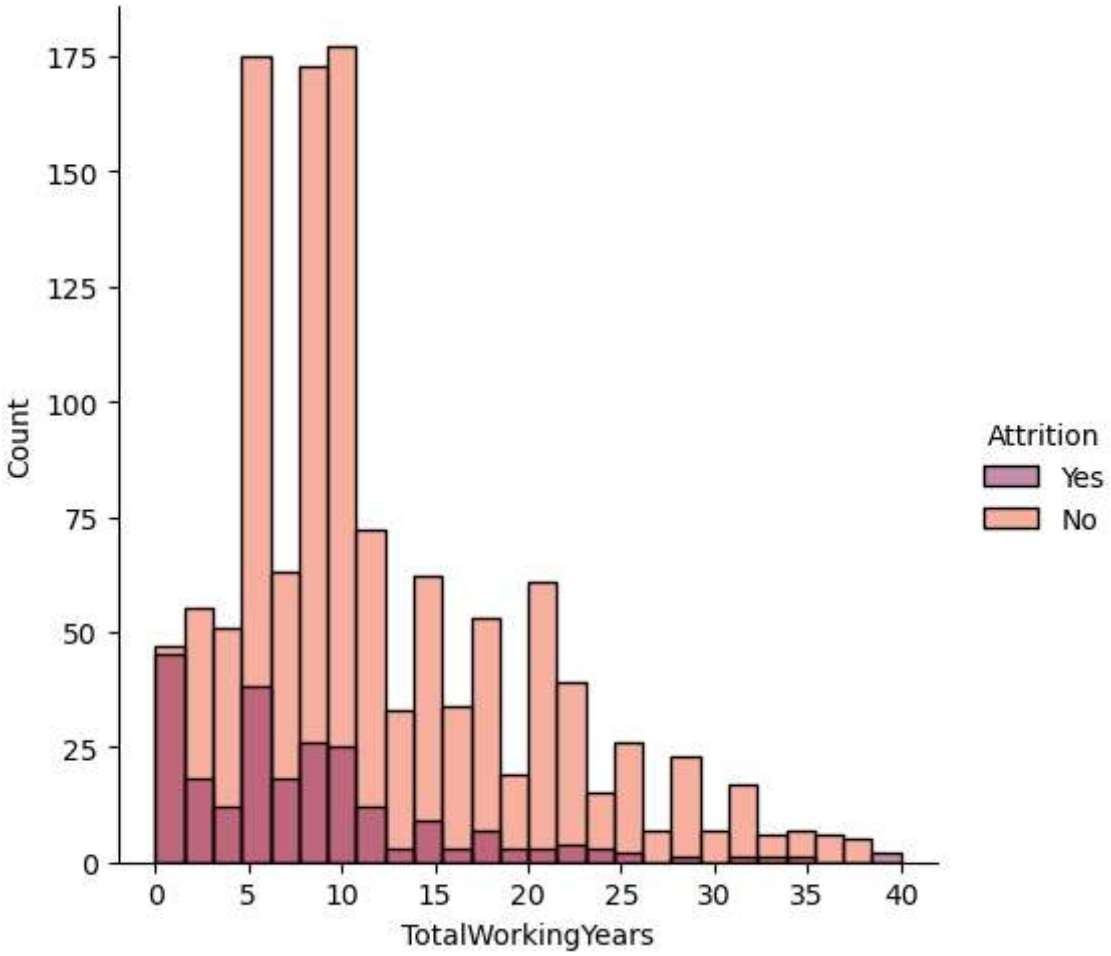

Count
NumCompaniesWorked
Attrition
Yes
No
0
50
100
150
200
250
300
350
400
0
2
4
6
8

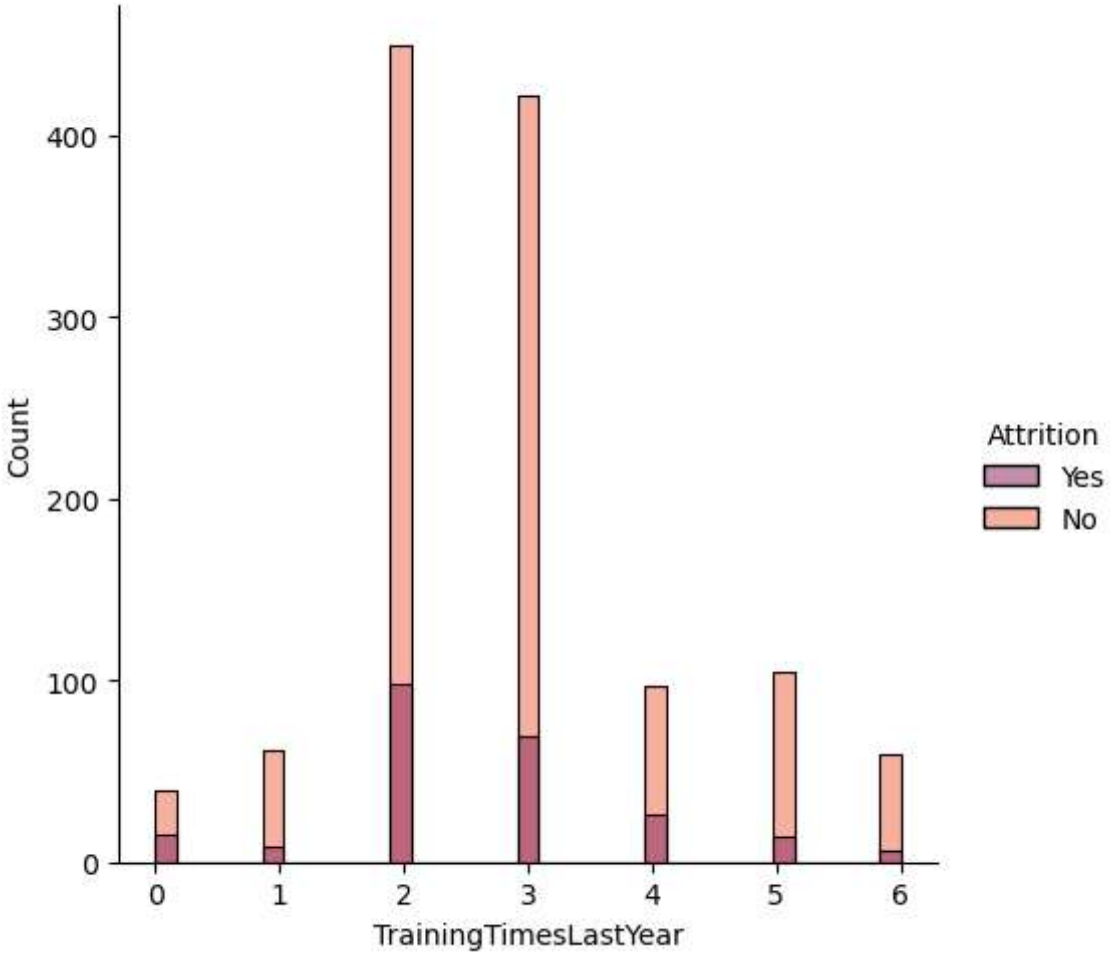

175
150
125
100
75
50
25
0
Count
12
14
16
18
20
22
24
PercentSalaryHike
Attrition
Yes
No

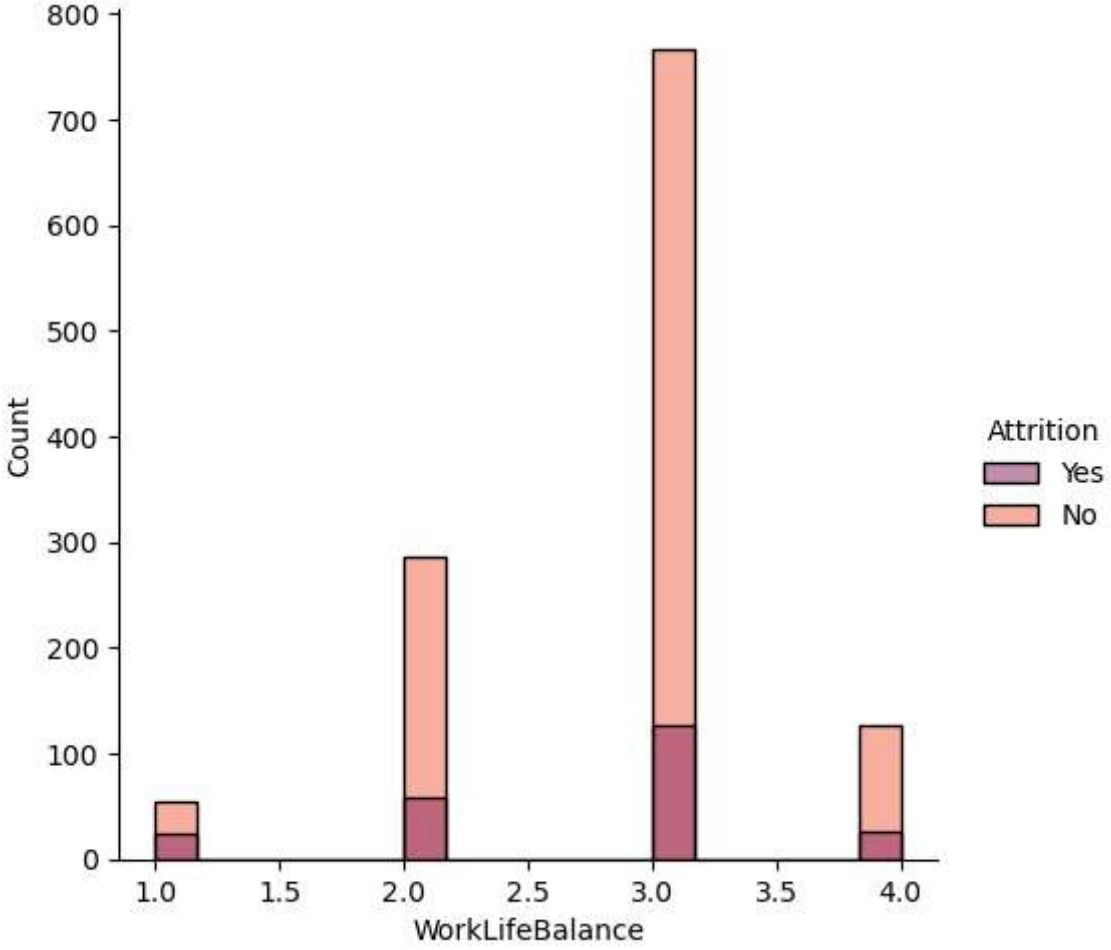

1000
800
600
400
200
0
Count
3.0
3.2
3.4
3.6
3.8
4.0
PerformanceRating
Attrition
Yes
No

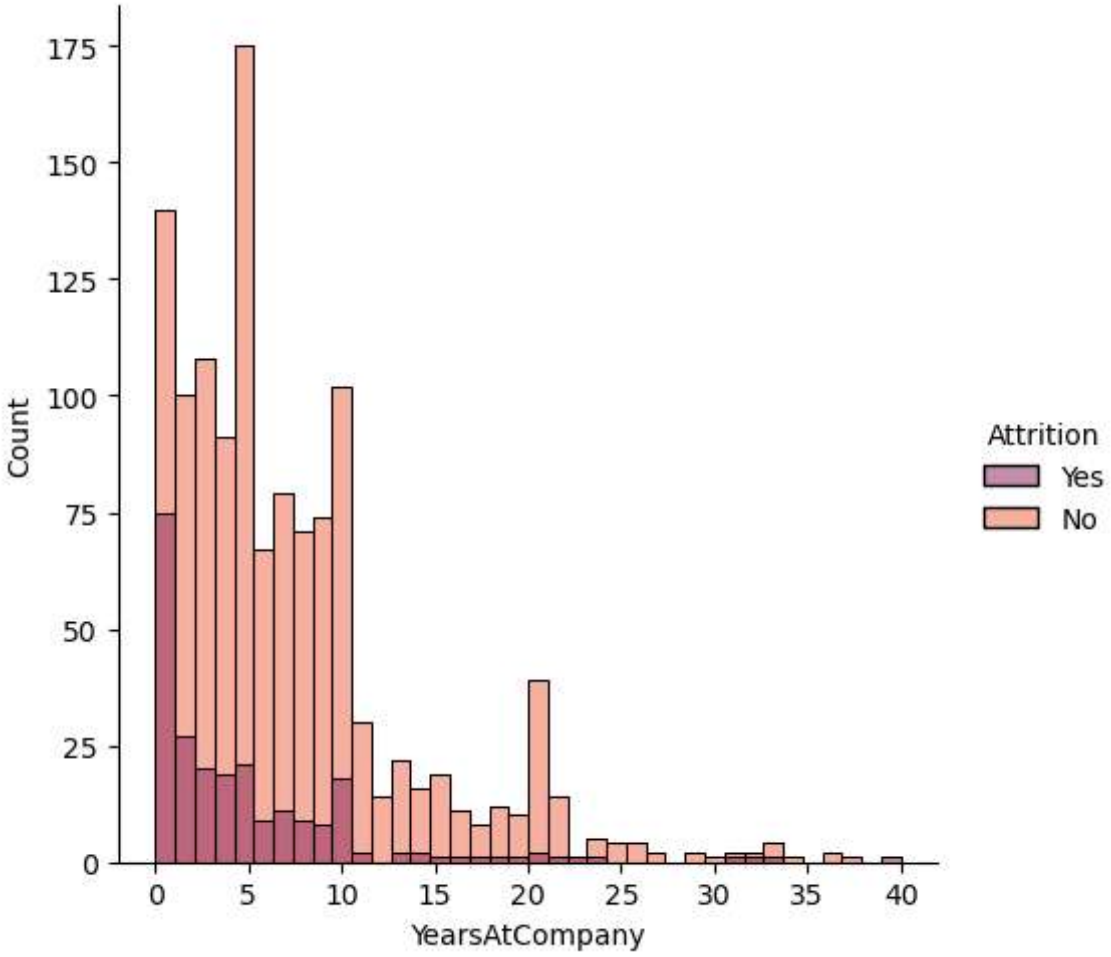

Count
RelationshipSatisfaction
Attrition
Yes
No

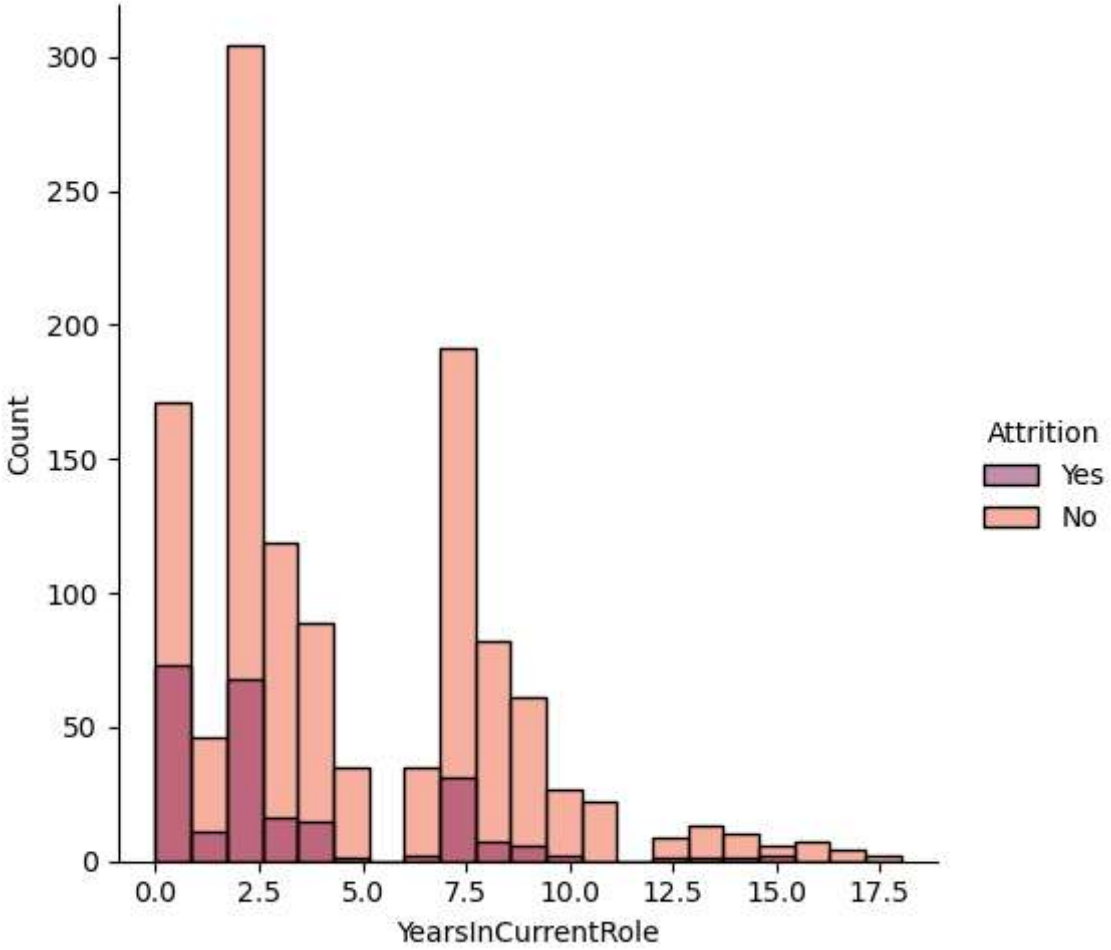

Count
StockOptionLevel
Attrition
Yes
No

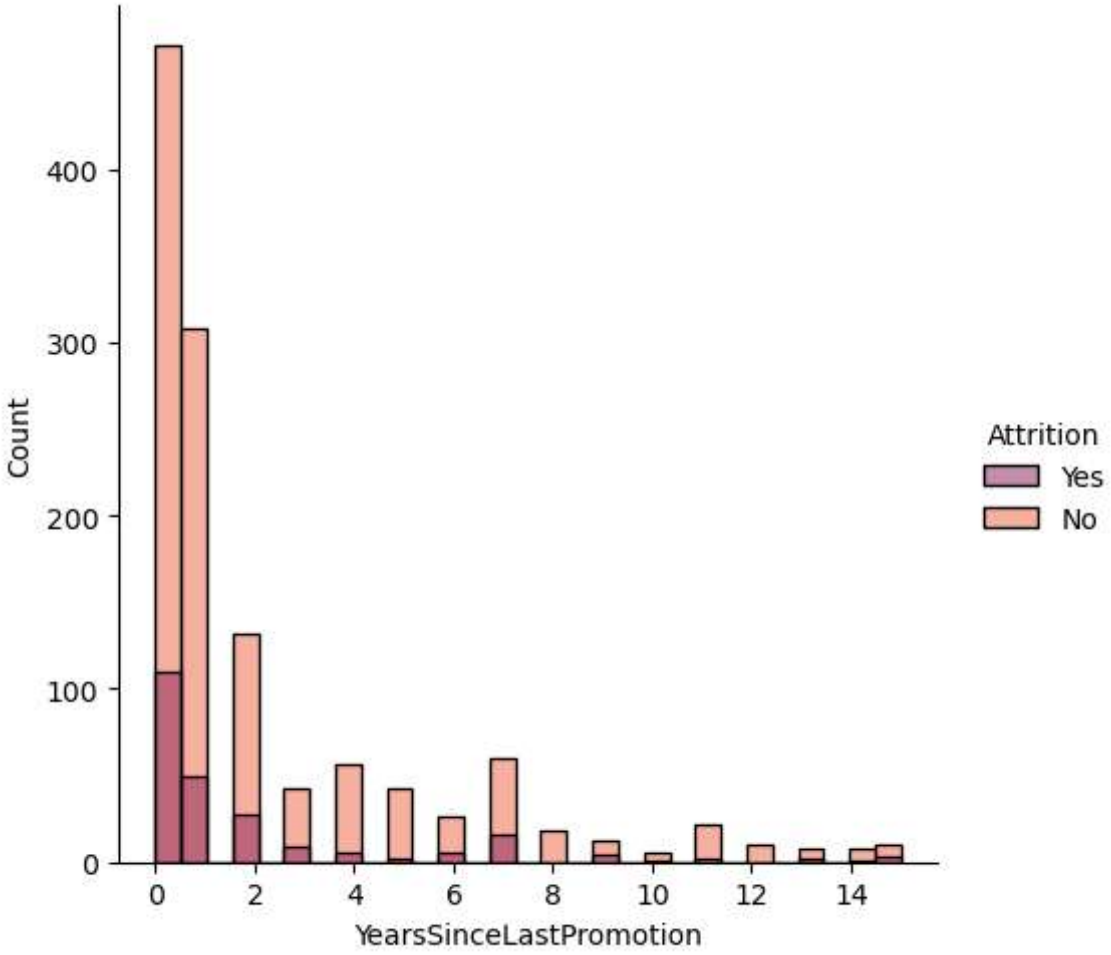

175
150
125
100
75
50
25
0
Count
0
5
10
15
20
25
30
35
40
TotalWorkingYears
Attrition
Yes
No

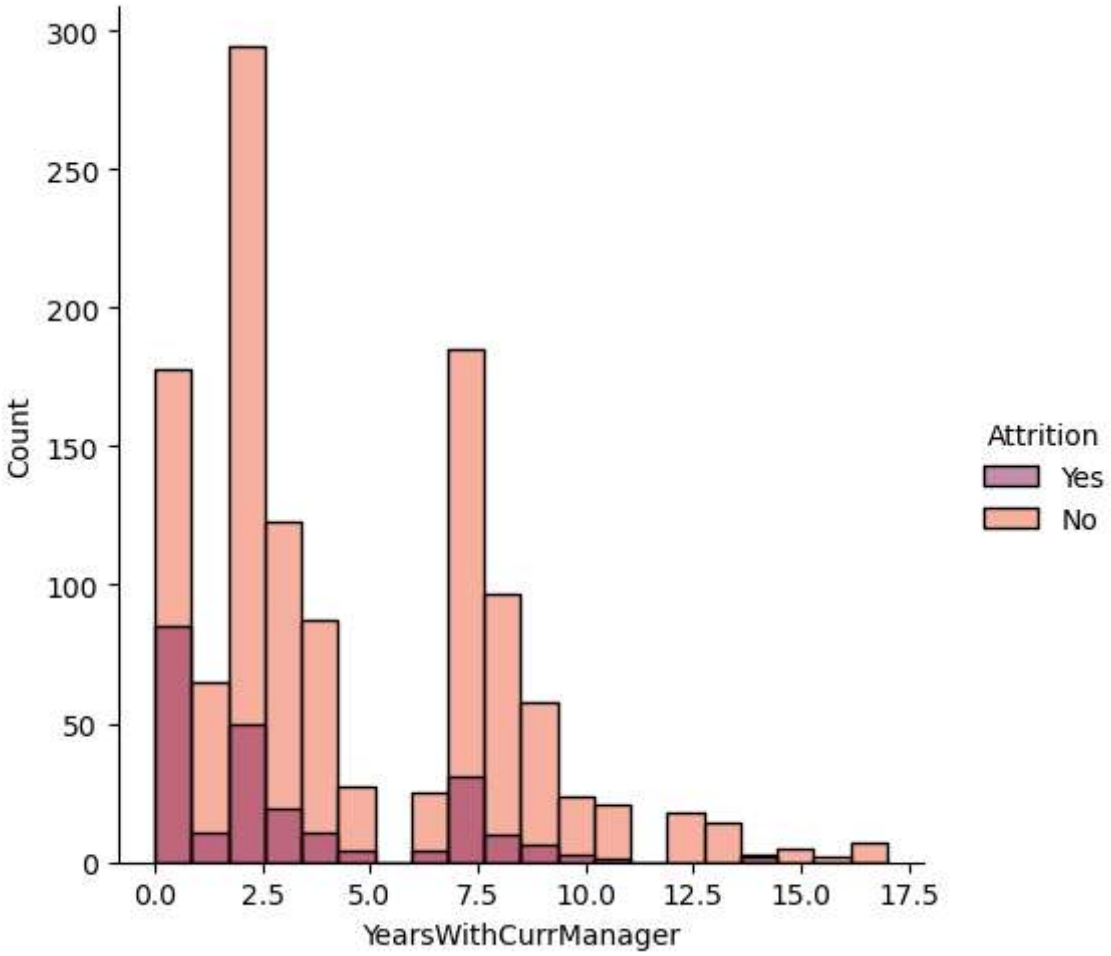

400
300
200
100
0
Count
0
1
2
3
4
5
6
TrainingTimesLastYear
Attrition
Yes
No


800
700
600
500
400
300
200
100
0
Count
1.0
1.5
2.0
2.5
3.0
3.5
4.0
WorkLifeBalance
Attrition
Yes
No


175
150
125
100
75
50
25
0
Count
0
5
10
15
20
25
30
35
40
YearsAtCompany
Attrition
Yes
No


Count
YearsInCurrentRole
Attrition
Yes
No


Count
YearsSinceLastPromotion
Attrition
Yes
No



inferences-

Age: The histogram shows that the age distribution of employees who have left the company is skewed towards the younger side. This indicates that younger employees are more likely to leave the company compared to older ones.

Business Travel: The plot shows that employees who travel frequently for business have a higher attrition rate compared to those who travel less often or not at all. This could be due to factors such as work-life balance and job satisfaction.

Daily Rate: The plot shows that there is no clear trend between daily rate and attrition. However, employees with higher daily rates have a slightly higher attrition rate compared to those with lower rates.

Department: The plot shows that employees in the Sales department have a higher attrition rate compared to those in the Research & Development and Human Resources departments. This could be due to factors such as job satisfaction, work-life balance, and career growth opportunities.

Distance from Home: The plot shows that employees who live farther away from their workplace have a higher attrition rate compared to those who live closer. This could be due to factors such as commute time, transportation costs, and work-life balance.

Education: The plot shows that employees with a higher level of education have a slightly lower attrition rate compared to those with a lower level of education. This could be due to

factors such as job satisfaction, career growth opportunities, and higher salaries.

Education Field: The plot shows that employees with degrees in Human Resources, Marketing, and Technical fields have a slightly higher attrition rate compared to those in other fields. This could be due to factors such as job satisfaction, career growth opportunities, and job demand.

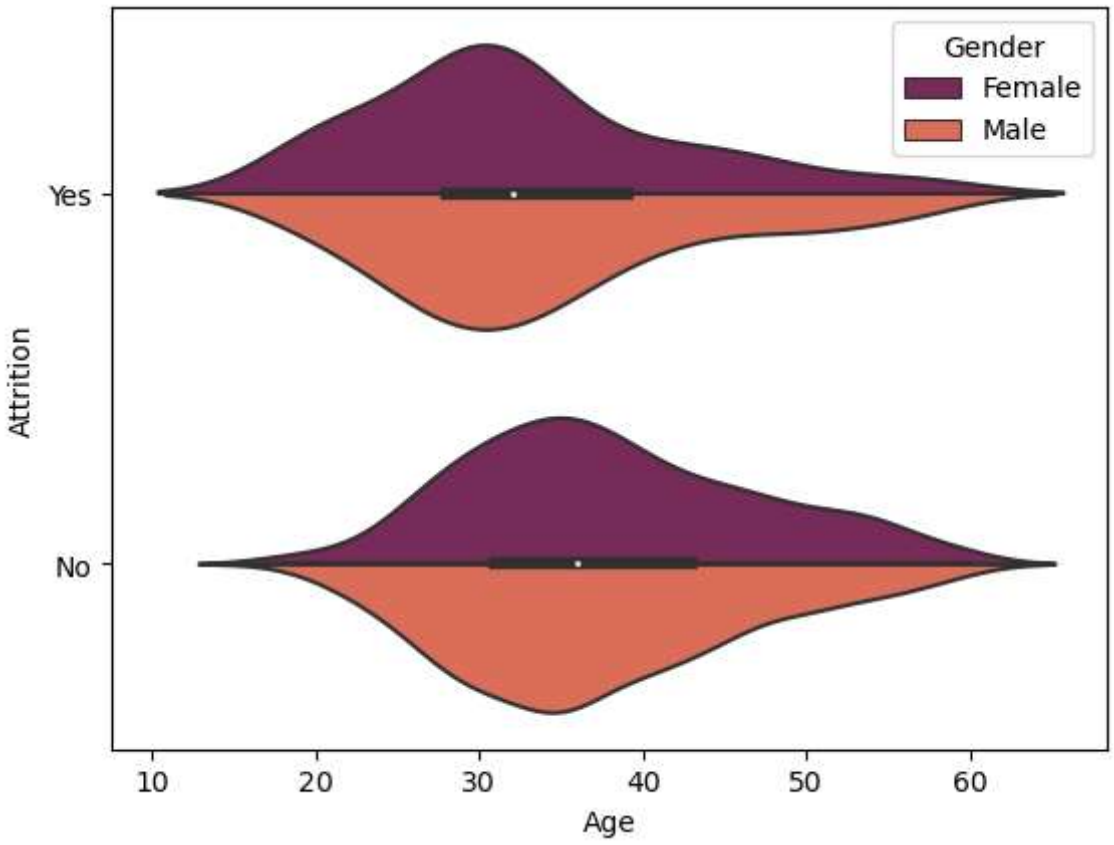
Environment Satisfaction: The plot shows that employees with lower environment satisfaction have a higher attrition rate compared to those with higher satisfaction. This indicates that a positive work environment is important in reducing employee turnover.

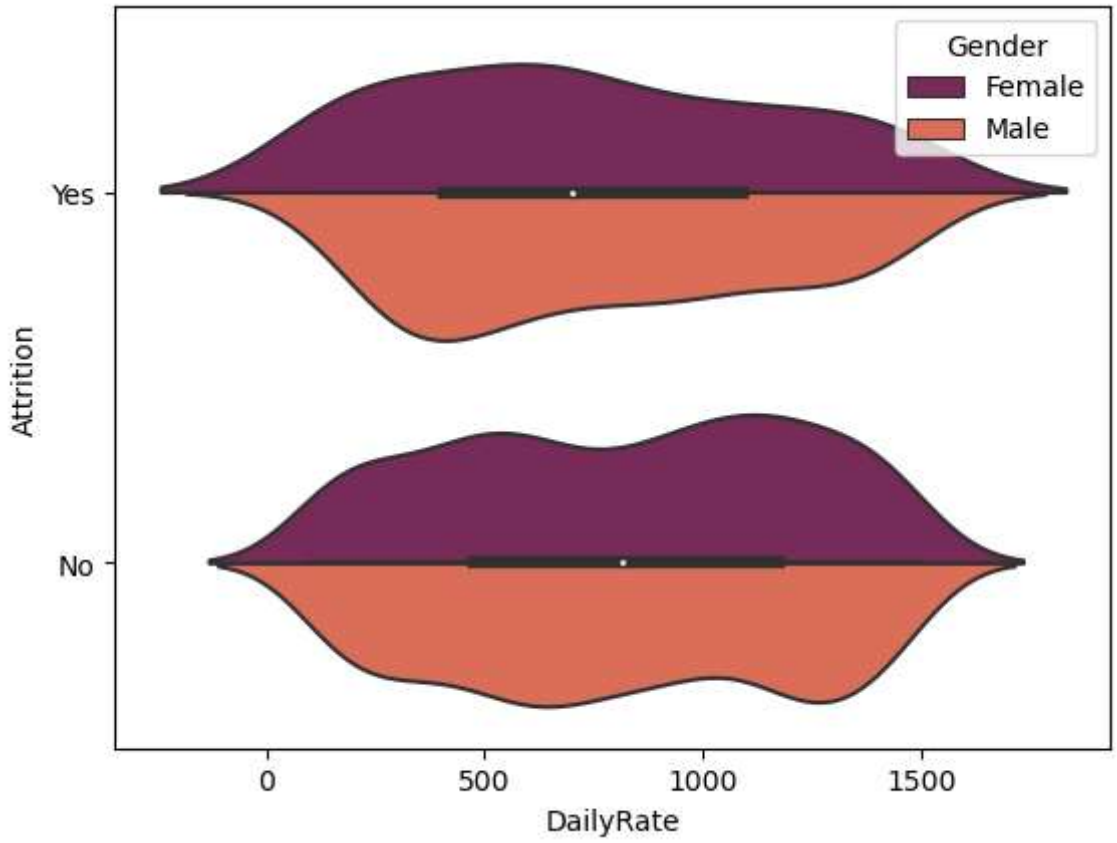
Gender: The plot shows that there is no significant difference in attrition rates between male and female employees.

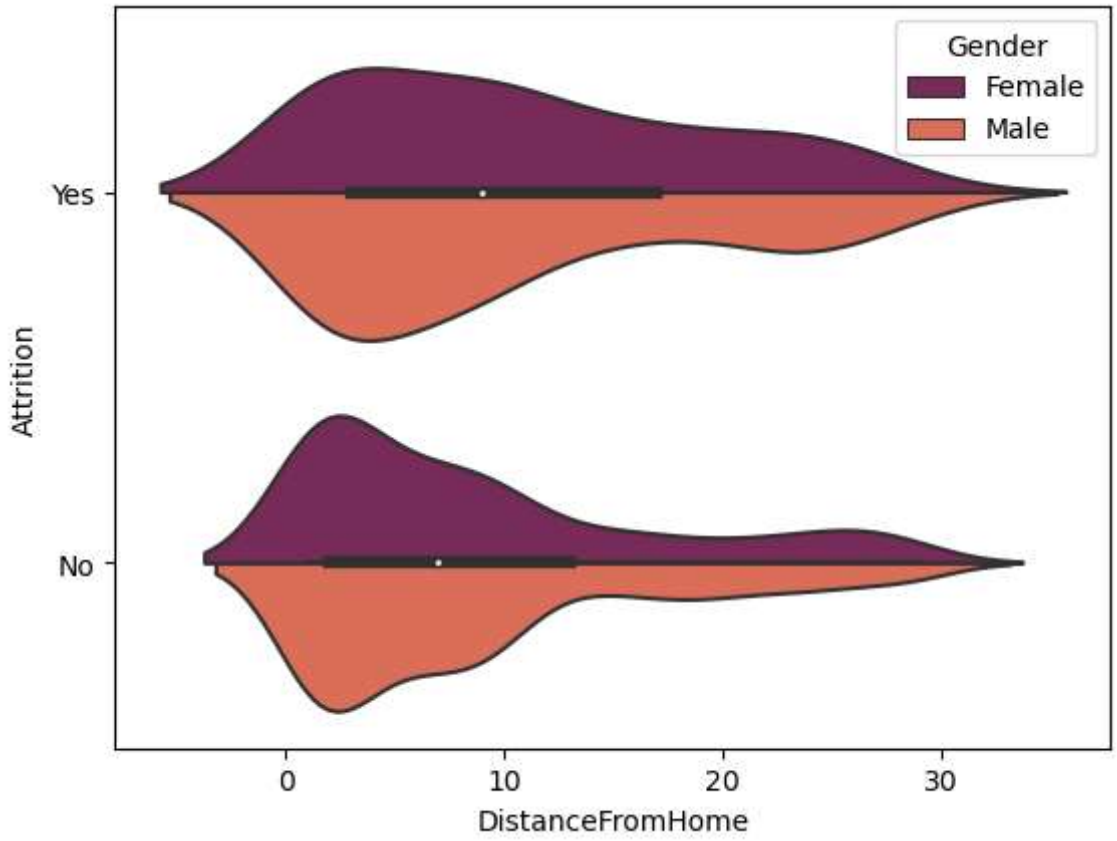
Job Involvement: The plot shows that employees with lower job involvement have a higher attrition rate compared to those with higher involvement. This indicates that employees who are more engaged and invested in their work are less likely to leave the company.

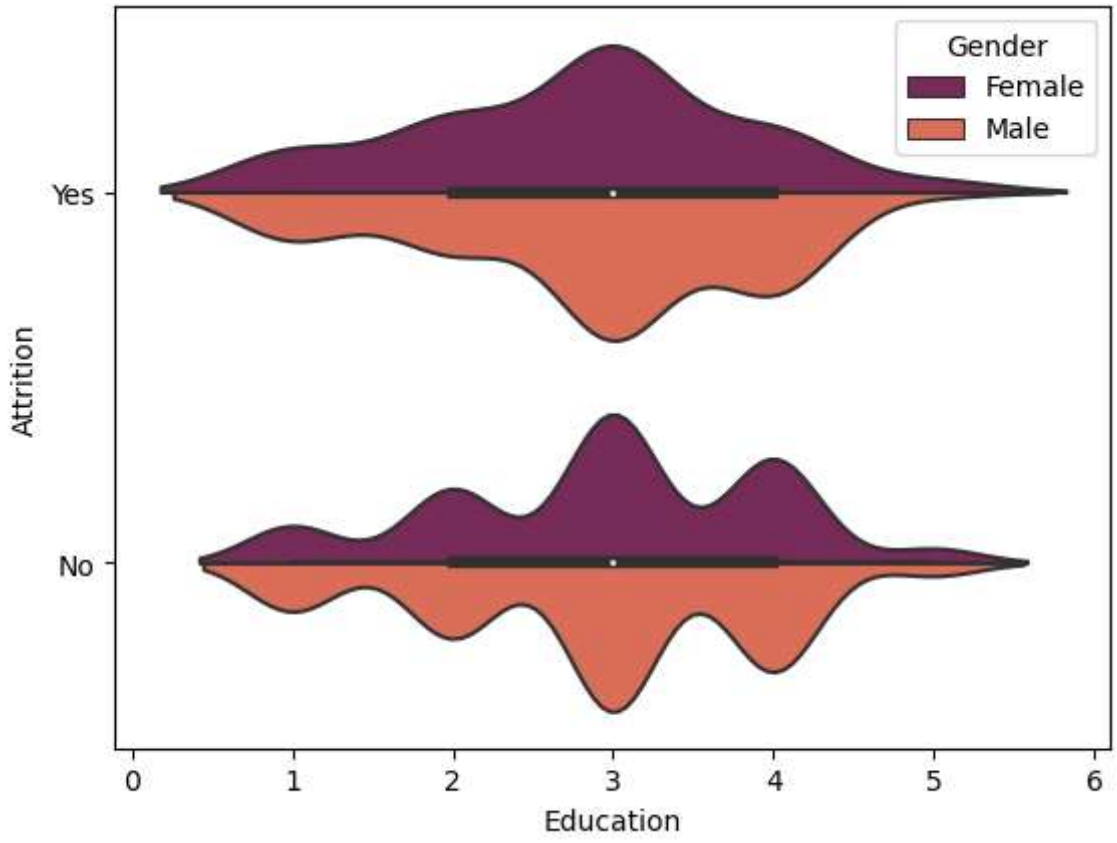
'Attrition','BusinessTravel','Department','EducationField','Gender','JobRole','MaritalStatus','OverTi

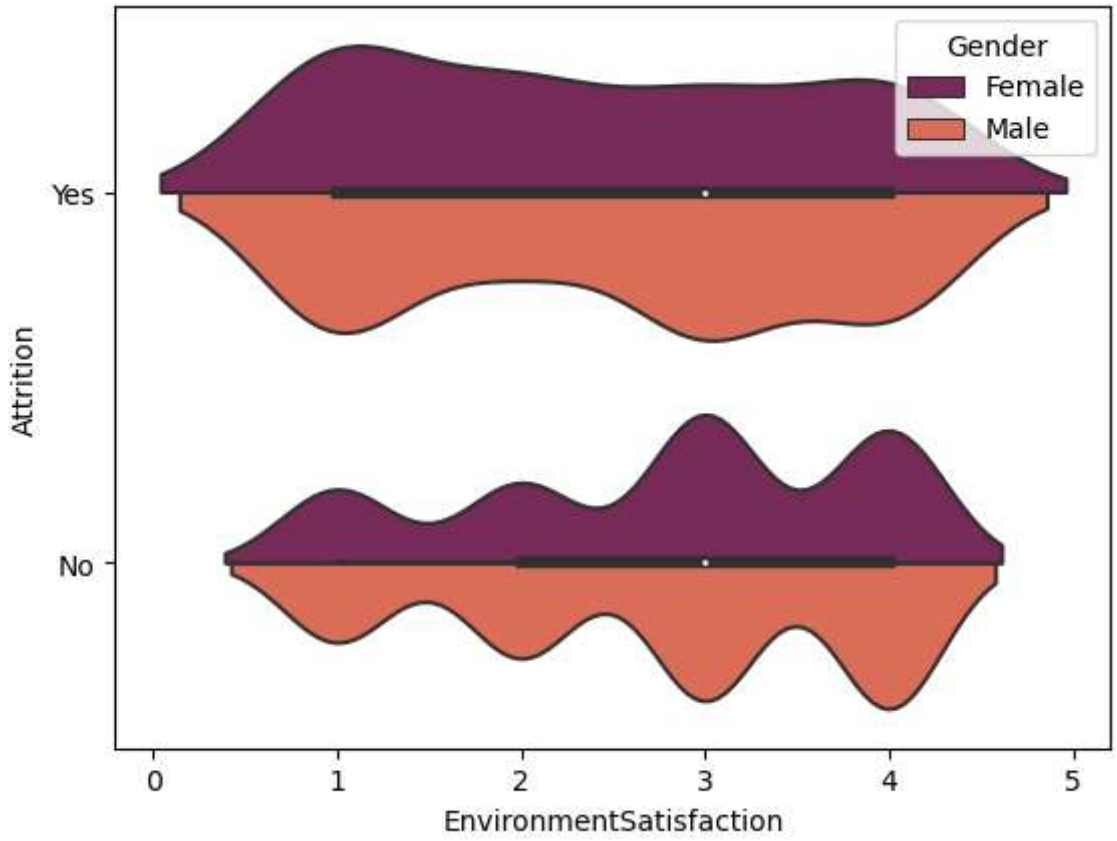
In [22]:
```
for x in columns:
    sns.violinplot(x=x,y='Attrition',hue='Gender',split=True,data=ibm,palette='roc
    plt.show()
```

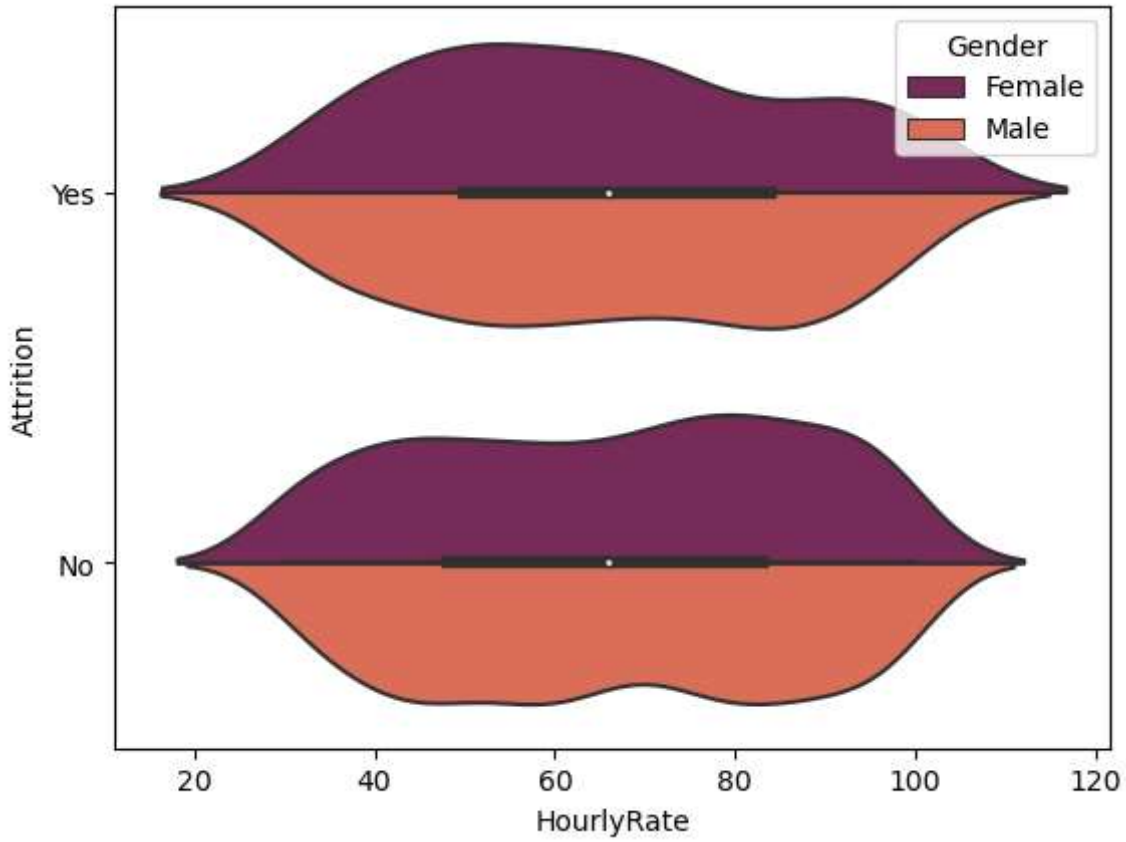


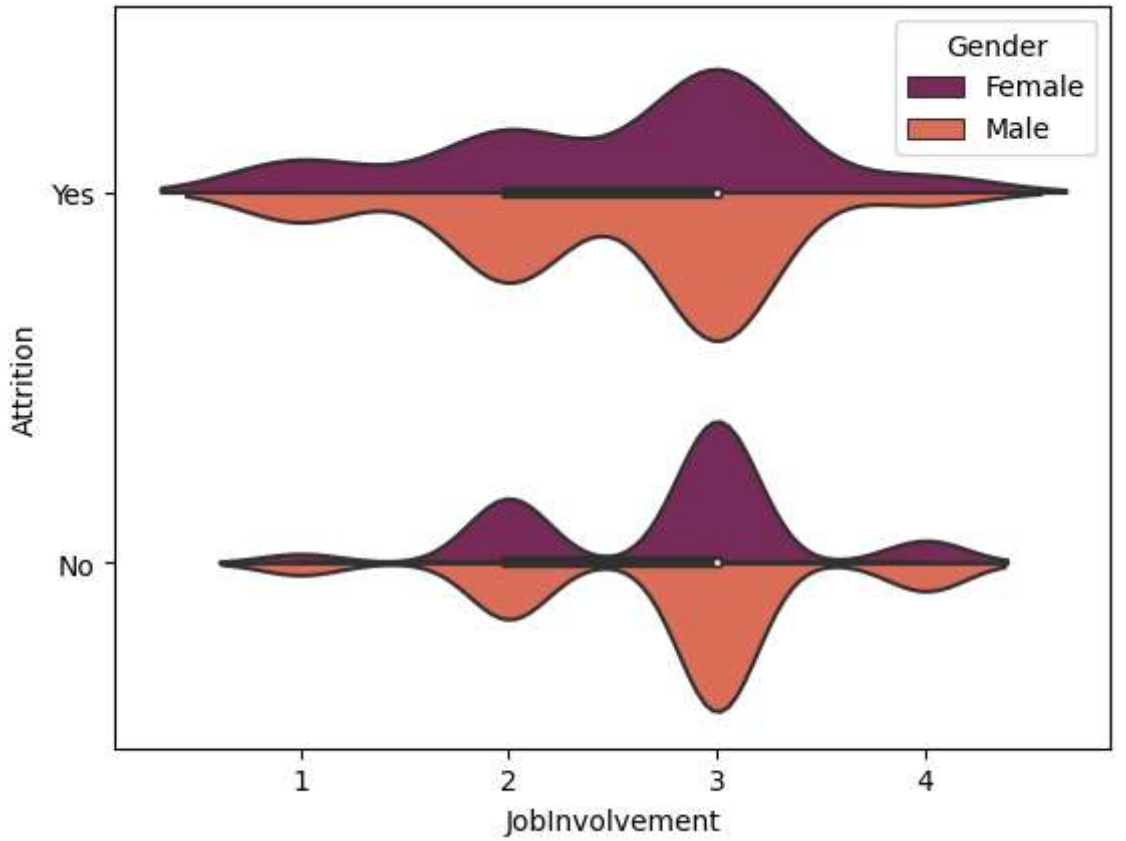

Gender
Female
Male
Yes
No
Attrition
0
500
1000
1500
DailyRate

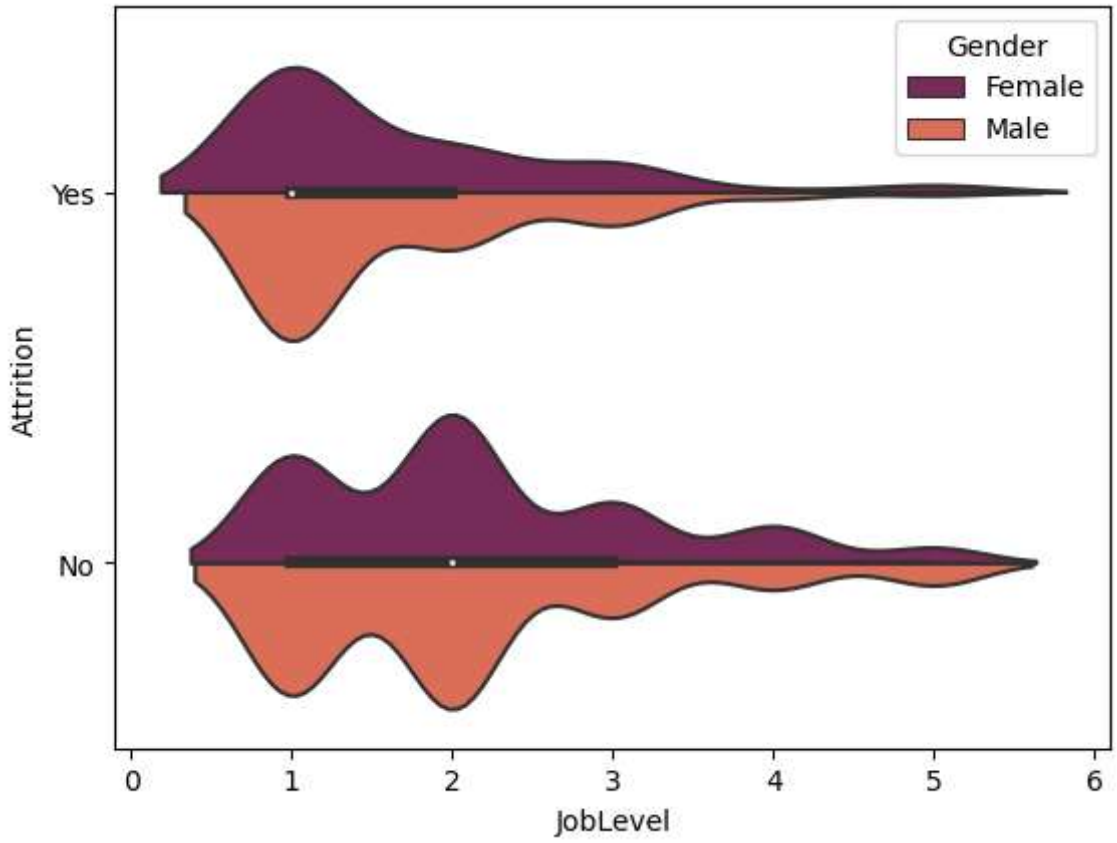

Gender
Female
Male
Yes
No
Attrition
0
10
20
30
DistanceFromHome

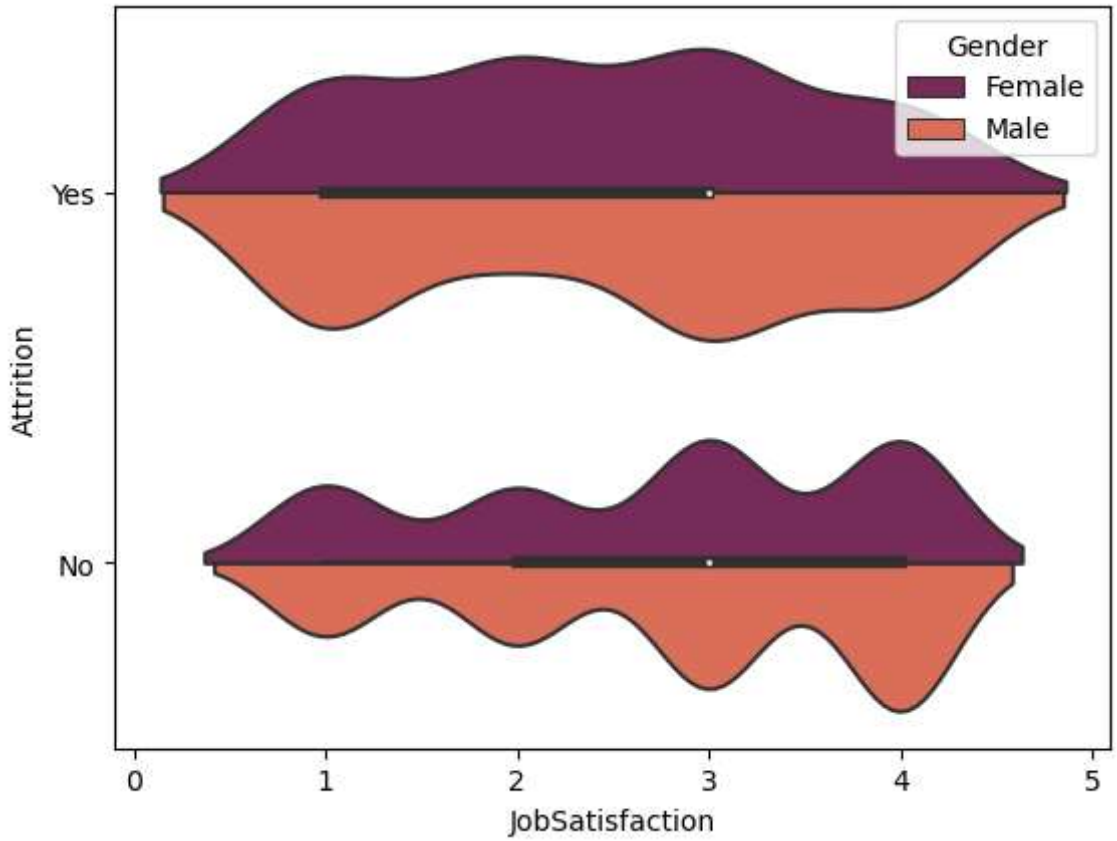

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
4
5
6
Education

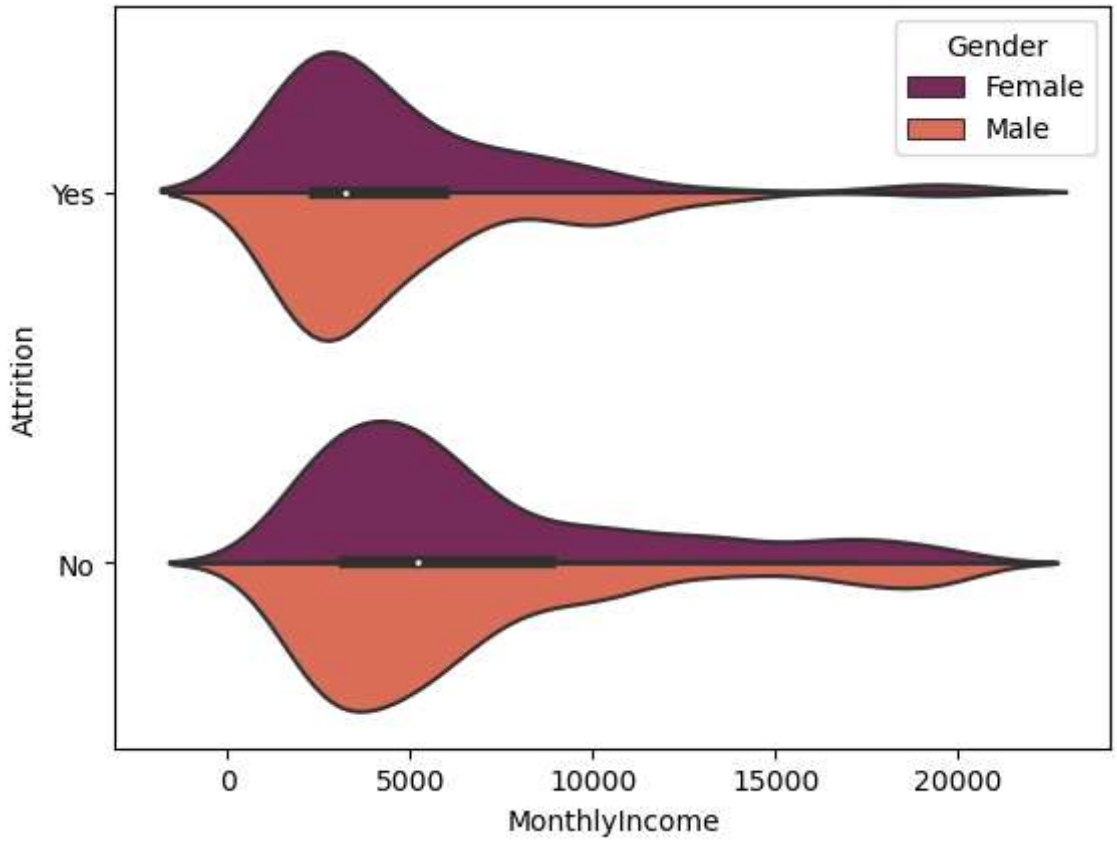

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
4
5
EnvironmentSatisfaction

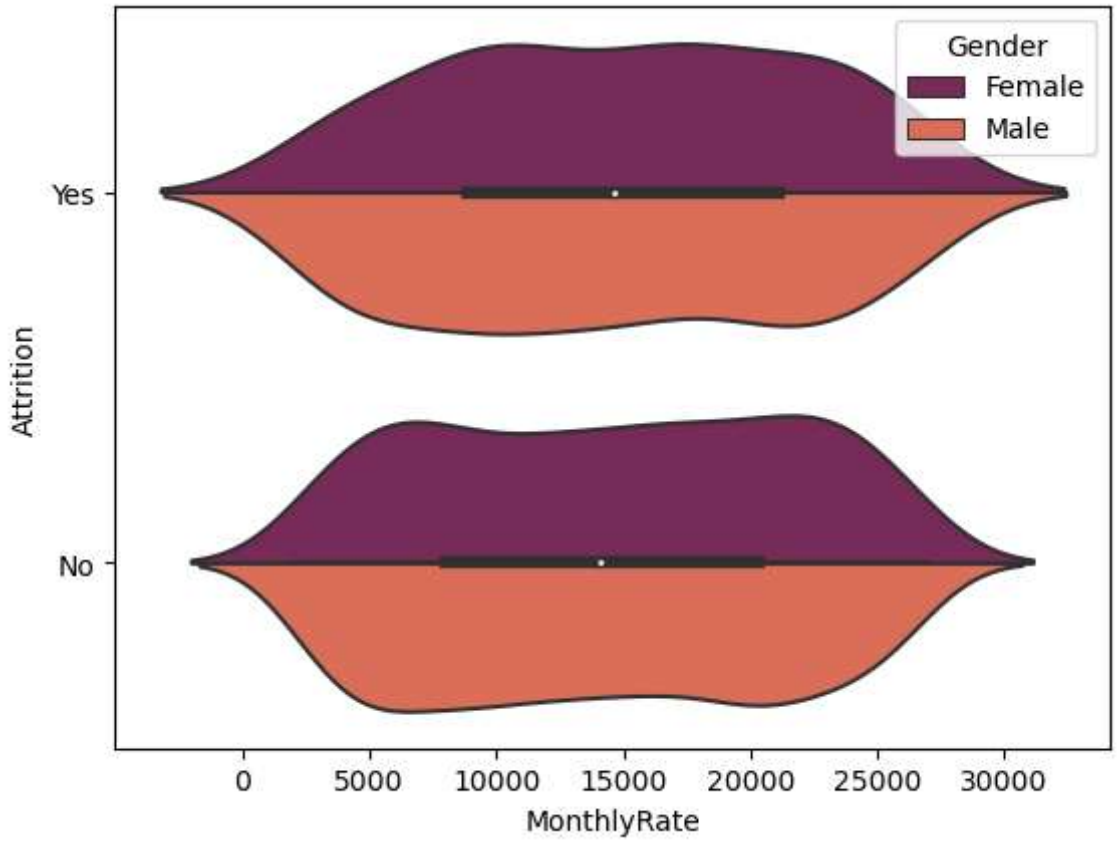

Gender
Female
Male
Yes
No
Attrition
20
40
60
80
100
120
HourlyRate

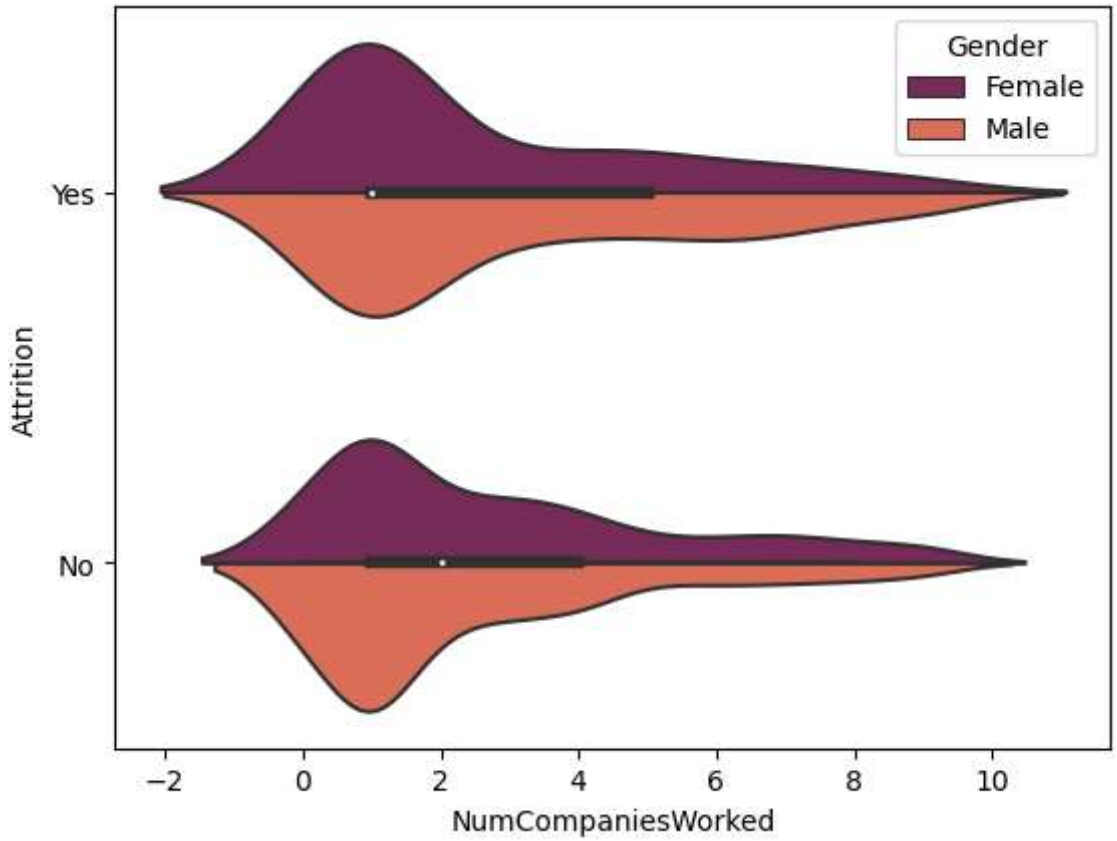

Gender
Female
Male
Yes
No
Attrition
1
2
3
4
JobInvolvement

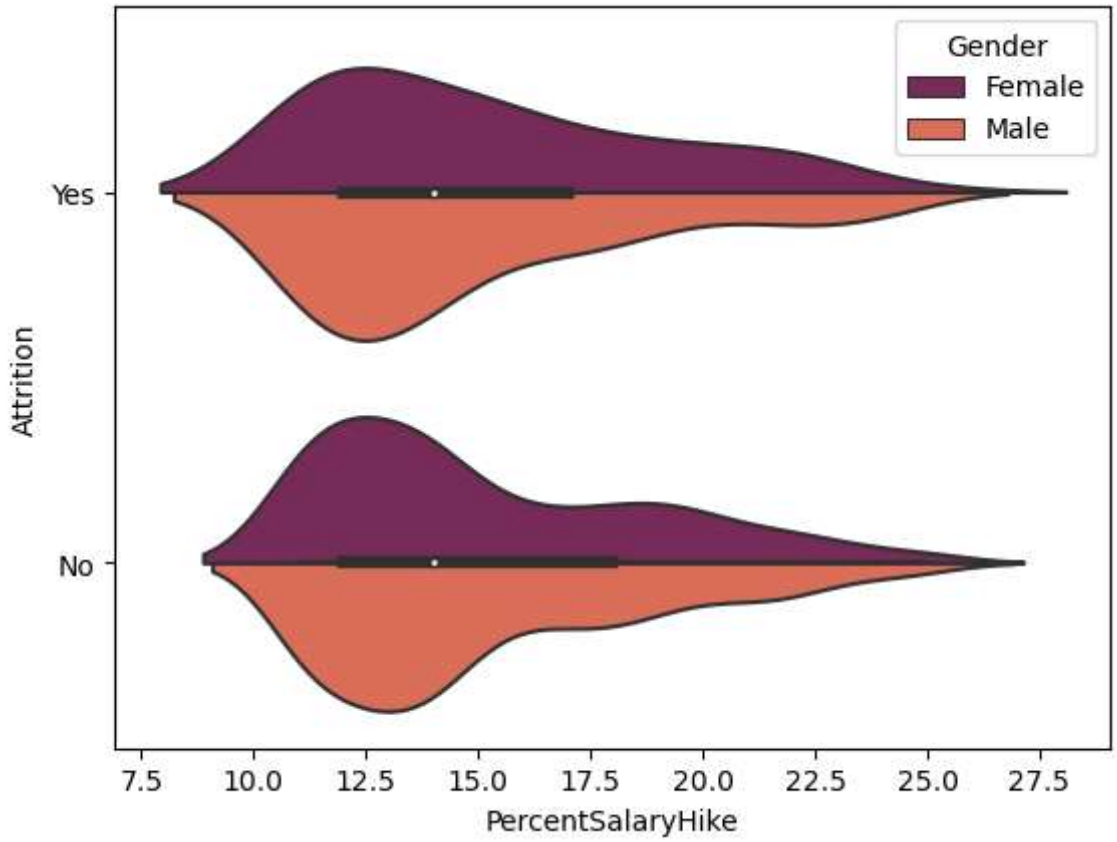

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
4
5
6
JobLevel

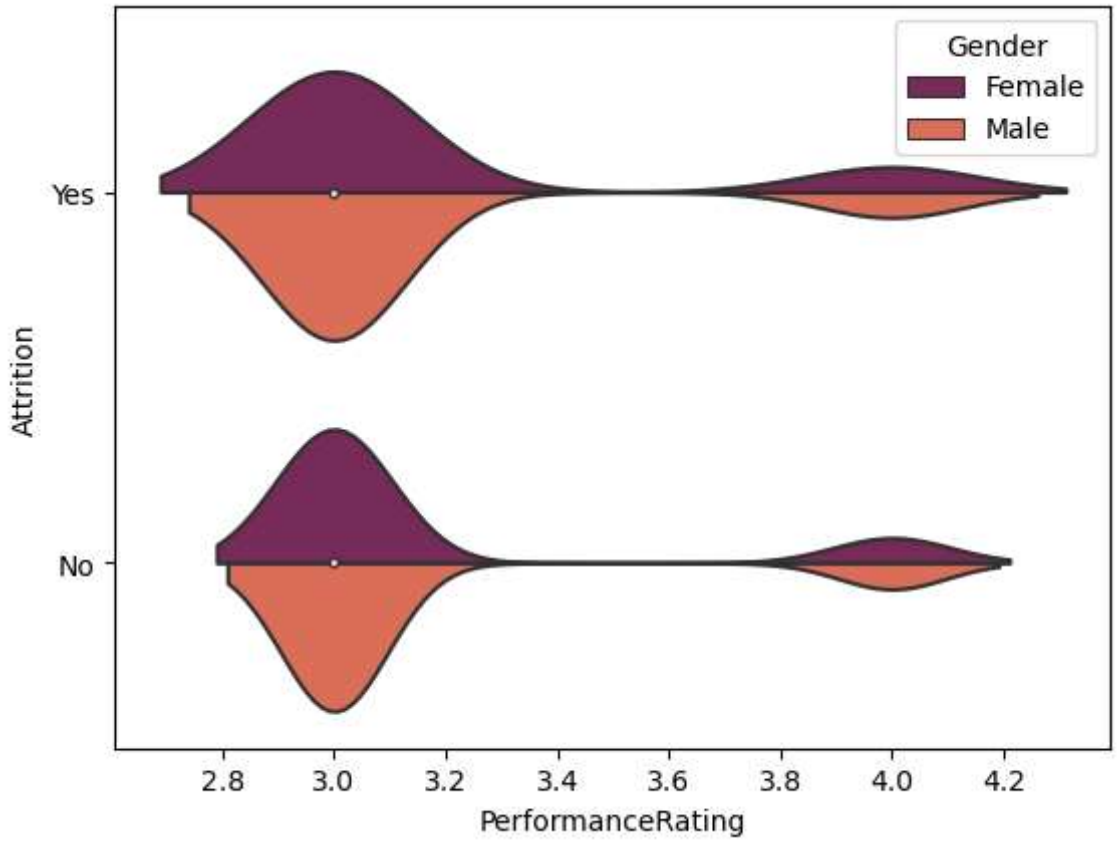

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
4
5
JobSatisfaction

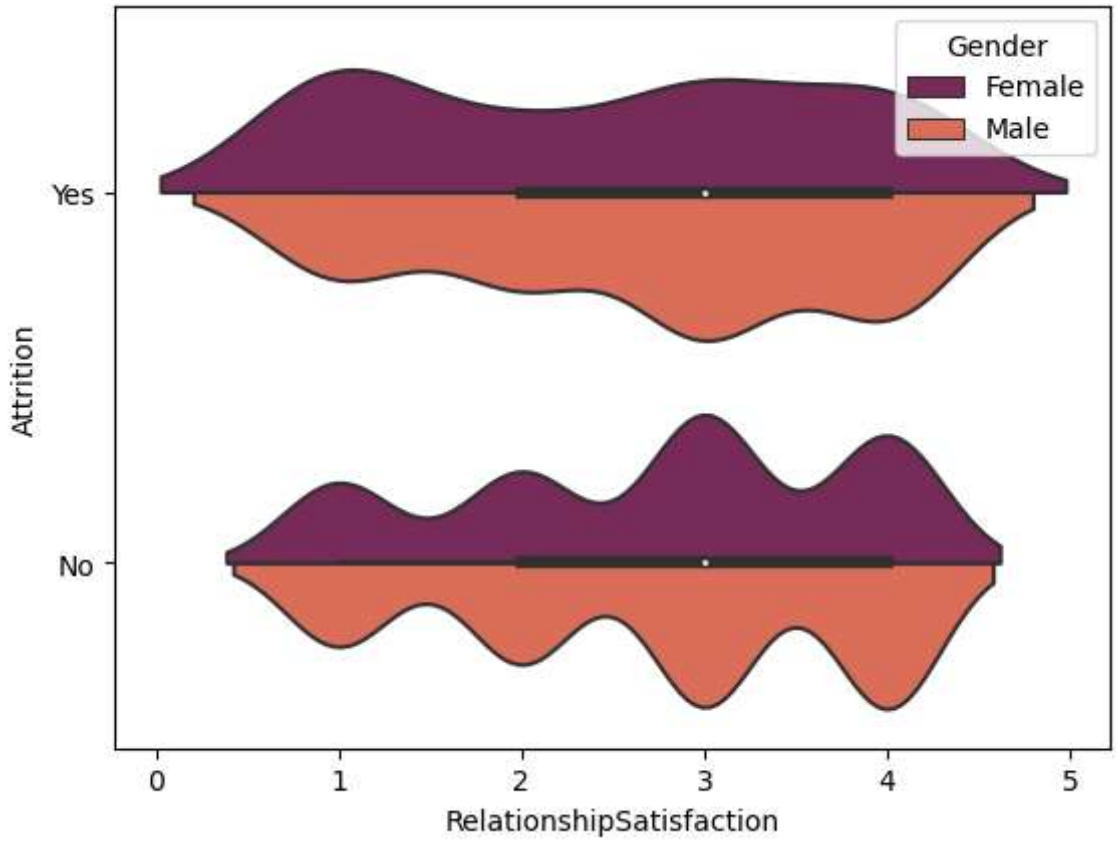

Gender
Female
Male
Yes
No
Attrition
0
5000
10000
15000
20000
MonthlyIncome

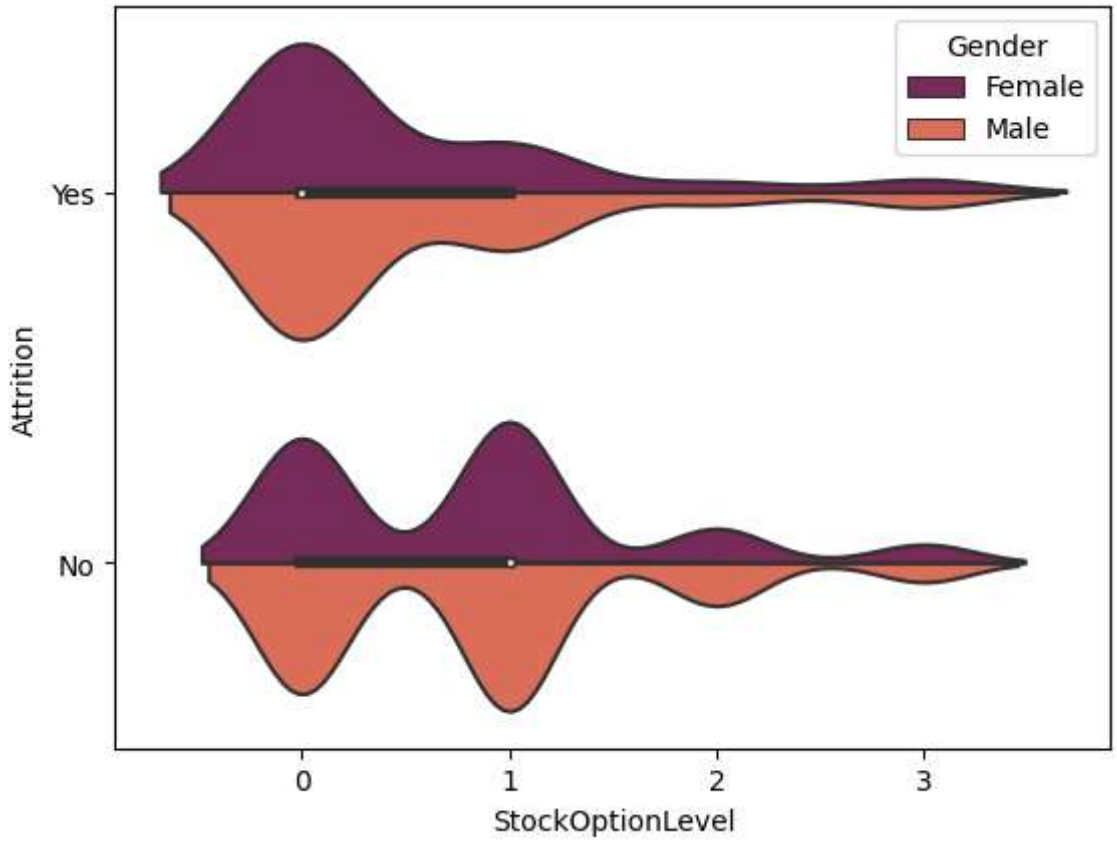

Gender
Female
Male
Yes
No
Attrition
0
5000
10000
15000
20000
25000
30000
MonthlyRate

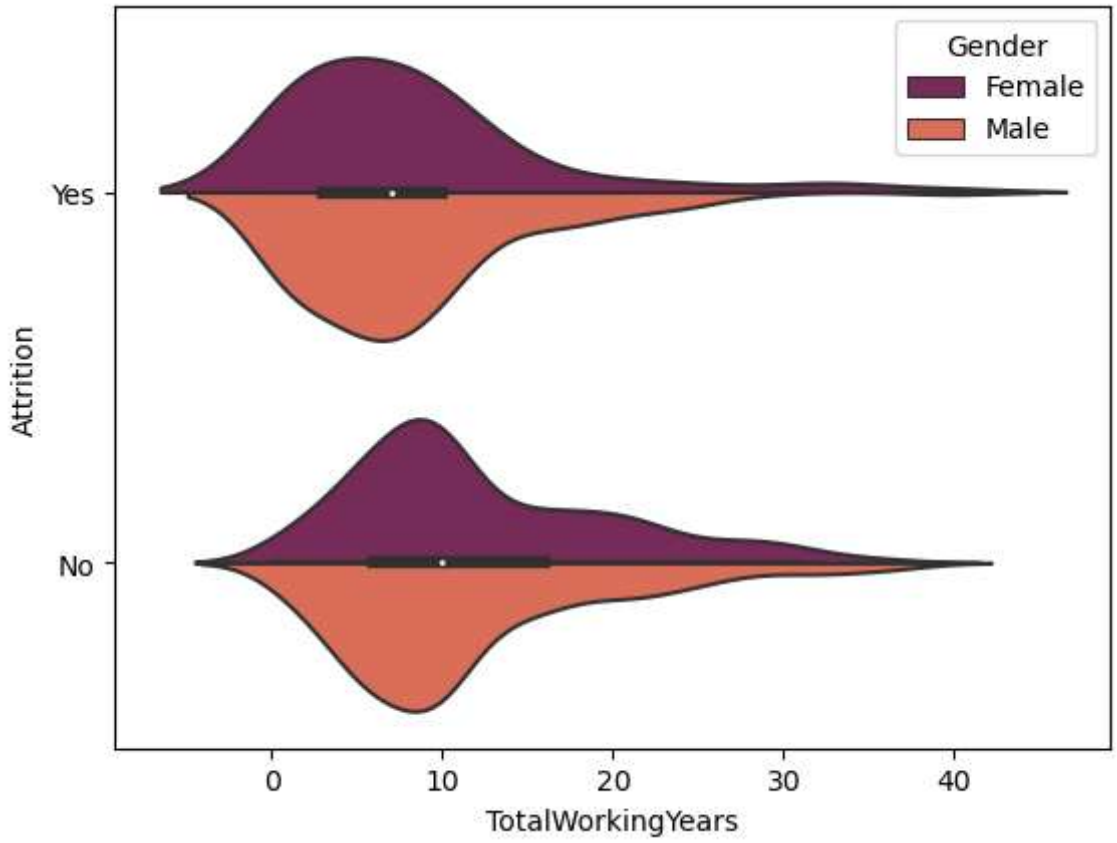

Gender
Female
Male
Yes
No
Attrition
−2
0
2
4
6
8
10
NumCompaniesWorked

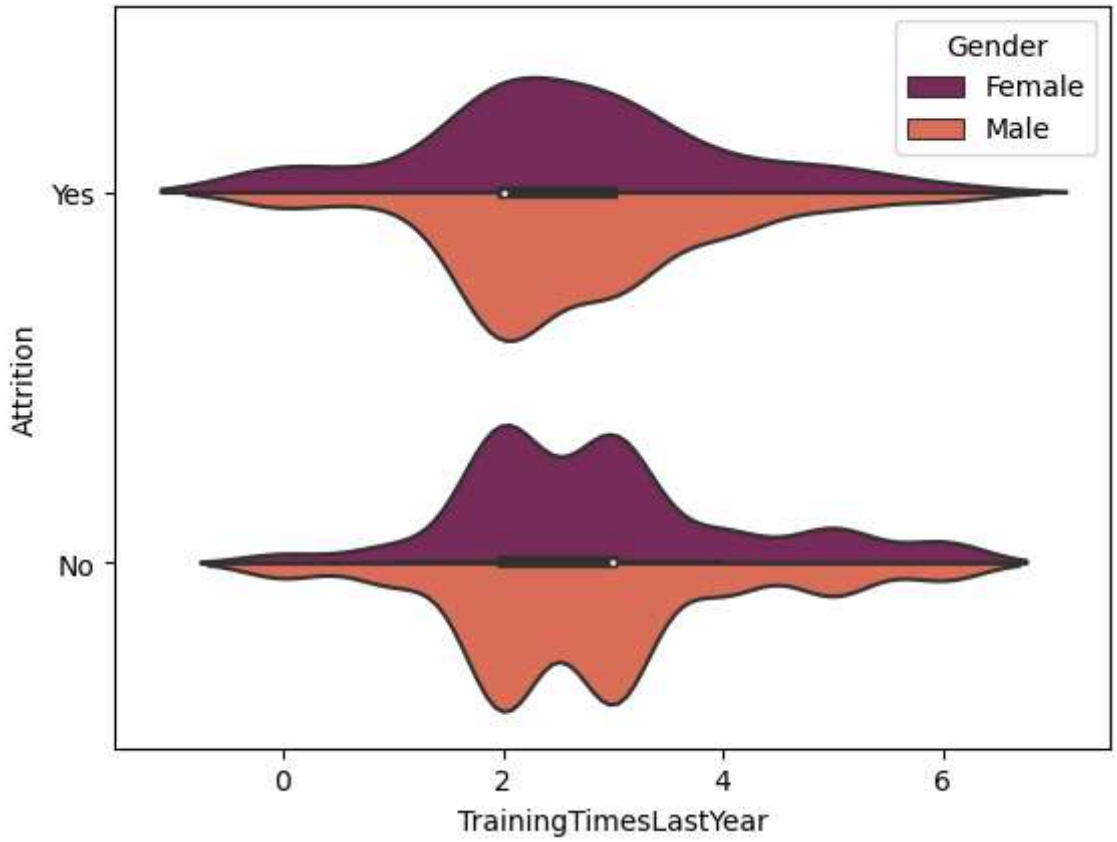

Gender
Female
Male
Yes
No
Attrition
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
PercentSalaryHike

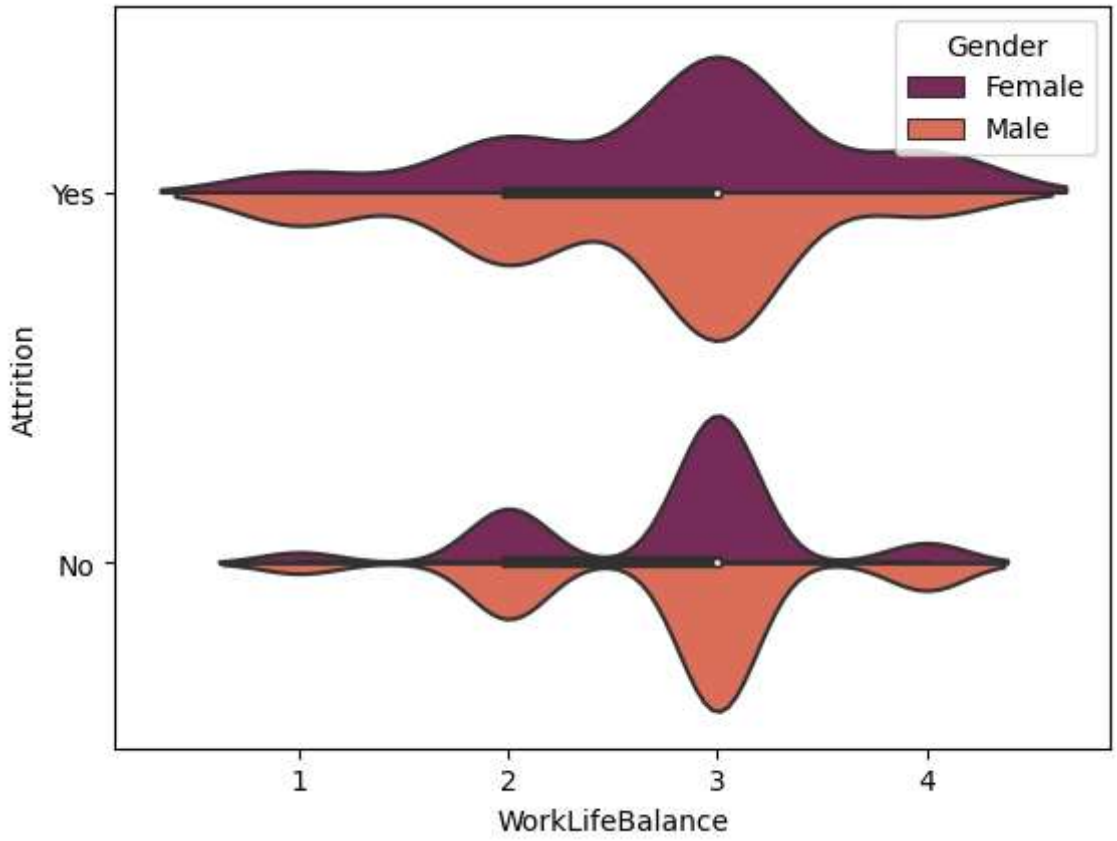

Gender
Female
Male
Yes
No
Attrition
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2
PerformanceRating

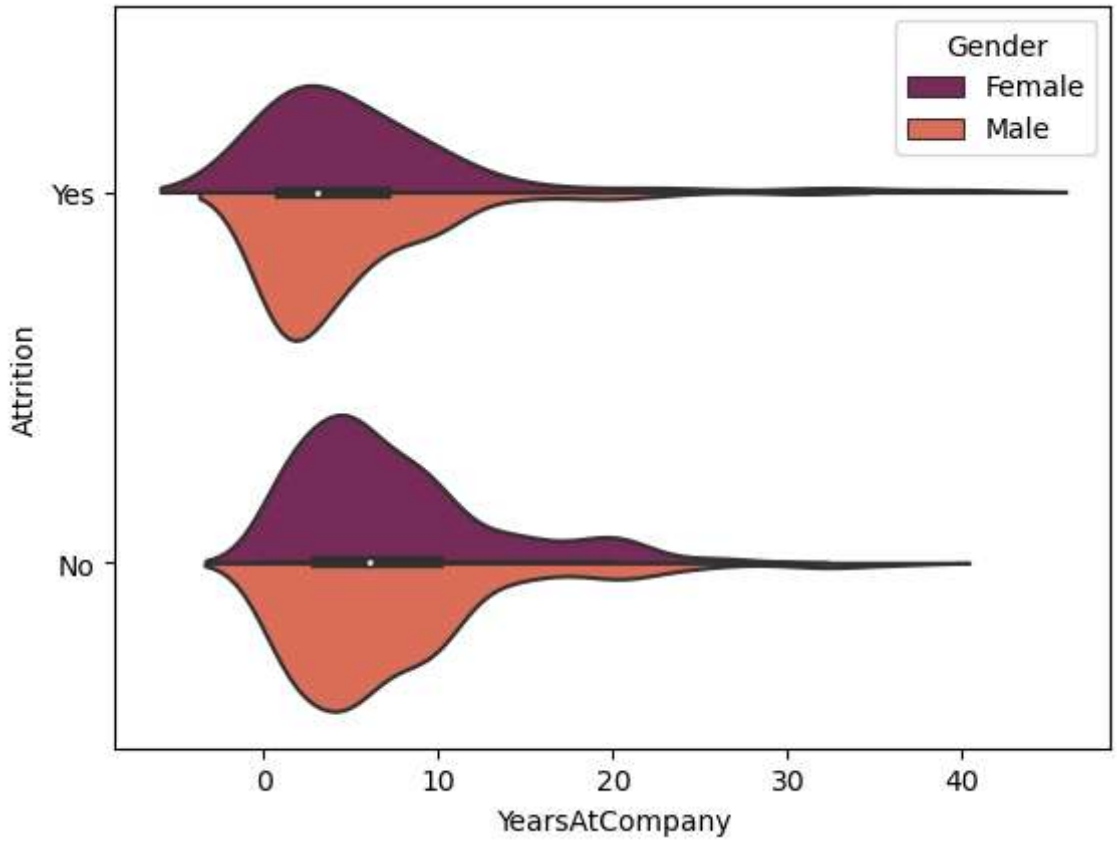

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
4
5
RelationshipSatisfaction

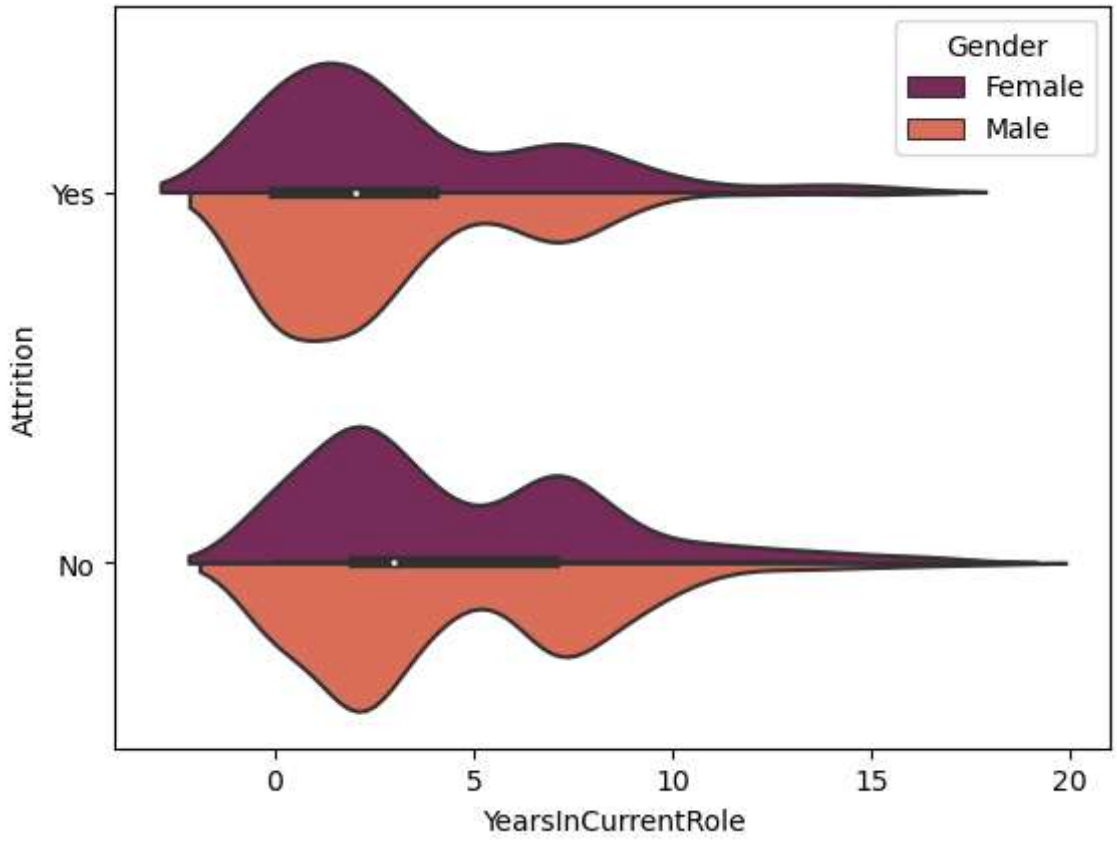

Gender
Female
Male
Yes
No
Attrition
0
1
2
3
StockOptionLevel

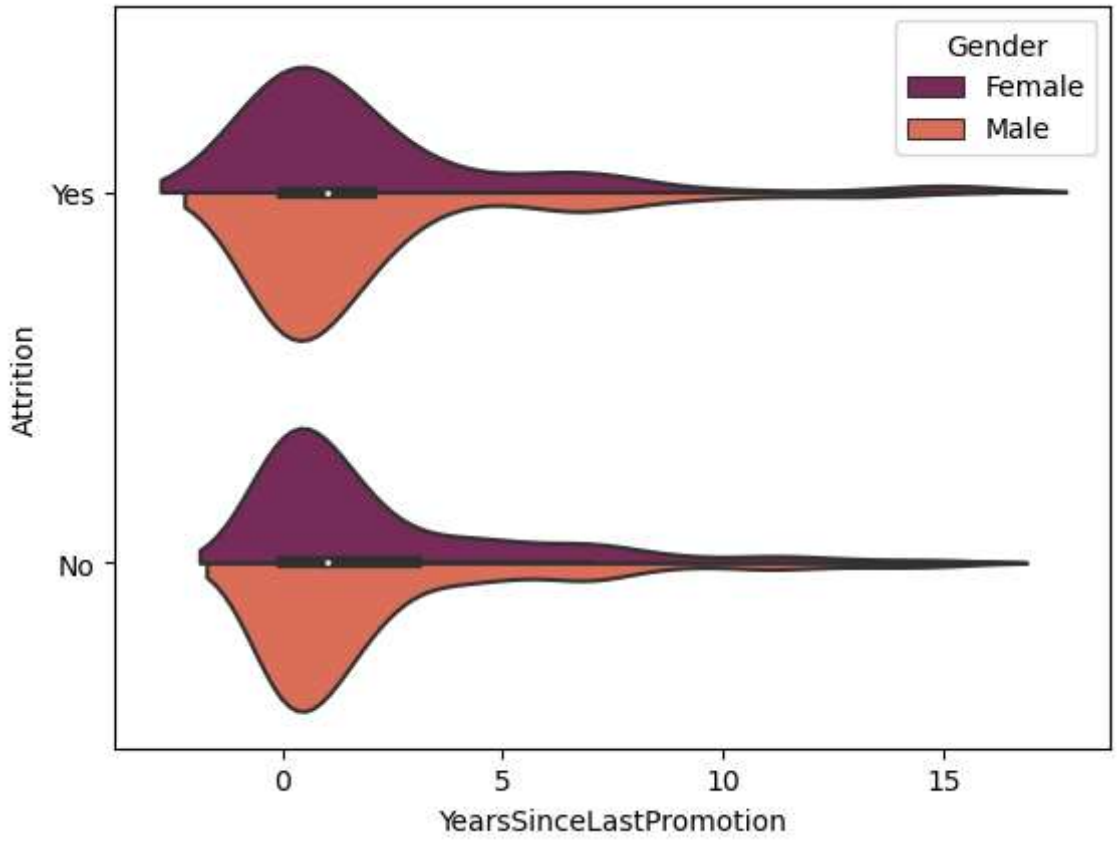

Gender
Female
Male
Yes
No
Attrition
0
10
20
30
40
TotalWorkingYears

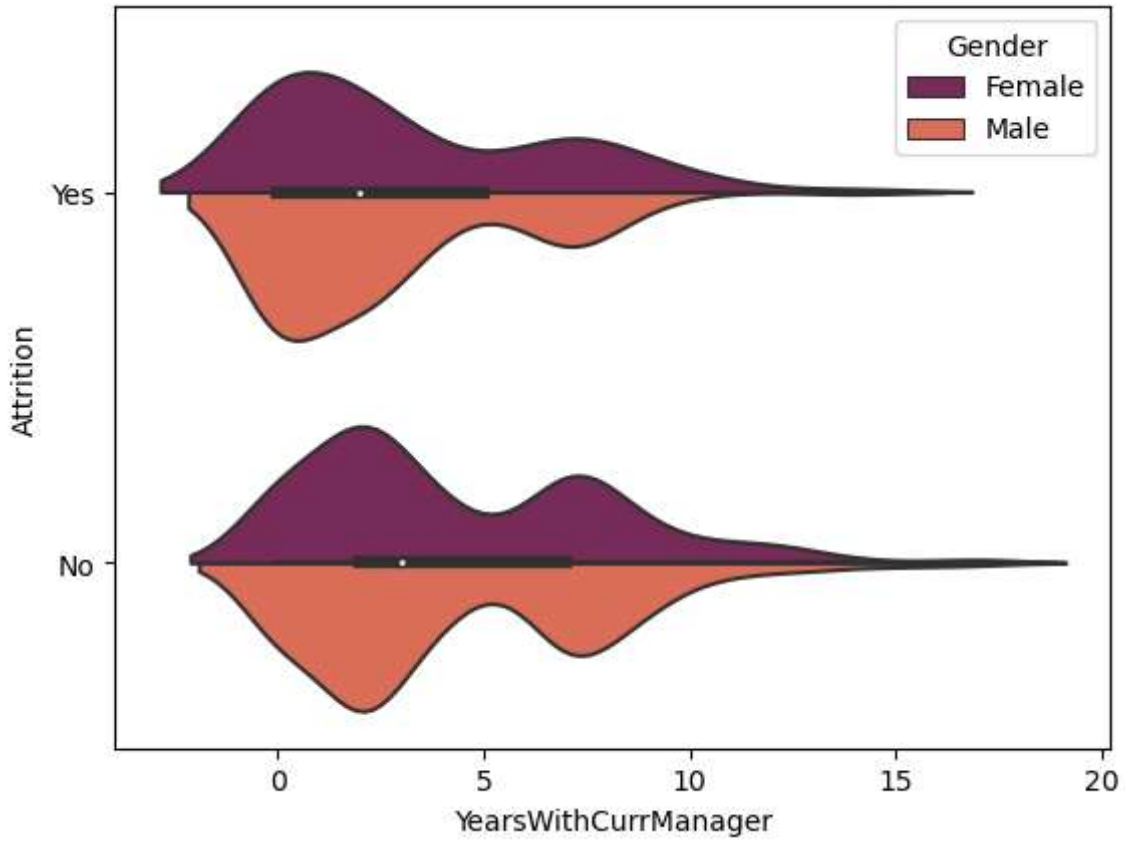

Gender
Female
Male
Yes
No
Attrition
0
2
4
6
TrainingTimesLastYear


Gender
Female
Male
Yes
No
Attrition
1
2
3
4
WorkLifeBalance

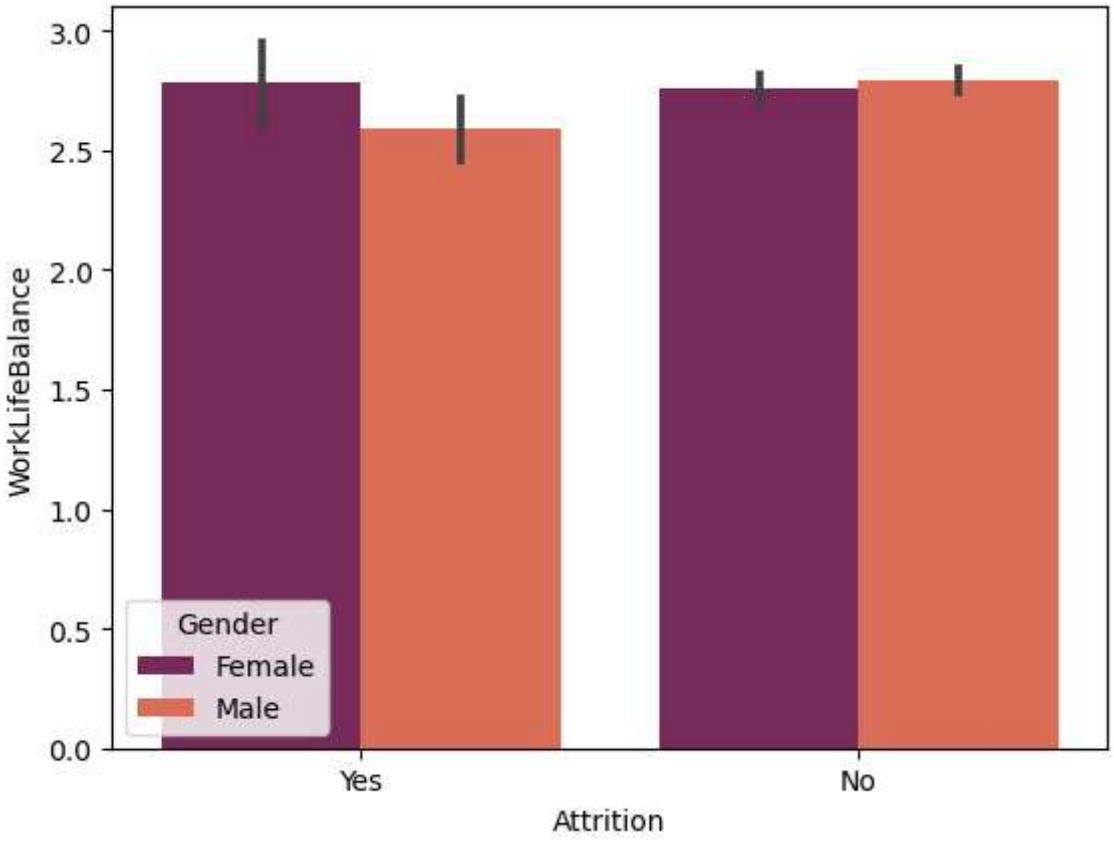

Gender
Female
Male
Yes
No
Attrition
0
10
20
30
40
YearsAtCompany

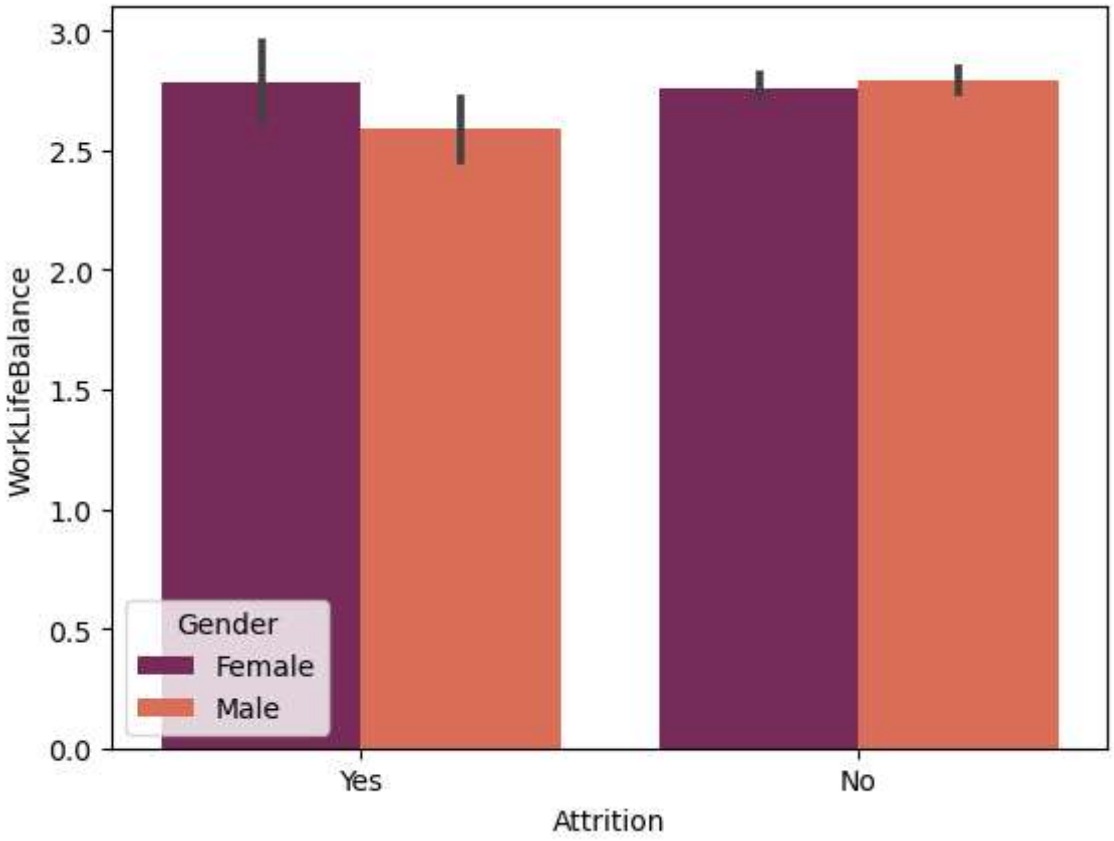

Gender
Female
Male
Yes
No
Attrition
0
5
10
15
20
YearsInCurrentRole

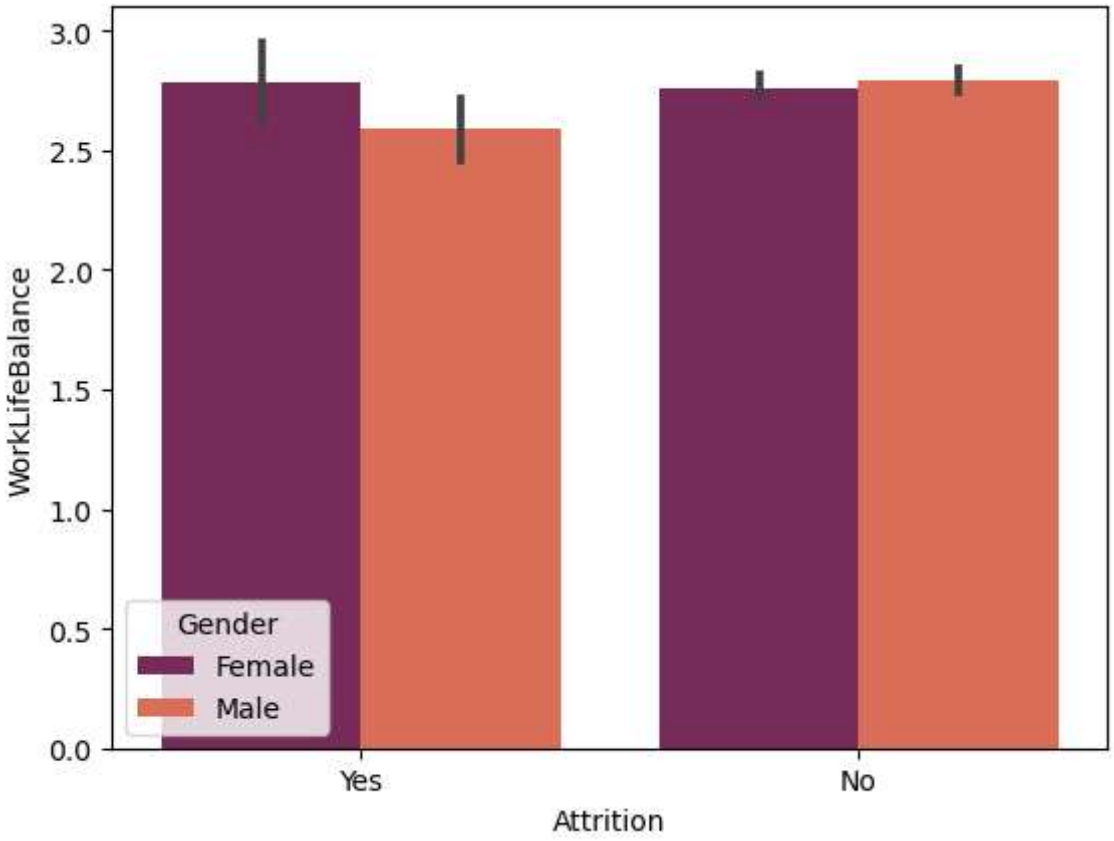


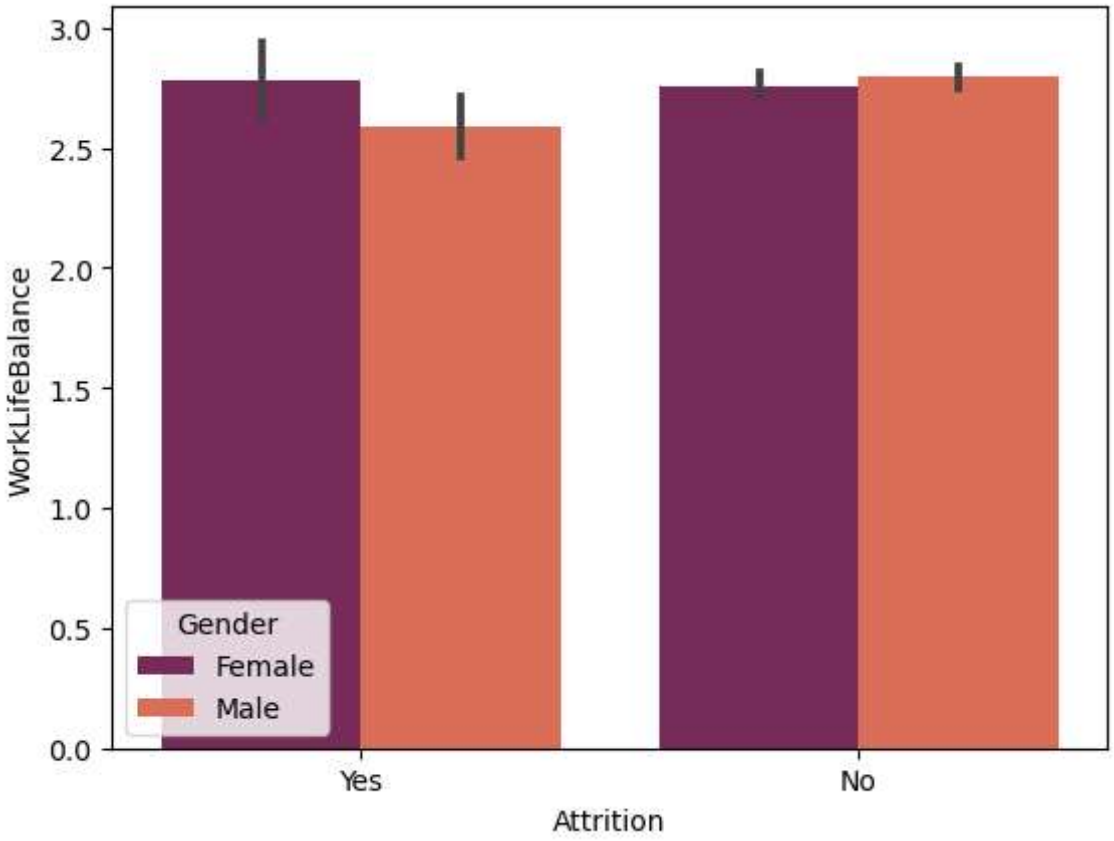


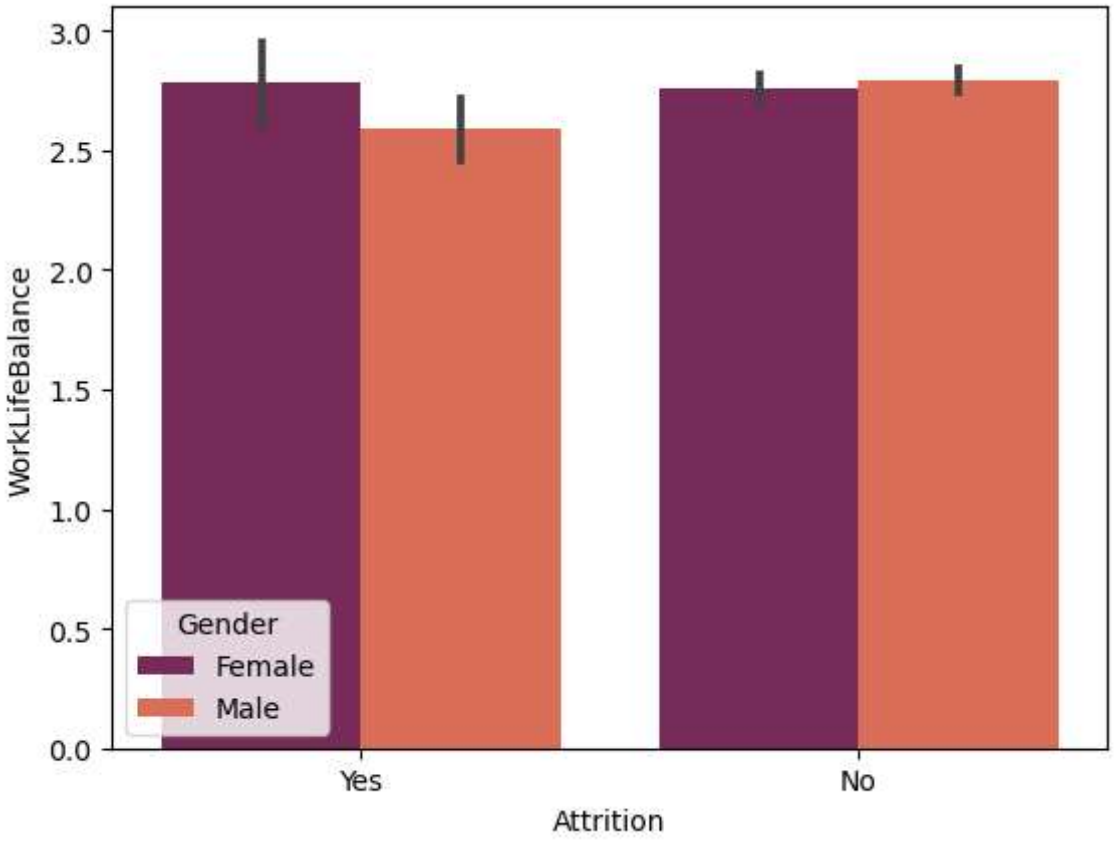
In [23]:
```
for x in columns:
    sns.barplot(x='Attrition',y=col,hue='Gender',data=ibm,palette='rocket')
    plt.show()
```

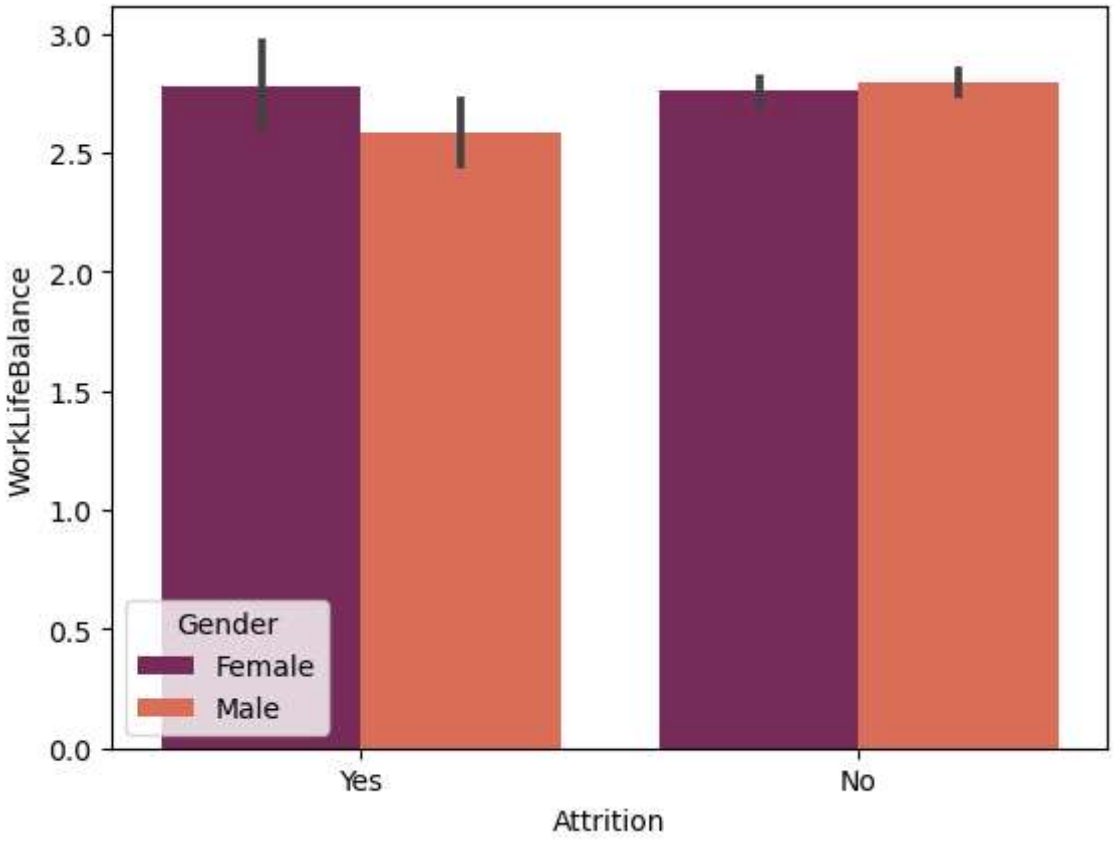

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

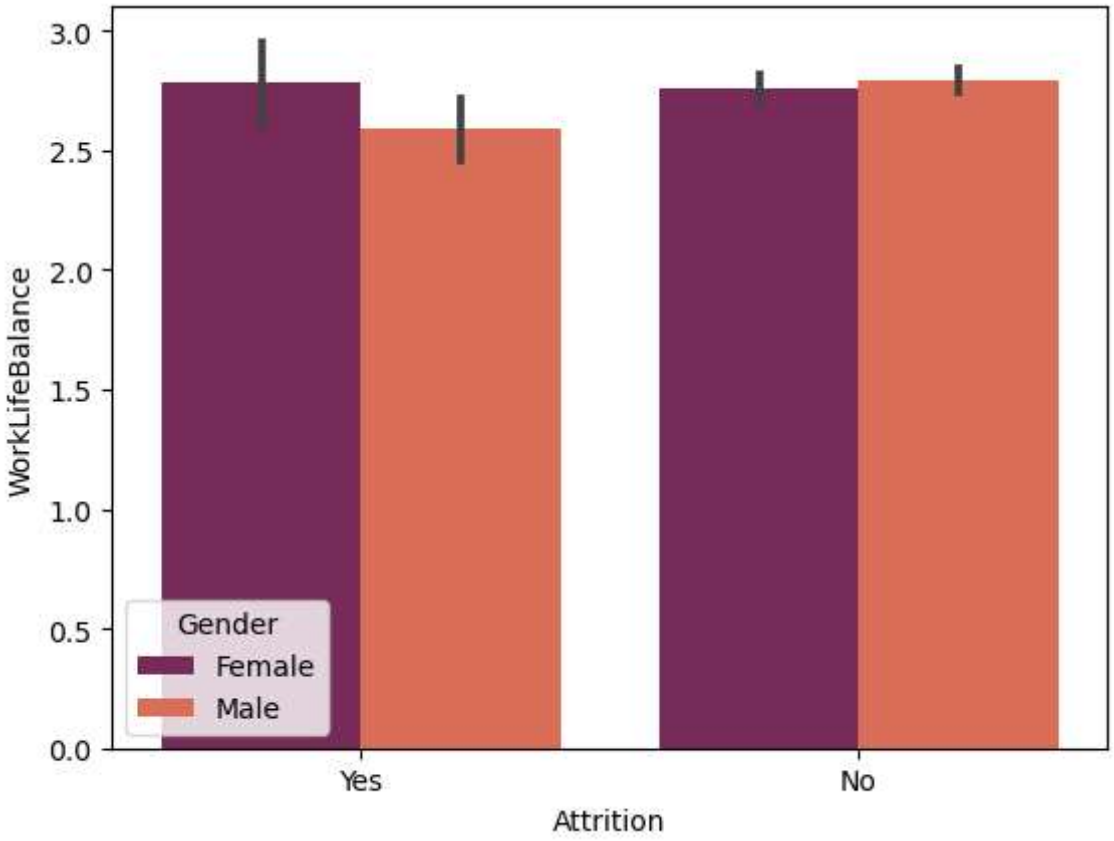

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

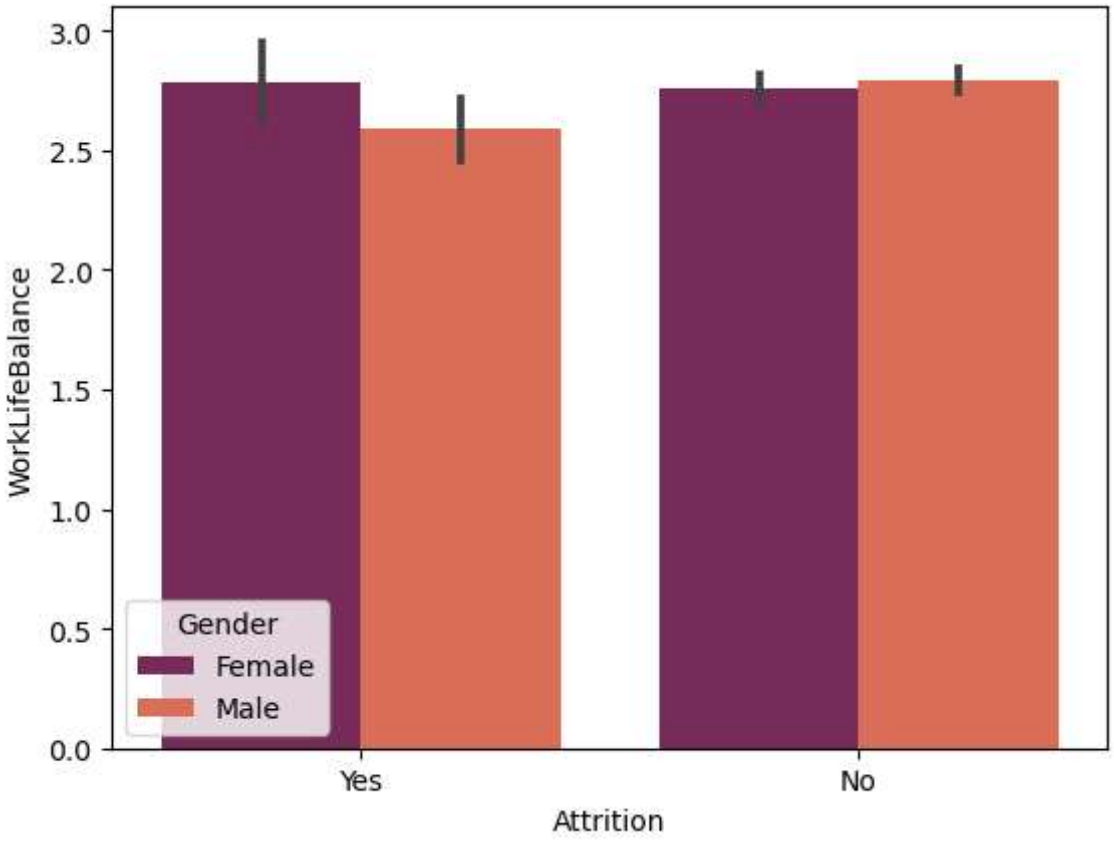

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

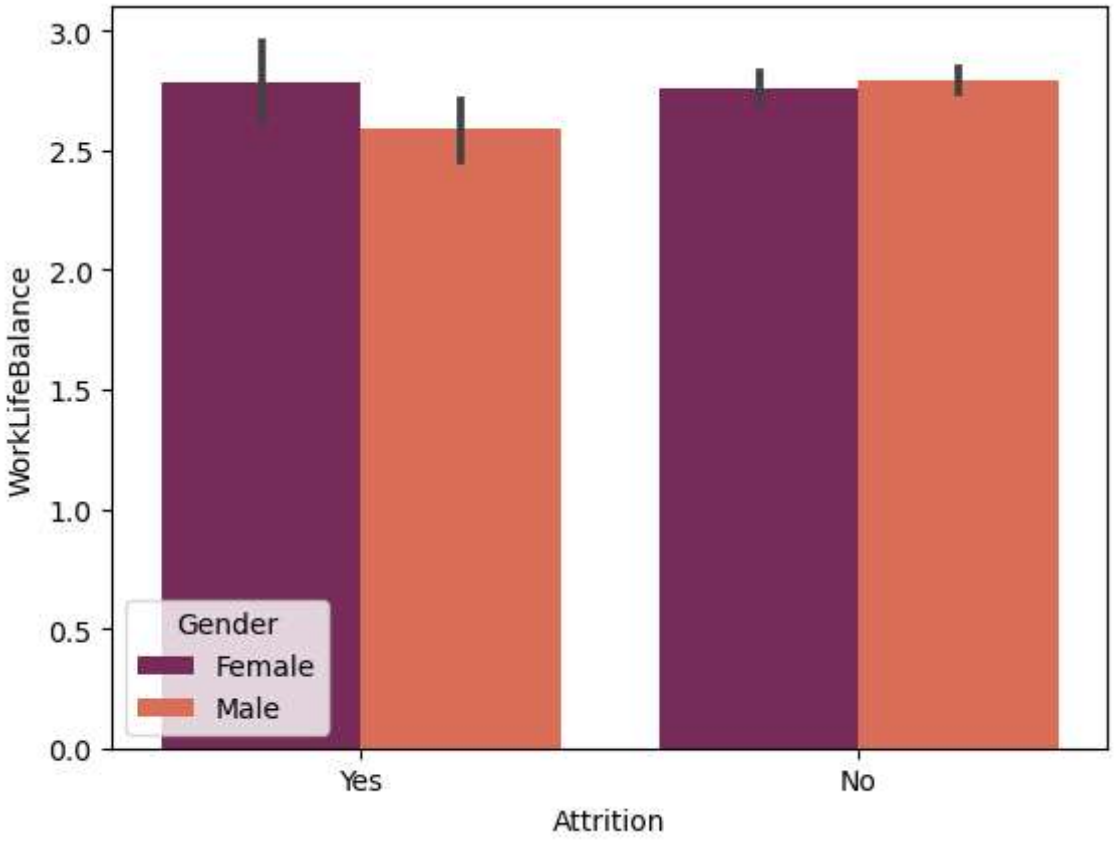

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

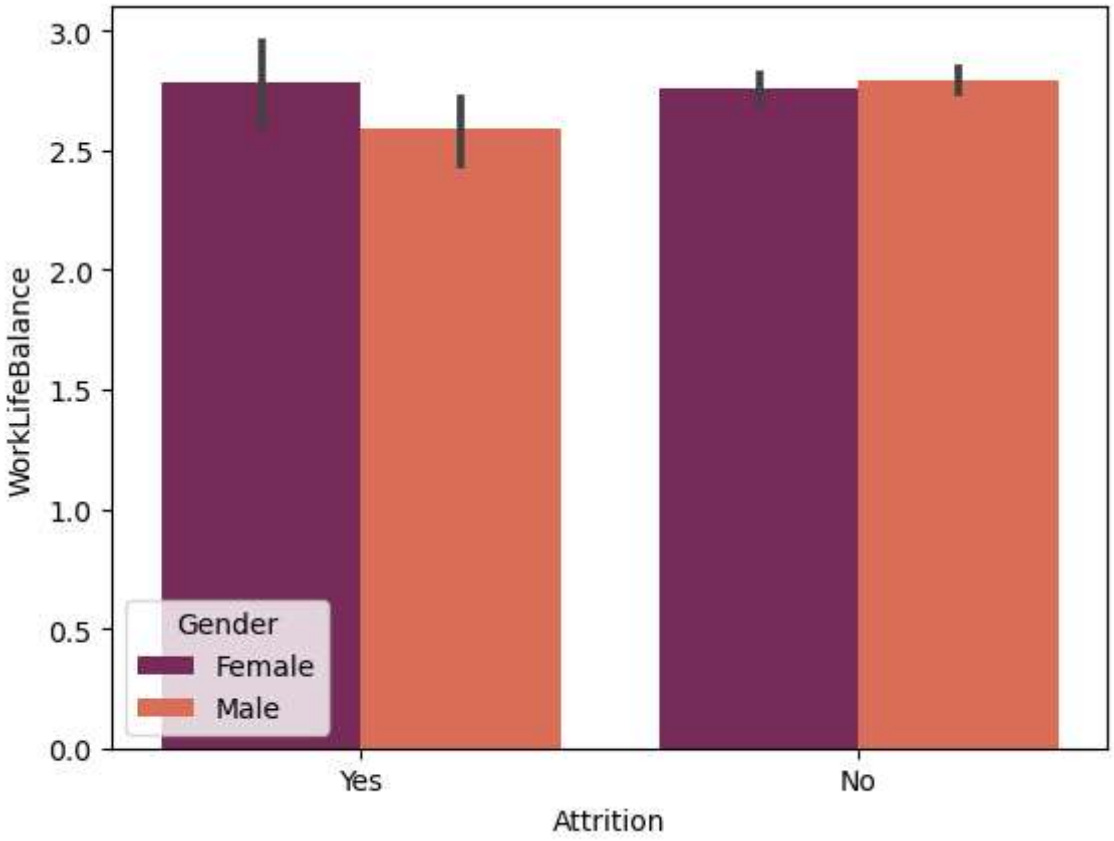

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

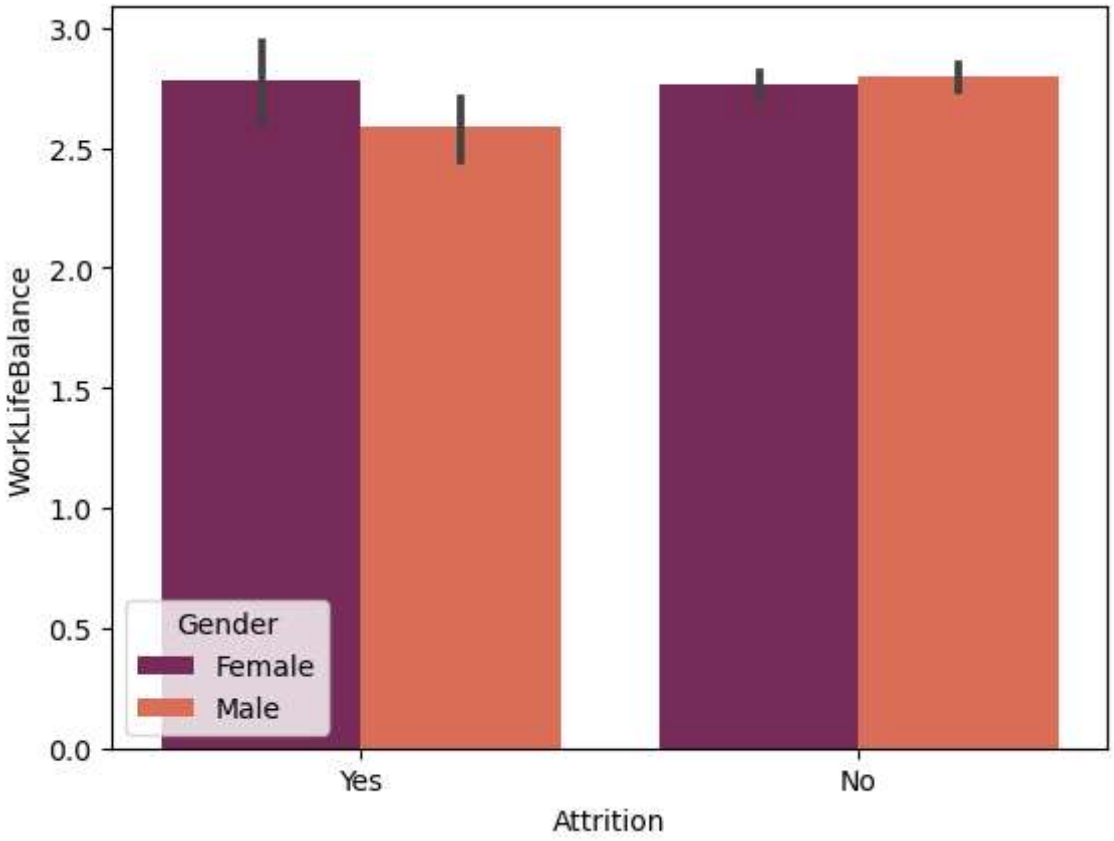

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

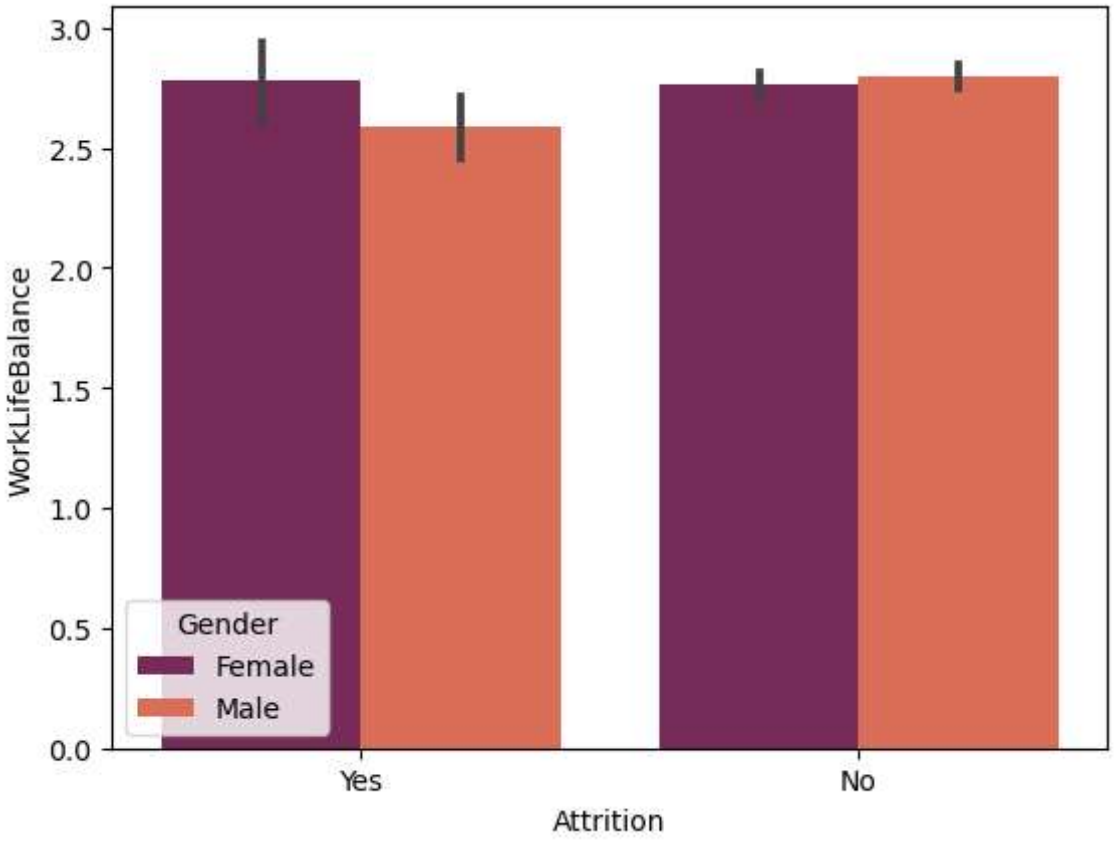

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

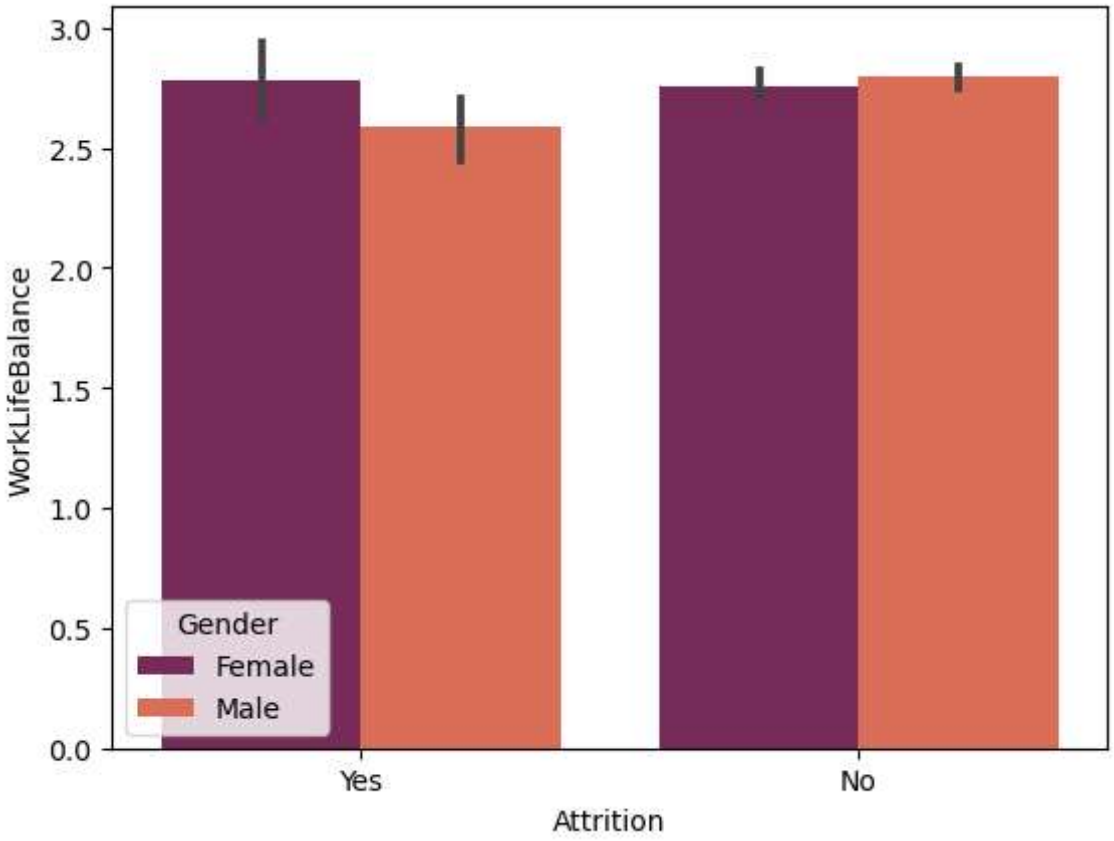

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

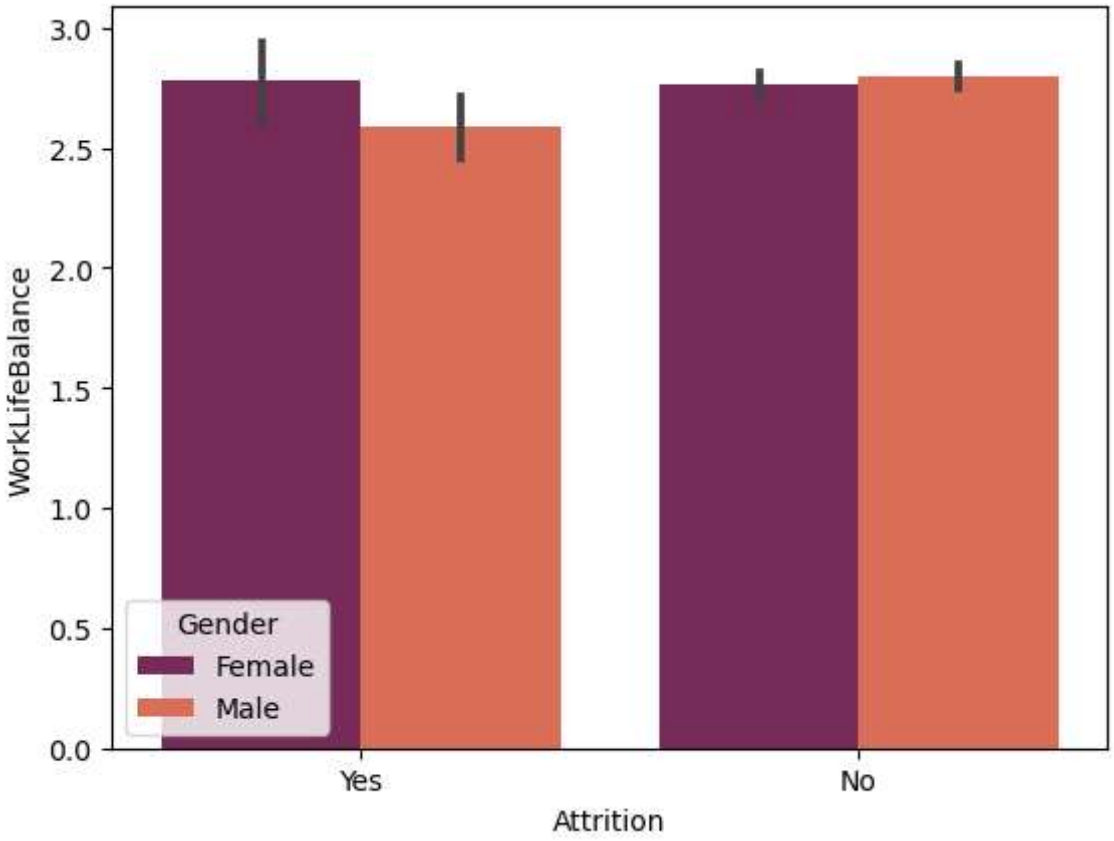

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

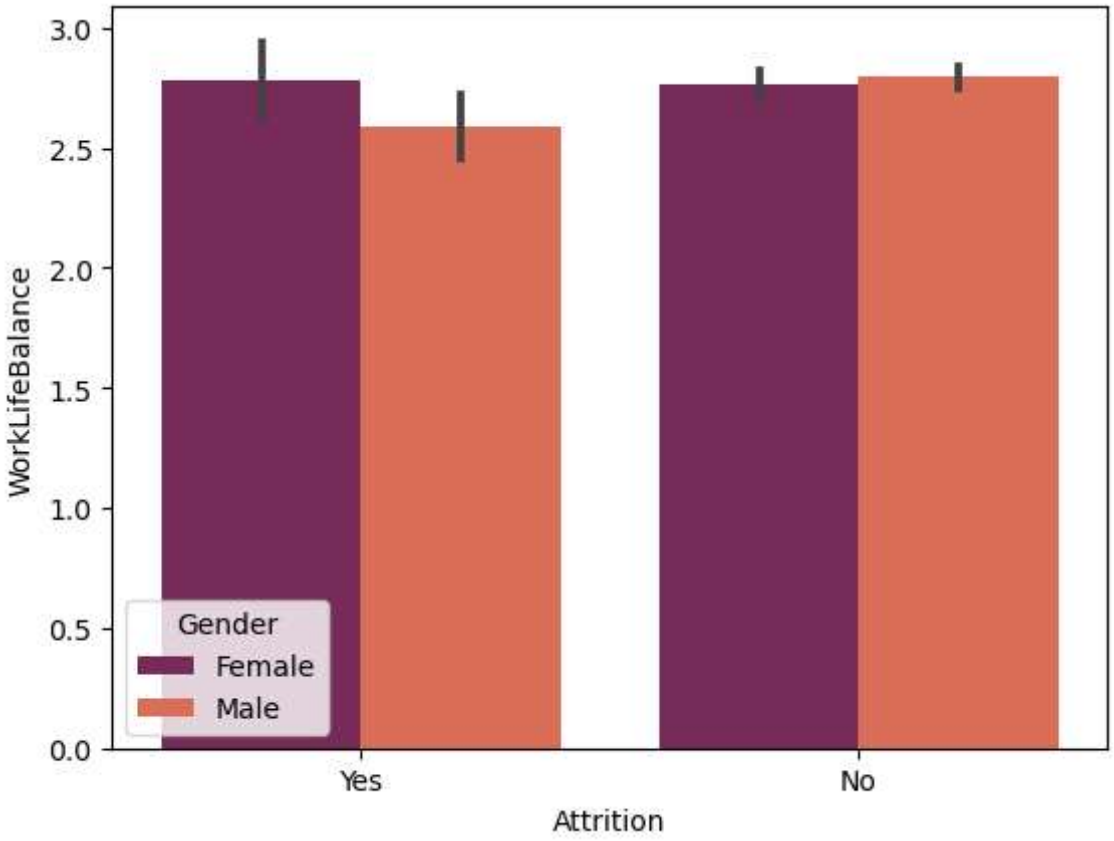

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

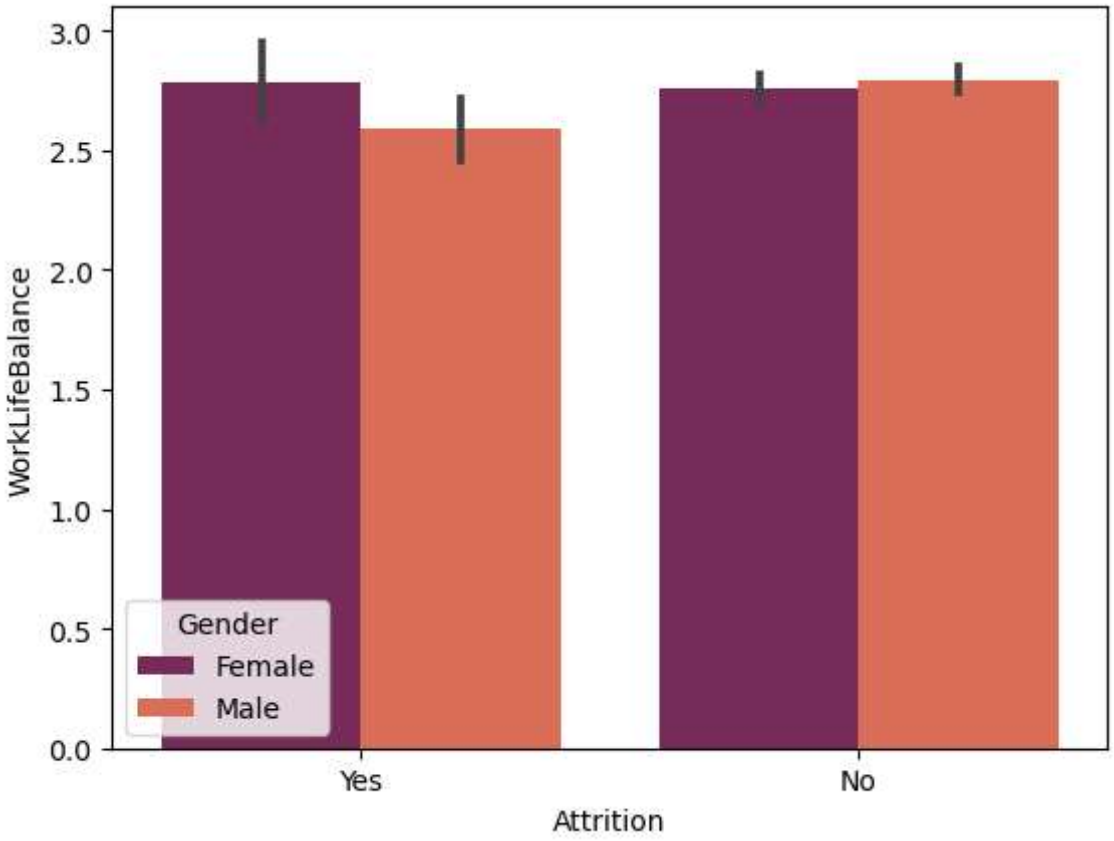

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

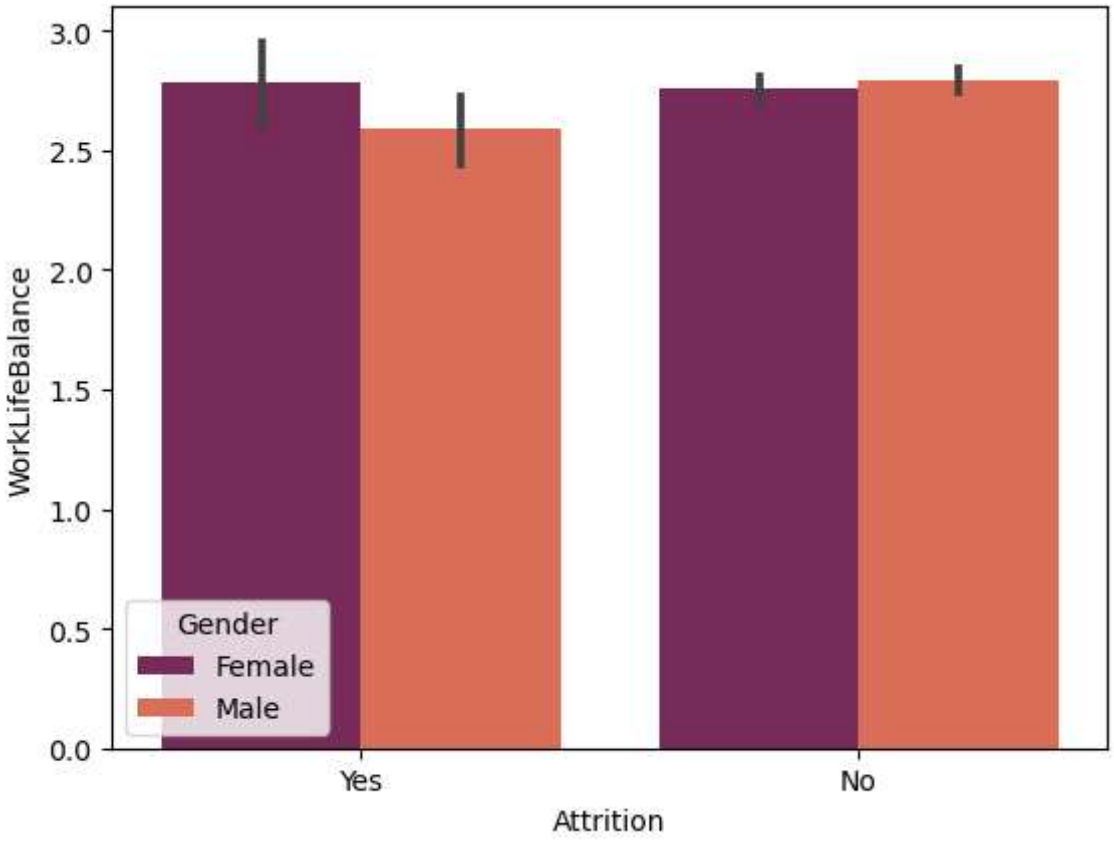

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

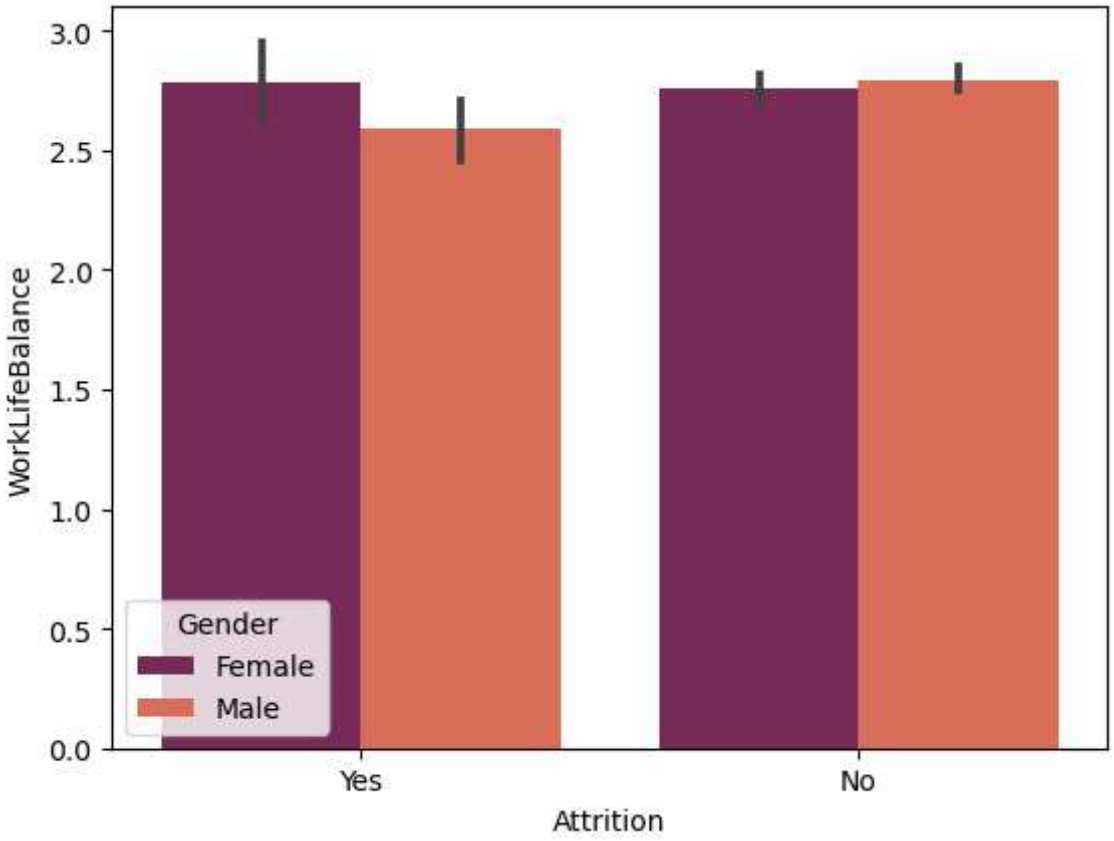

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

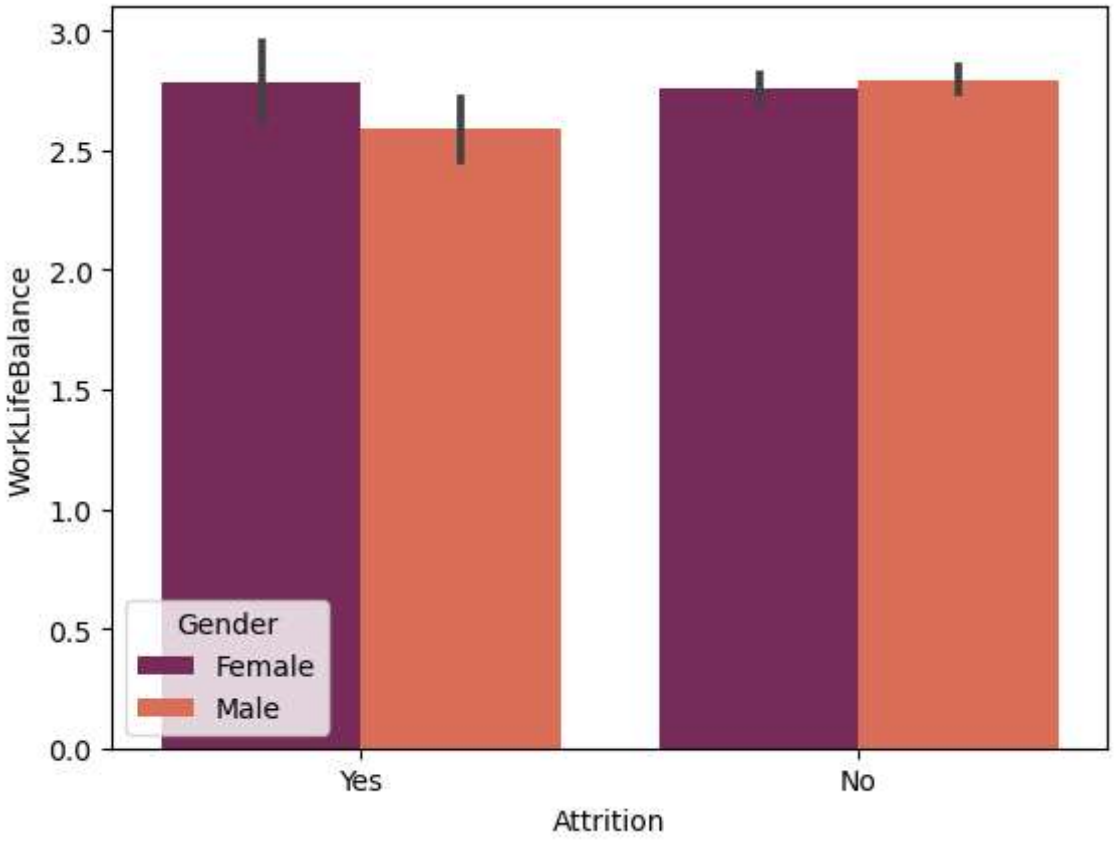

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

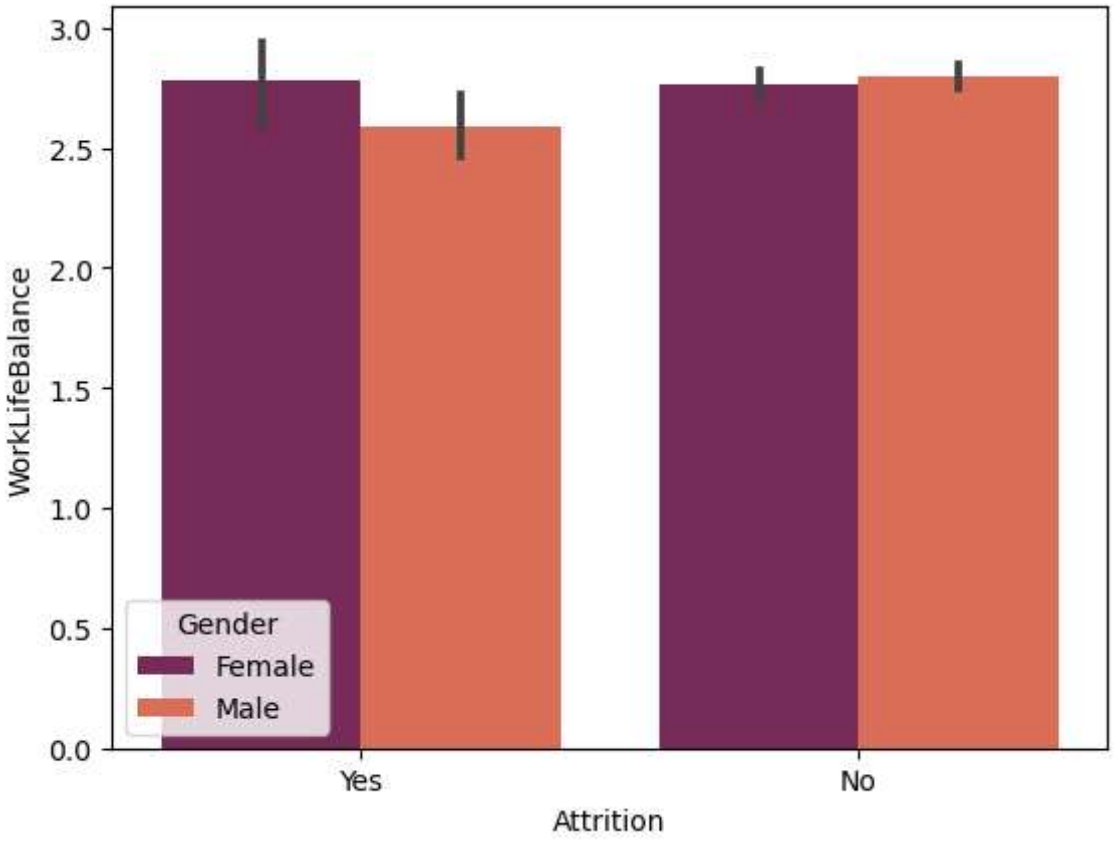

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

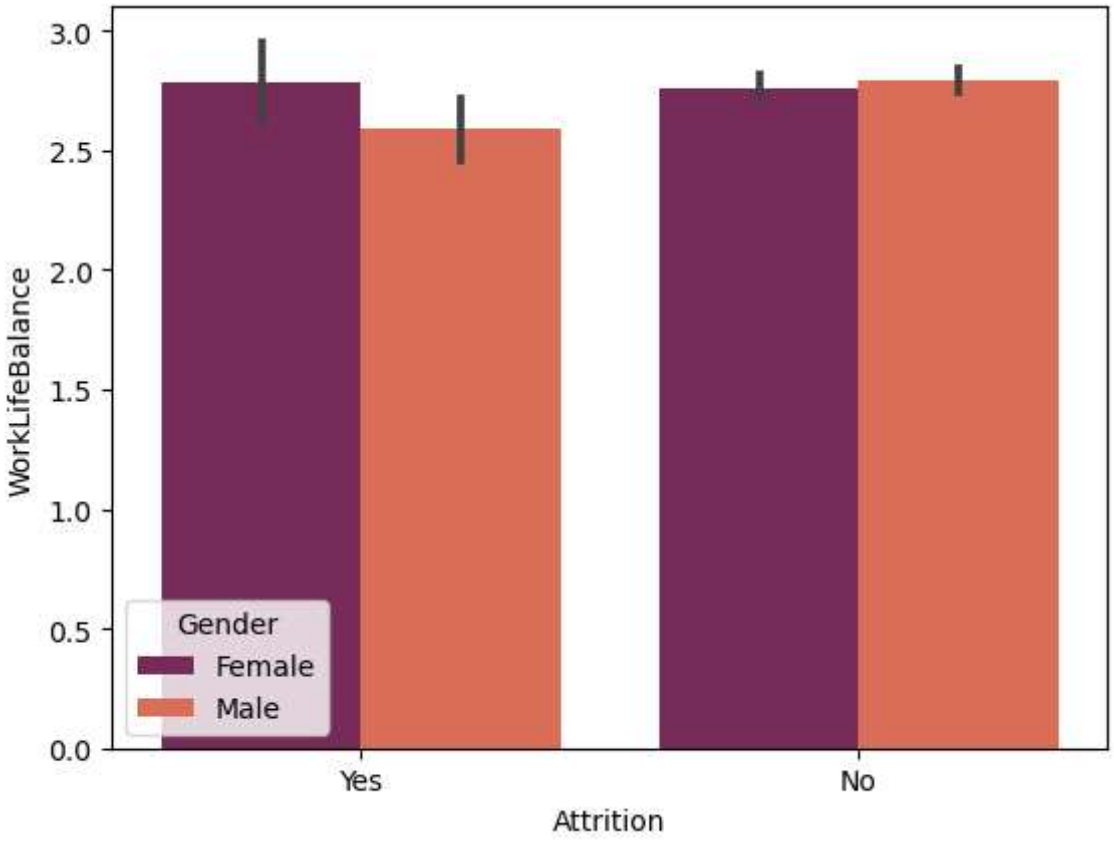

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

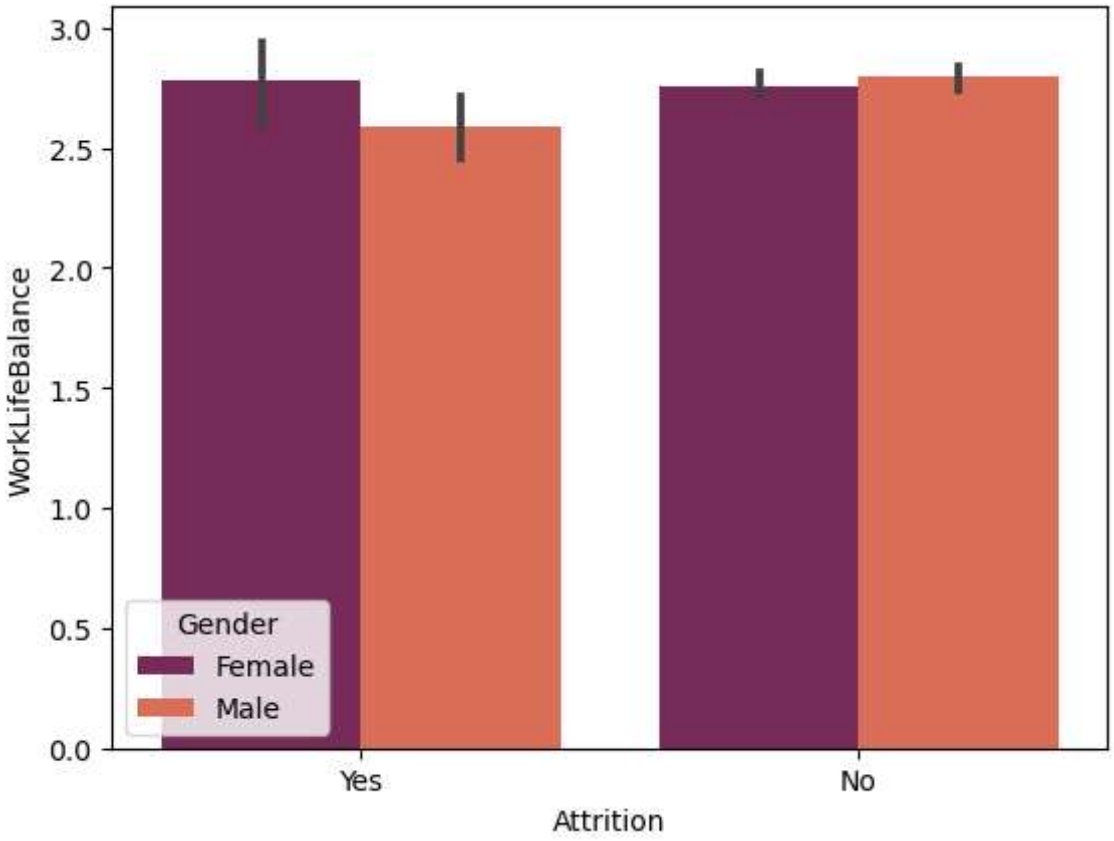

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition

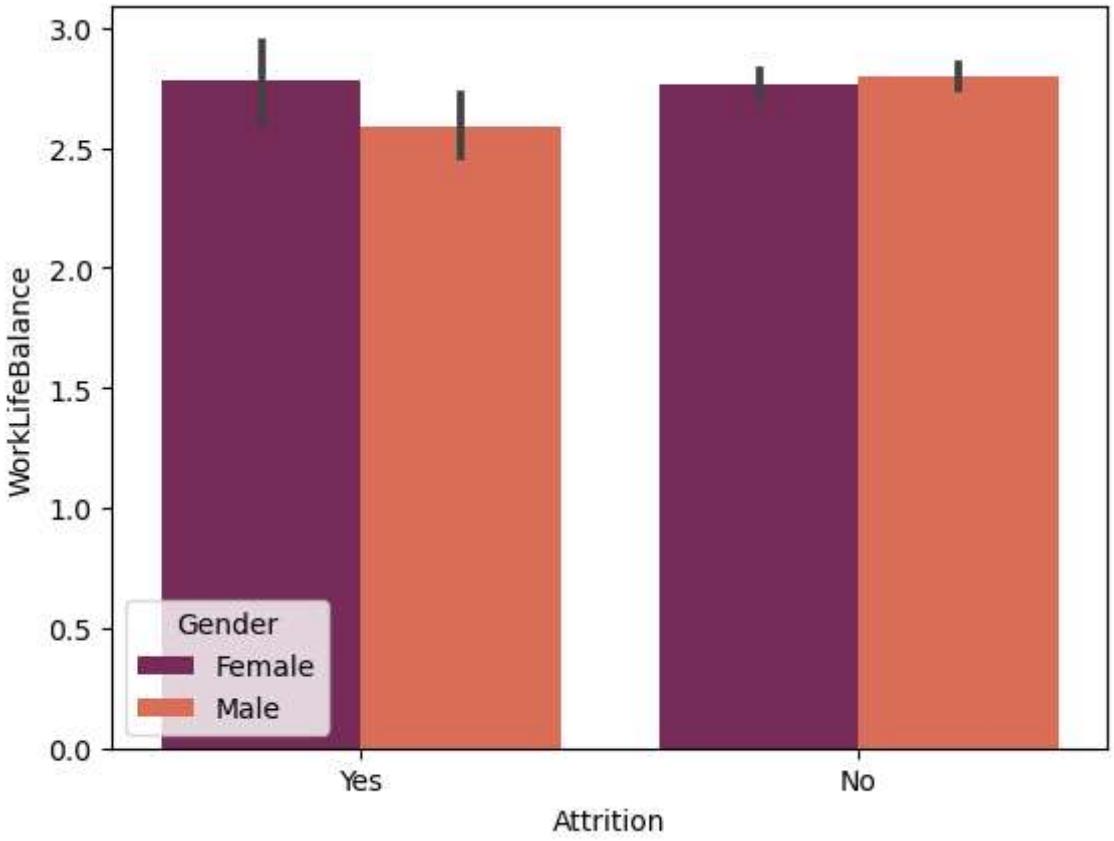

3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition


3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition


3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition


3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition


3.0
2.5
2.0
1.5
1.0
0.5
0.0
WorkLifeBalance
Gender
Female
Male
Yes
No
Attrition



In [24]:
```
ibm.groupby(['Attrition']).sum()
```

Out[24]:

Attrition	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	HourlyRate
No	46313	1001818	10993	3609	3417	81319
Yes	7965	177836	2520	673	584	15541

2 rows × 23 columns

In [25]:
```
col = ['Attrition','BusinessTravel','Department','EducationField','Gender','JobRol
for x in col:
    ibm[f'{x}_codes']=ibm[x].astype('category').cat.codes
```

In [26]:
```
ibm.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 39 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Attrition                 1470 non-null   object
 1   BusinessTravel            1470 non-null   object
 2   Department                1470 non-null   object
 3   EducationField            1470 non-null   object
 4   Gender                    1470 non-null   object
 5   JobRole                   1470 non-null   object
 6   MaritalStatus             1470 non-null   object
 7   OverTime                  1470 non-null   object
 8   Age                       1470 non-null   int64
 9   DailyRate                 1470 non-null   int64
 10  DistanceFromHome          1470 non-null   int64
 11  Education                 1470 non-null   int64
 12  EnvironmentSatisfaction   1470 non-null   int64
 13  HourlyRate                1470 non-null   int64
 14  JobInvolvement            1470 non-null   int64
 15  JobLevel                  1470 non-null   int64
 16  JobSatisfaction           1470 non-null   int64
 17  MonthlyIncome             1470 non-null   int64
 18  MonthlyRate               1470 non-null   int64
 19  NumCompaniesWorked        1470 non-null   int64
 20  PercentSalaryHike         1470 non-null   int64
 21  PerformanceRating         1470 non-null   int64
 22  RelationshipSatisfaction  1470 non-null   int64
 23  StockOptionLevel          1470 non-null   int64
 24  TotalWorkingYears         1470 non-null   int64
 25  TrainingTimesLastYear     1470 non-null   int64
 26  WorkLifeBalance           1470 non-null   int64
 27  YearsAtCompany            1470 non-null   int64
 28  YearsInCurrentRole        1470 non-null   int64
 29  YearsSinceLastPromotion   1470 non-null   int64
 30  YearsWithCurrManager      1470 non-null   int64
 31  Attrition_codes           1470 non-null   int8
 32  BusinessTravel_codes      1470 non-null   int8
 33  Department_codes          1470 non-null   int8
 34  EducationField_codes      1470 non-null   int8
 35  Gender_codes              1470 non-null   int8
 36  JobRole_codes             1470 non-null   int8
 37  MaritalStatus_codes       1470 non-null   int8
 38  OverTime_codes            1470 non-null   int8
dtypes: int64(23), int8(8), object(8)
memory usage: 367.6+ KB
```

In [27]:
```
ibm.iloc[:,[0,-8,1,-7,2,-6]]
```

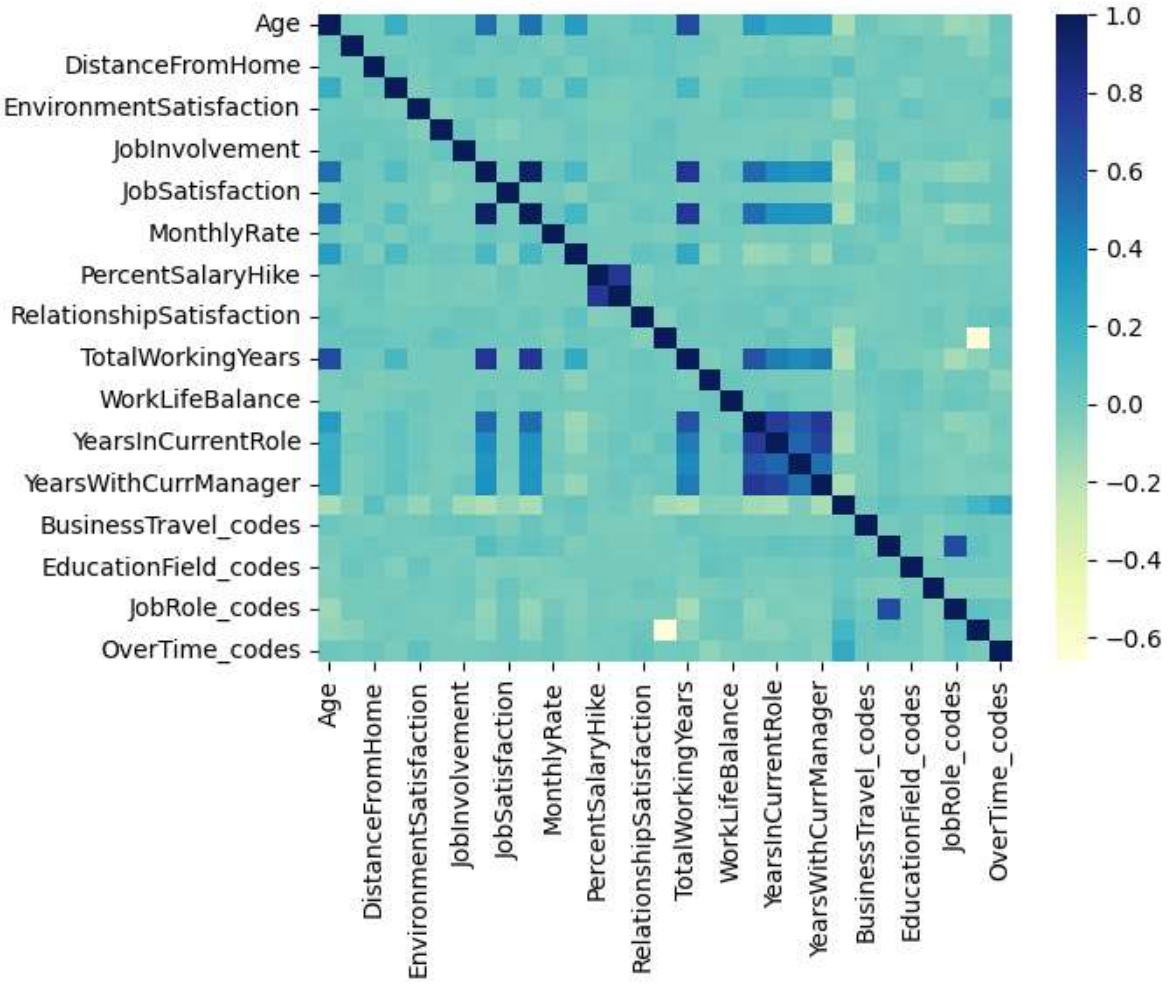
Out[27]:

	Attrition	Attrition_codes	BusinessTravel	BusinessTravel_codes	Department	Department_
0	Yes	1	Travel_Rarely	2	Sales	
1	No	0	Travel_Frequently	1	Research & Development	
2	Yes	1	Travel_Rarely	2	Research & Development	
3	No	0	Travel_Frequently	1	Research & Development	
4	No	0	Travel_Rarely	2	Research & Development	
...	...	...	...	...	...	
1465	No	0	Travel_Frequently	1	Research & Development	
1466	No	0	Travel_Rarely	2	Research & Development	
1467	No	0	Travel_Rarely	2	Research & Development	
1468	No	0	Travel_Frequently	1	Sales	
1469	No	0	Travel_Rarely	2	Research & Development	

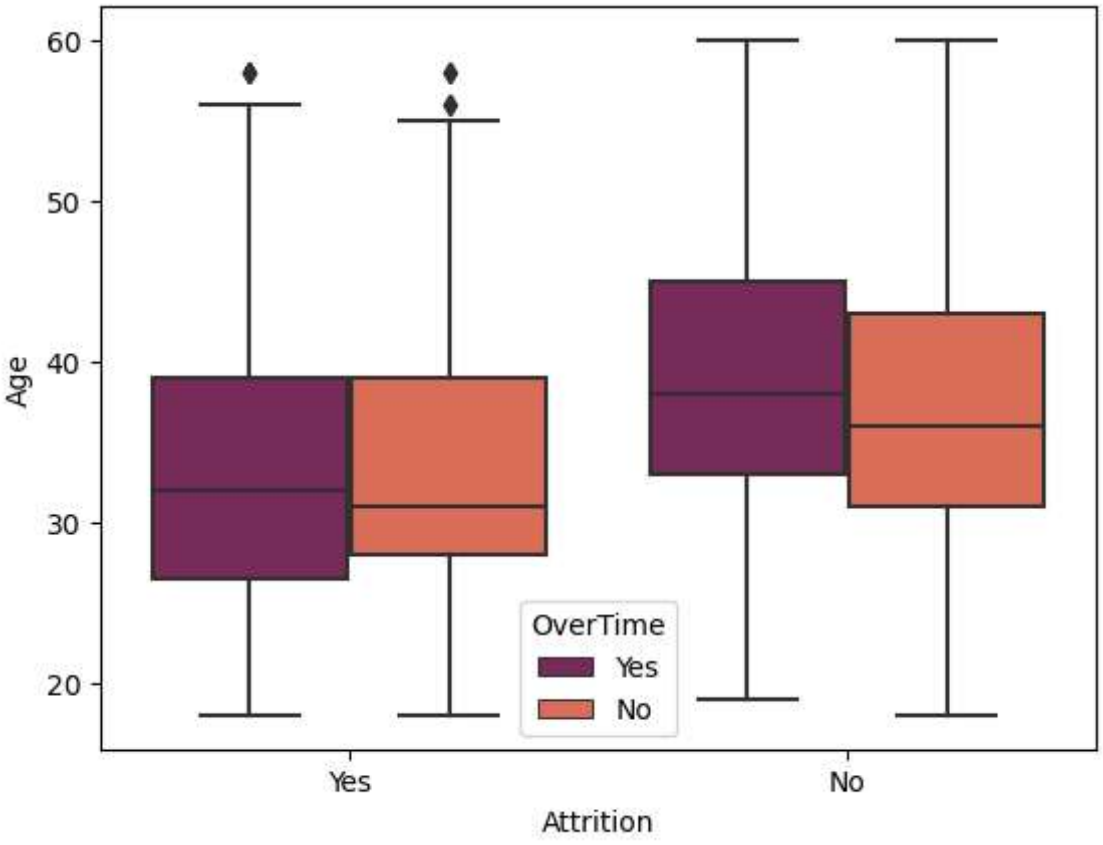
1470 rows × 6 columns

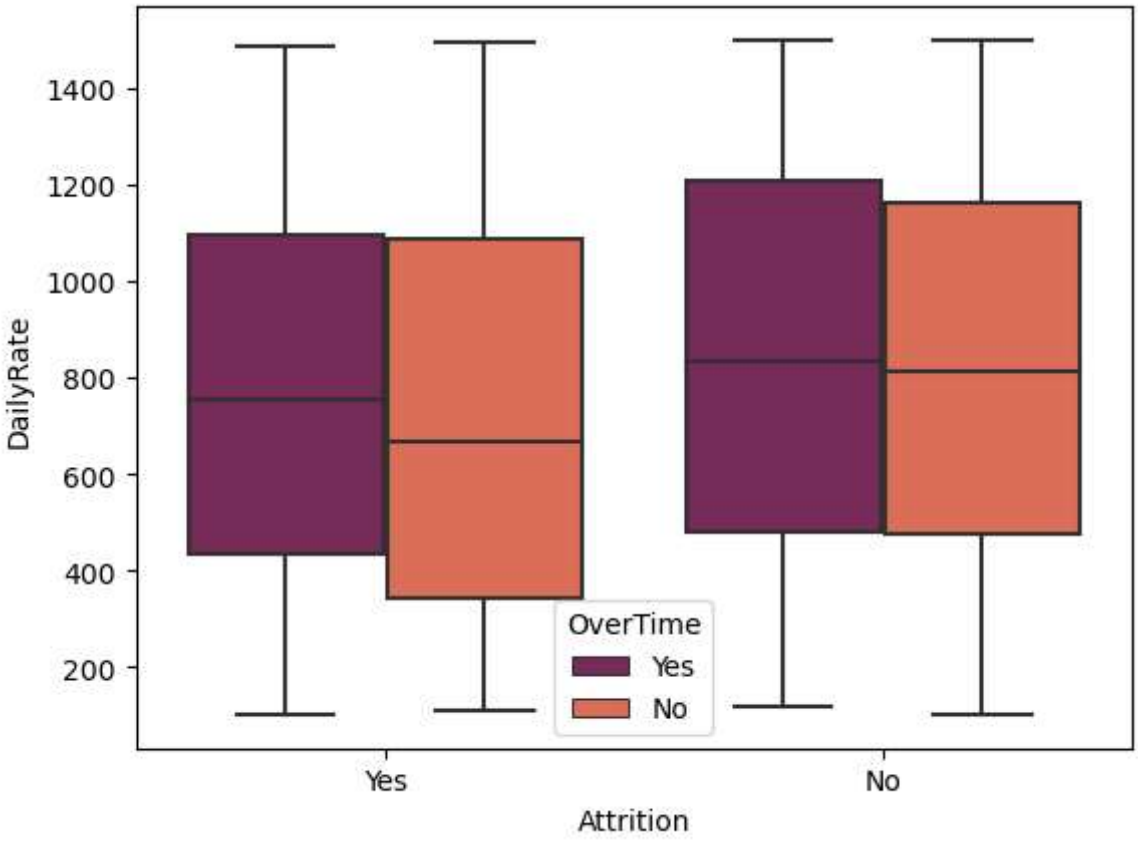
In [28]:
```
sns.heatmap(ibm.corr(),cmap='YlGnBu')
```

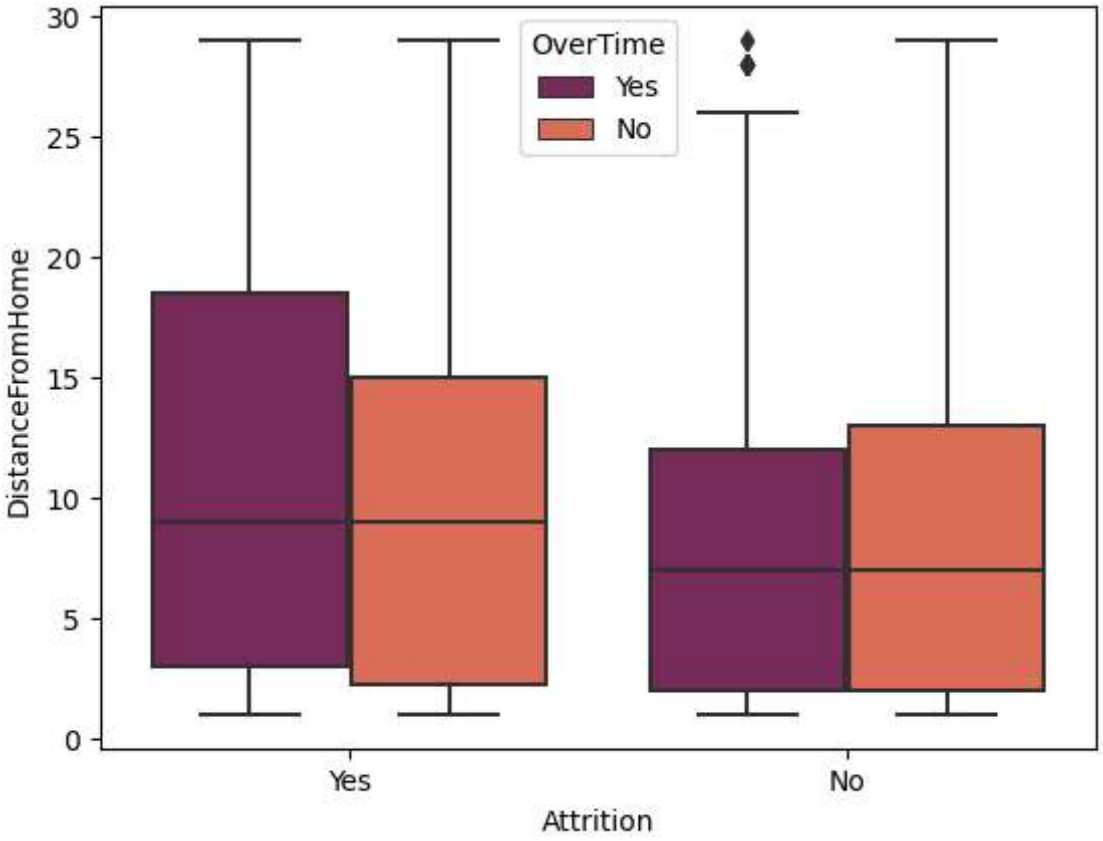
Out[28]:
```
<AxesSubplot:>
```

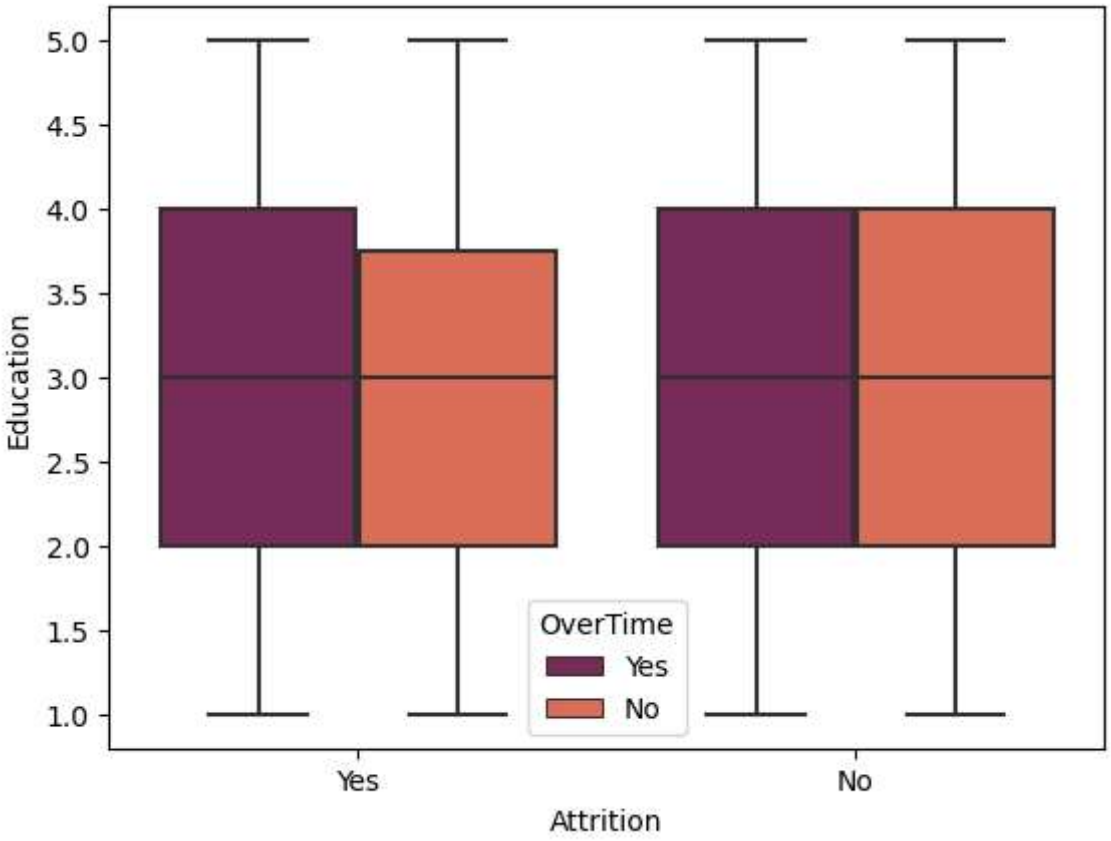


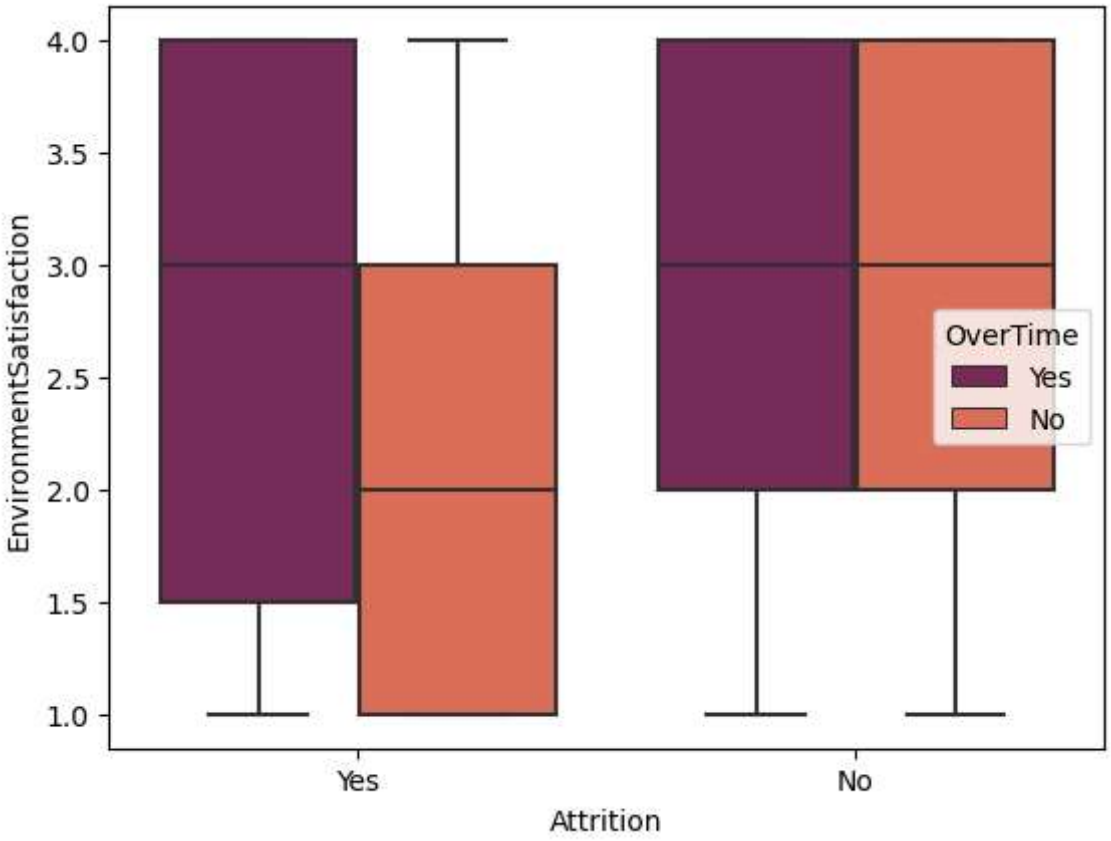
In [29]:
```
for AA in ibm.columns[8:]:
    sns.boxplot(x='Attrition',y=AA,data=ibm,hue='OverTime',palette='rocket')
    plt.show()
```

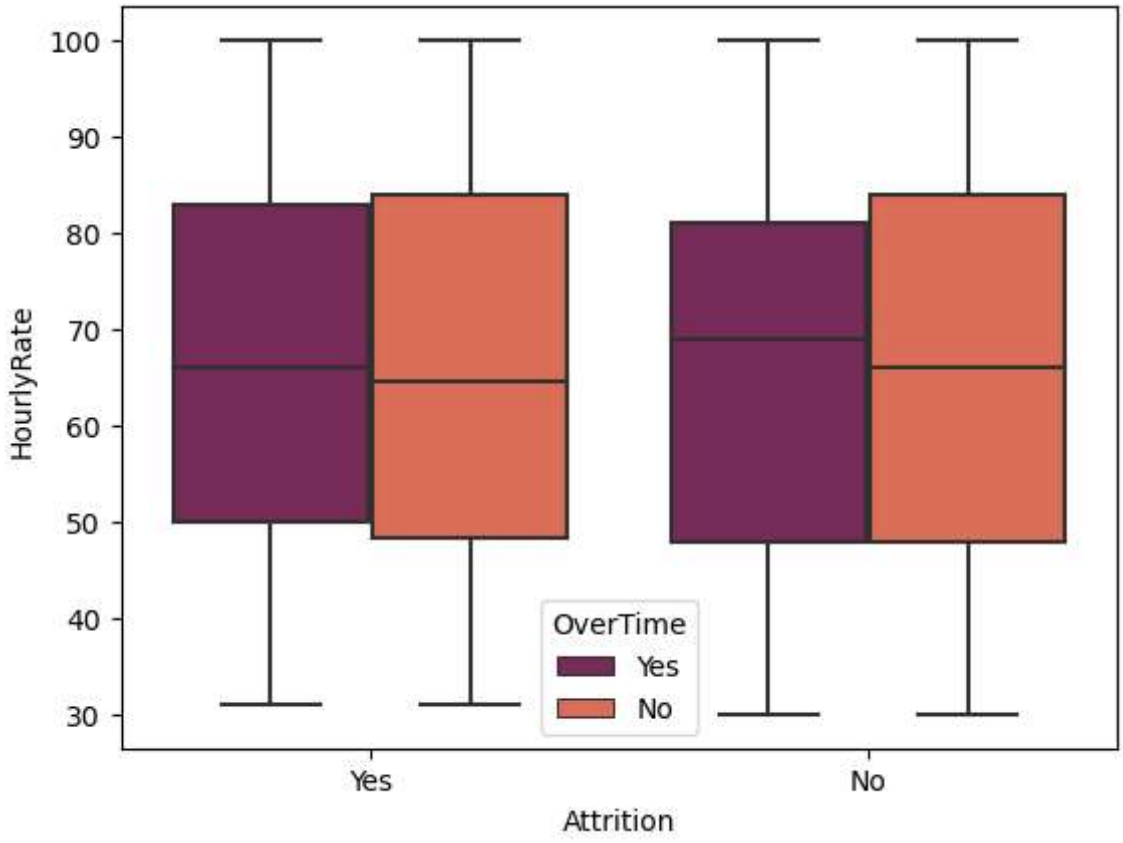


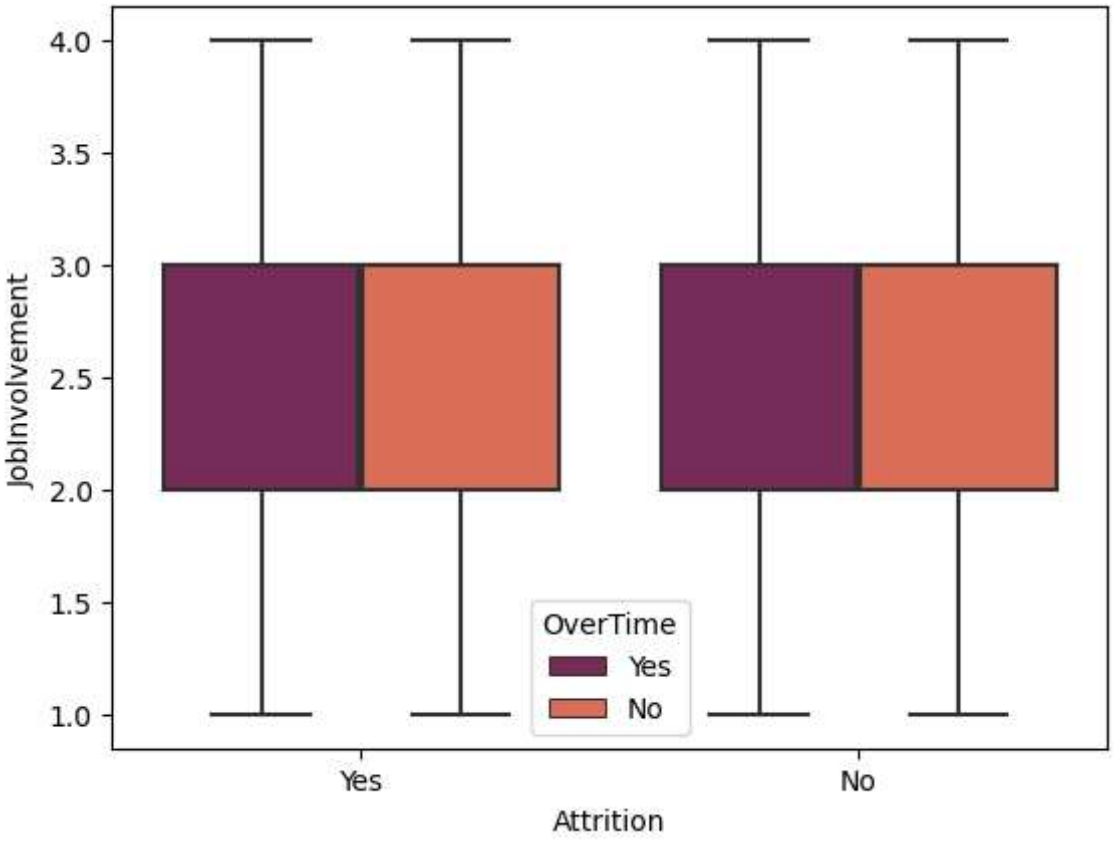

1400
1200
1000
800
600
400
200
DailyRate
OverTime
Yes
No
Yes
No
Attrition

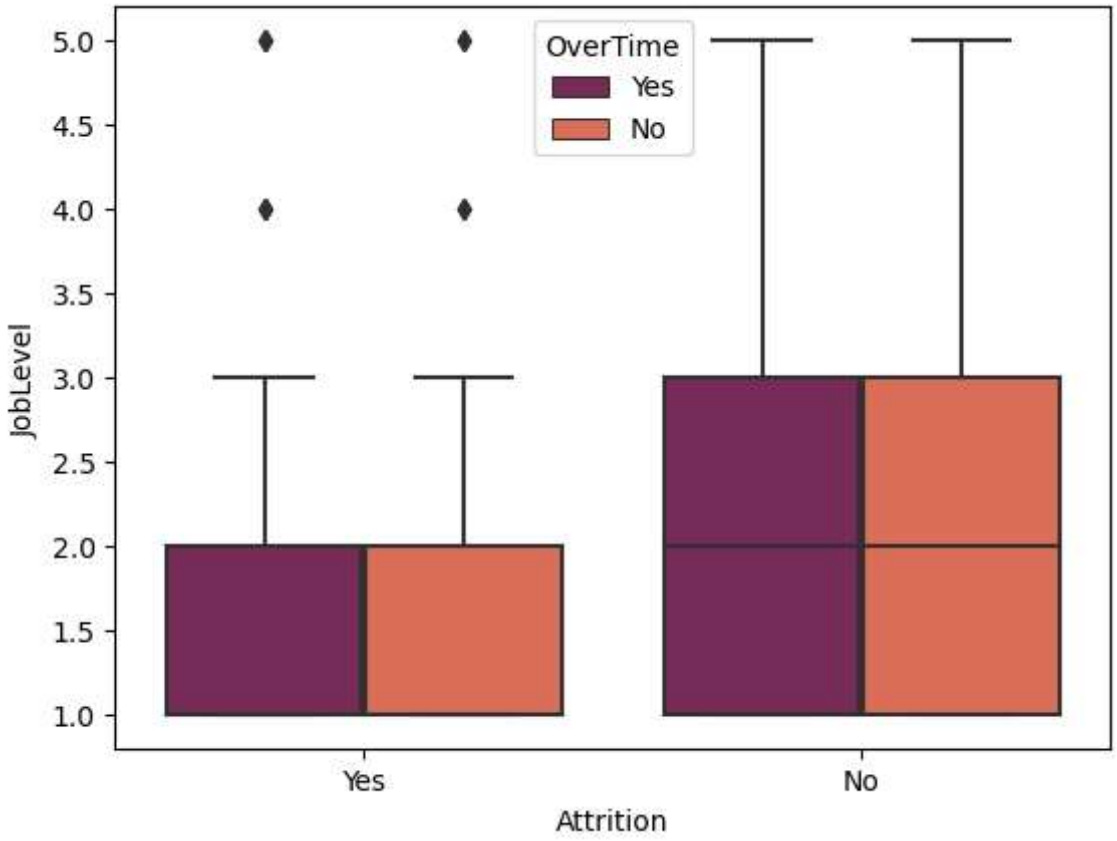

30
25
20
15
10
5
0
DistanceFromHome
OverTime
Yes
No
Yes
No
Attrition

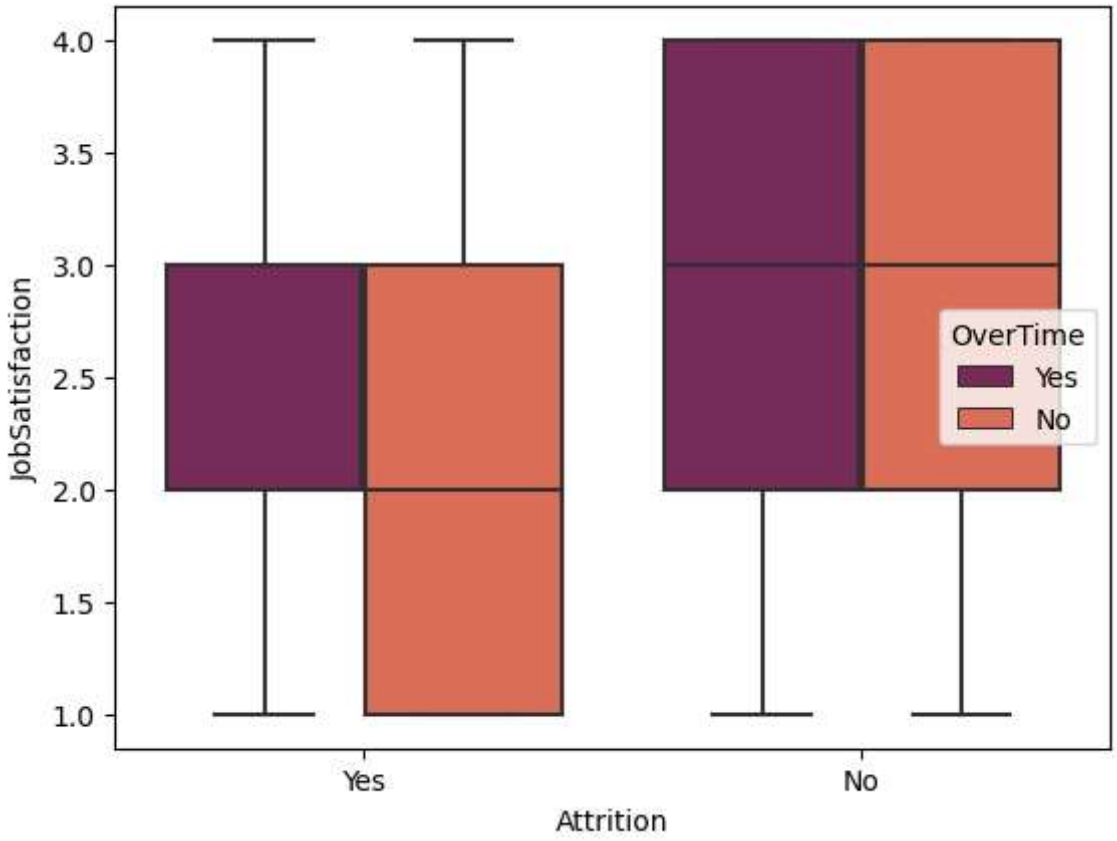

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Education
OverTime
Yes
No
Yes
No
Attrition

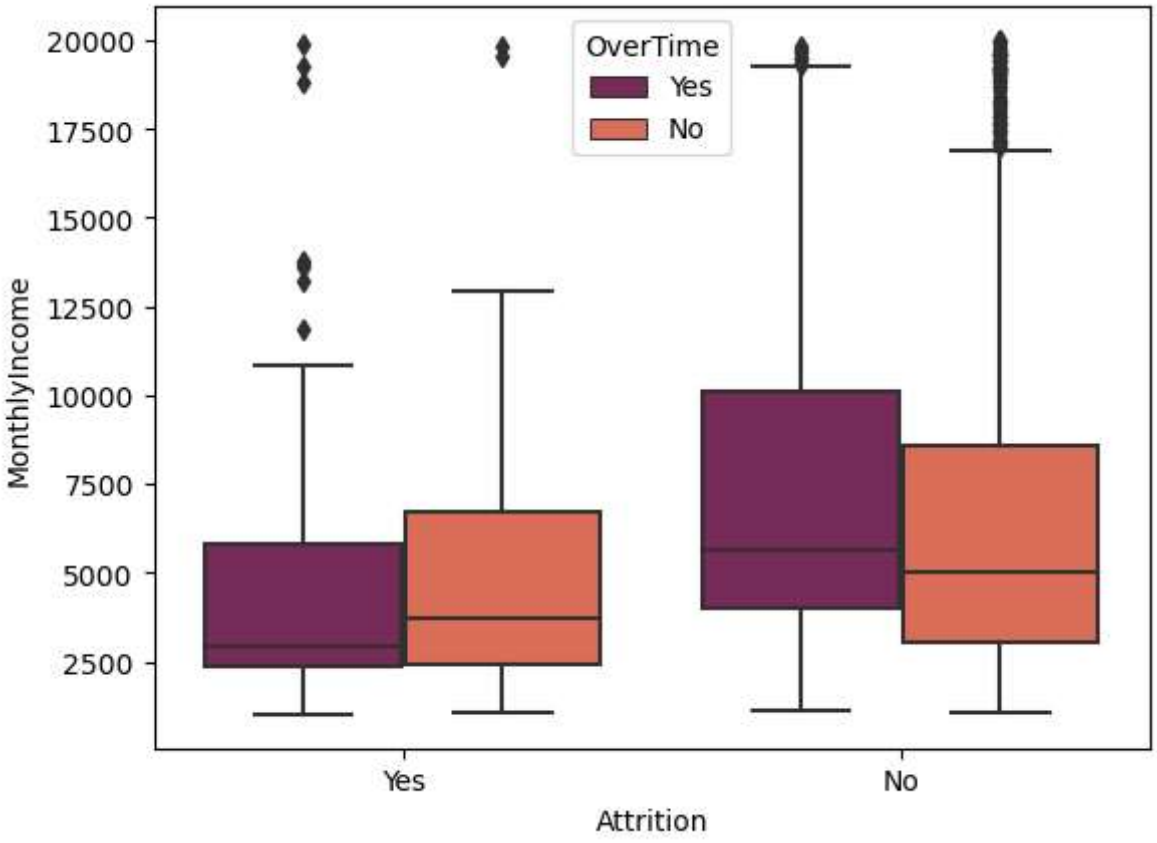

4.0
3.5
3.0
2.5
2.0
1.5
1.0
EnvironmentSatisfaction
OverTime
Yes
No
Yes
No
Attrition

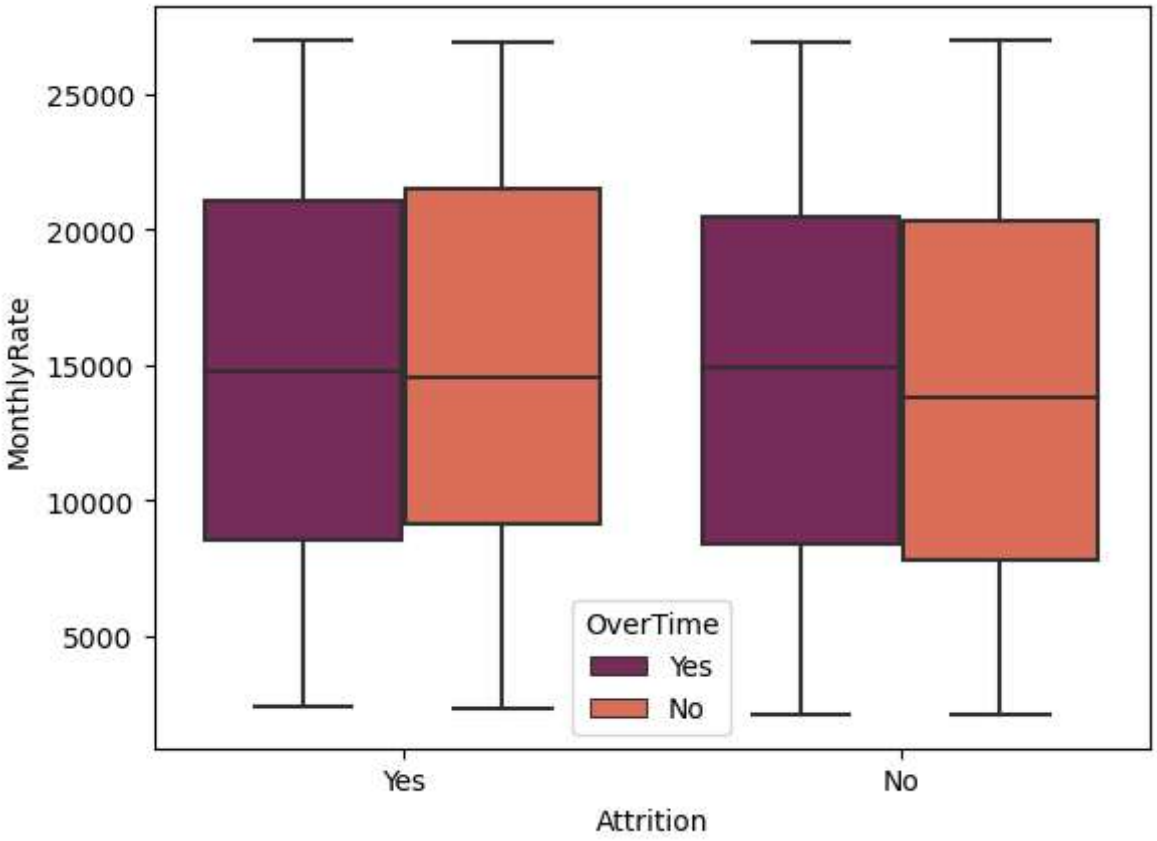

100
90
80
70
60
50
40
30
HourlyRate
OverTime
Yes
No
Yes
No
Attrition

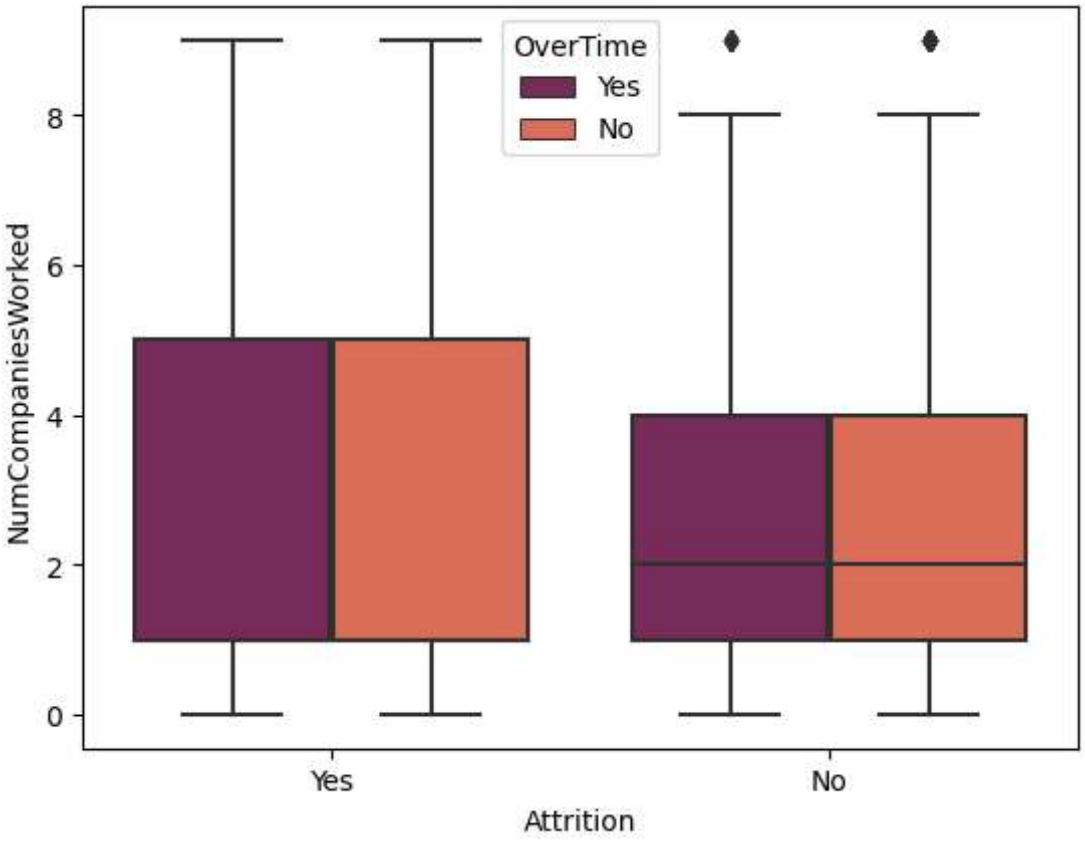

4.0
3.5
3.0
2.5
2.0
1.5
1.0
JobInvolvement
OverTime
Yes
No
Yes
No
Attrition

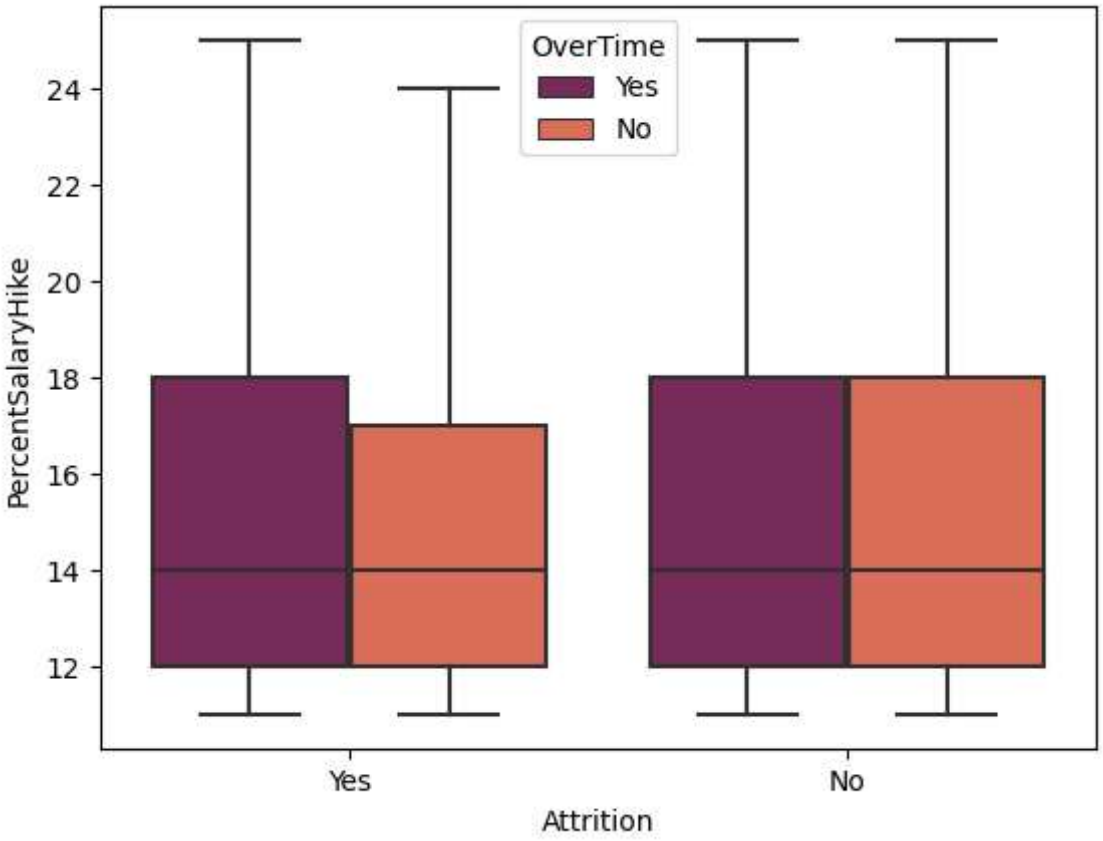

OverTime
Yes
No
JobLevel
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Yes
No
Attrition

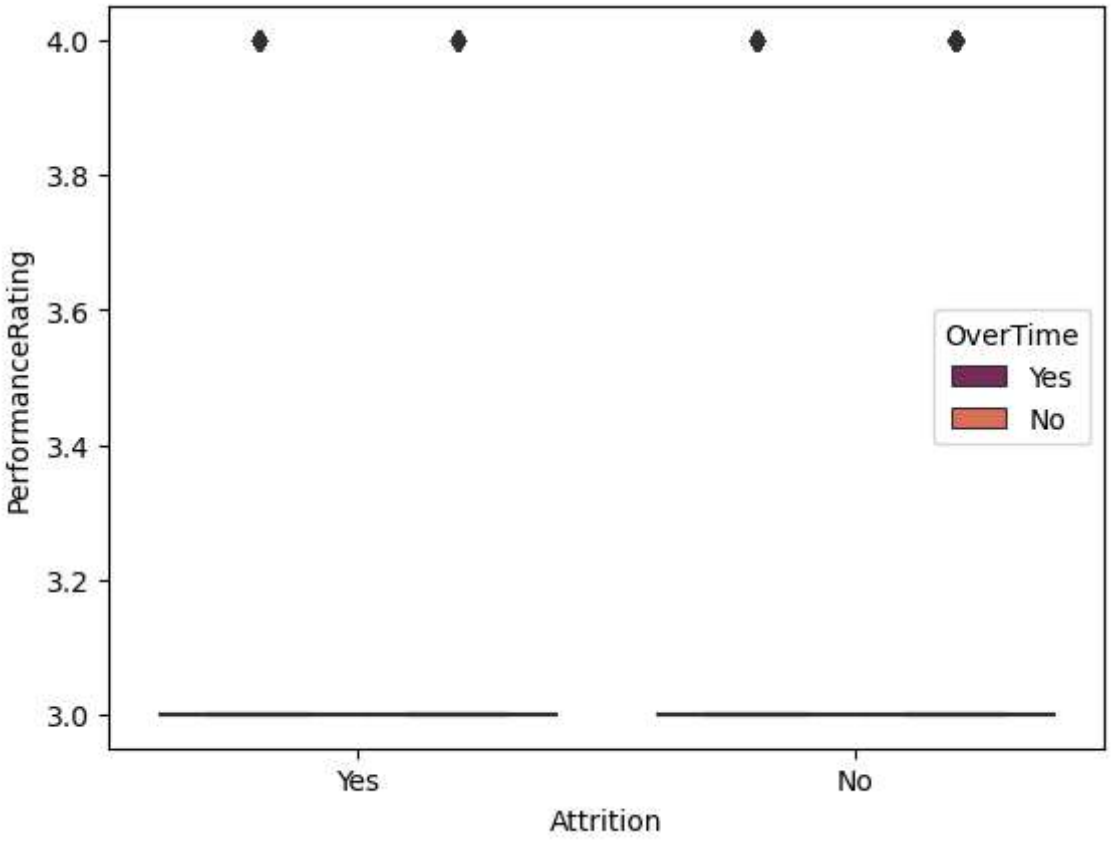

JobSatisfaction
4.0
3.5
3.0
2.5
2.0
1.5
1.0
OverTime
Yes
No
Yes
No
Attrition

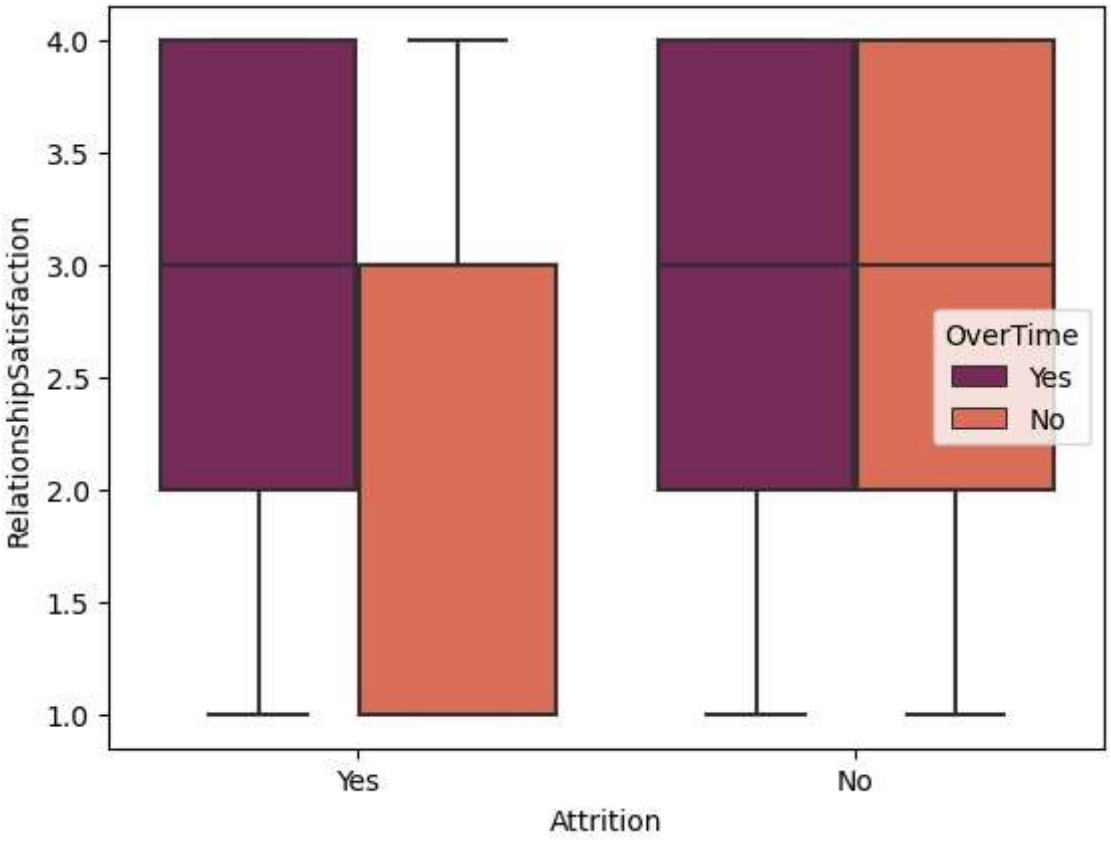

OverTime
Yes
No
MonthlyIncome
20000
17500
15000
12500
10000
7500
5000
2500
Yes
No
Attrition

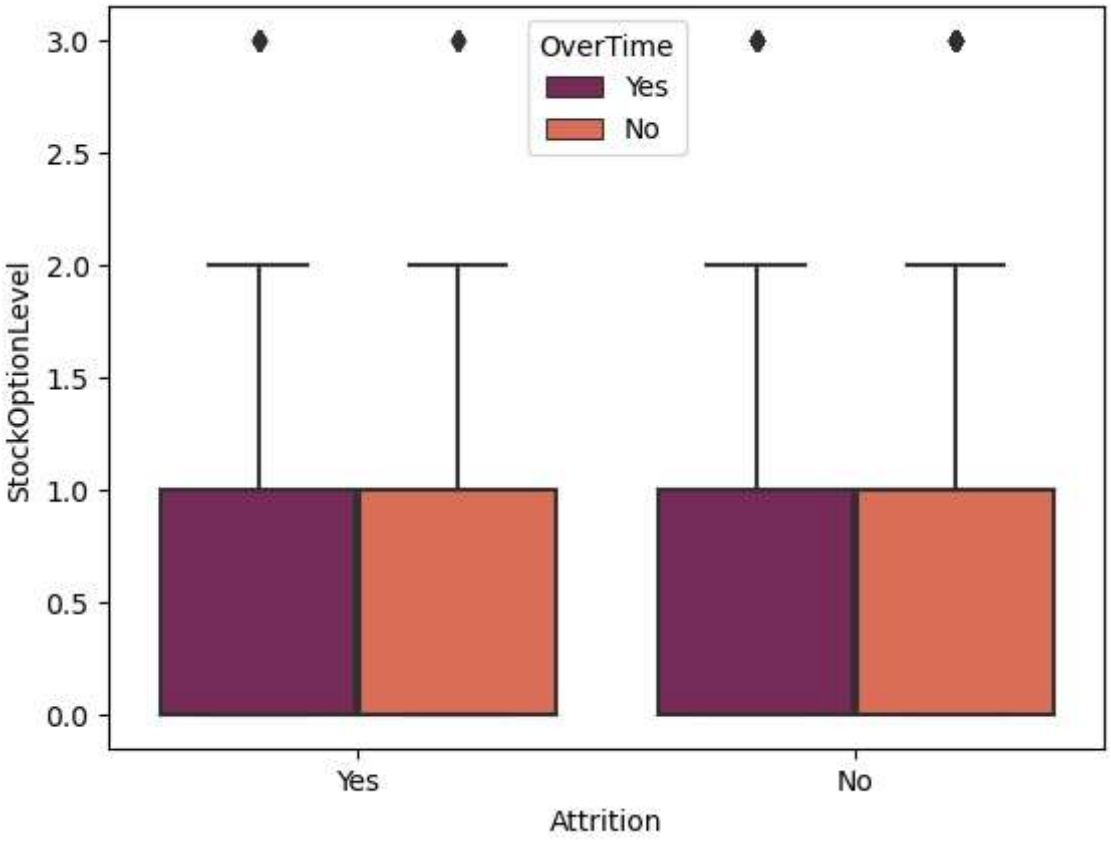

MonthlyRate
25000
20000
15000
10000
5000
OverTime
Yes
No
Yes
No
Attrition

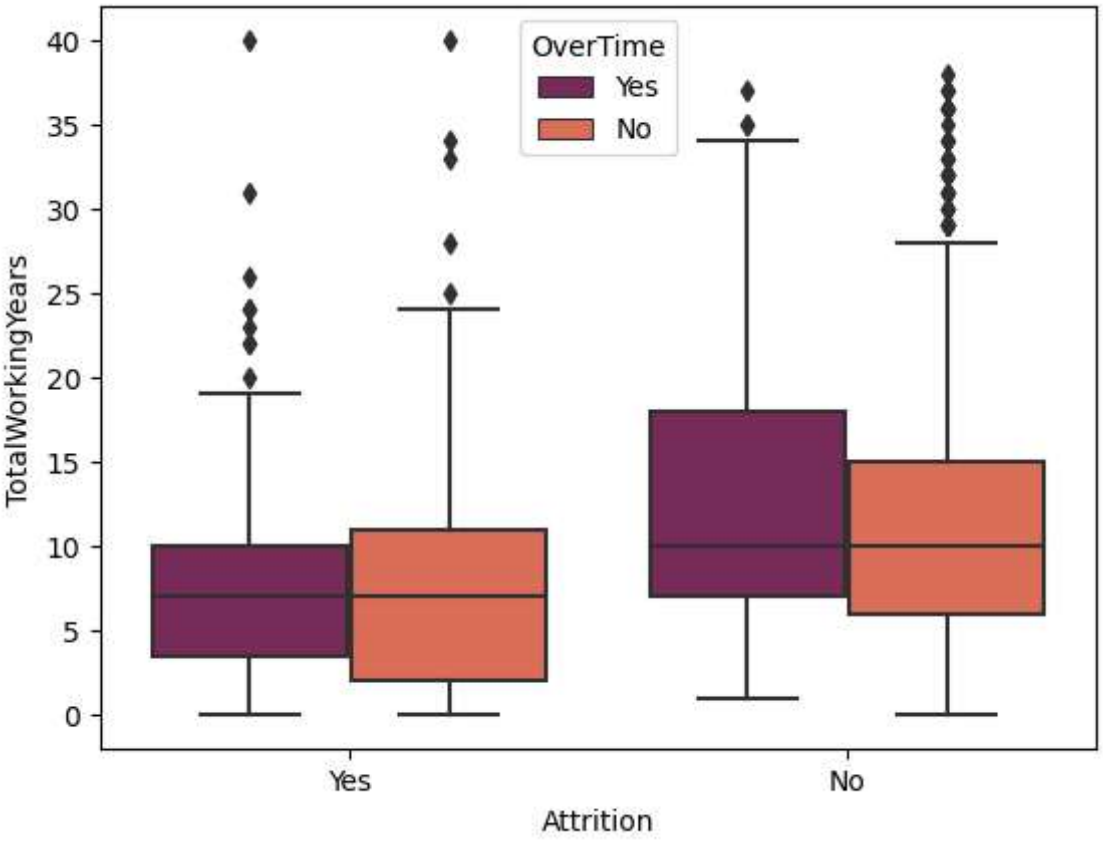

OverTime
Yes
No
NumCompaniesWorked
0
2
4
6
8
Yes
No
Attrition

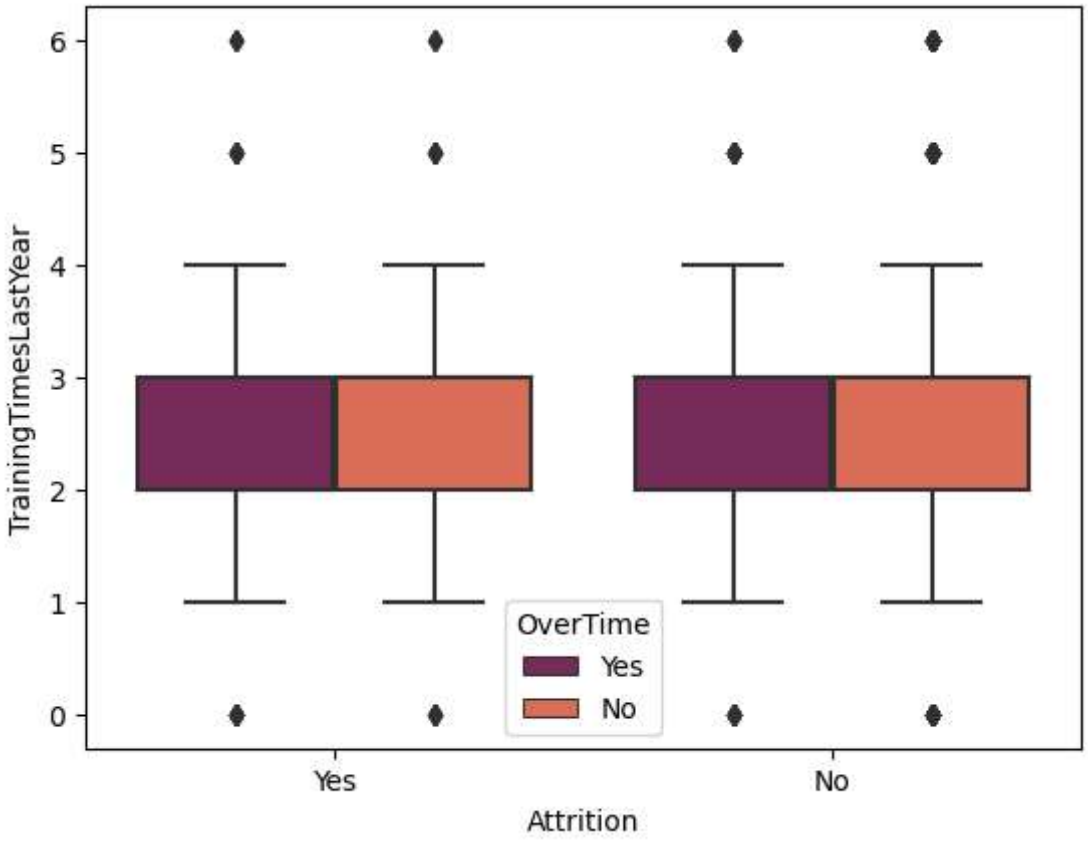

OverTime
Yes
No
PercentSalaryHike
12
14
16
18
20
22
24
Yes
No
Attrition

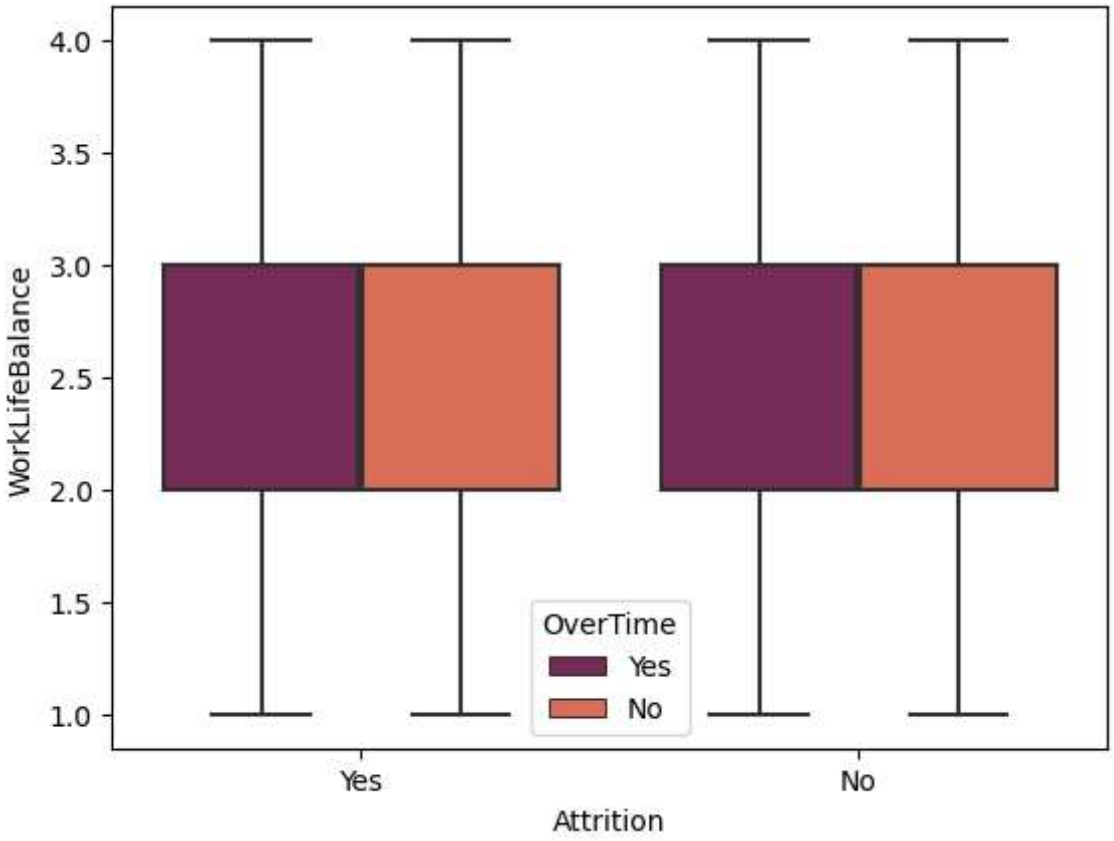

4.0
3.8
3.6
3.4
3.2
3.0
PerformanceRating
OverTime
Yes
No
Yes
No
Attrition

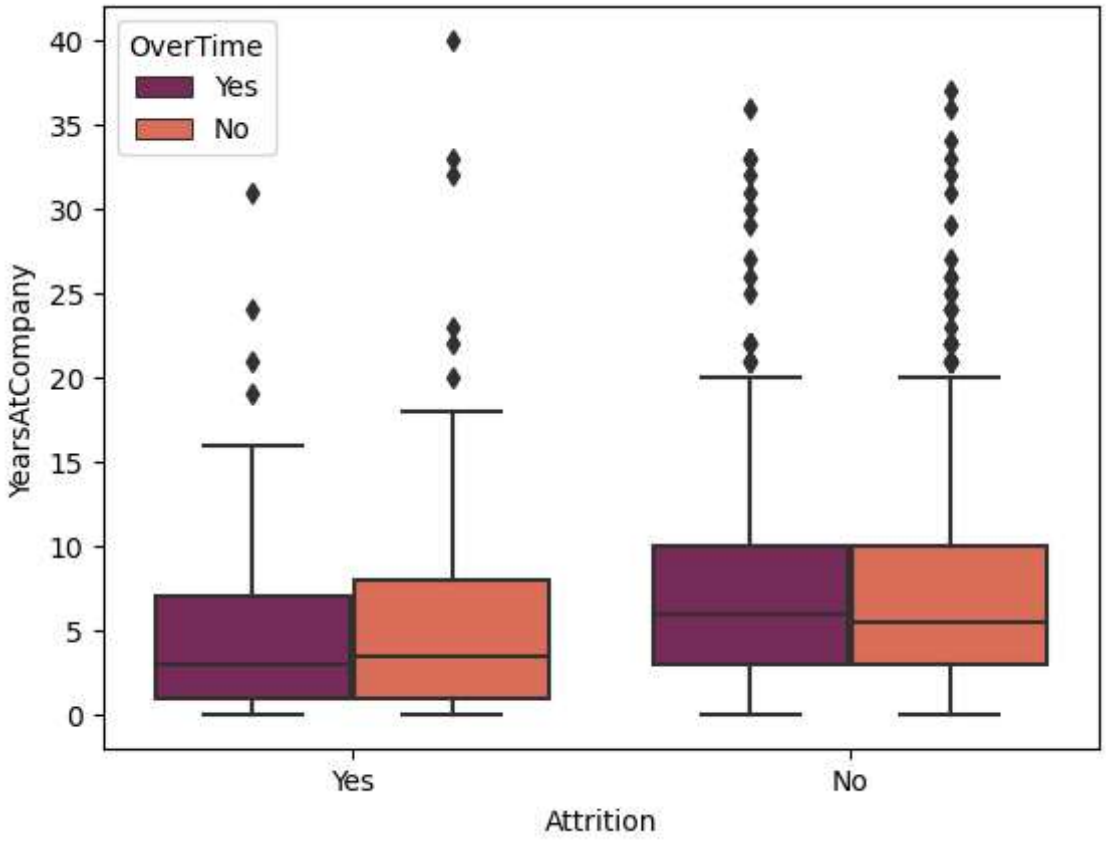

4.0
3.5
3.0
2.5
2.0
1.5
1.0
RelationshipSatisfaction
OverTime
Yes
No
Yes
No
Attrition

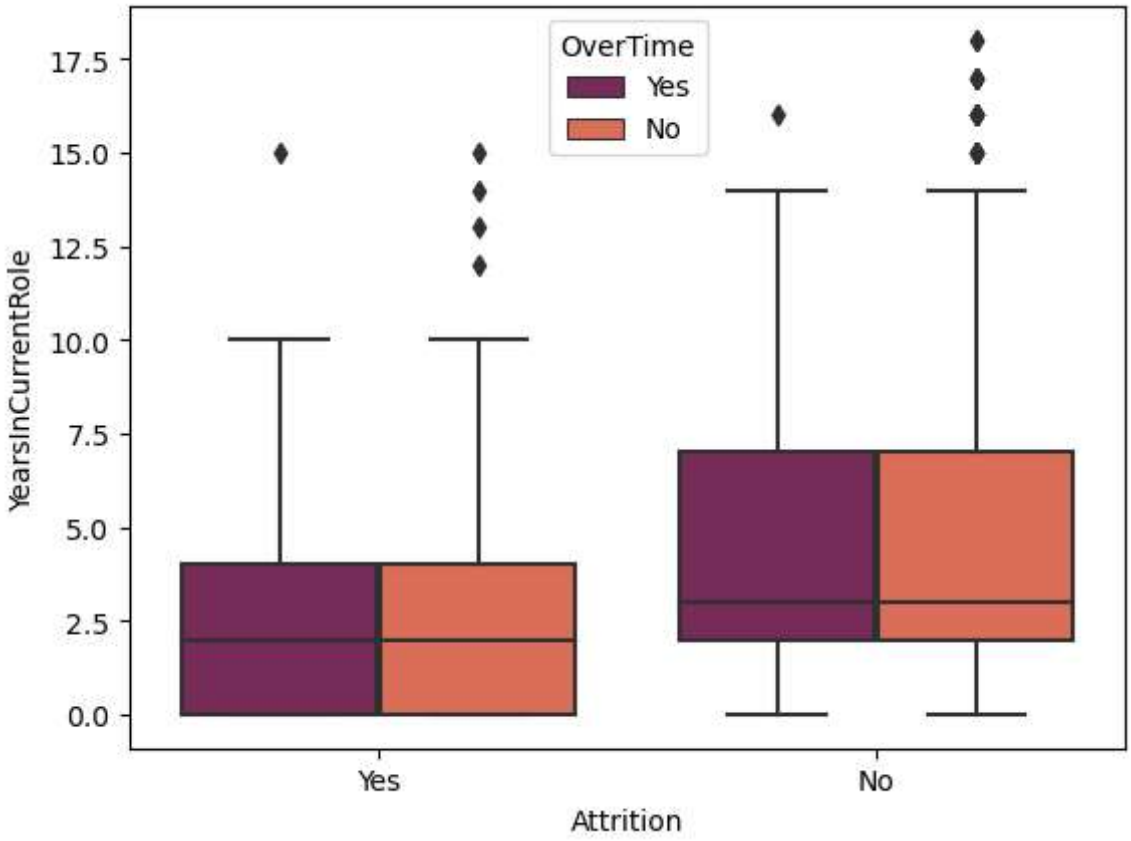

OverTime
Yes
No
StockOptionLevel
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Yes
No
Attrition

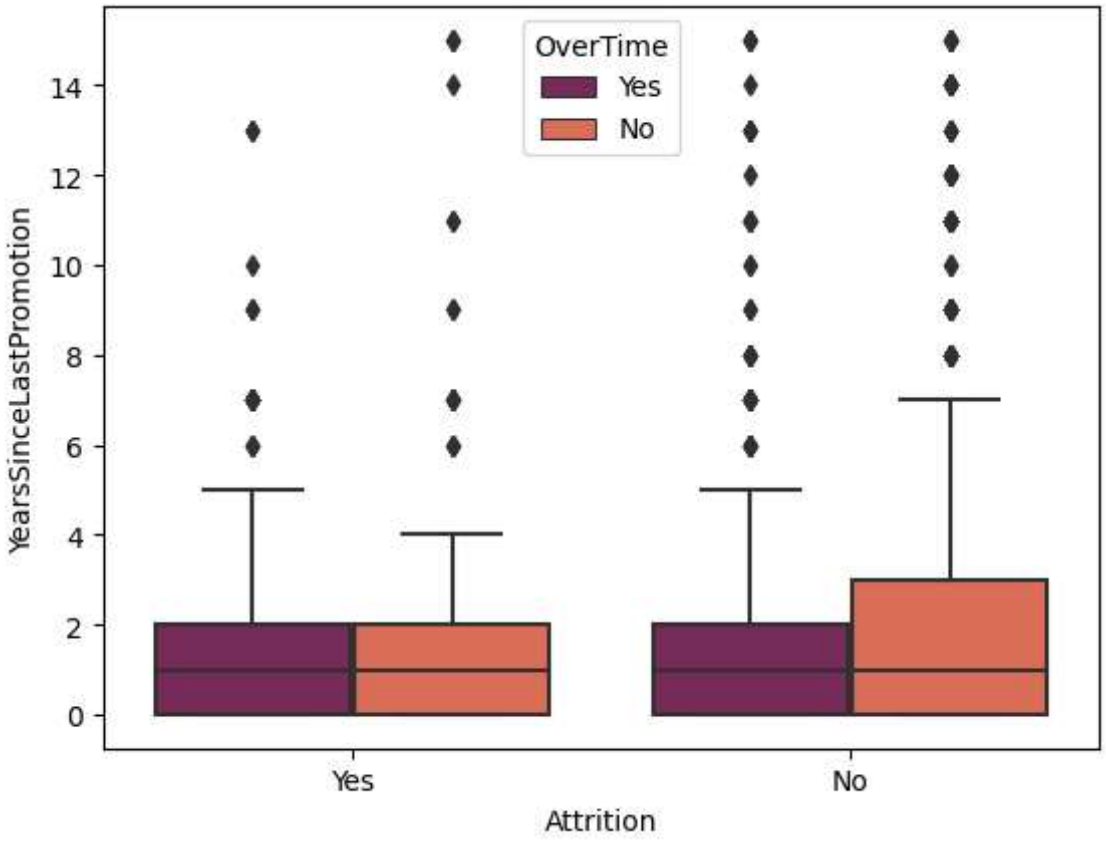

OverTime
Yes
No
TotalWorkingYears
40
35
30
25
20
15
10
5
0
Yes
No
Attrition

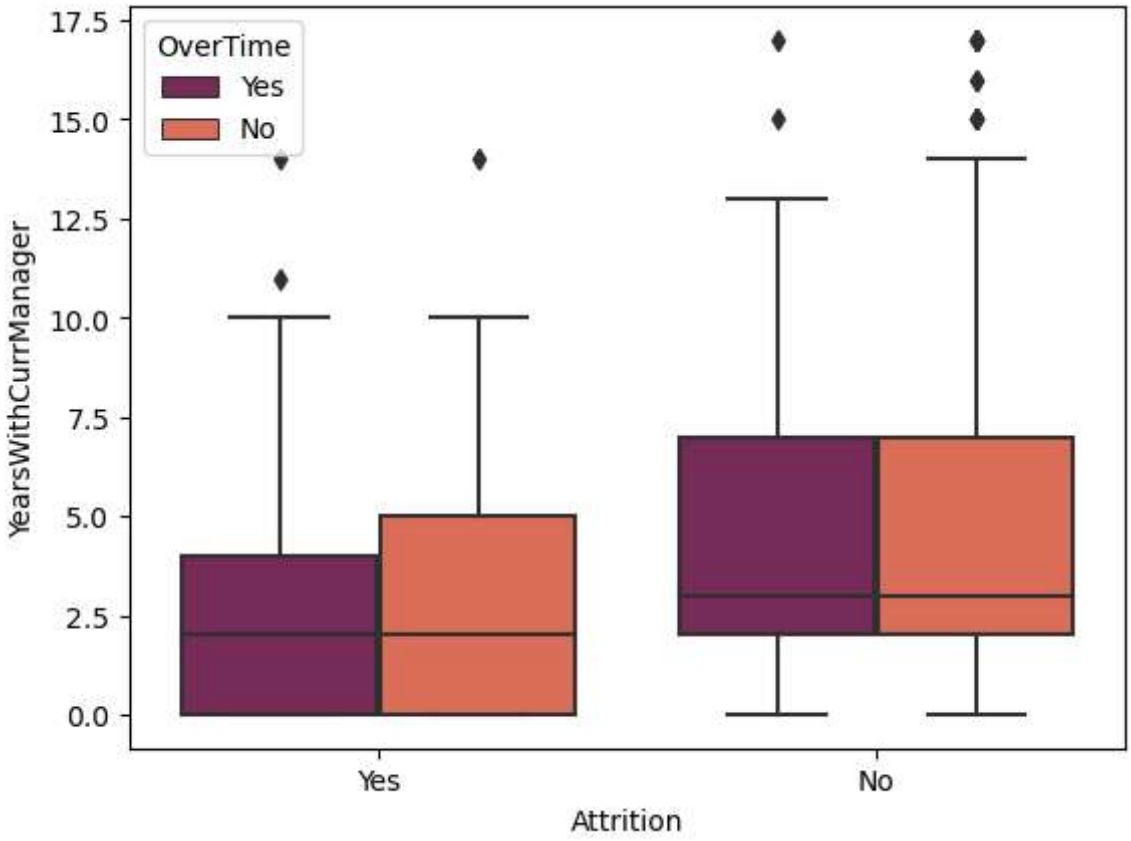

TrainingTimesLastYear
0
1
2
3
4
5
6
Yes
No
Attrition
OverTime
Yes
No

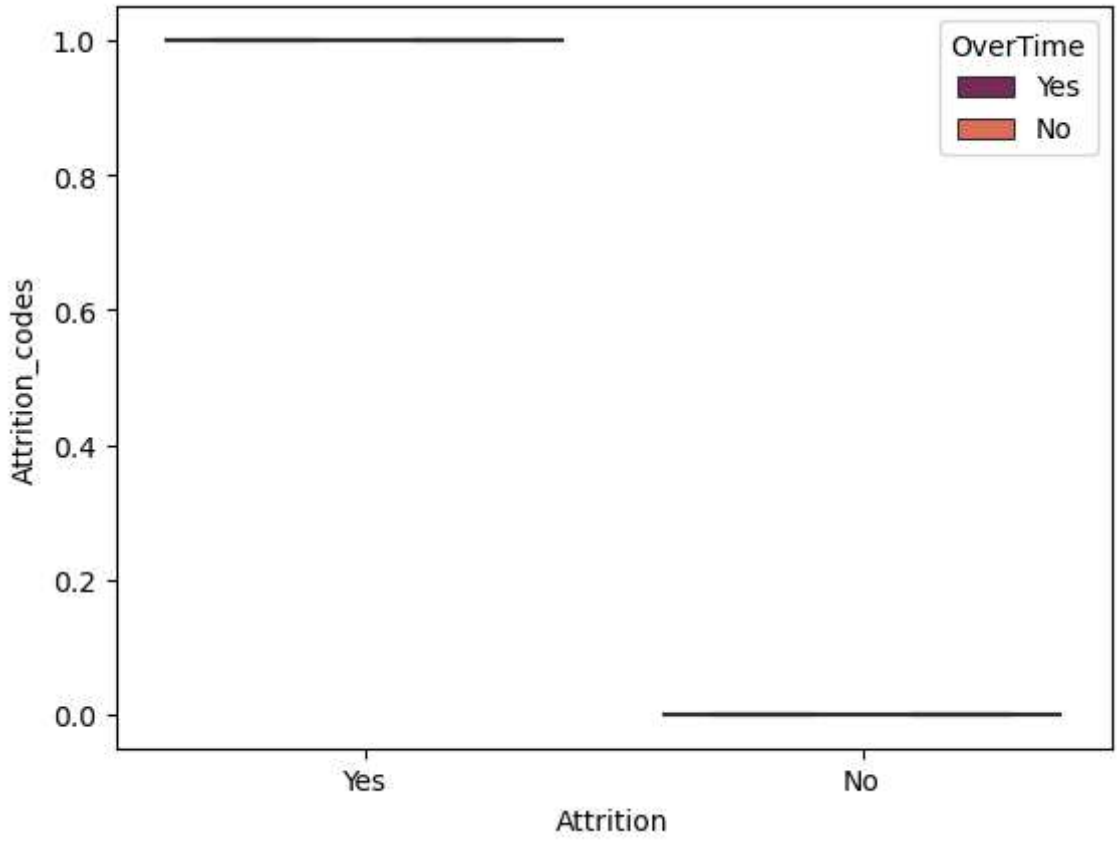

WorkLifeBalance
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Yes
No
Attrition
OverTime
Yes
No

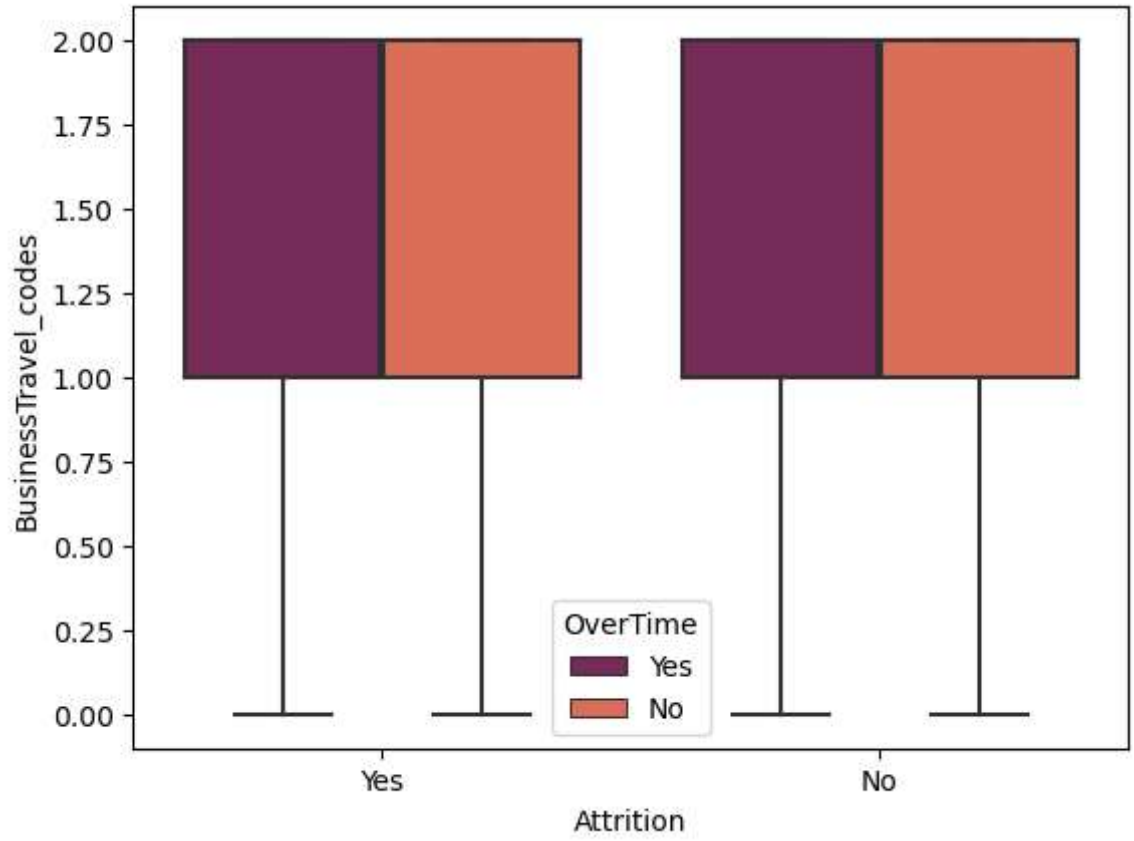

OverTime
Yes
No
YearsAtCompany
40
35
30
25
20
15
10
5
0
Yes
No
Attrition

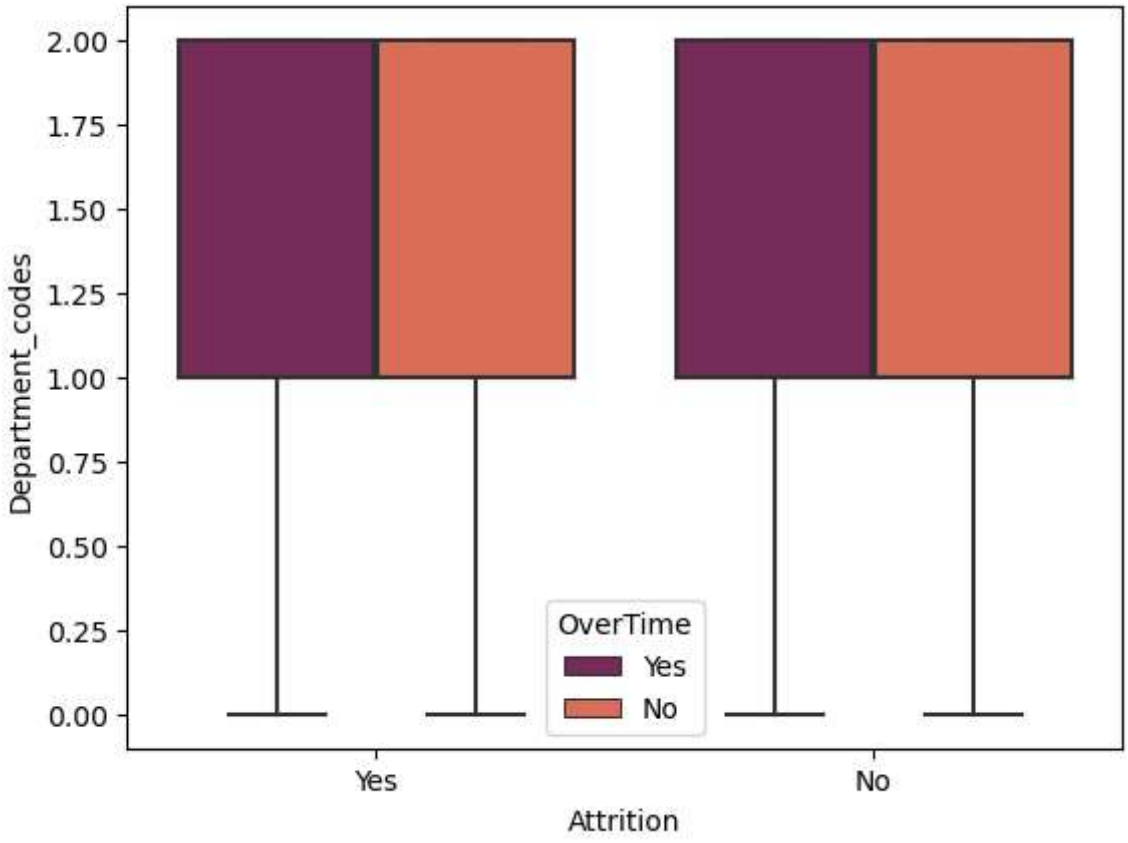

OverTime
Yes
No
YearsInCurrentRole
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Yes
No
Attrition

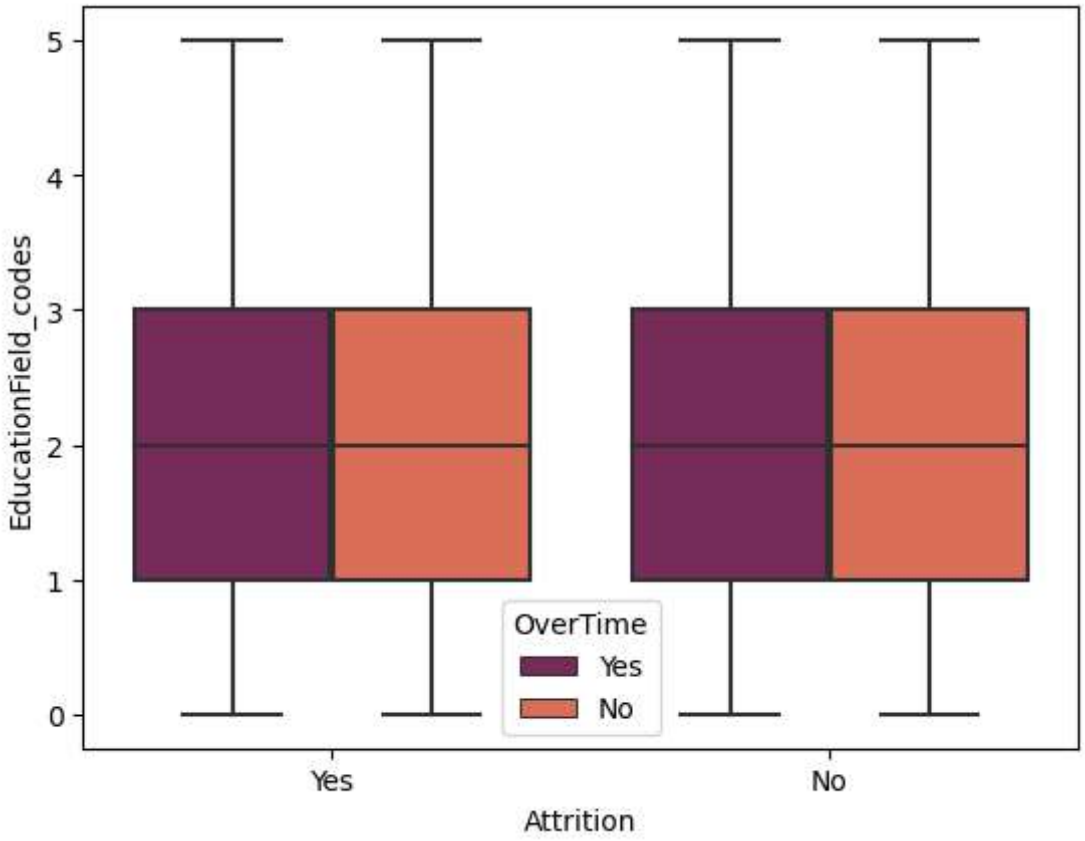

OverTime
Yes
No
YearsSinceLastPromotion
0
2
4
6
8
10
12
14
Yes
No
Attrition

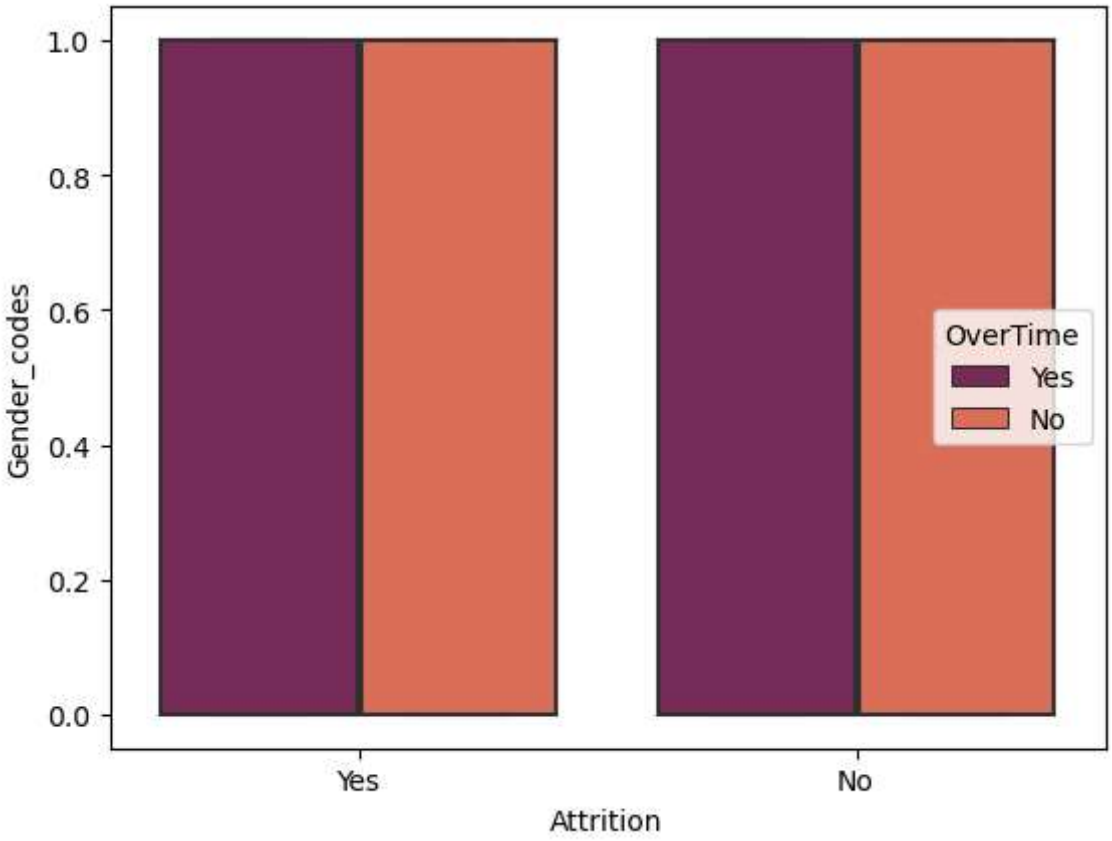

OverTime
Yes
No
YearsWithCurrManager
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Yes
No
Attrition

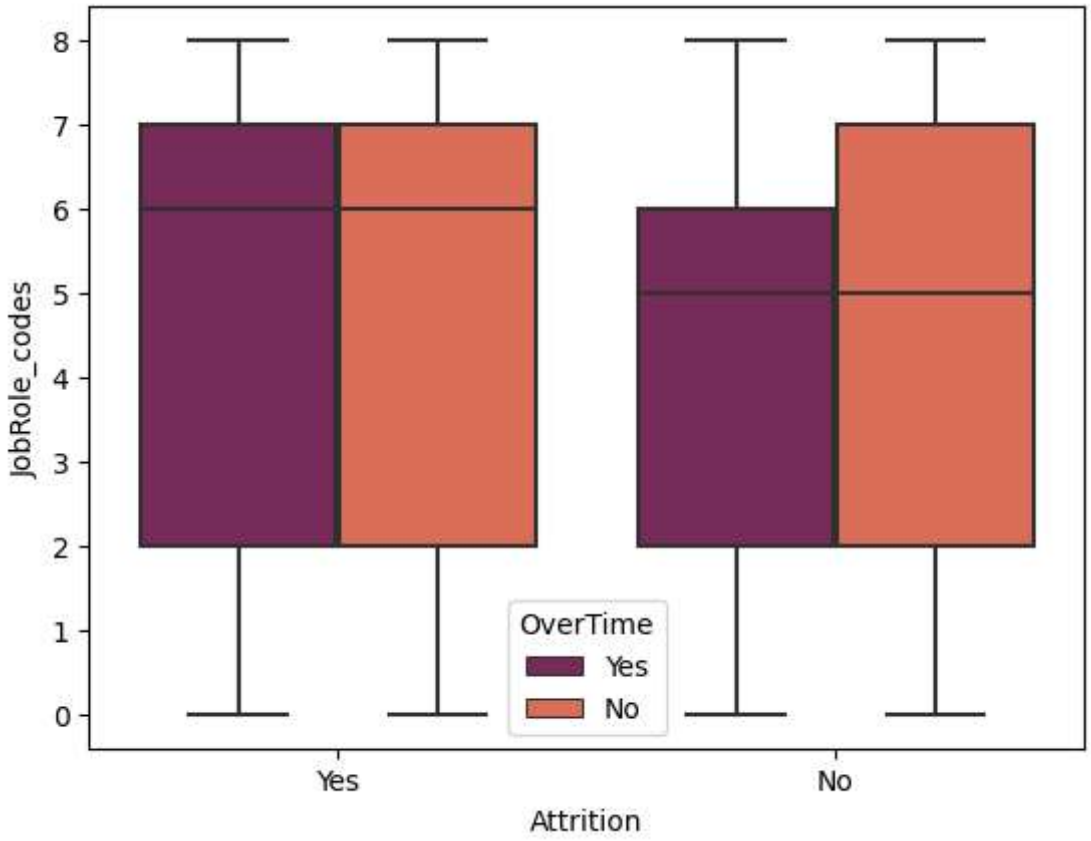

OverTime
Yes
No
1.0
0.8
0.6
0.4
0.2
0.0
Attrition_codes
Yes
No
Attrition

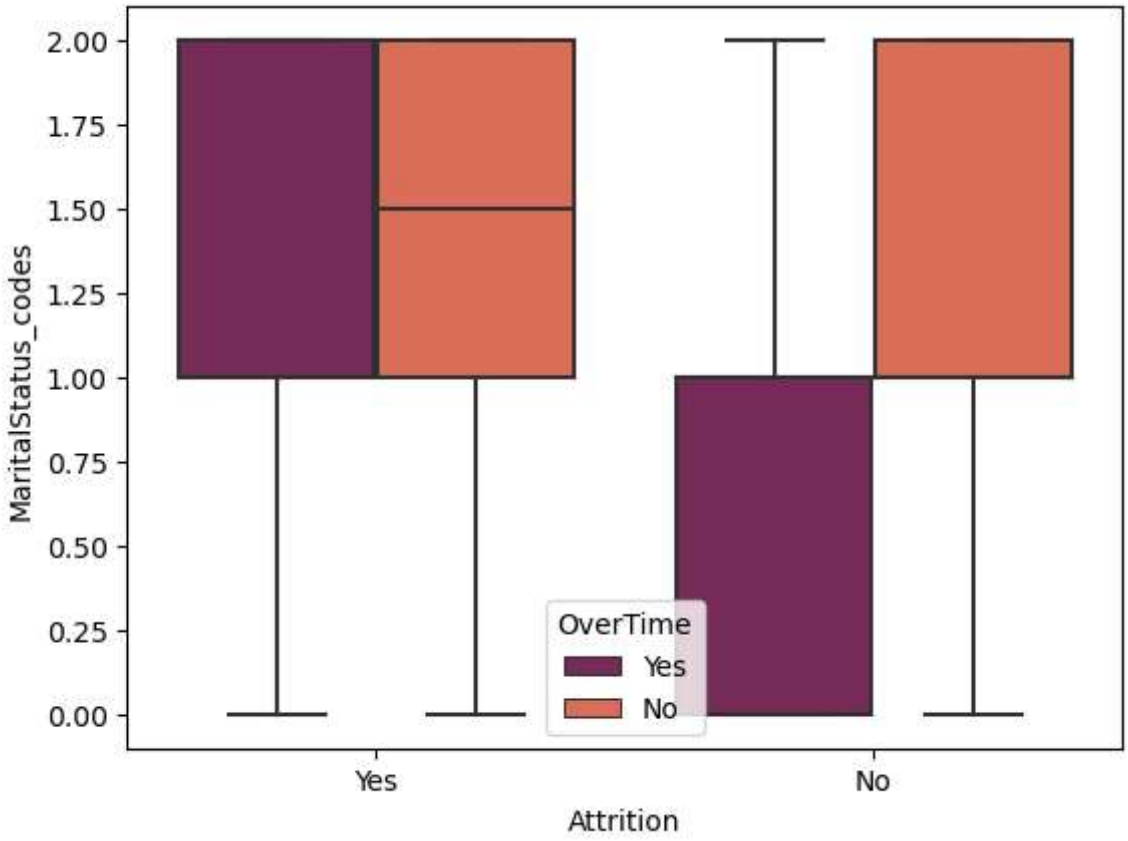

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
BusinessTravel_codes
OverTime
Yes
No
Yes
No
Attrition

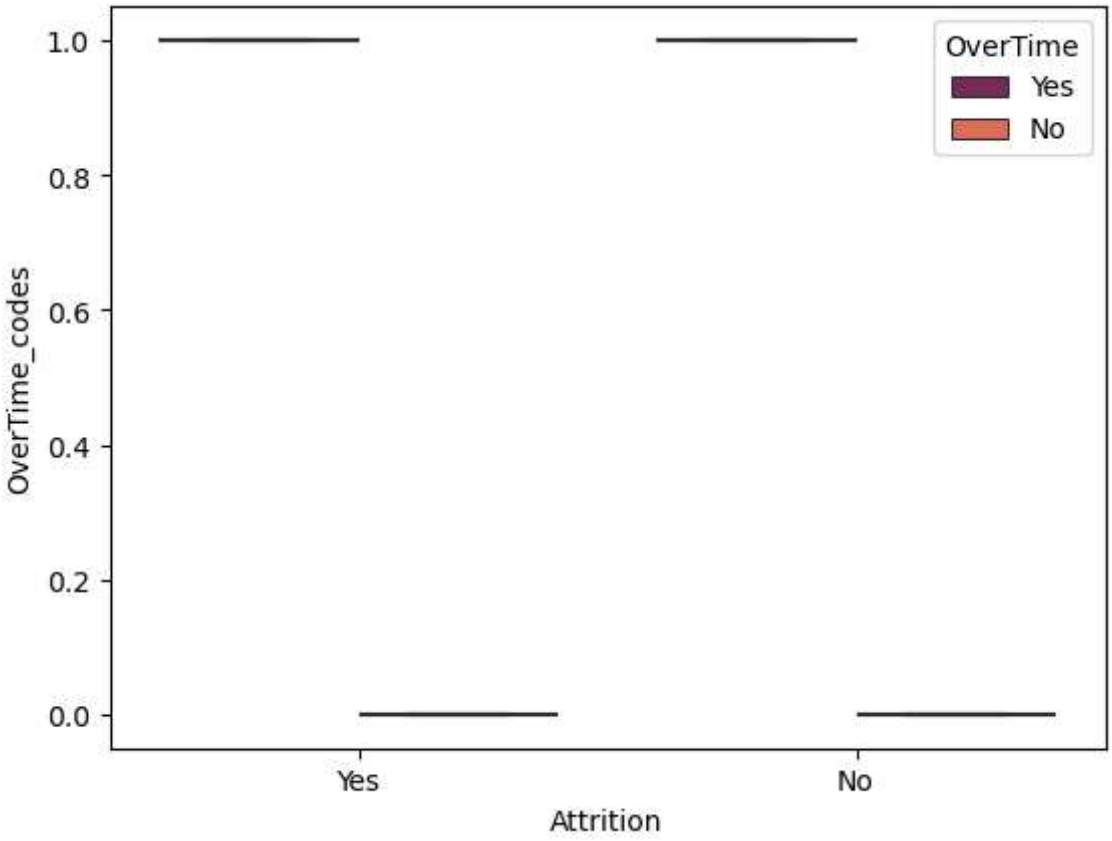

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Department_codes
OverTime
Yes
No
Yes
No
Attrition


5
4
3
2
1
0
EducationField_codes
OverTime
Yes
No
Yes
No
Attrition

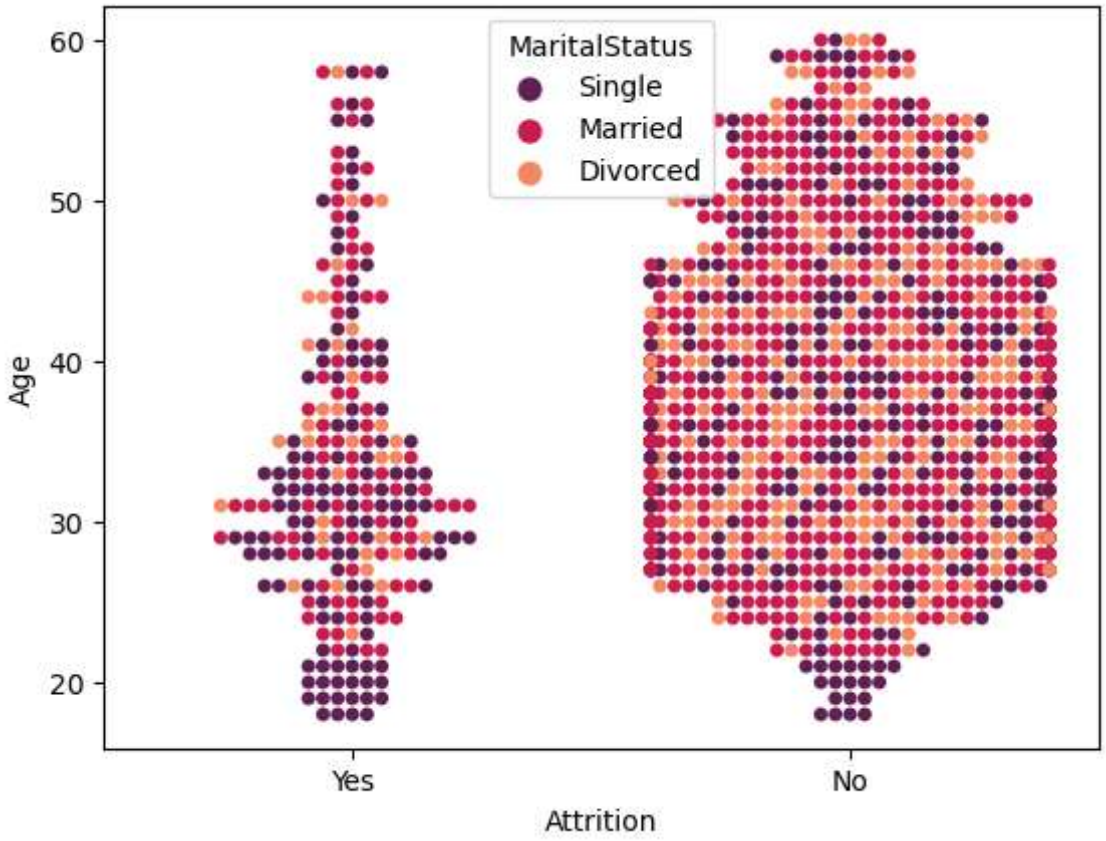

1.0
0.8
0.6
0.4
0.2
0.0
Gender_codes
OverTime
Yes
No
Yes
No
Attrition

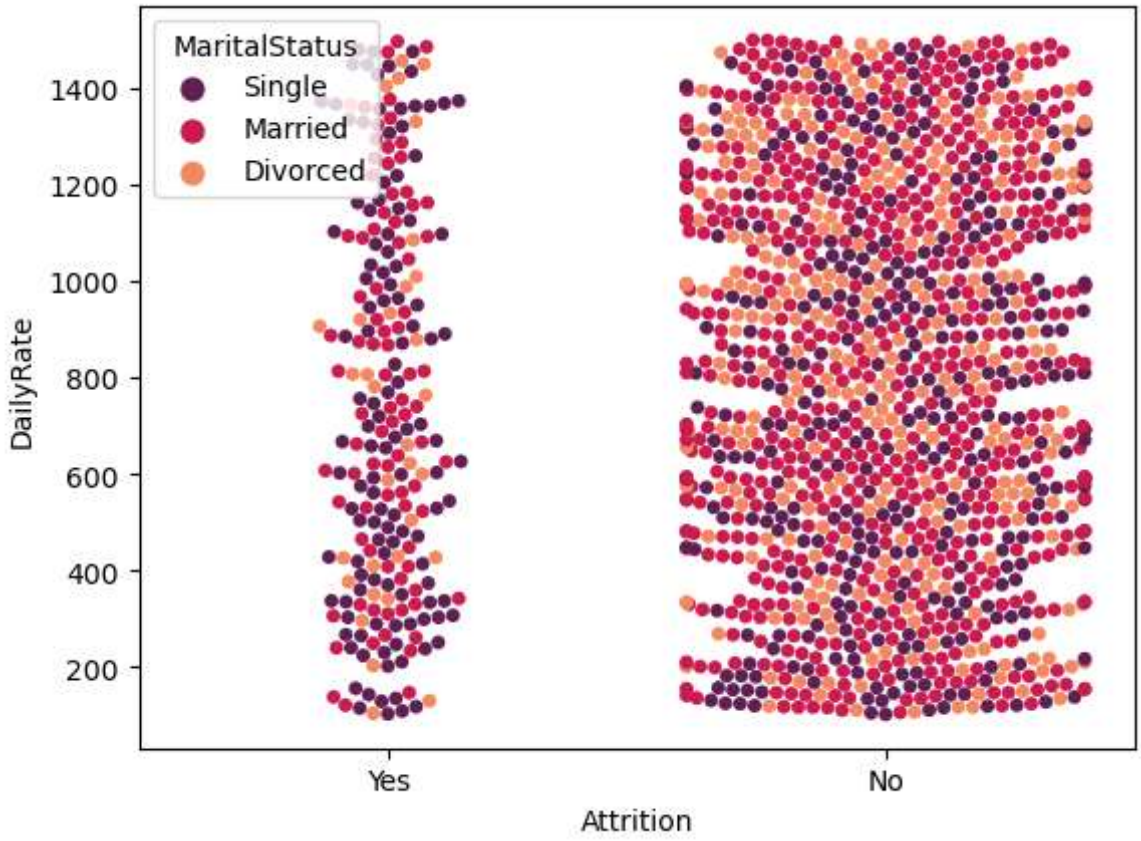

8
7
6
5
4
3
2
1
0
JobRole_codes
OverTime
Yes
No
Yes
No
Attrition

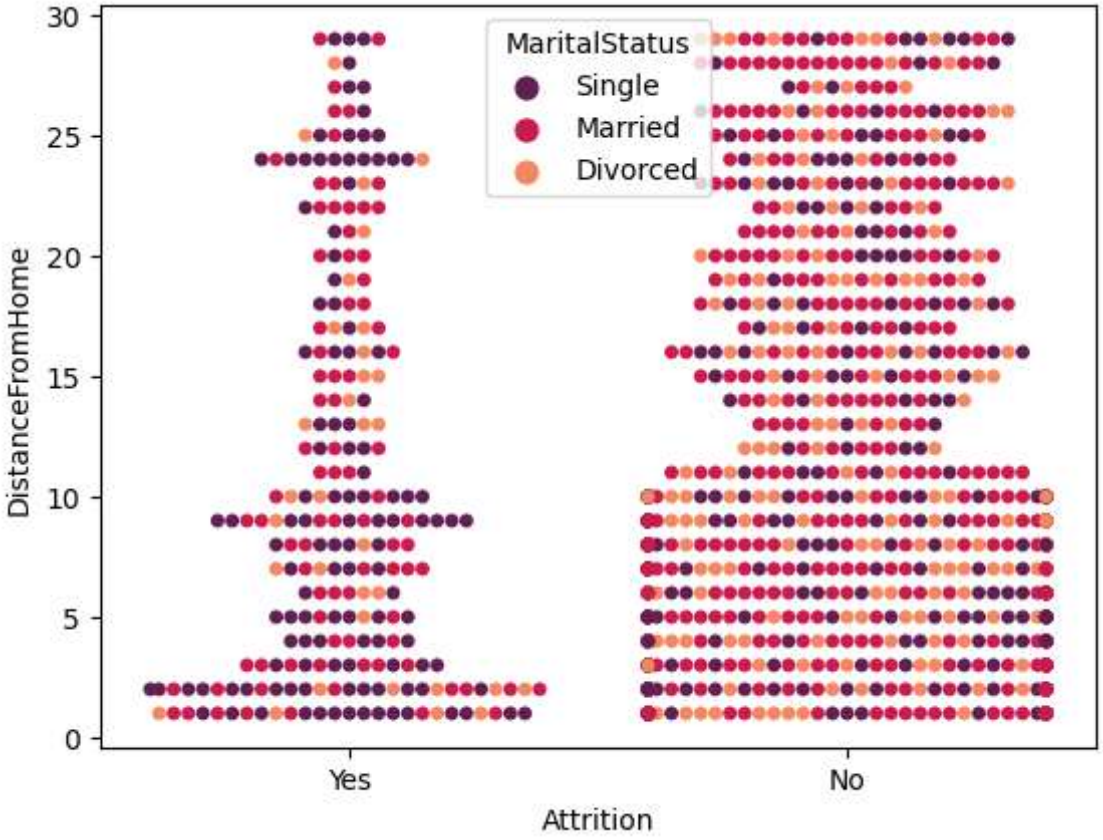


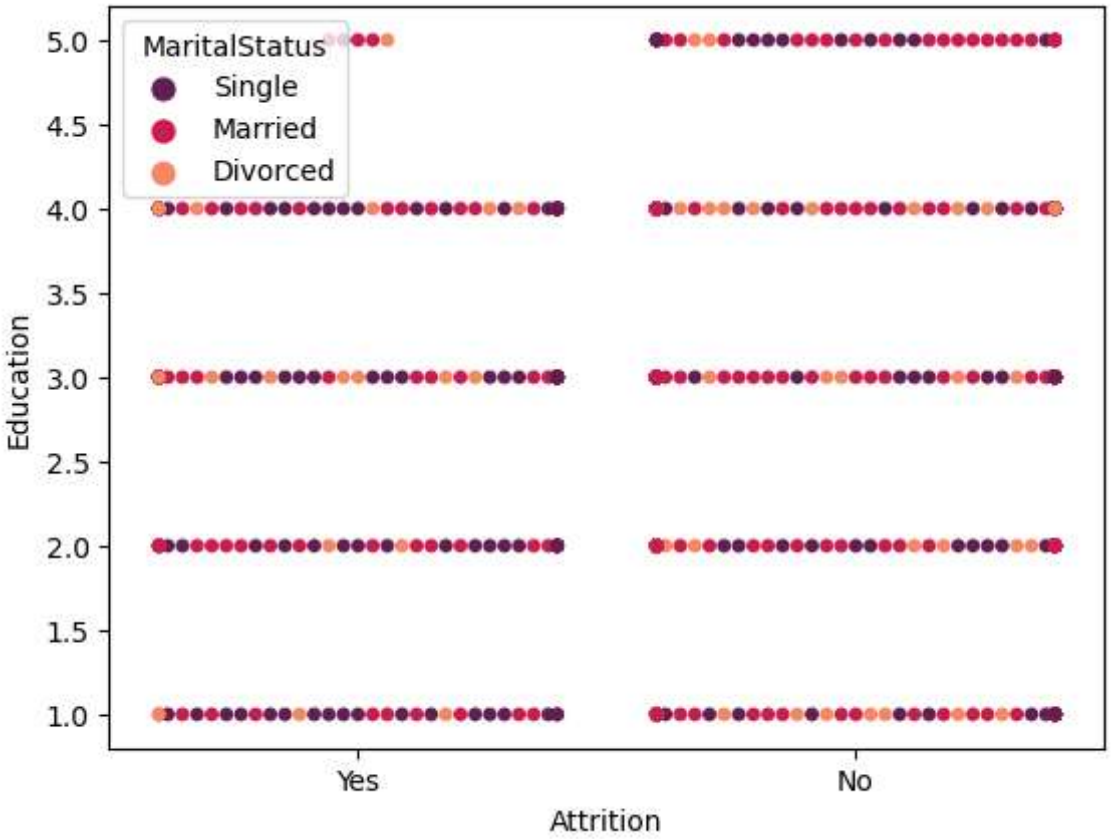


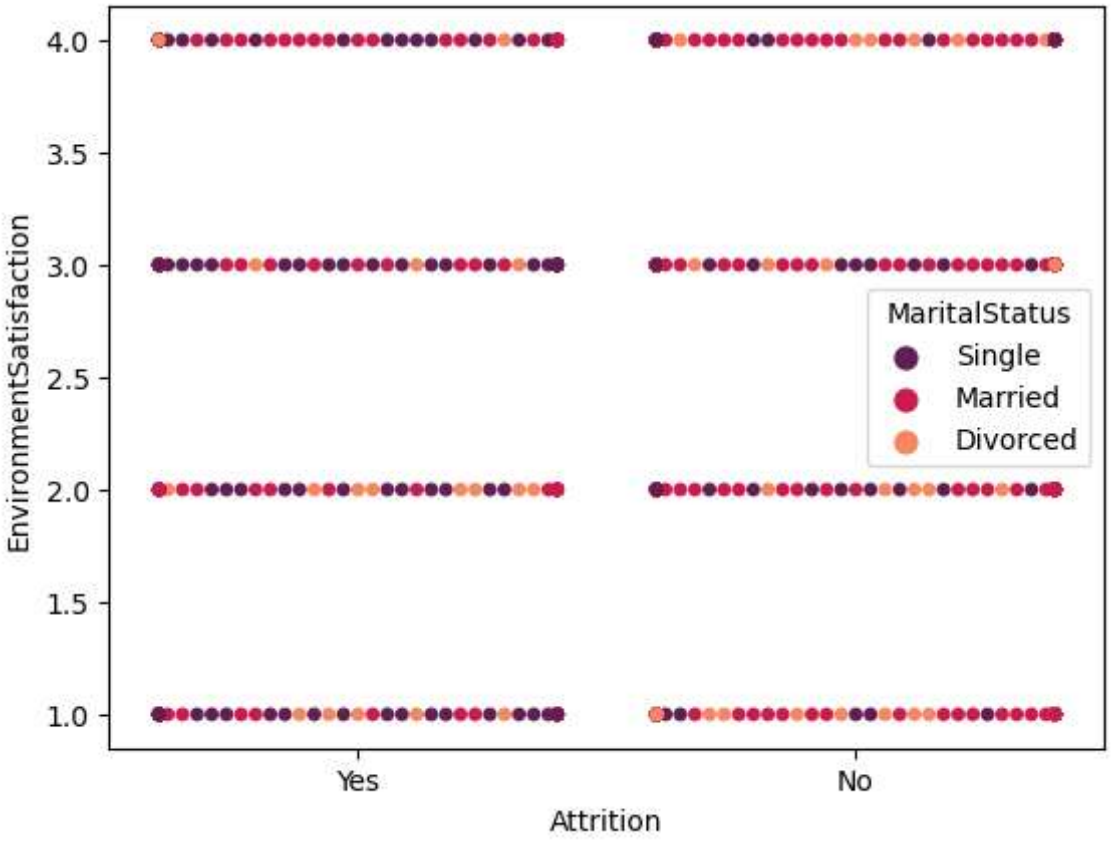
In [30]:
```
for AA in ibm.columns[8:]:
    sns.swarmplot(x='Attrition',y=AA,data=ibm,hue='MaritalStatus',palette='rocket'
    plt.show()
```

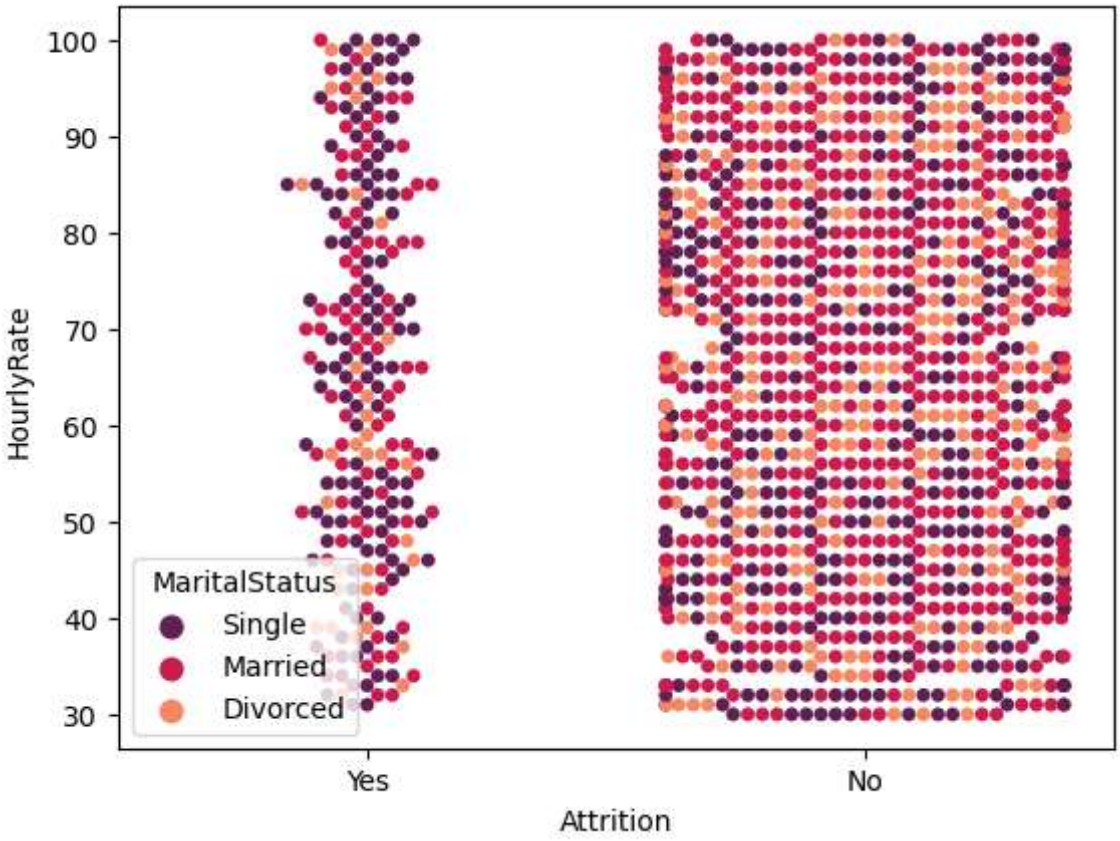

MaritalStatus
Single
Married
Divorced
Age
60
50
40
30
20
Yes
No
Attrition

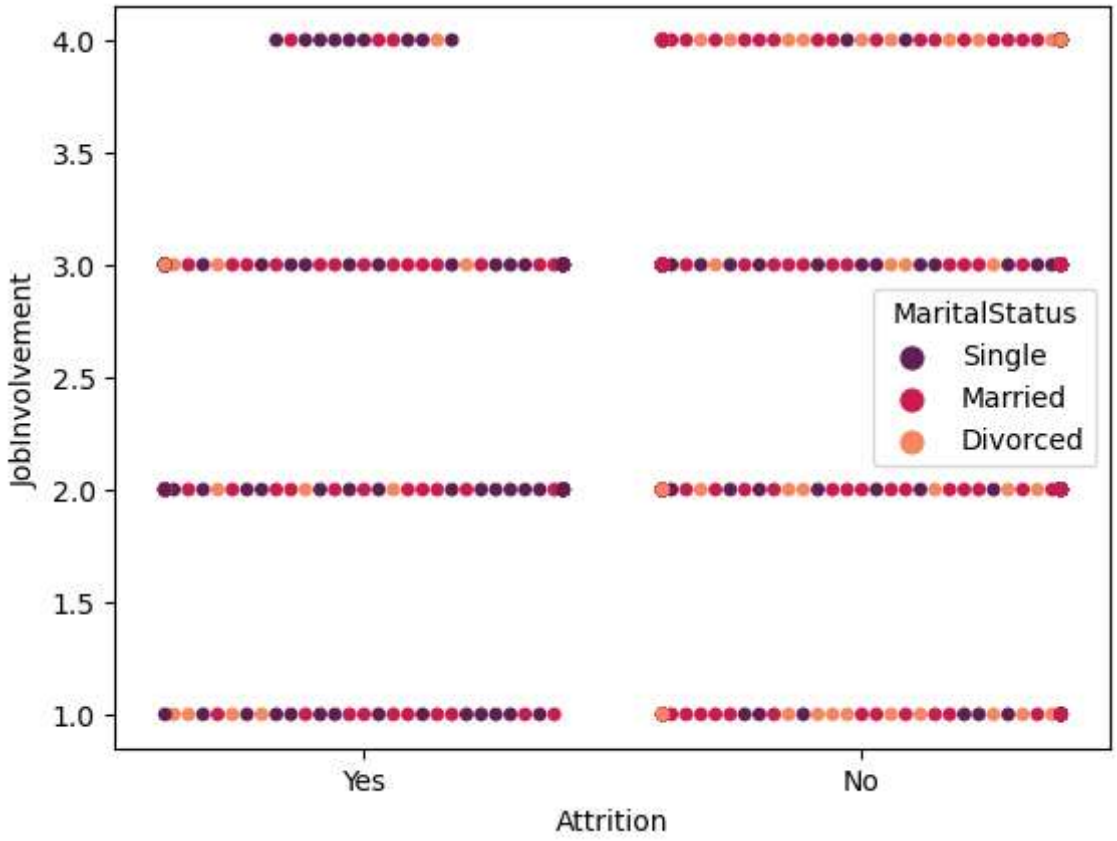

MaritalStatus
Single
Married
Divorced
DailyRate
1400
1200
1000
800
600
400
200
Yes
No
Attrition

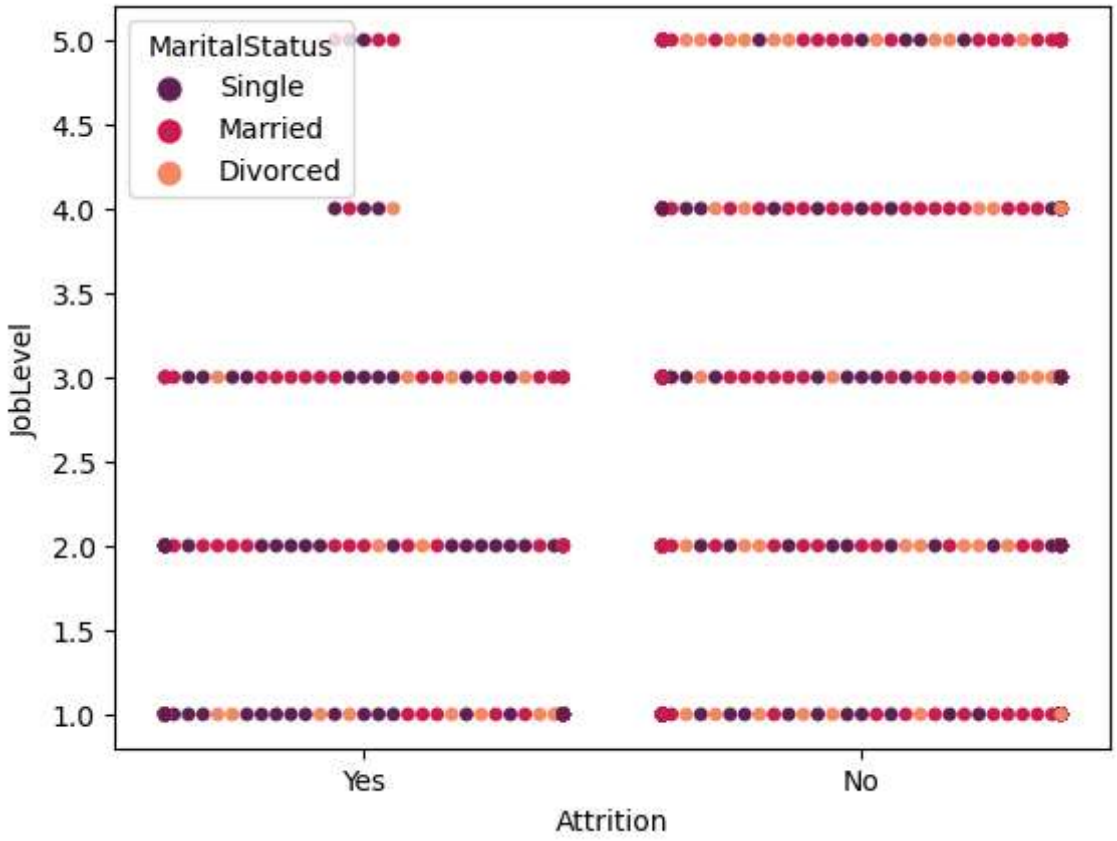

MaritalStatus
Single
Married
Divorced
DistanceFromHome
0
5
10
15
20
25
30
Yes
No
Attrition

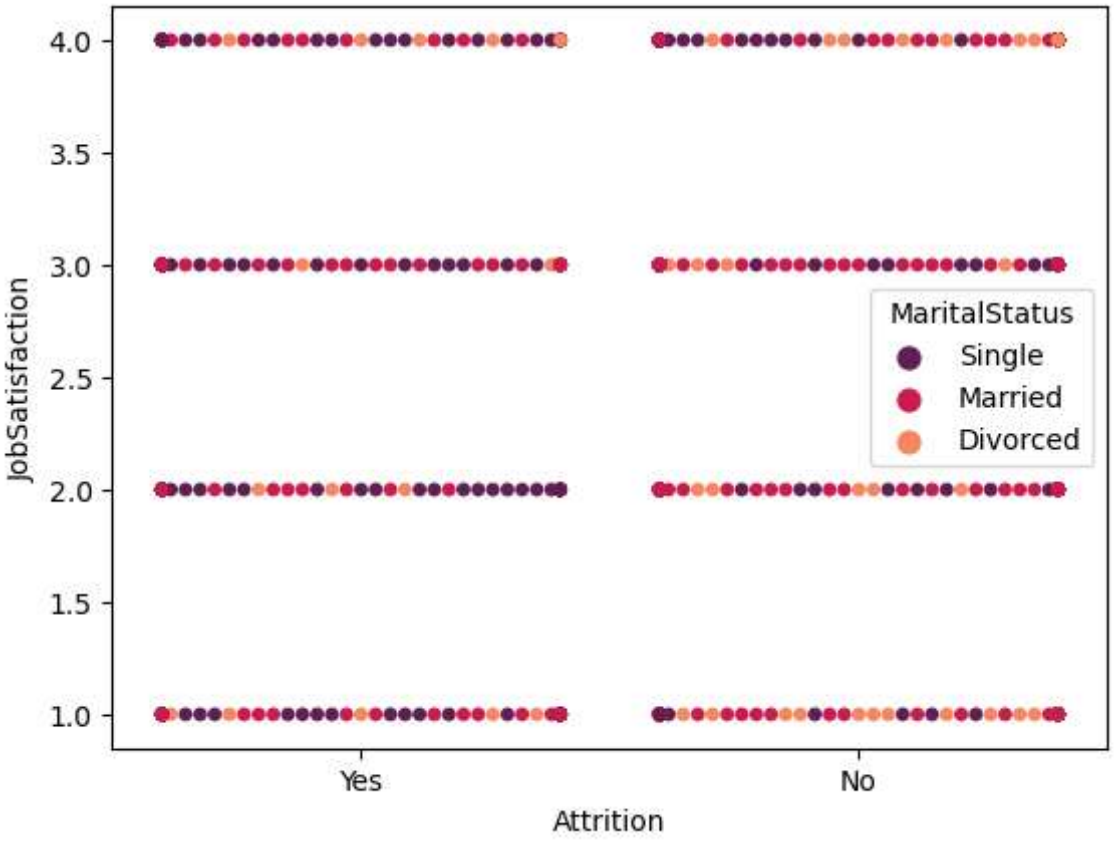

MaritalStatus
Single
Married
Divorced
Education
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Yes
No
Attrition

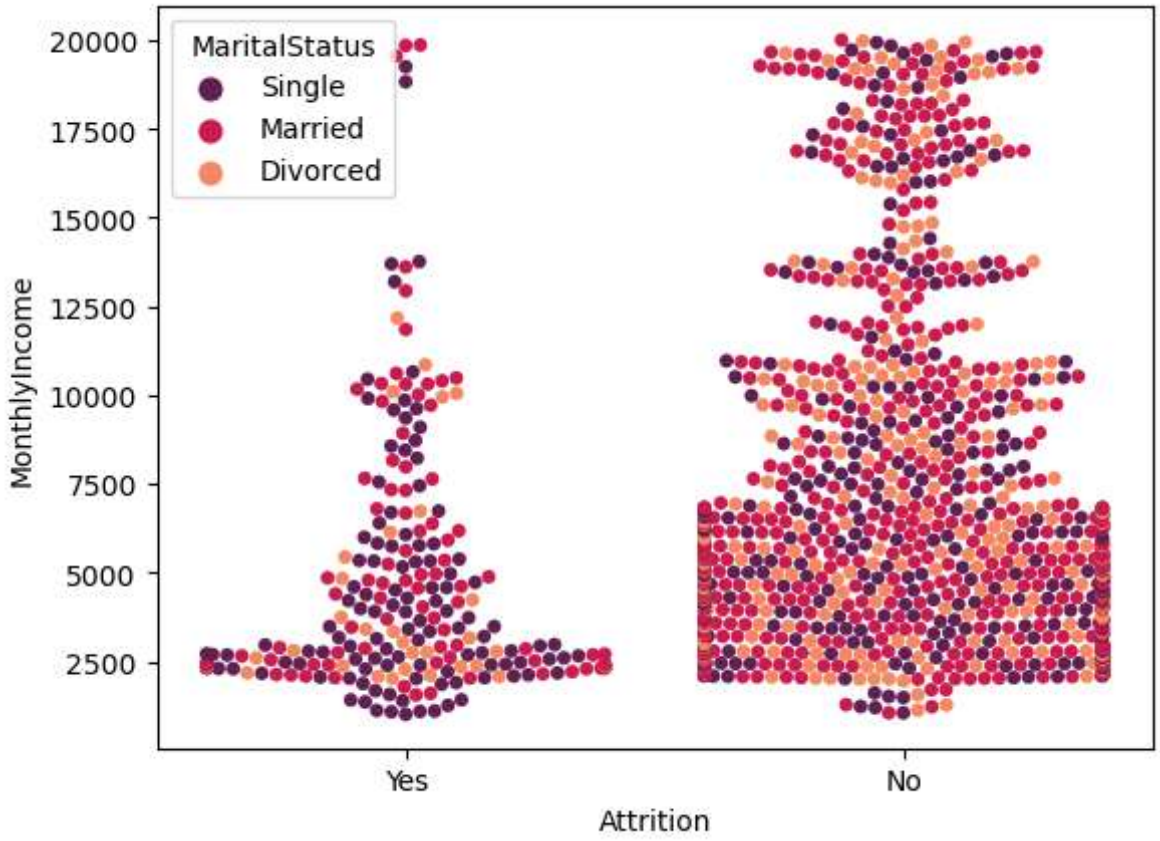

4.0
3.5
3.0
2.5
2.0
1.5
1.0
EnvironmentSatisfaction
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

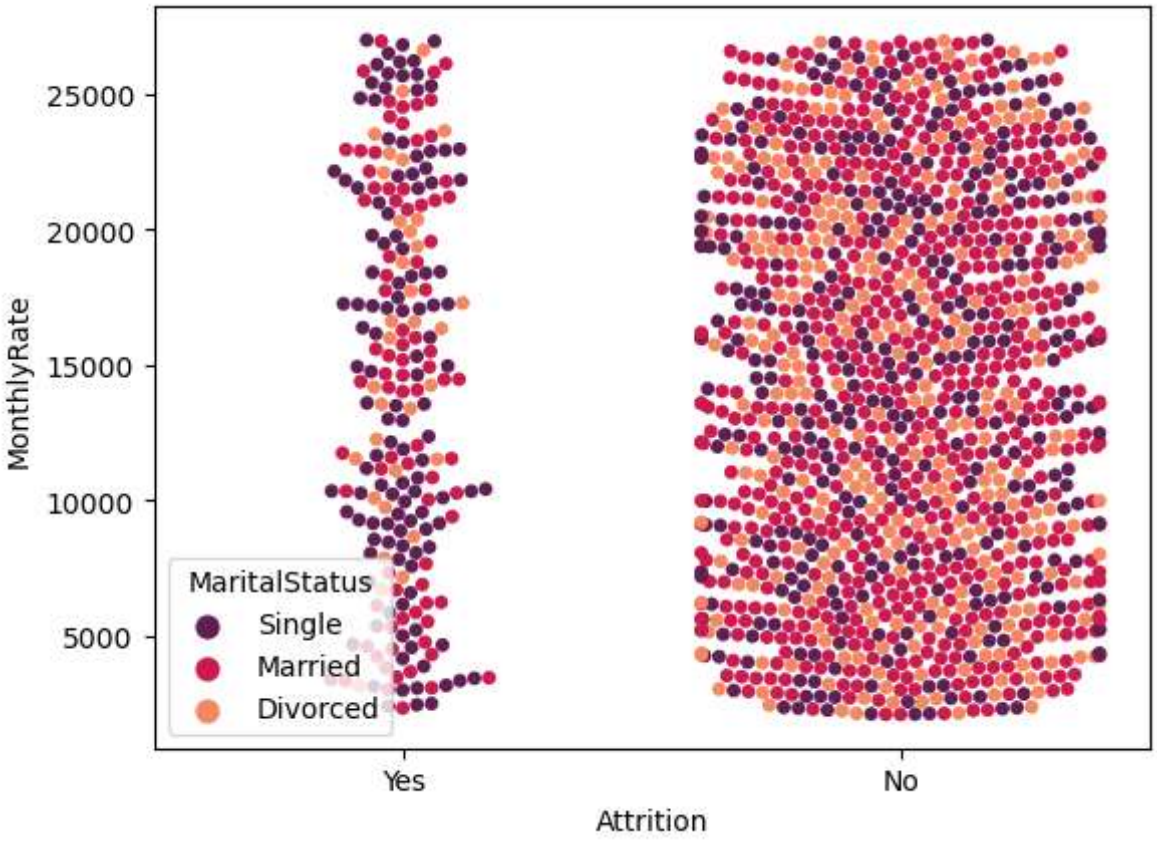

100
90
80
70
60
50
40
30
HourlyRate
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

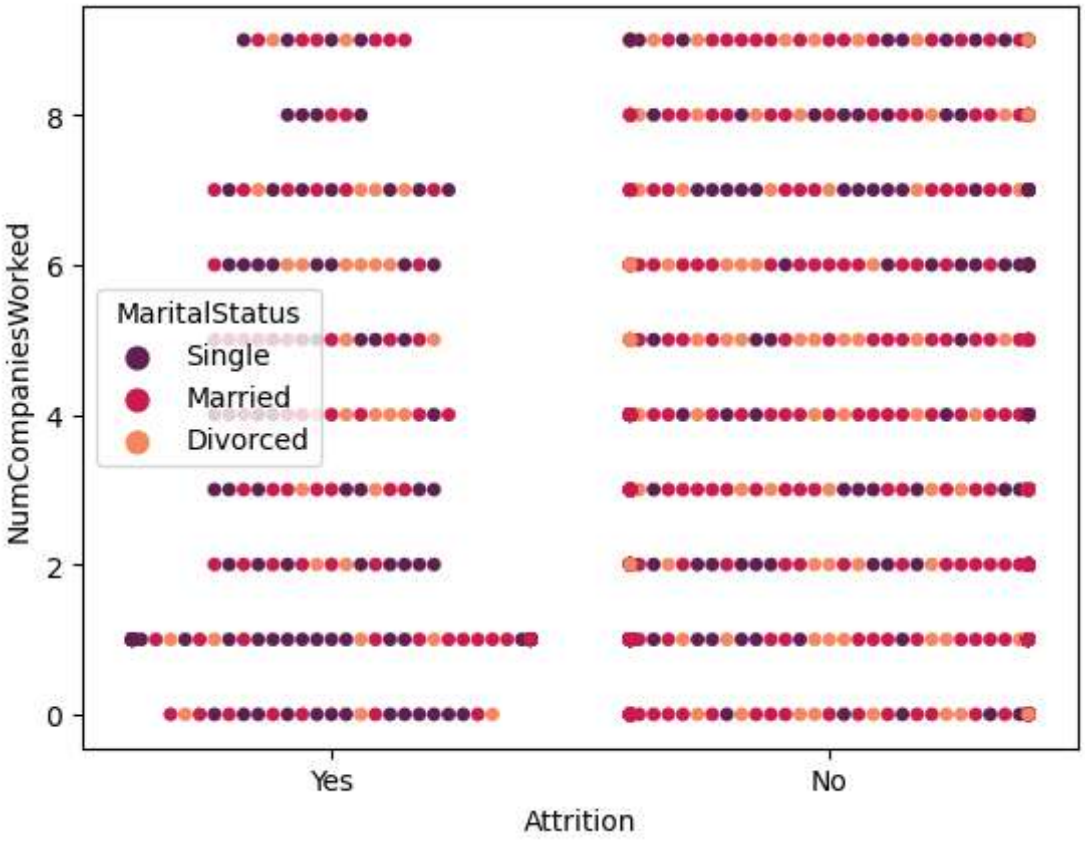

JobInvolvement
4.0
3.5
3.0
2.5
2.0
1.5
1.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

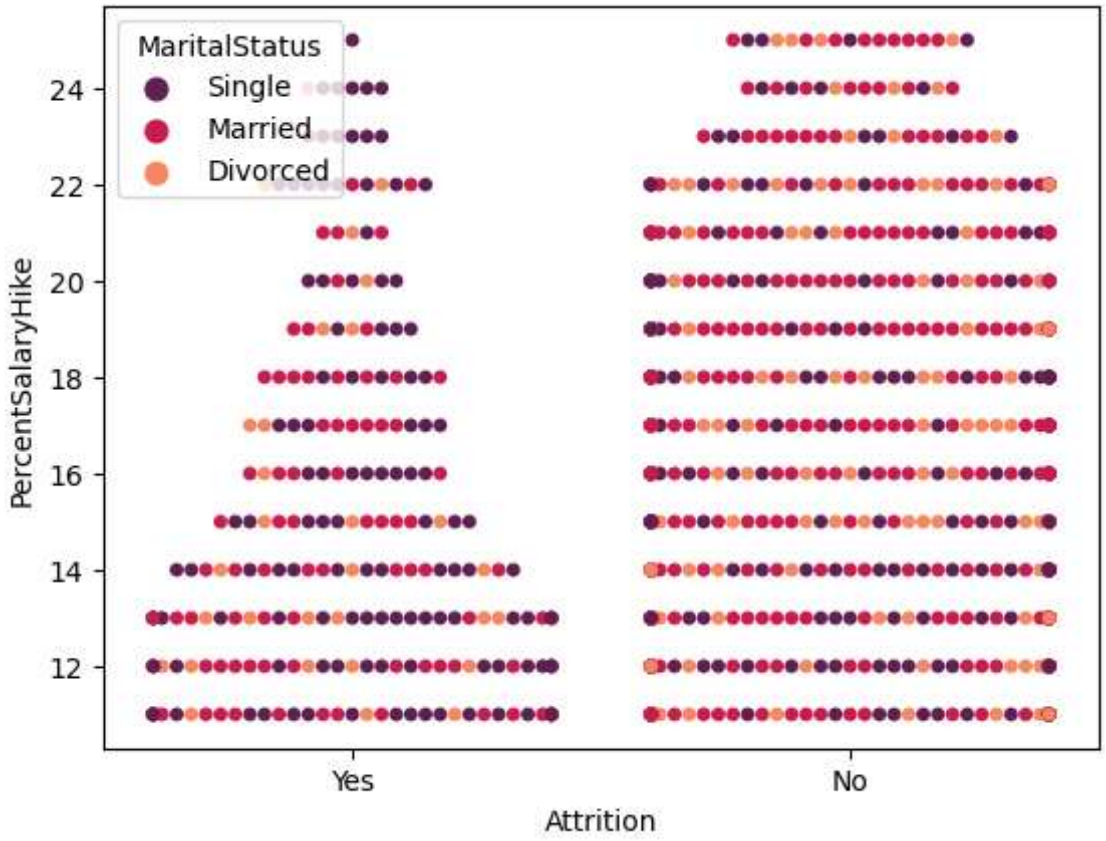

JobLevel
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

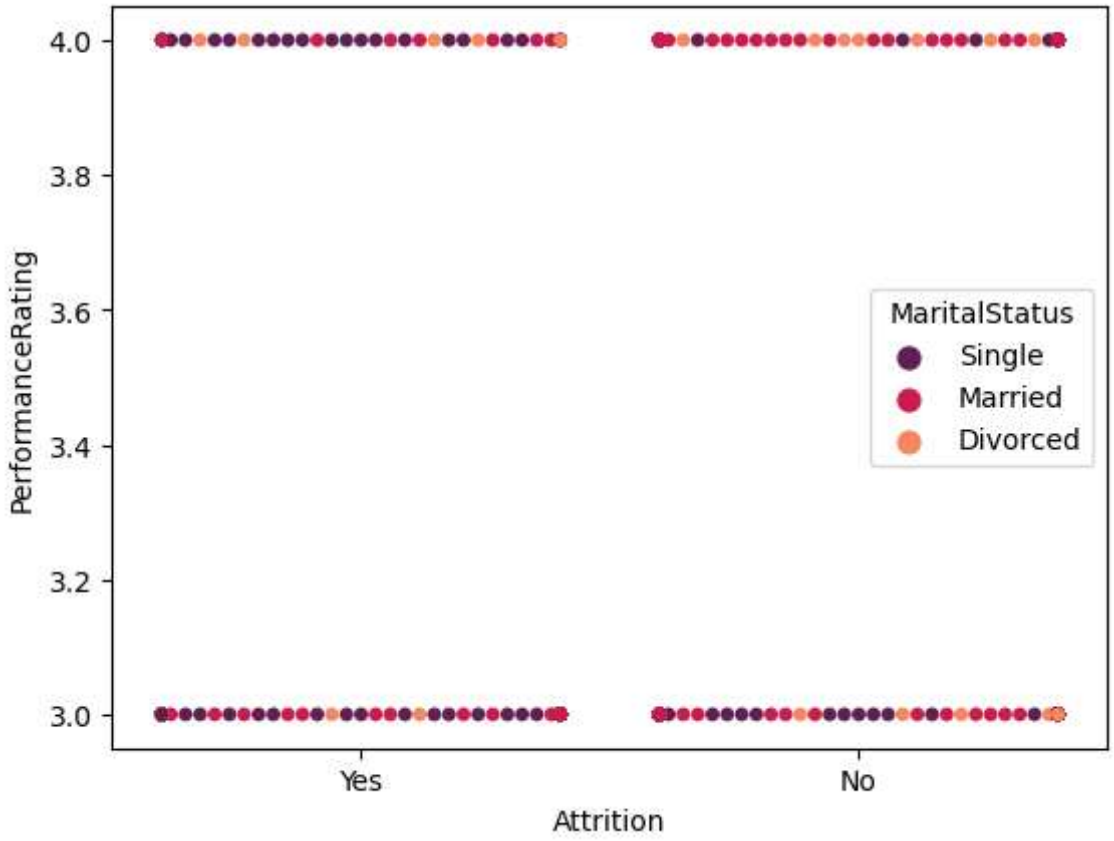

4.0
3.5
3.0
2.5
2.0
1.5
1.0
JobSatisfaction
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

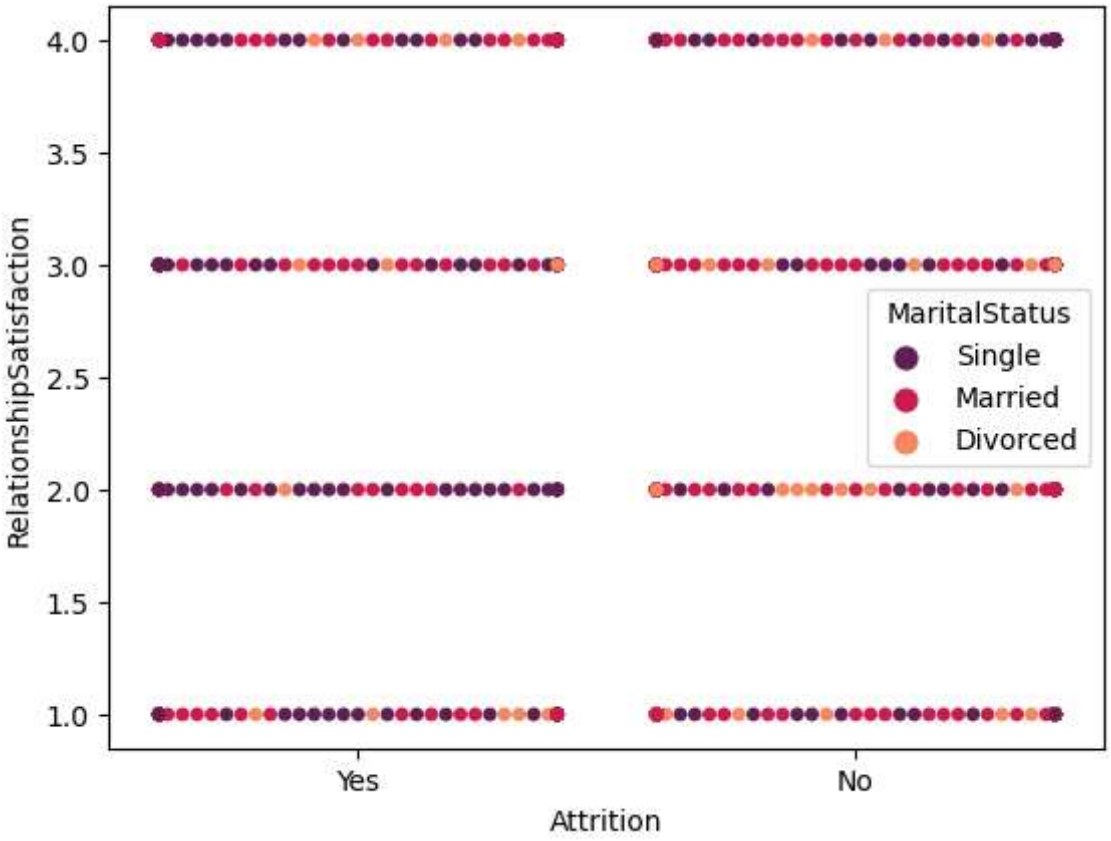

20000
17500
15000
12500
10000
7500
5000
2500
MonthlyIncome
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

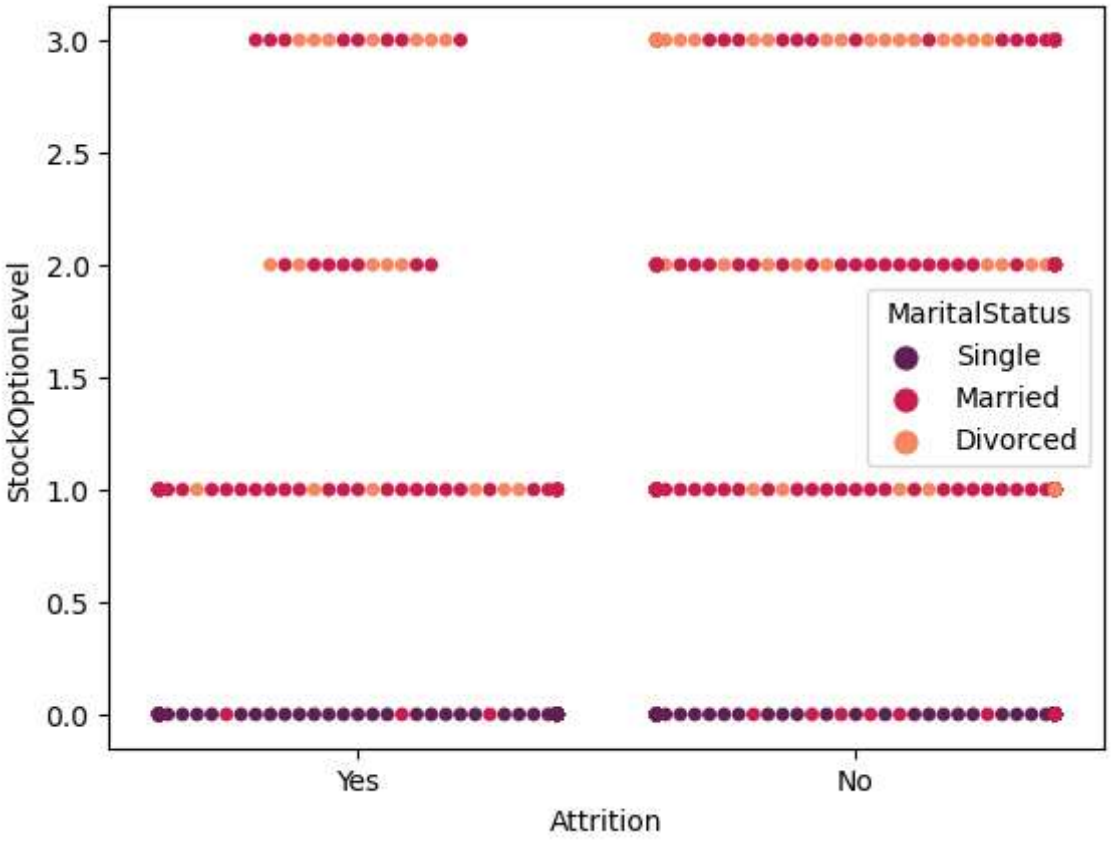

25000
20000
15000
10000
5000
MonthlyRate
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

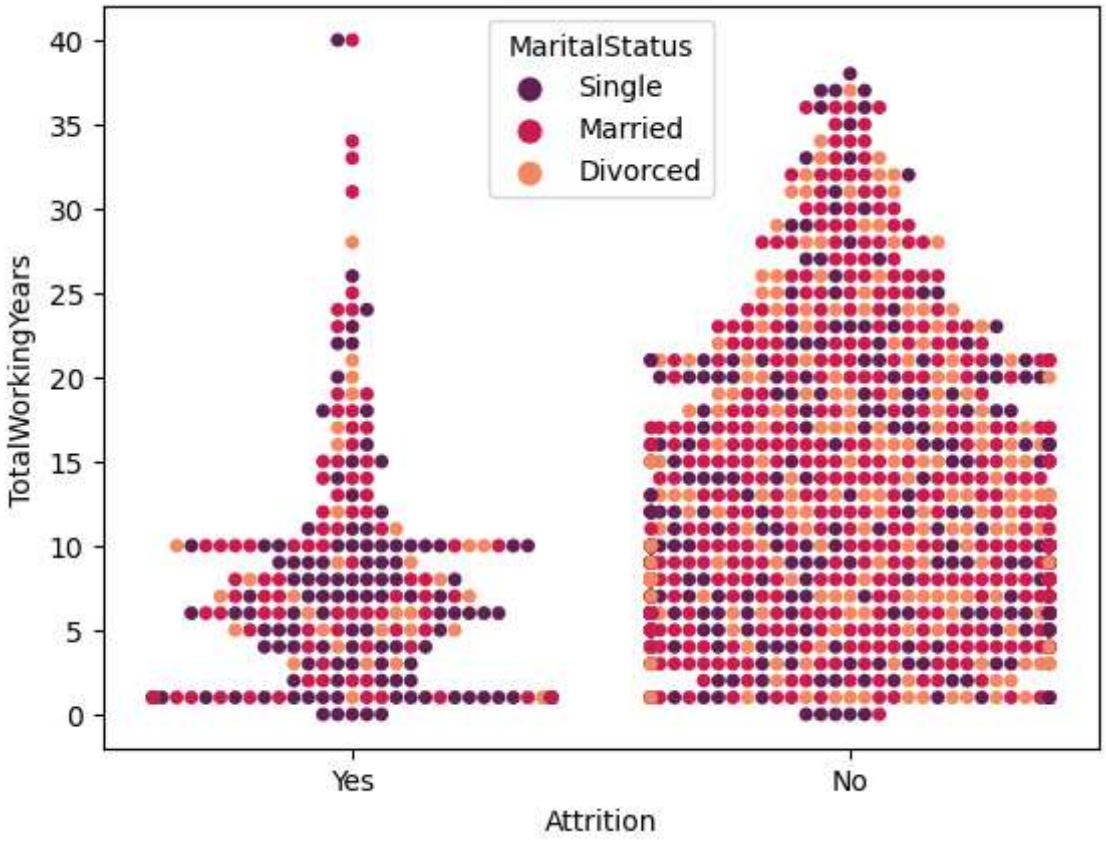

8
6
4
2
0
NumCompaniesWorked
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

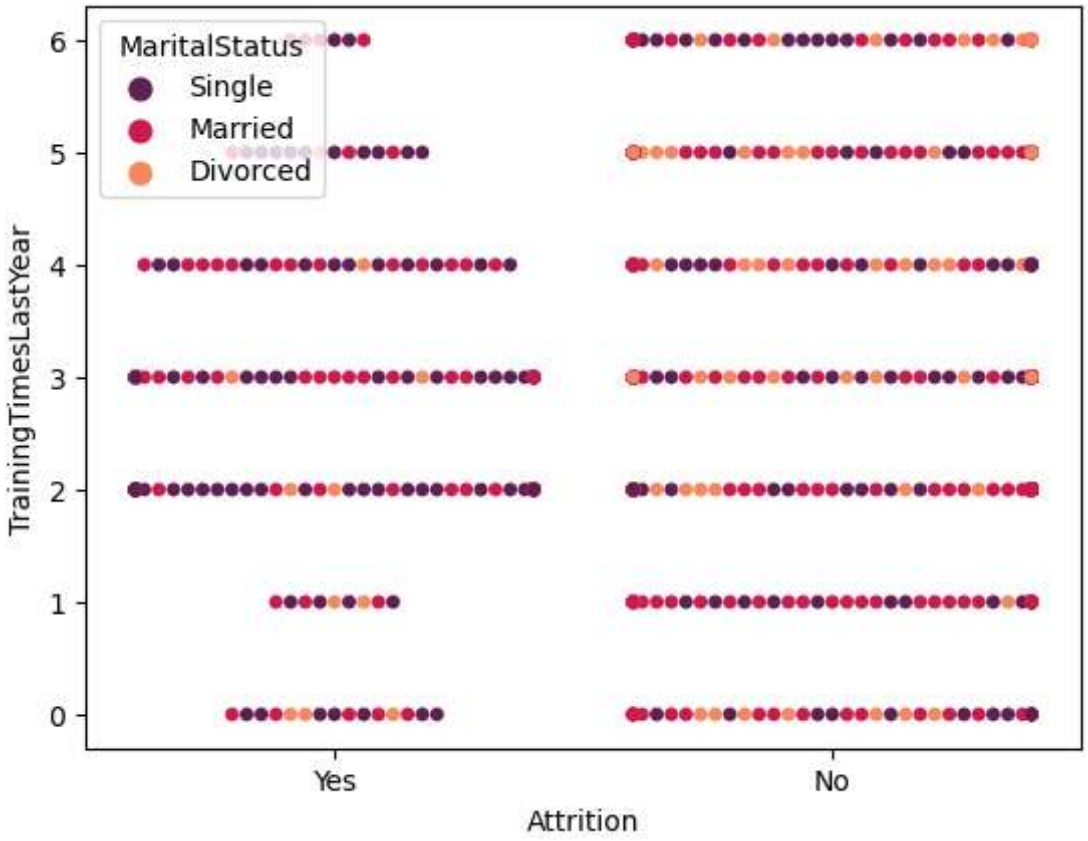

MaritalStatus
Single
Married
Divorced
PercentSalaryHike
24
22
20
18
16
14
12
Yes
No
Attrition

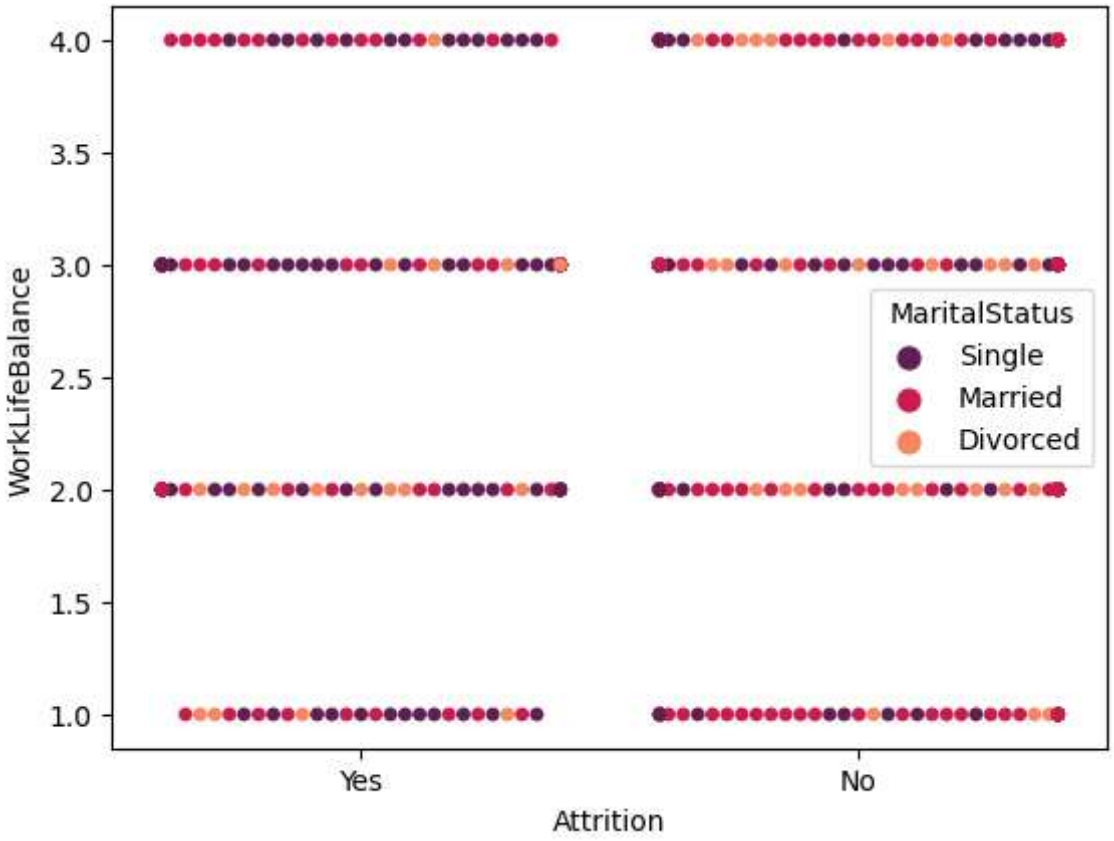

PerformanceRating
4.0
3.8
3.6
3.4
3.2
3.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

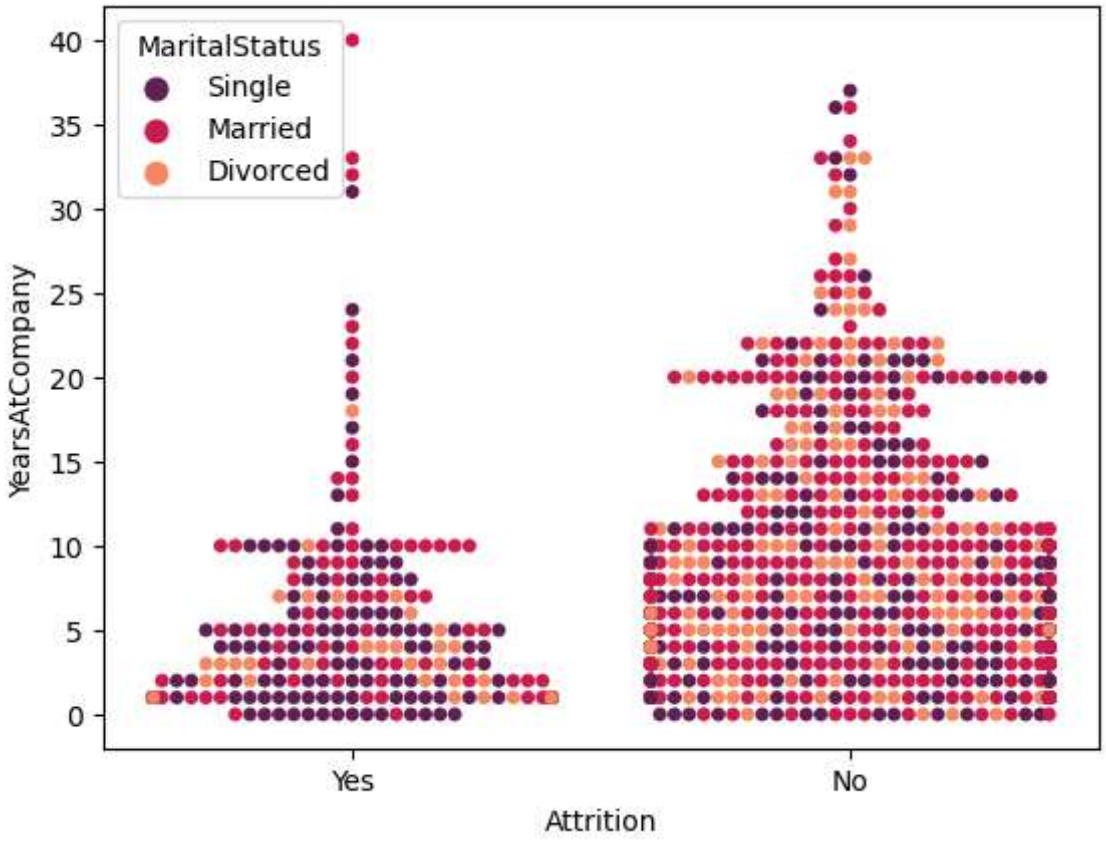

RelationshipSatisfaction
4.0
3.5
3.0
2.5
2.0
1.5
1.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

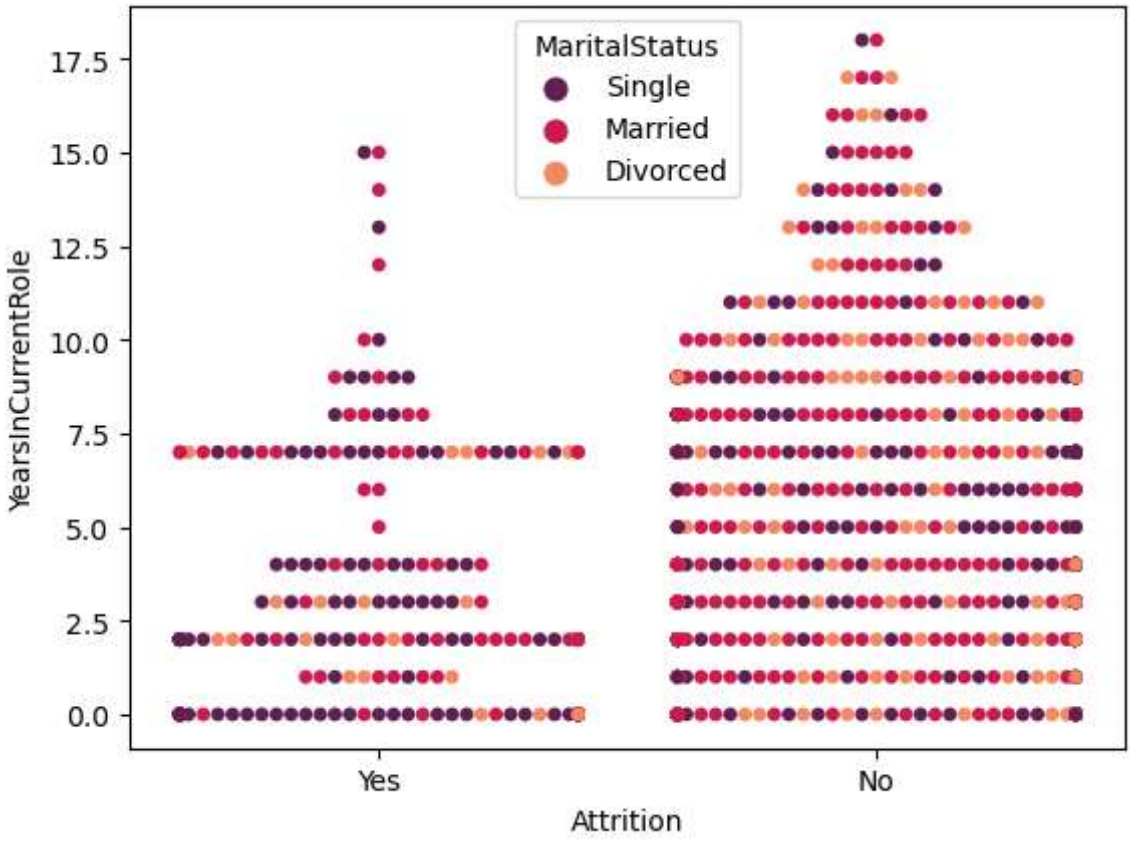

StockOptionLevel
3.0
2.5
2.0
1.5
1.0
0.5
0.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

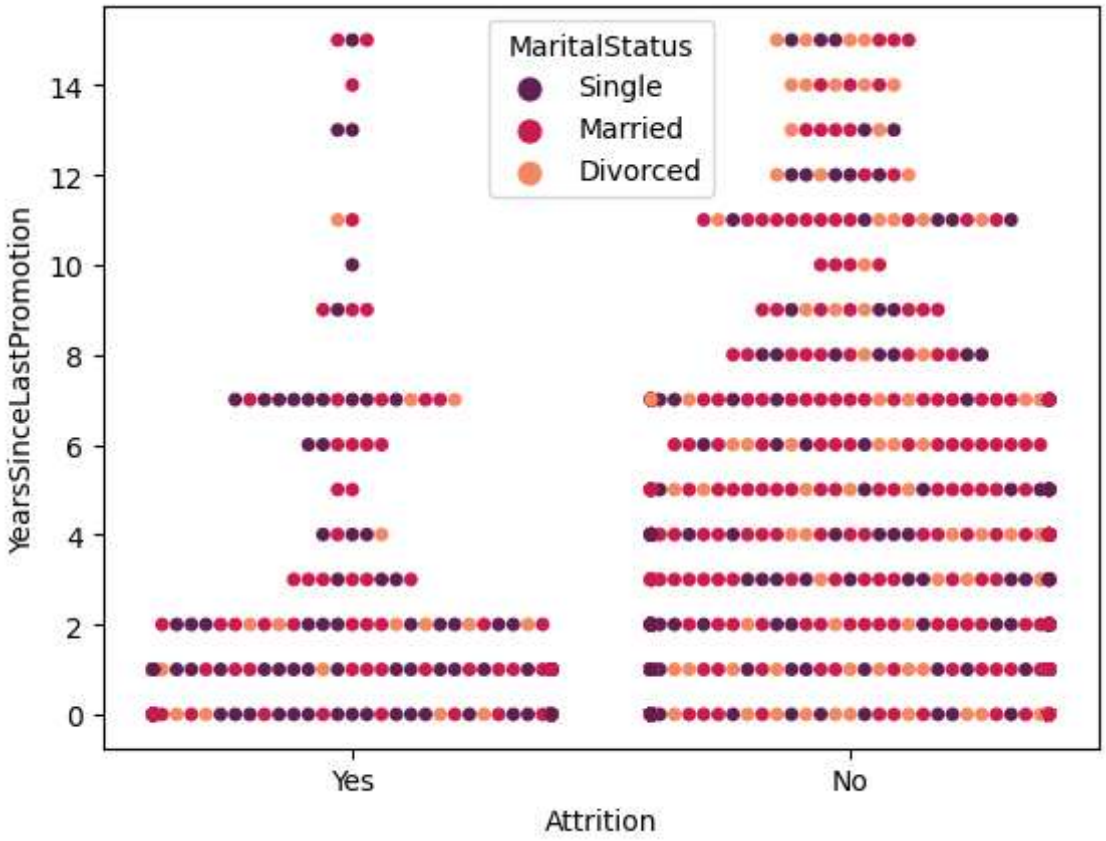

MaritalStatus
Single
Married
Divorced
TotalWorkingYears
0
5
10
15
20
25
30
35
40
Yes
No
Attrition

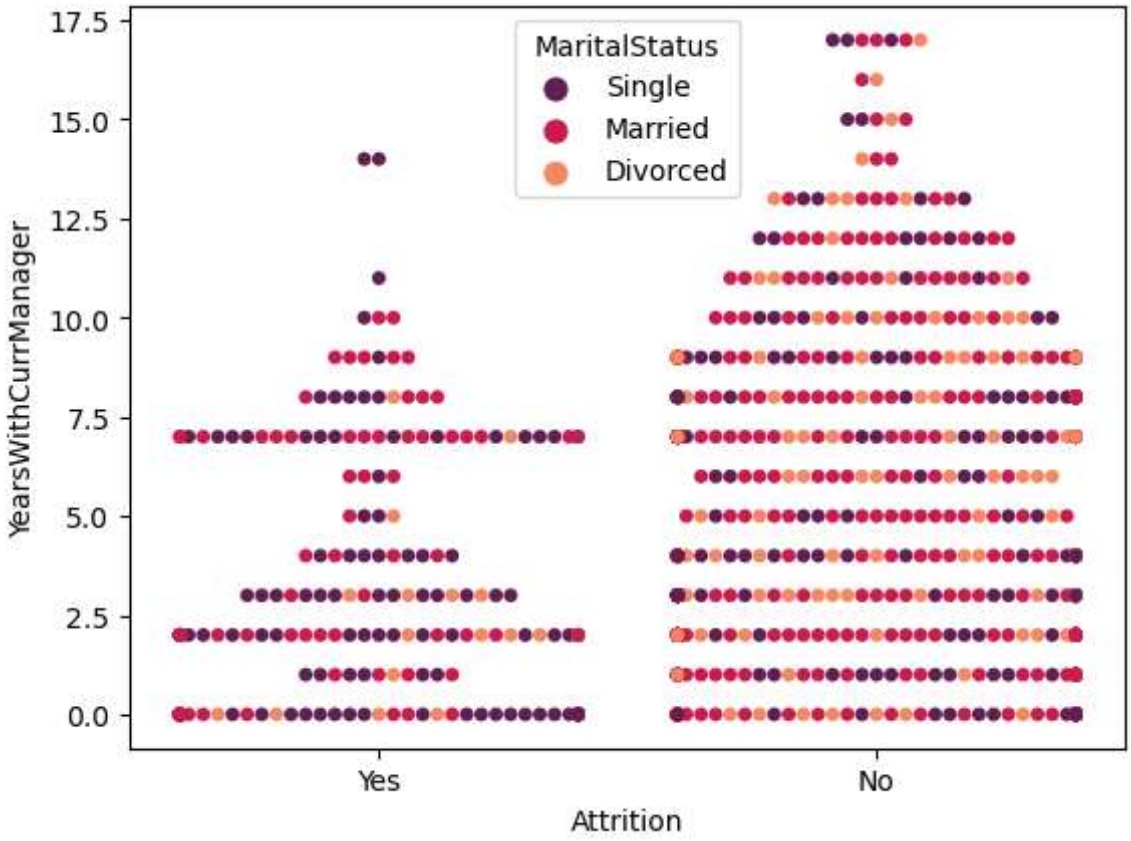

MaritalStatus
Single
Married
Divorced
TrainingTimesLastYear
0
1
2
3
4
5
6
Yes
No
Attrition

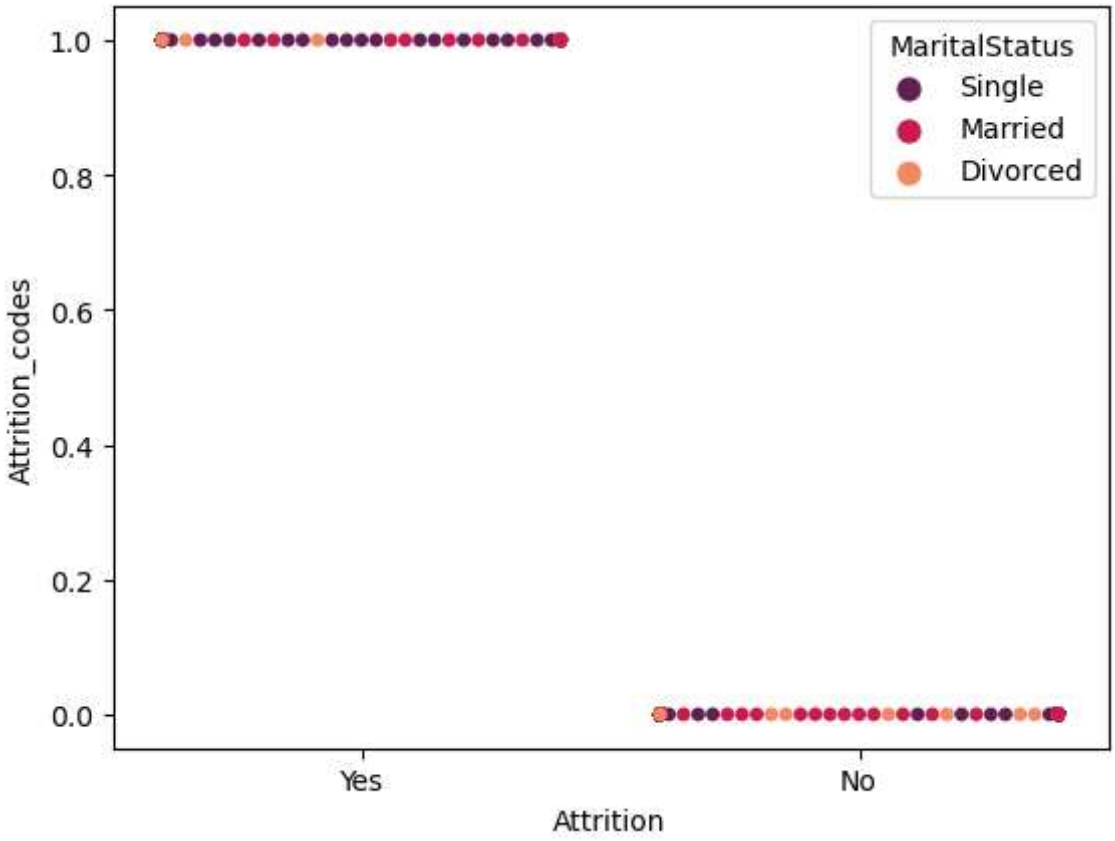

WorkLifeBalance
4.0
3.5
3.0
2.5
2.0
1.5
1.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

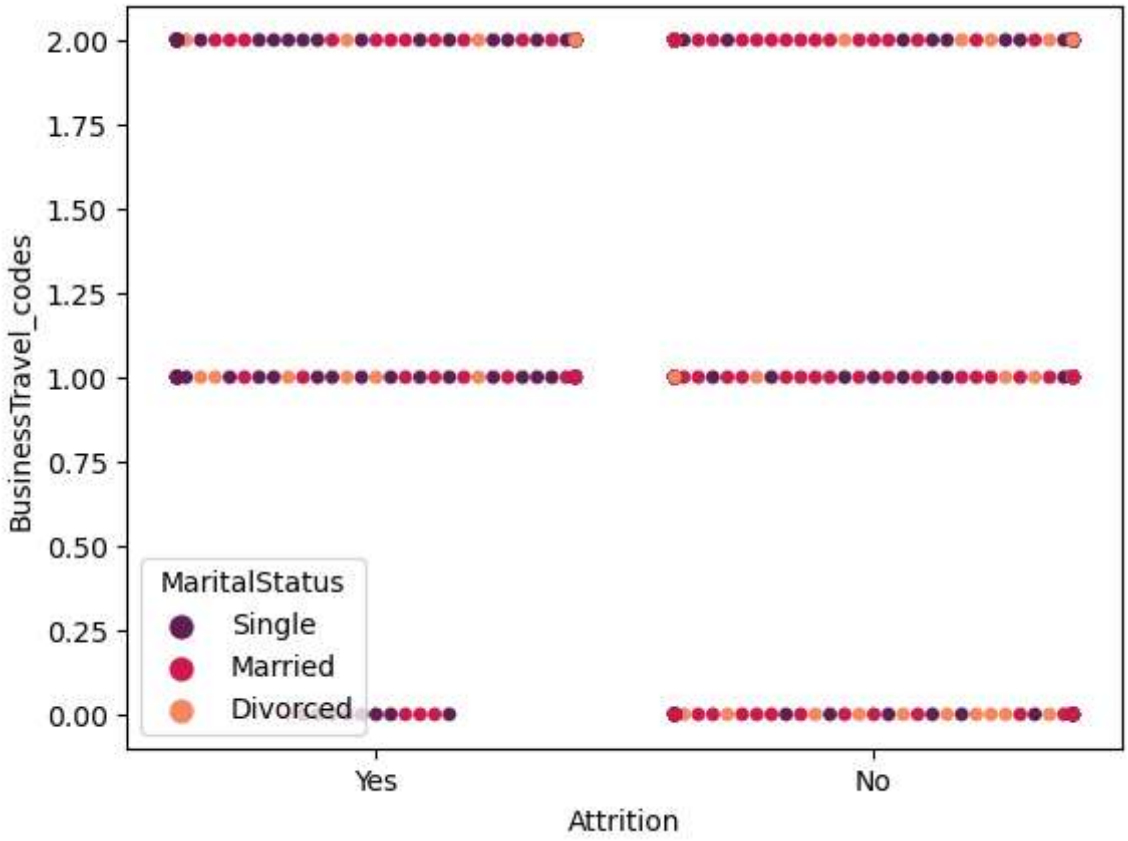

YearsAtCompany
40
35
30
25
20
15
10
5
0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

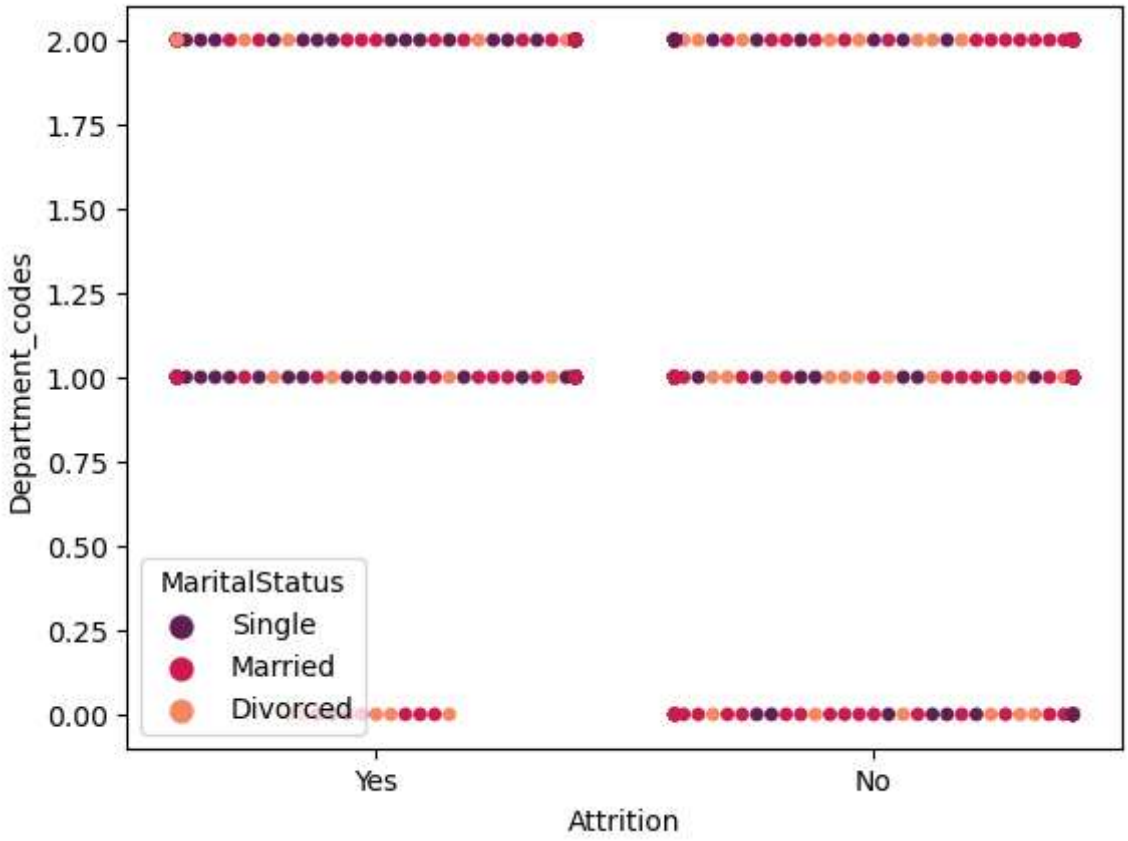

MaritalStatus
Single
Married
Divorced
YearsInCurrentRole
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Yes
No
Attrition

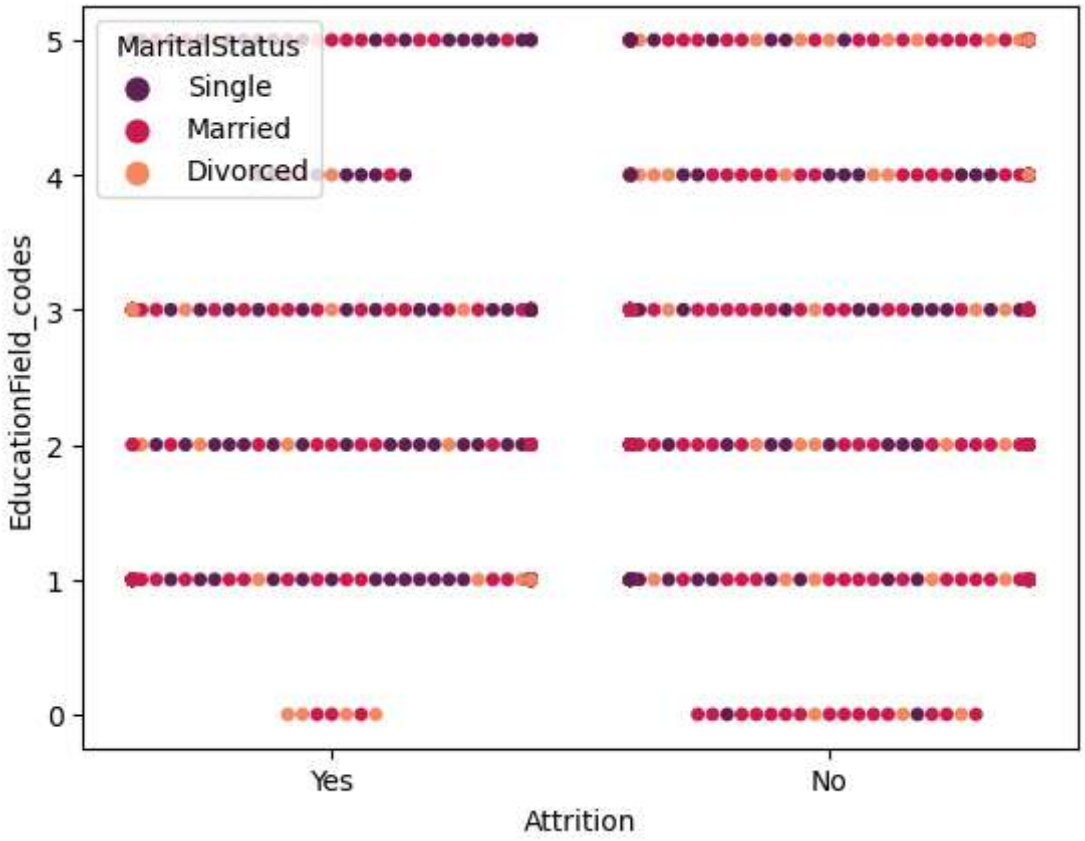

MaritalStatus
Single
Married
Divorced
YearsSinceLastPromotion
14
12
10
8
6
4
2
0
Yes
No
Attrition

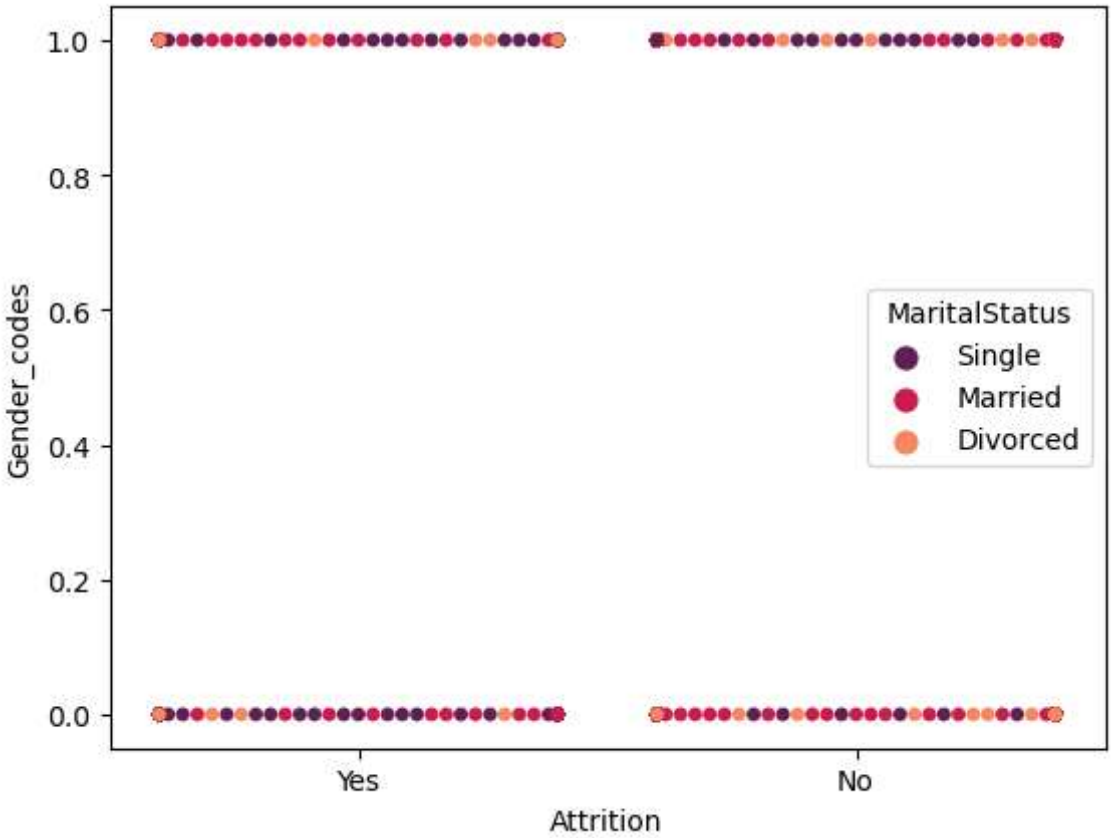

MaritalStatus
Single
Married
Divorced
YearsWithCurrManager
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Yes
No
Attrition

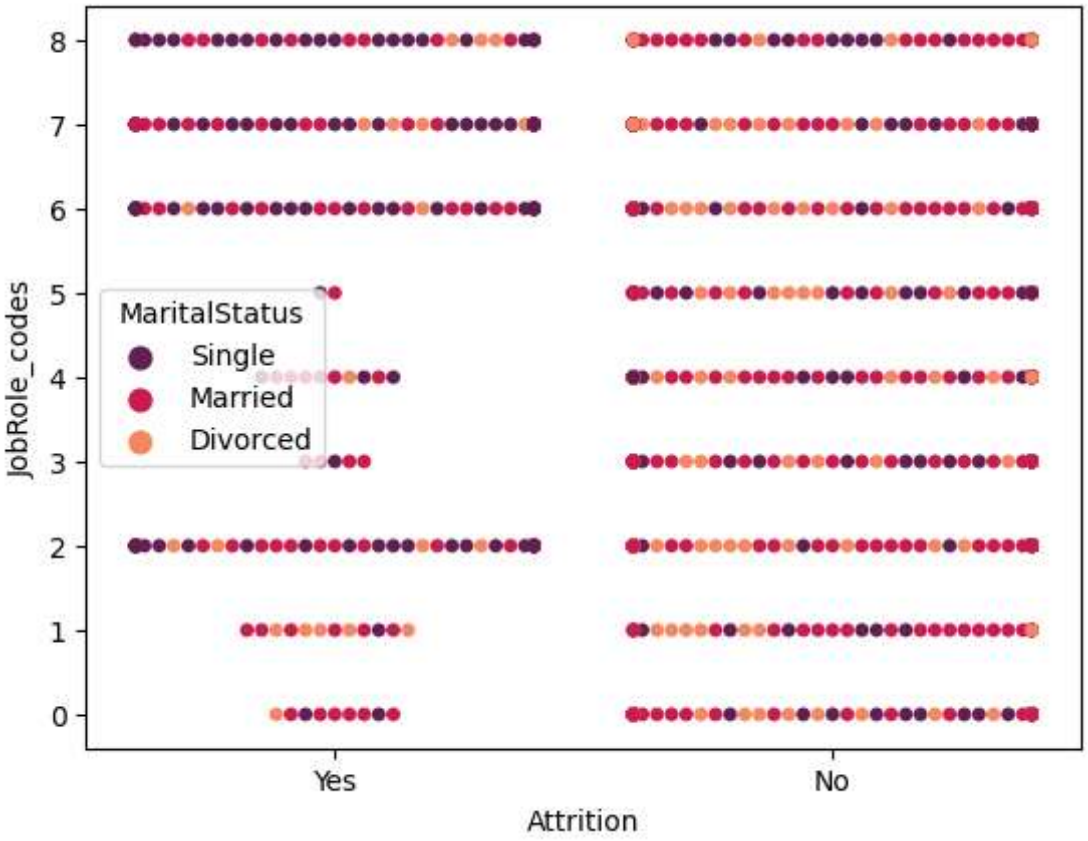

MaritalStatus
Single
Married
Divorced
Attrition_codes
1.0
0.8
0.6
0.4
0.2
0.0
Yes
No
Attrition

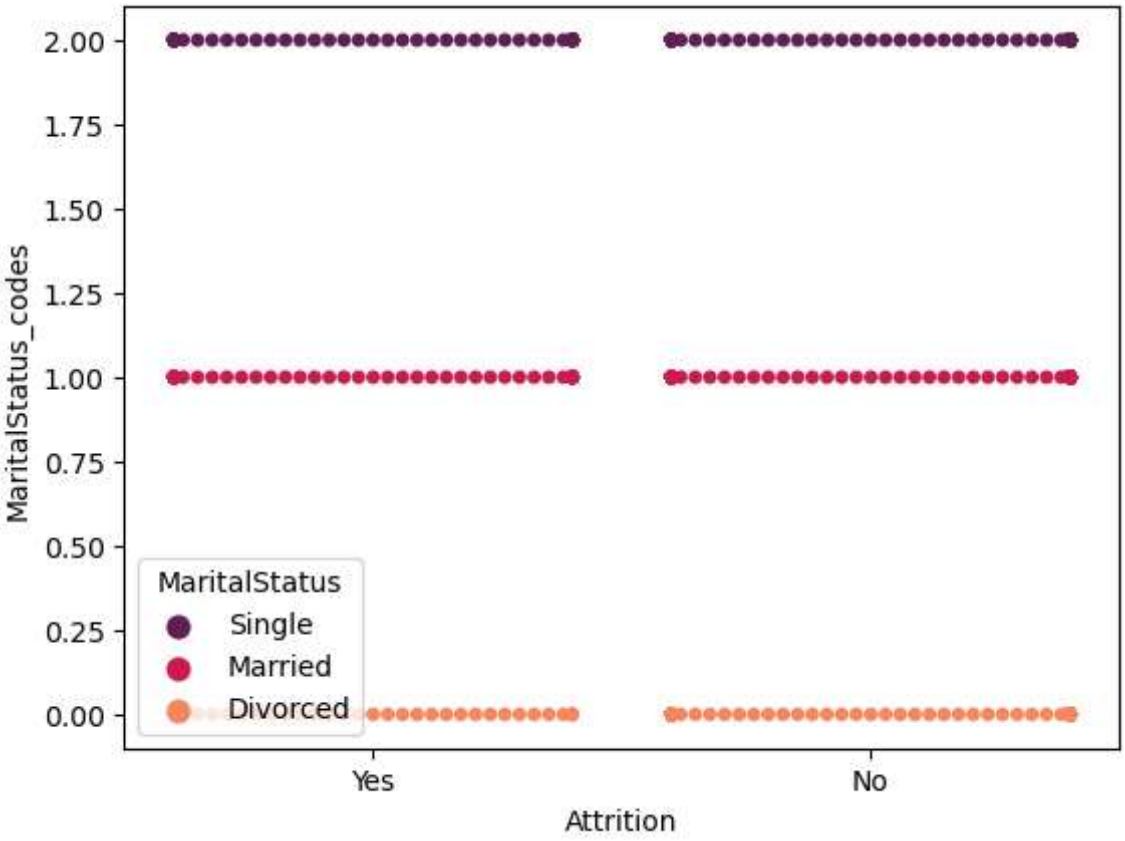

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
BusinessTravel_codes
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

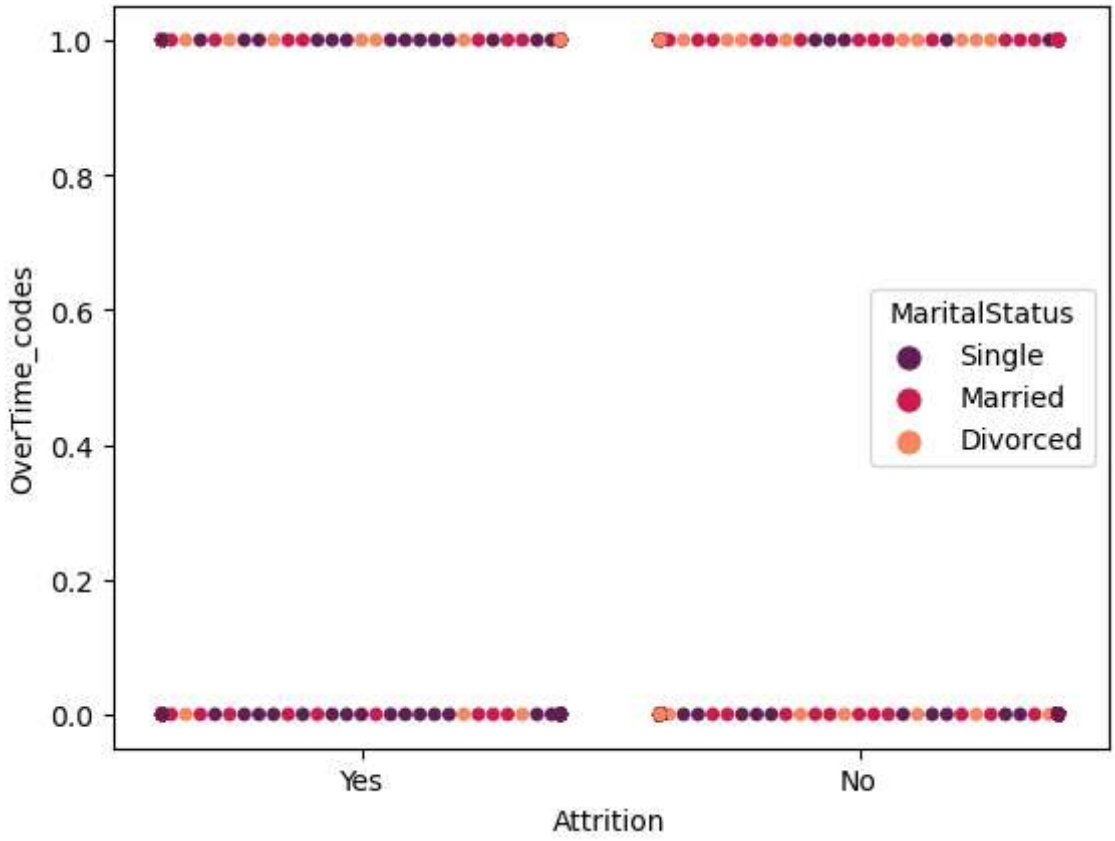

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Department_codes
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition


MaritalStatus
Single
Married
Divorced
EducationField_codes
0
1
2
3
4
5
Yes
No
Attrition

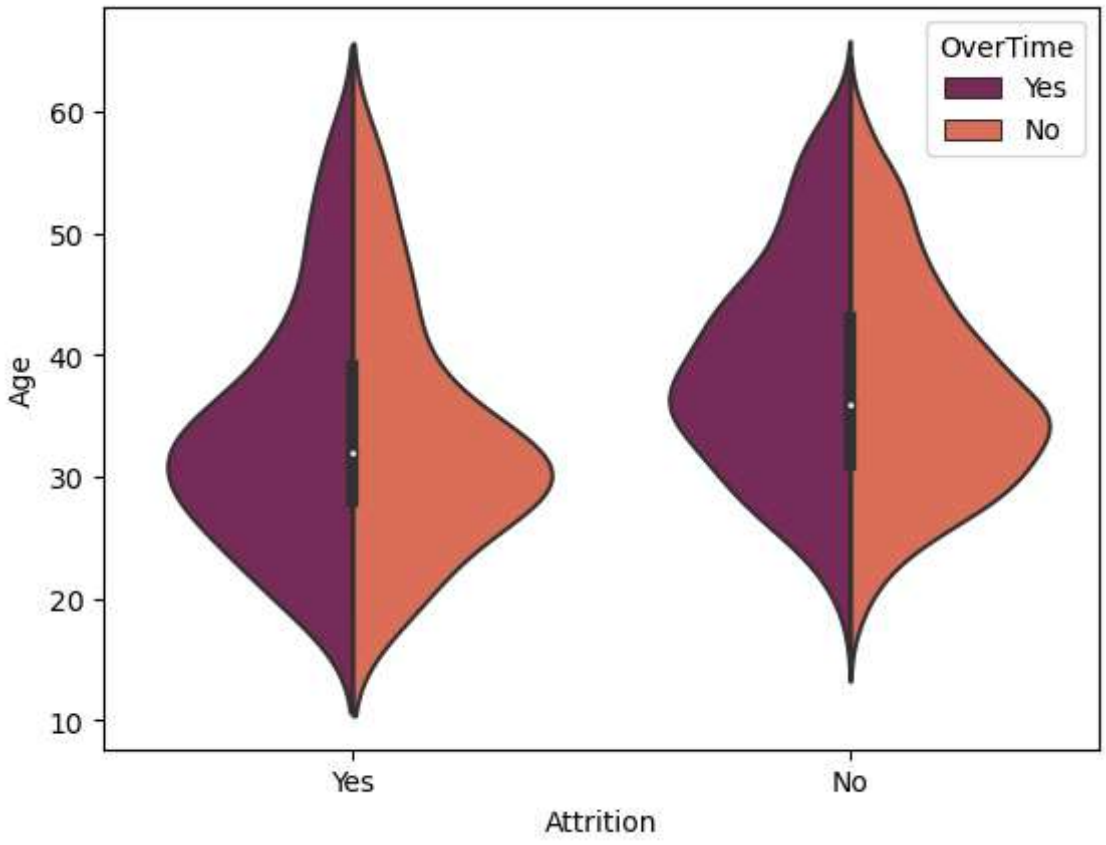

Gender_codes
0.0
0.2
0.4
0.6
0.8
1.0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

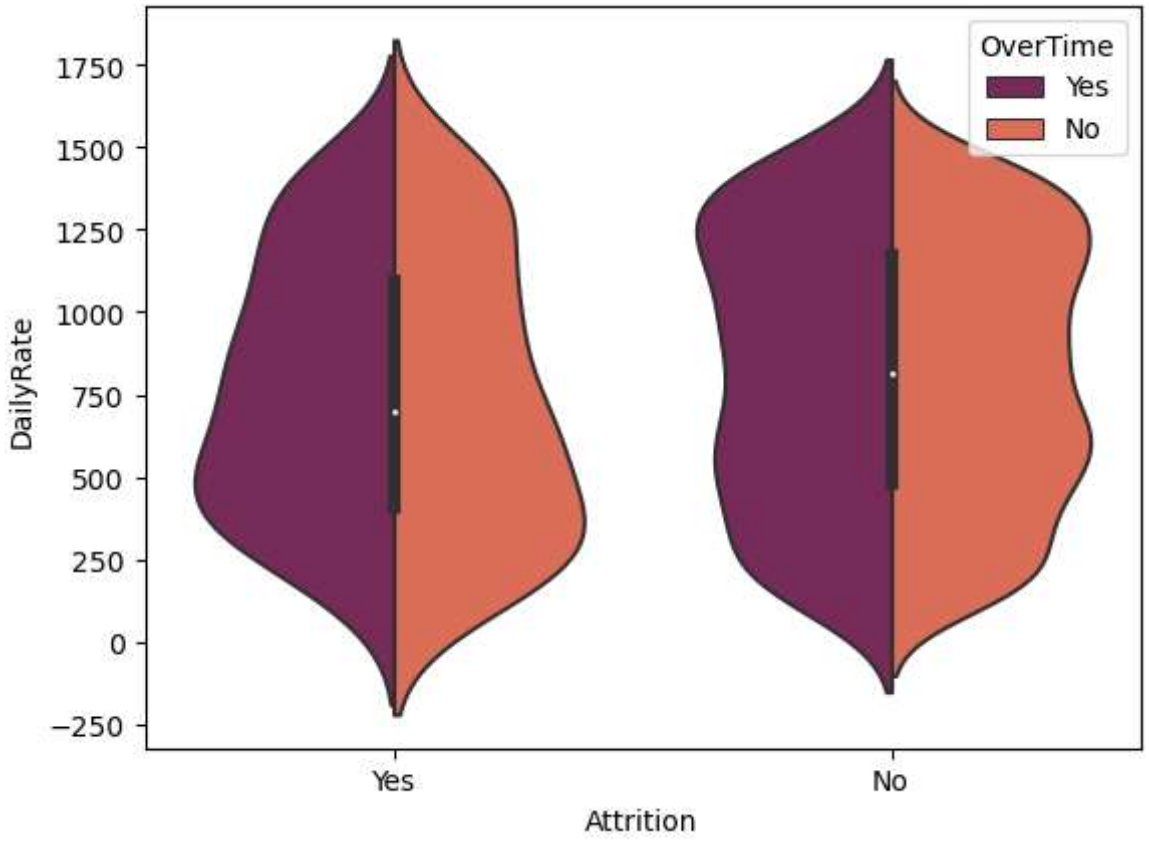

JobRole_codes
8
7
6
5
4
3
2
1
0
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

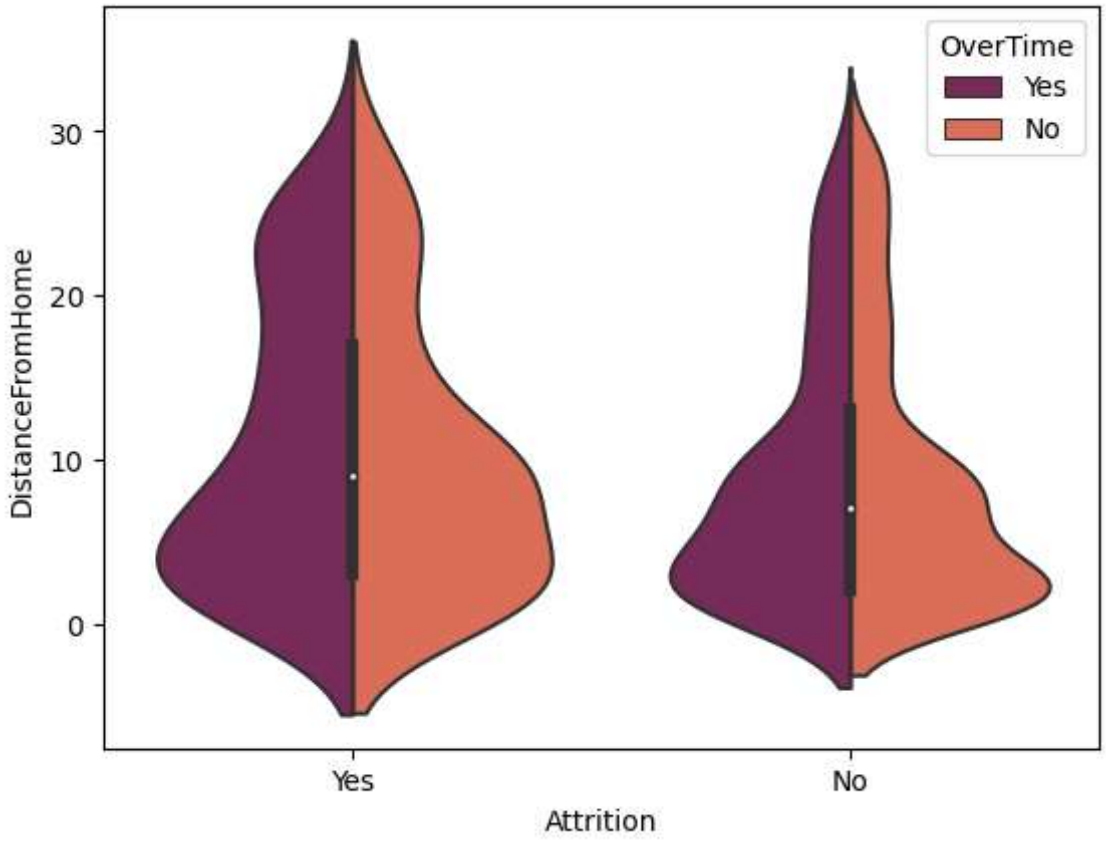

MaritalStatus_codes
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
MaritalStatus
Single
Married
Divorced
Yes
No
Attrition

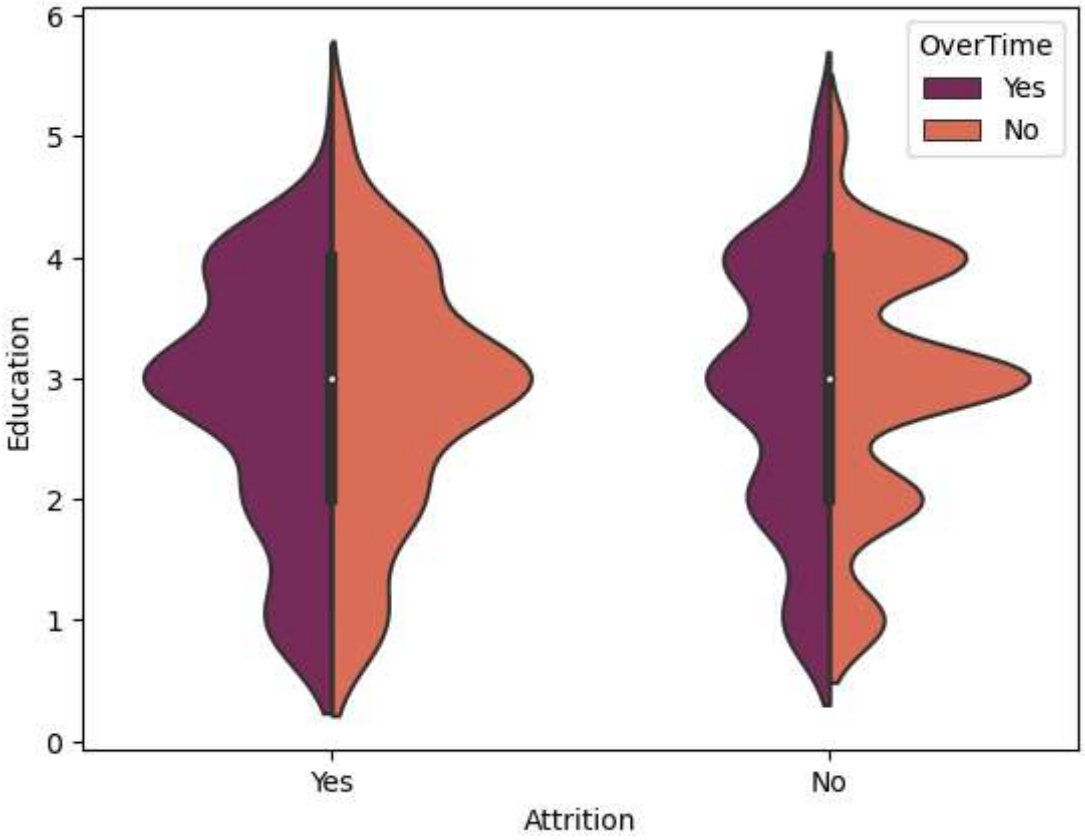


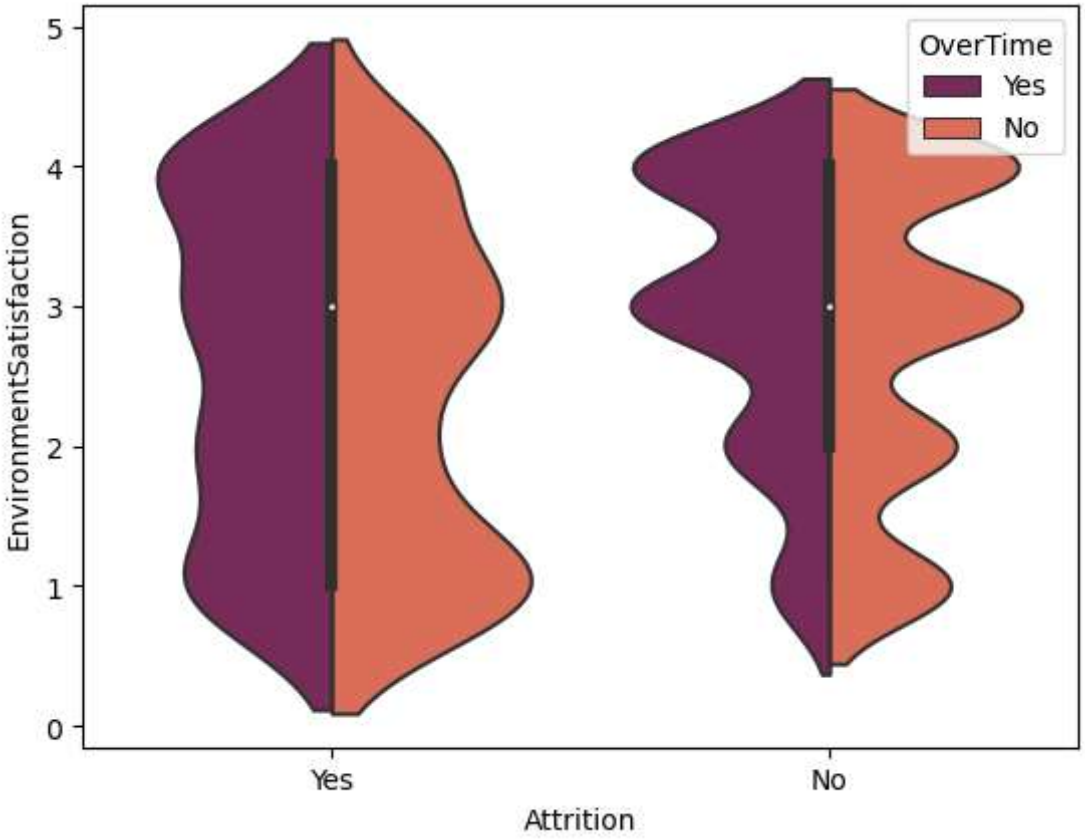
In [31]:
```
for AA in ibm.columns[8:]:
    sns.violinplot(x='Attrition',y=AA,data=ibm,hue='OverTime',split=True,palette='
    plt.show()
```

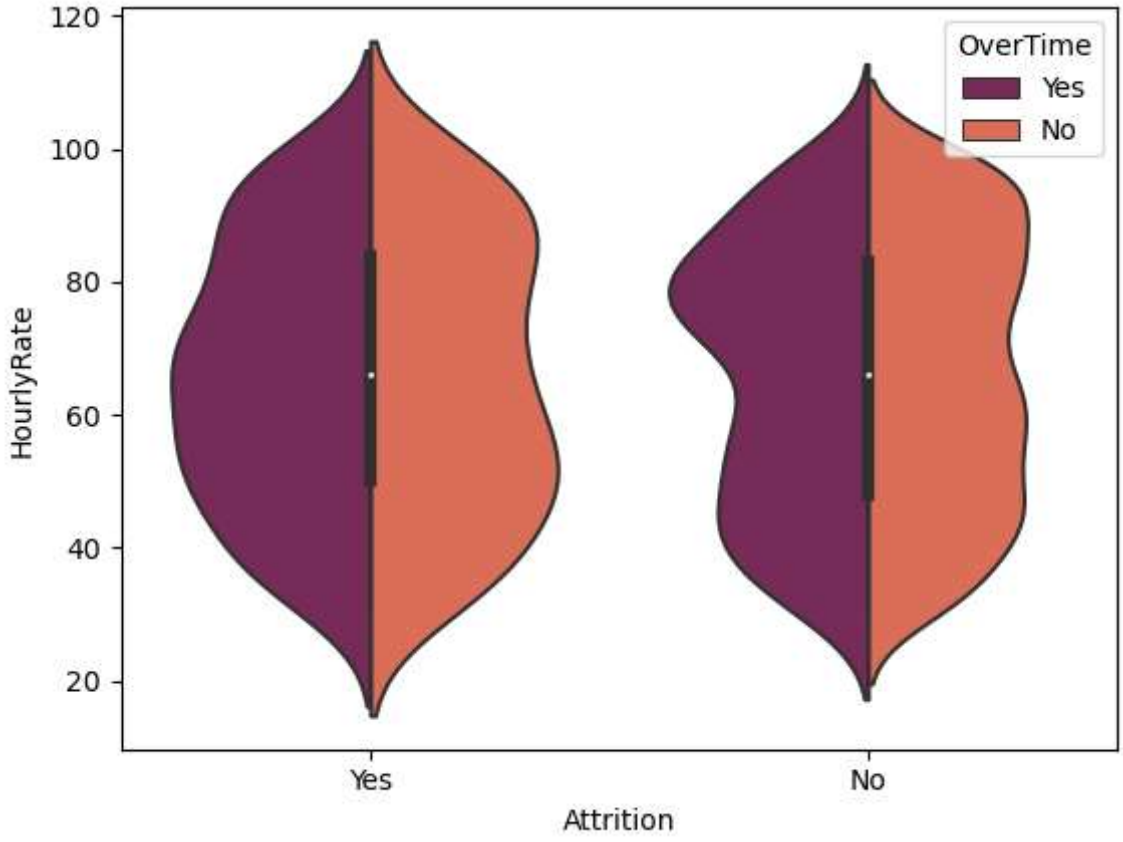


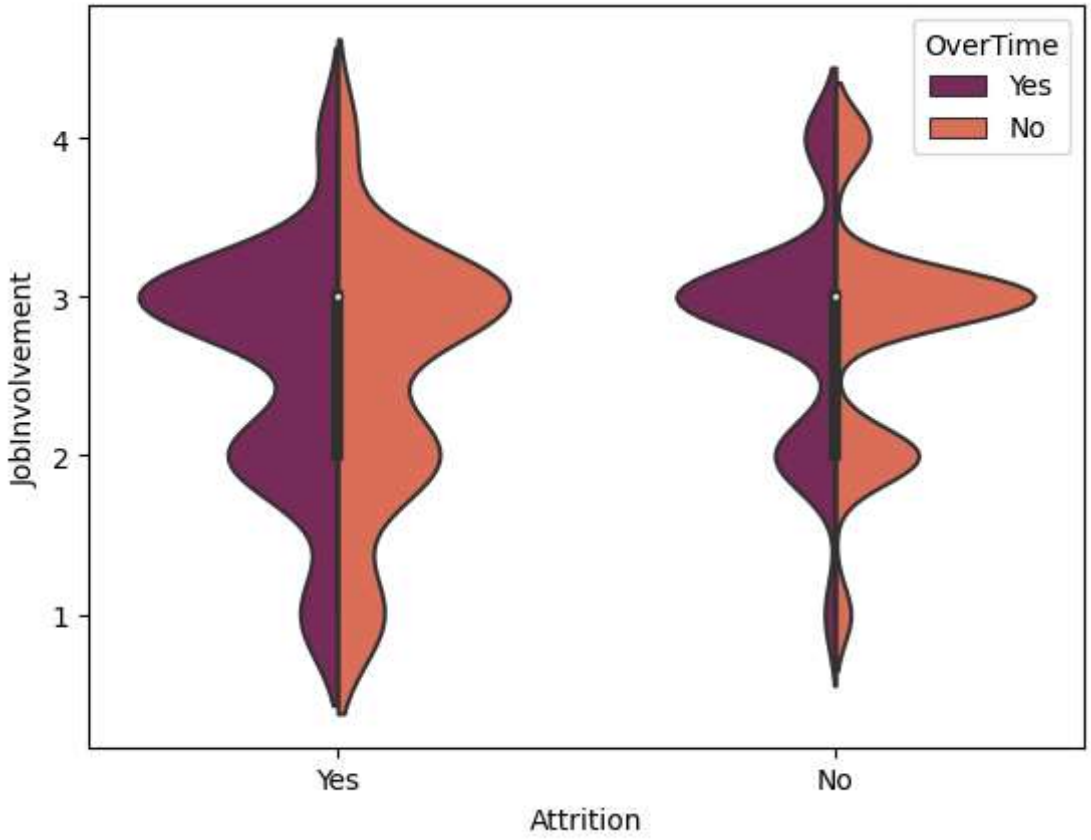

OverTime
Yes
No
1750
1500
1250
1000
750
500
250
0
−250
DailyRate
Yes
No
Attrition

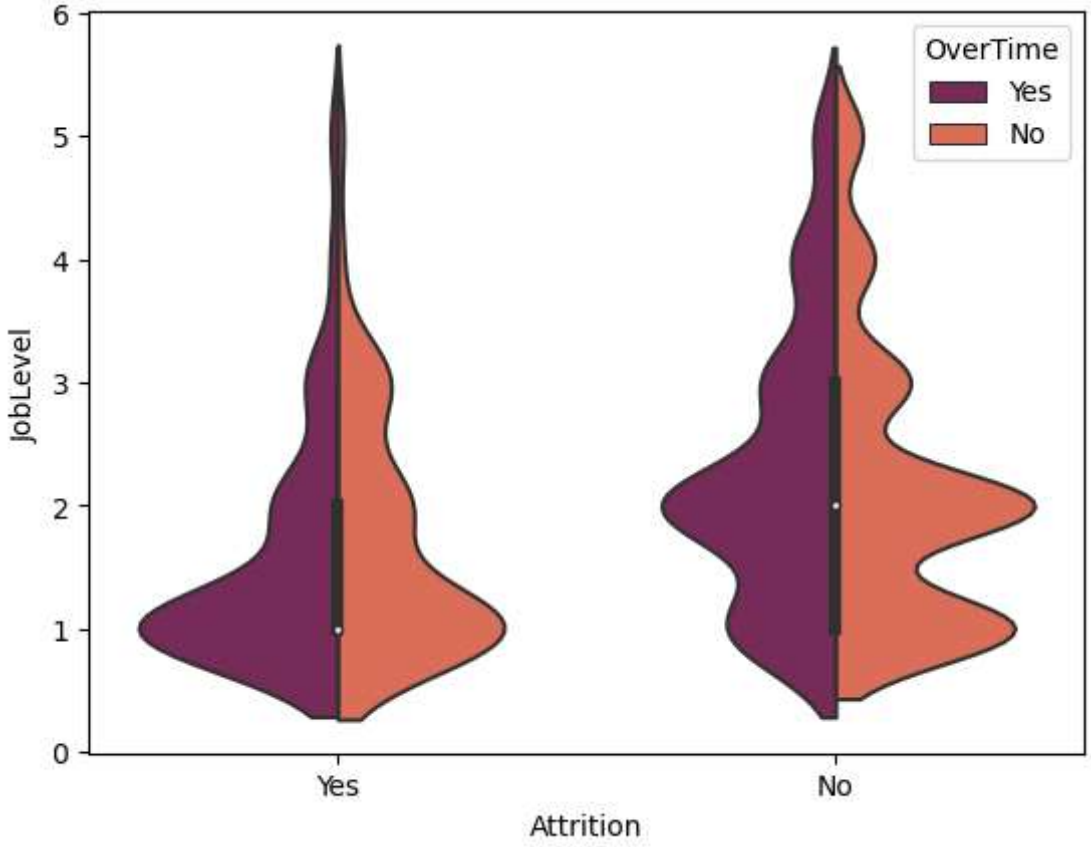

OverTime
Yes
No
30
20
10
0
DistanceFromHome
Yes
No
Attrition

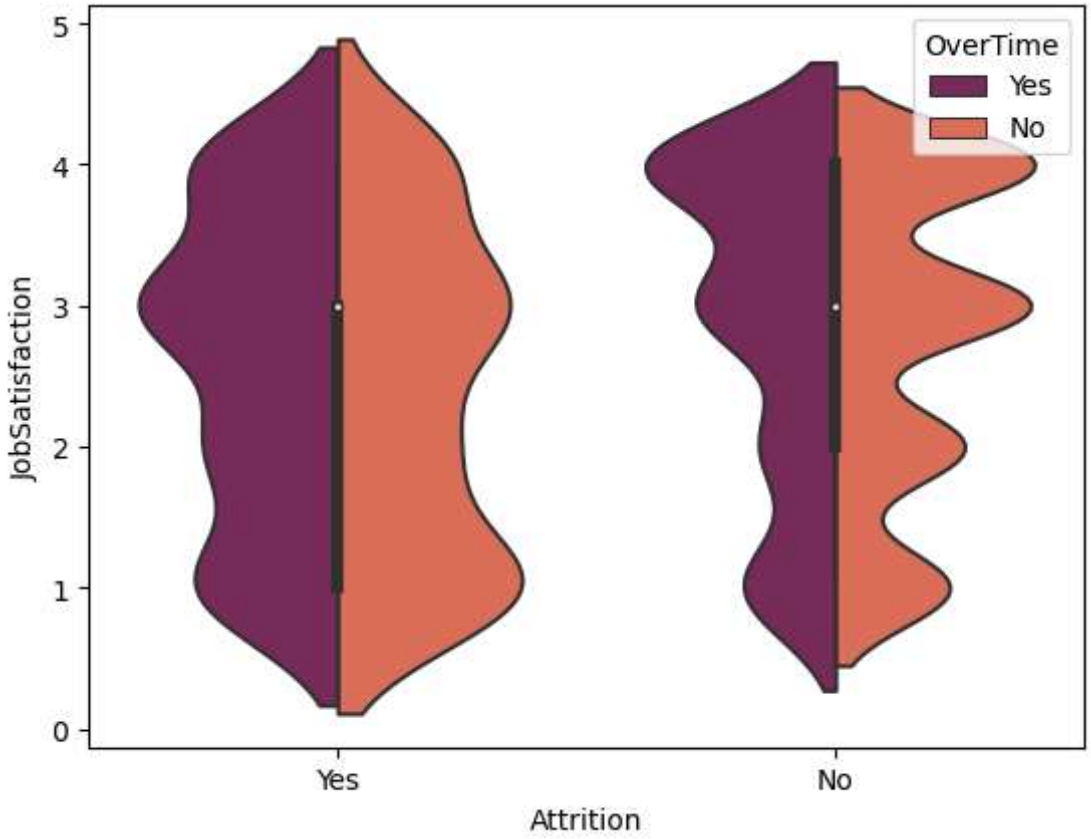

OverTime
Yes
No
Education
0
1
2
3
4
5
6
Yes
No
Attrition

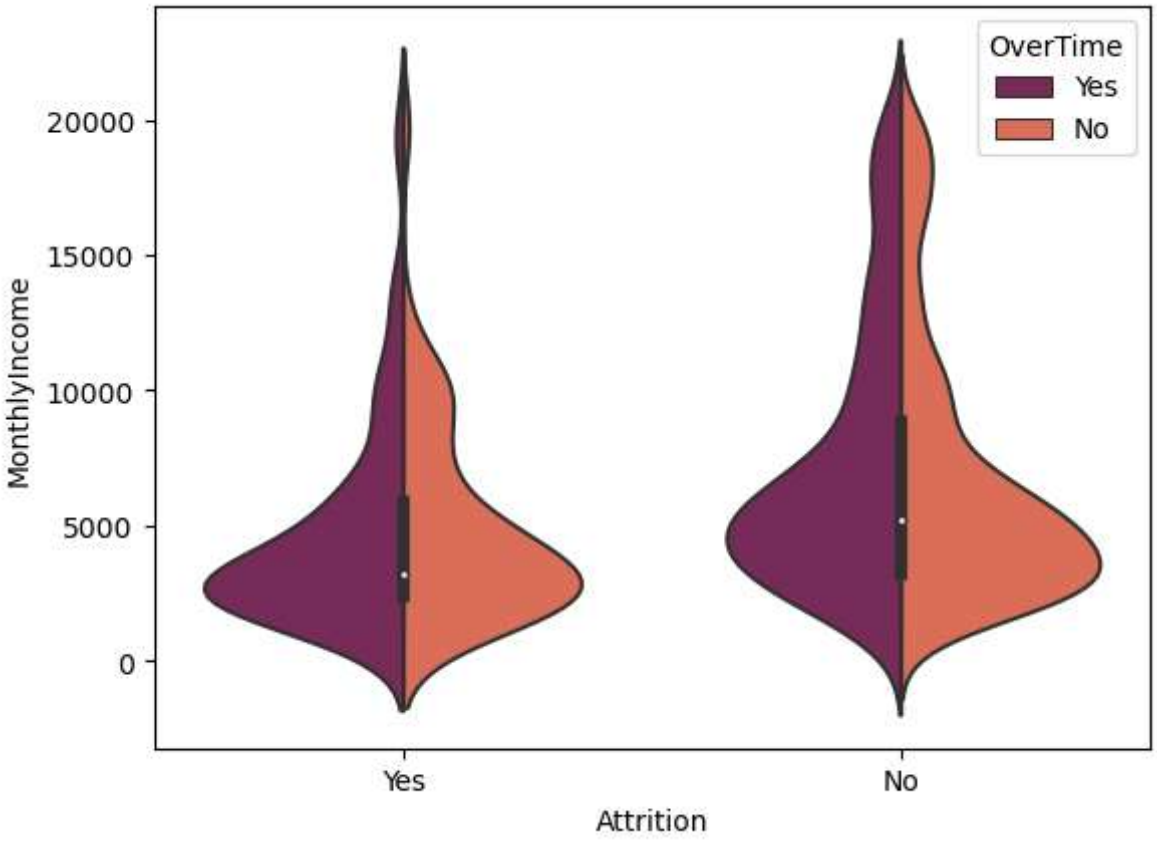

OverTime
Yes
No
EnvironmentSatisfaction
0
1
2
3
4
5
Yes
No
Attrition

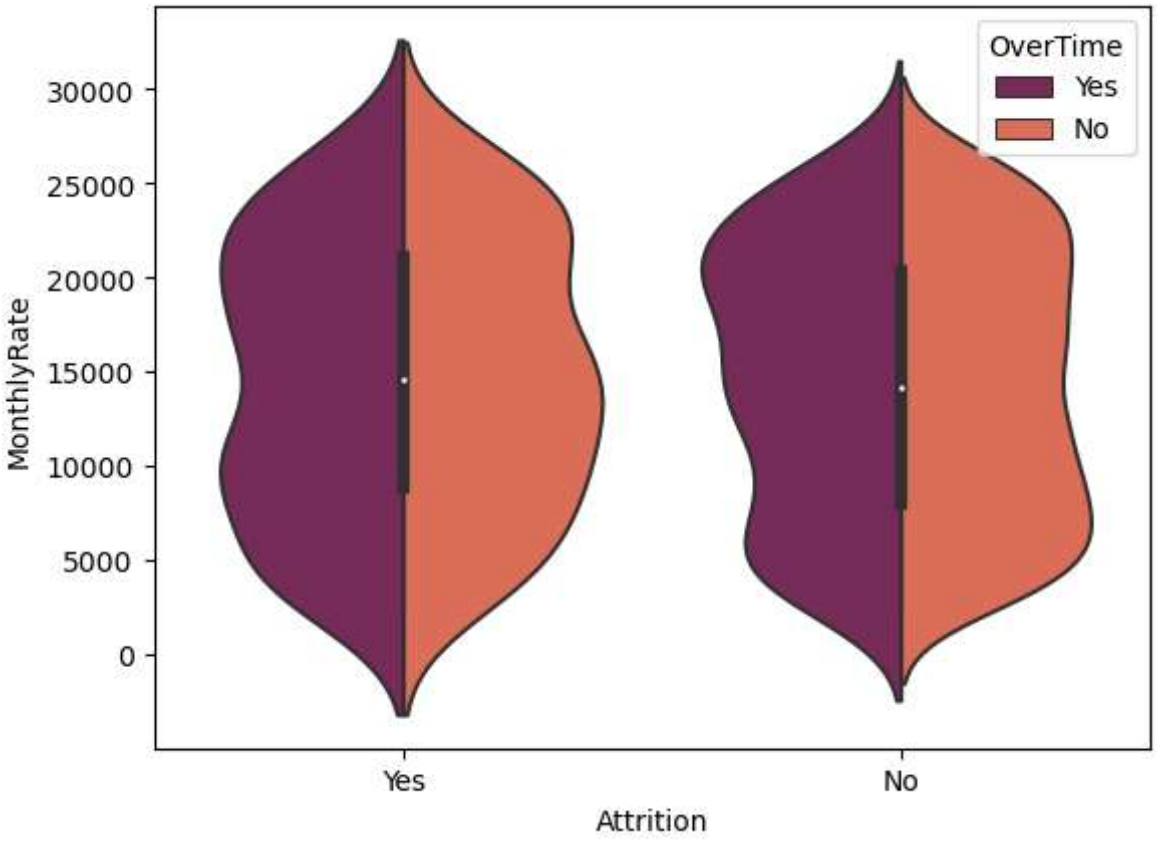

OverTime
Yes
No
HourlyRate
120
100
80
60
40
20
Yes
No
Attrition

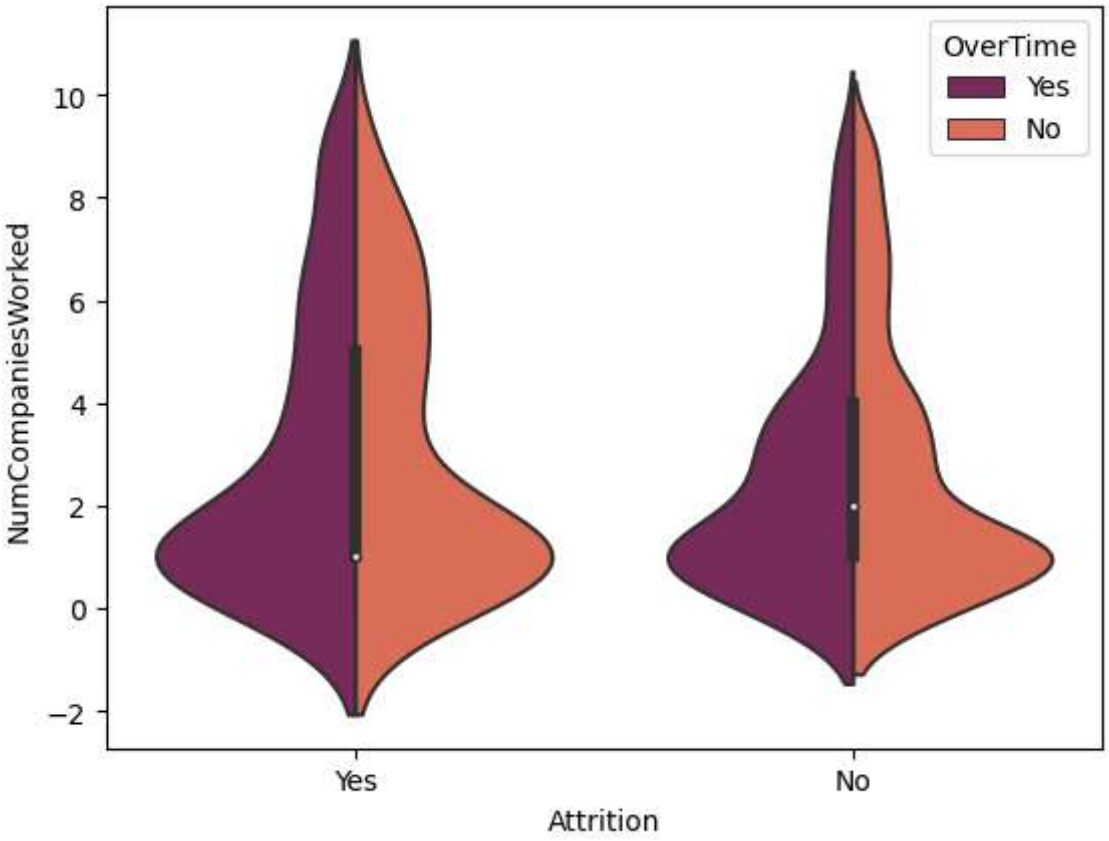

OverTime
Yes
No
JobInvolvement
4
3
2
1
Yes
No
Attrition

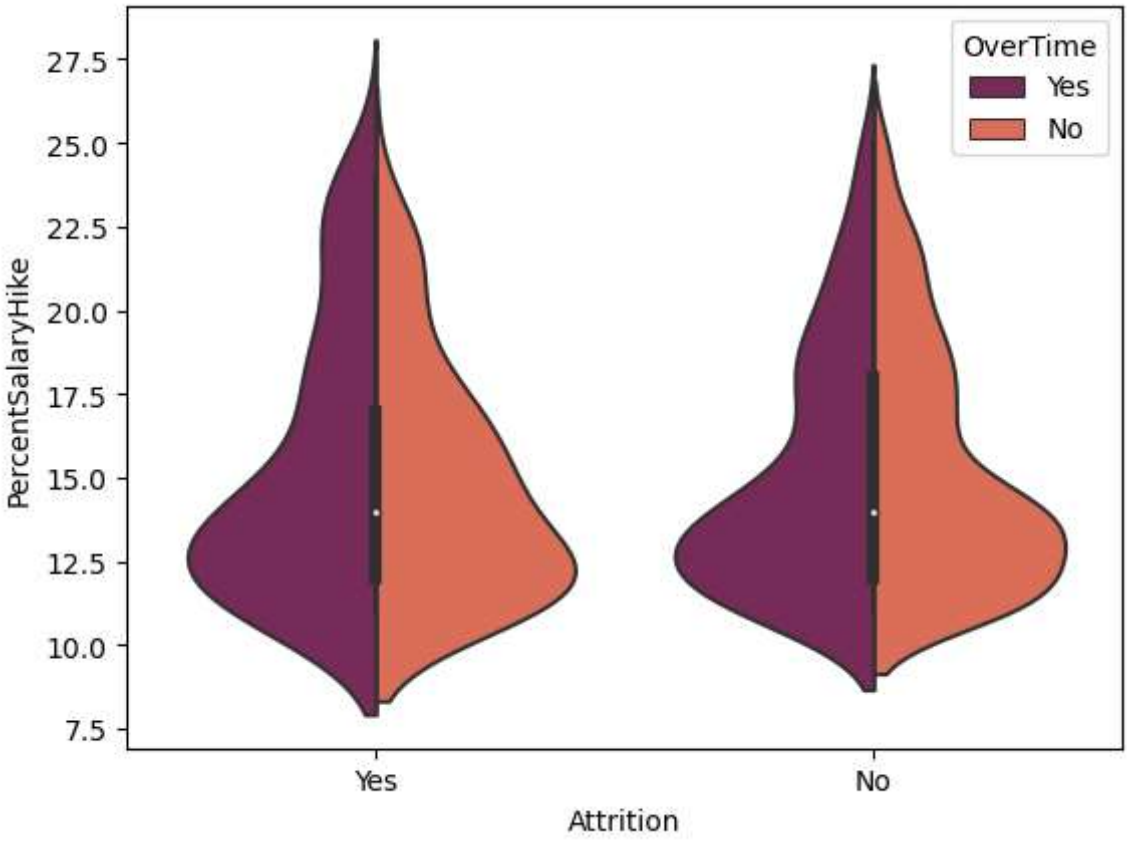

OverTime
Yes
No
JobLevel
0
1
2
3
4
5
6
Yes
No
Attrition

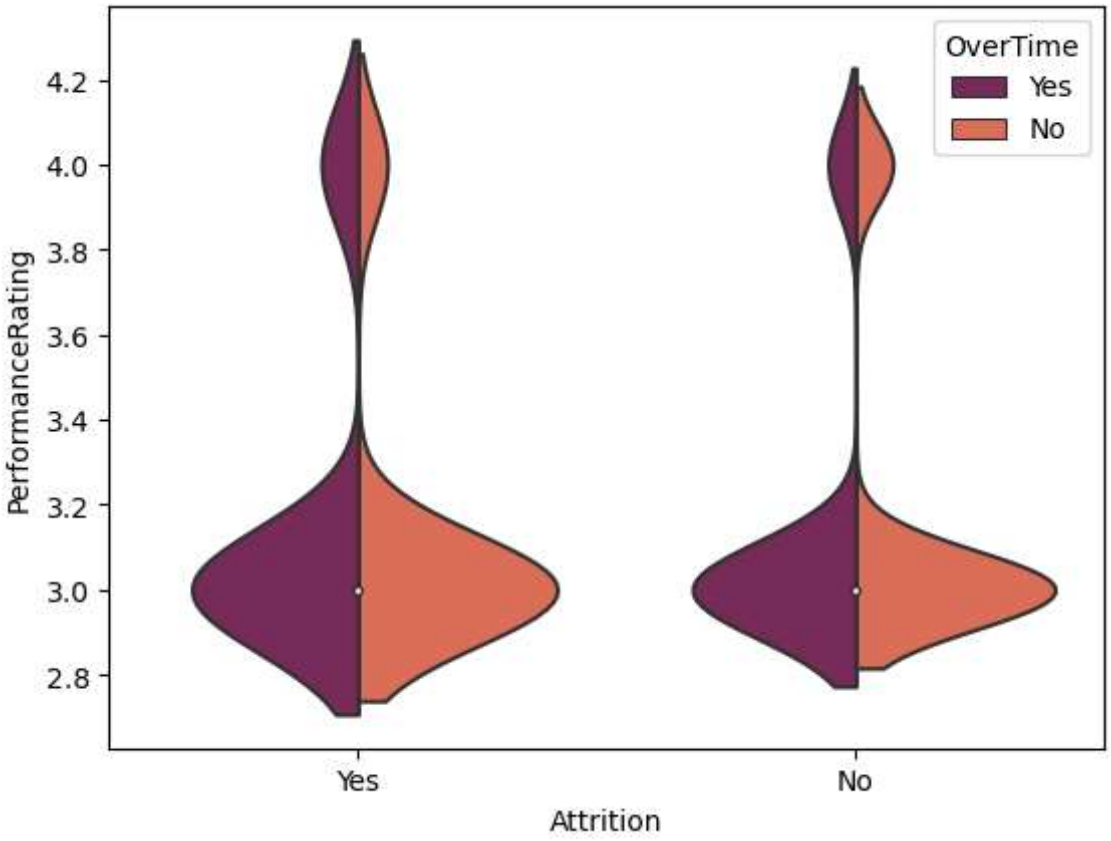

OverTime
Yes
No
JobSatisfaction
0
1
2
3
4
5
Yes
No
Attrition

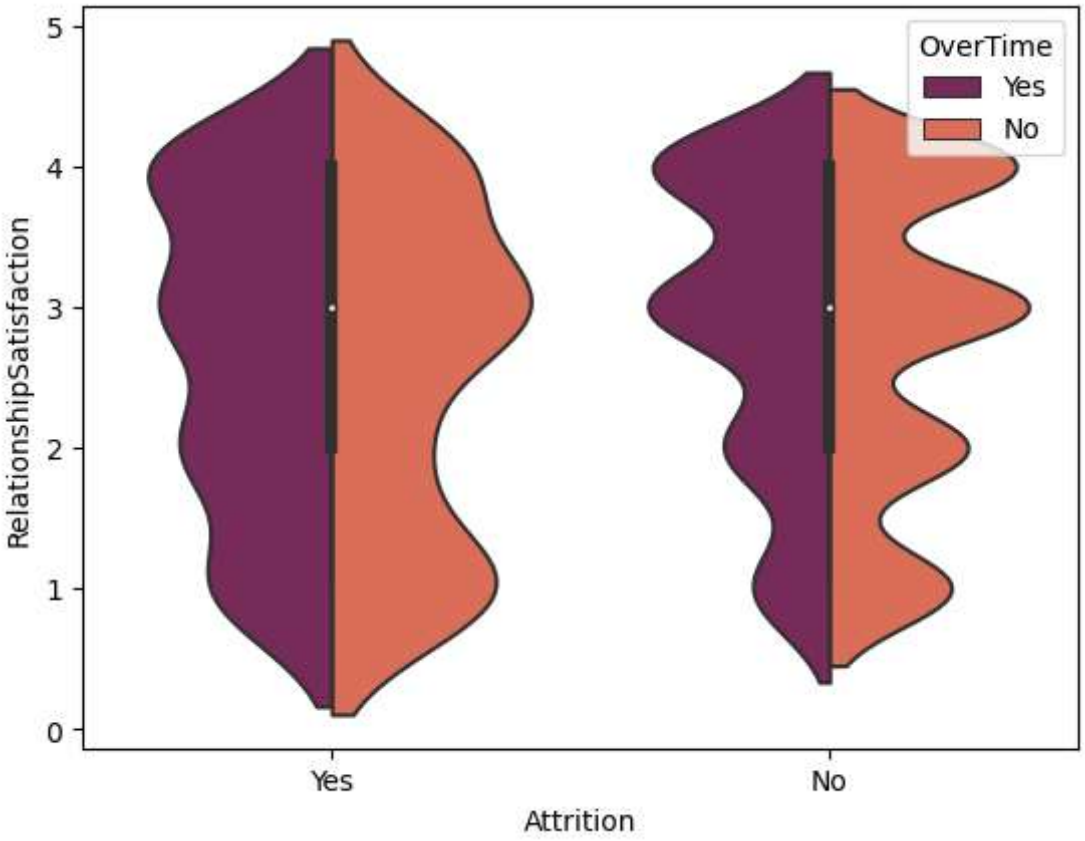

OverTime
Yes
No
20000
15000
10000
5000
0
MonthlyIncome
Yes
No
Attrition

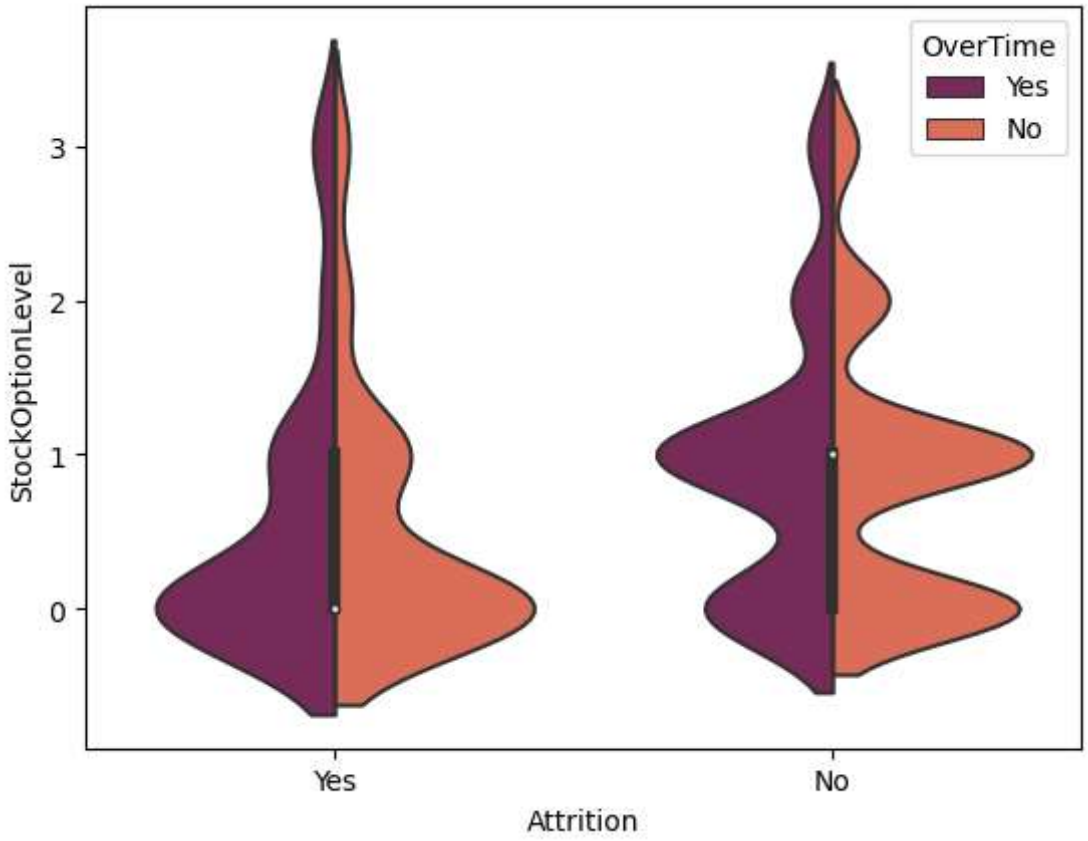

OverTime
Yes
No
30000
25000
20000
15000
10000
5000
0
MonthlyRate
Yes
No
Attrition

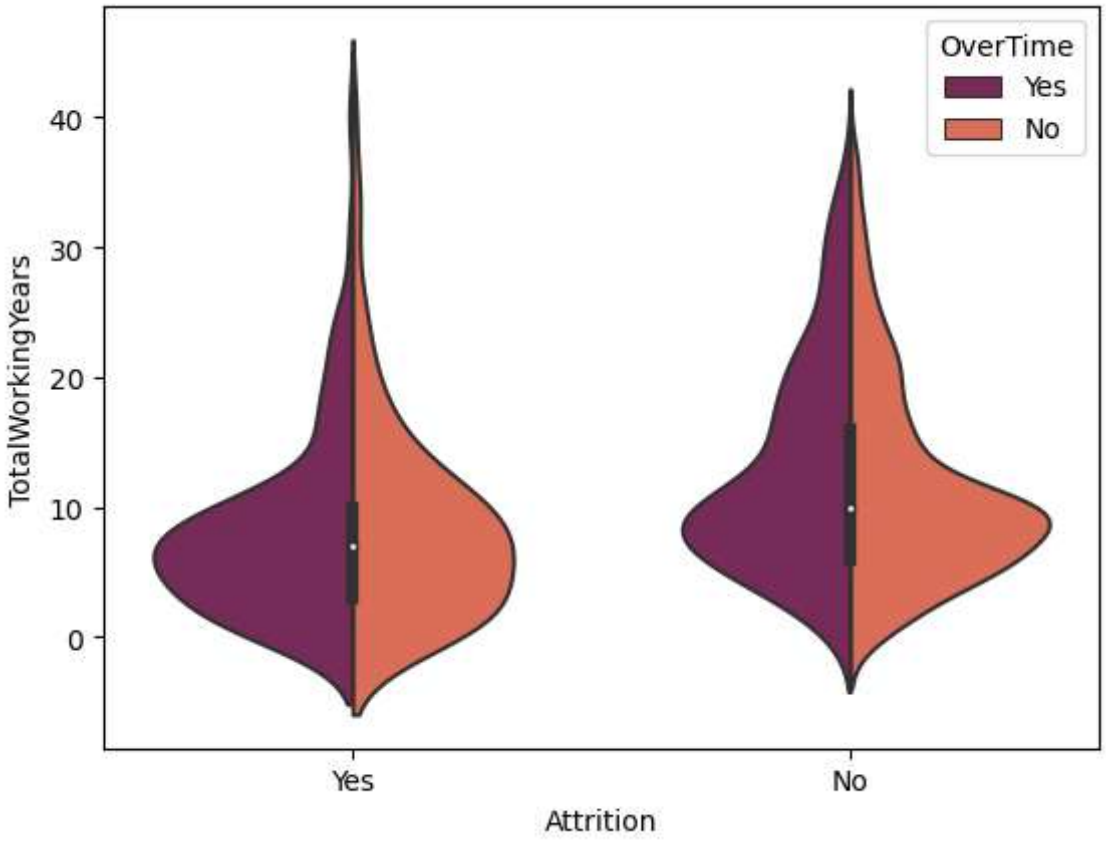

OverTime
Yes
No
NumCompaniesWorked
10
8
6
4
2
0
−2
Yes
No
Attrition

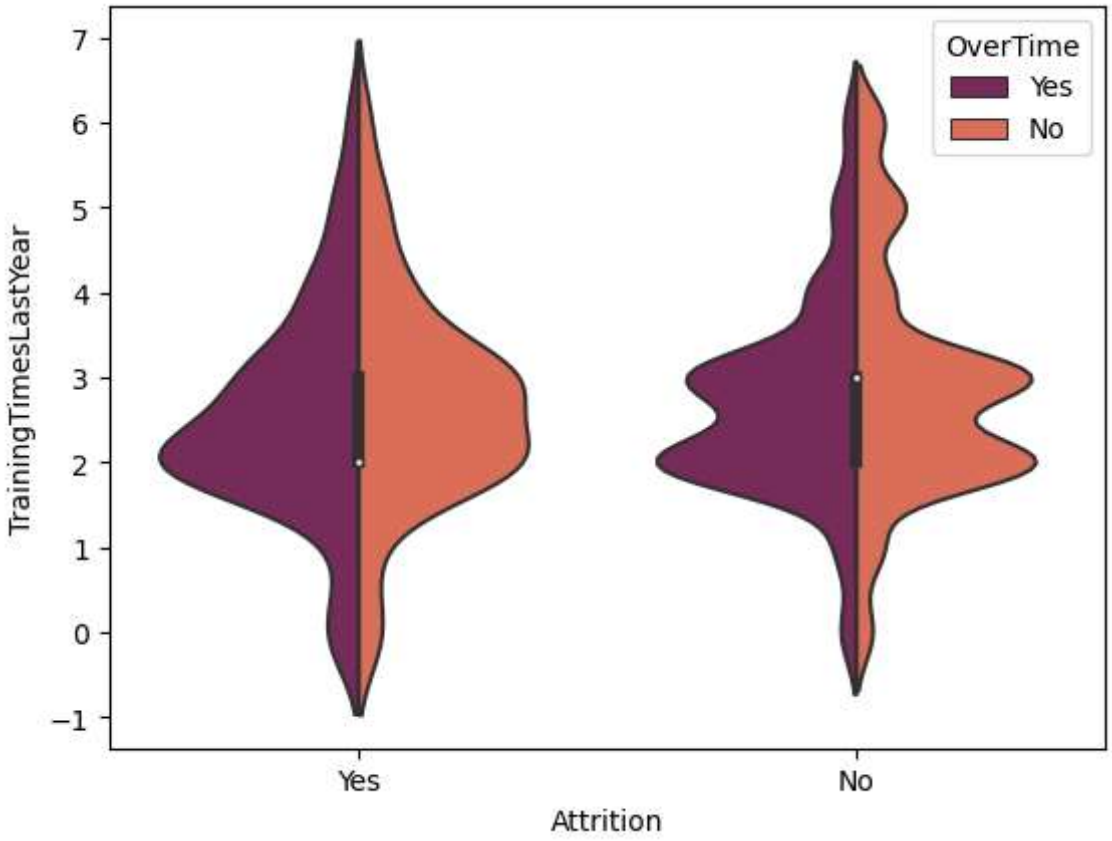

OverTime
Yes
No
PercentSalaryHike
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
Yes
No
Attrition

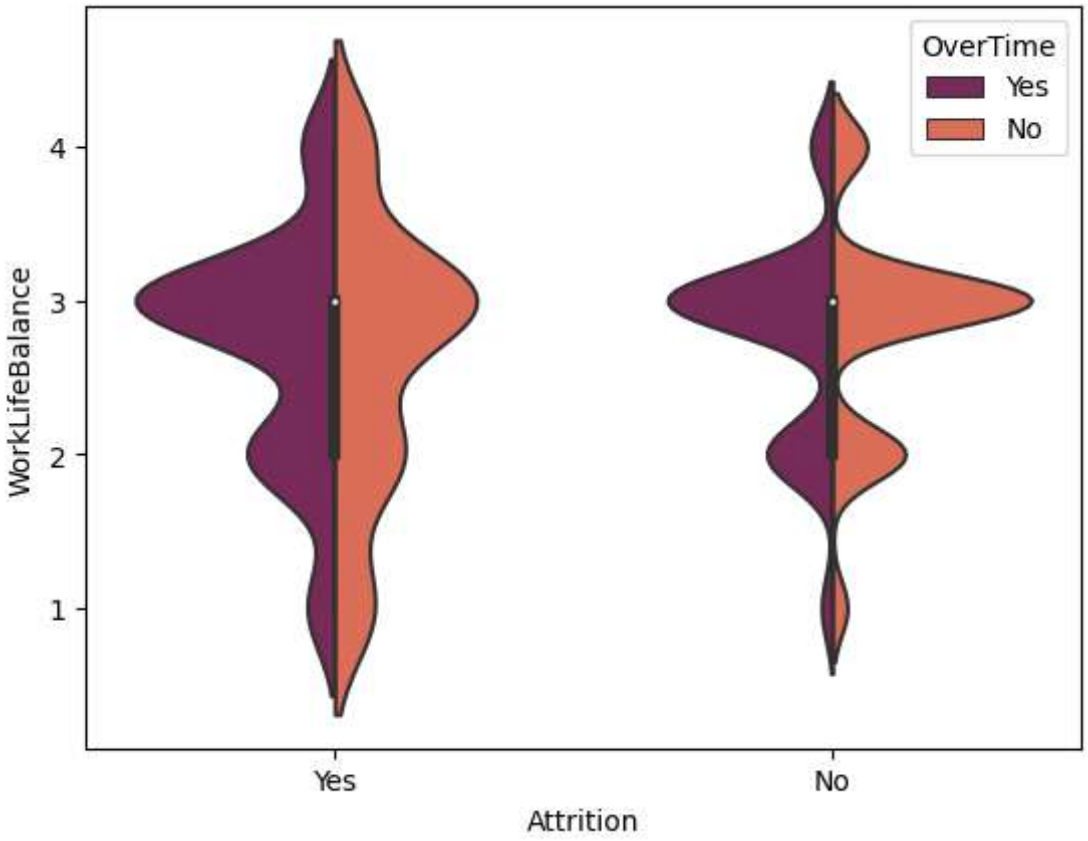

OverTime
Yes
No
PerformanceRating
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
Yes
No
Attrition

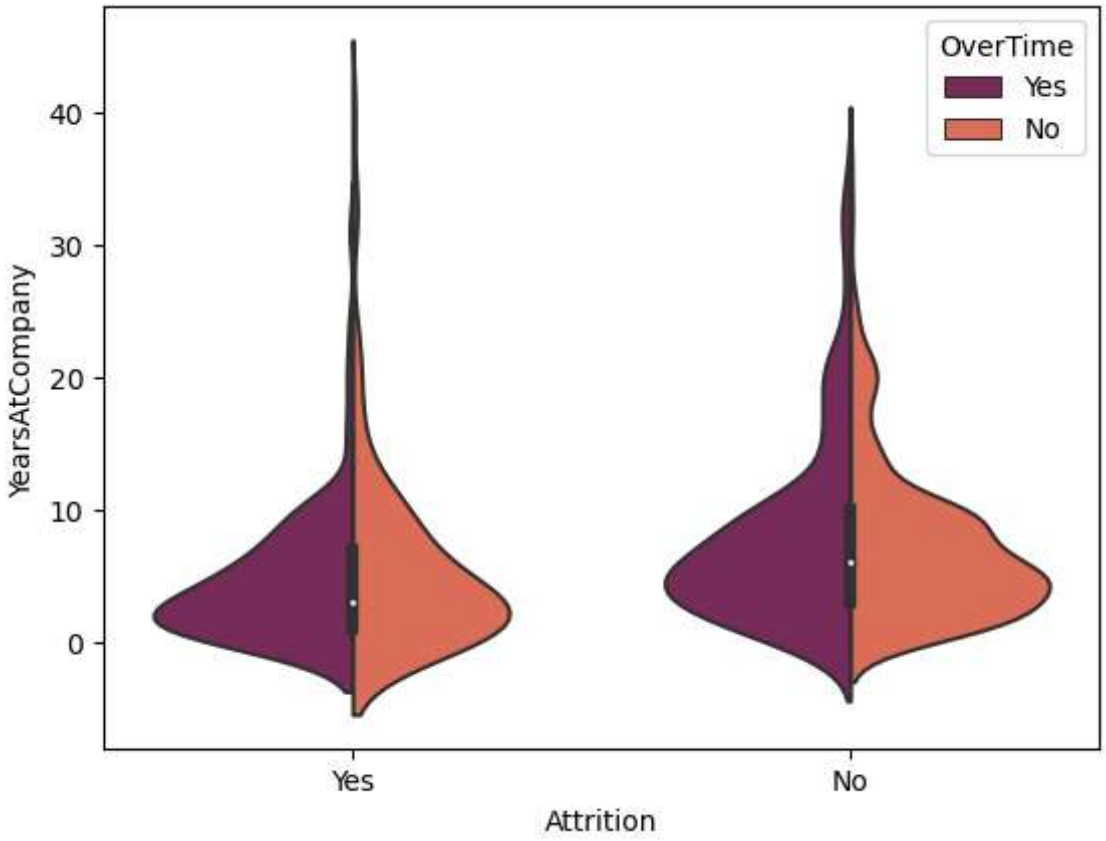

OverTime
Yes
No
RelationshipSatisfaction
5
4
3
2
1
0
Yes
No
Attrition

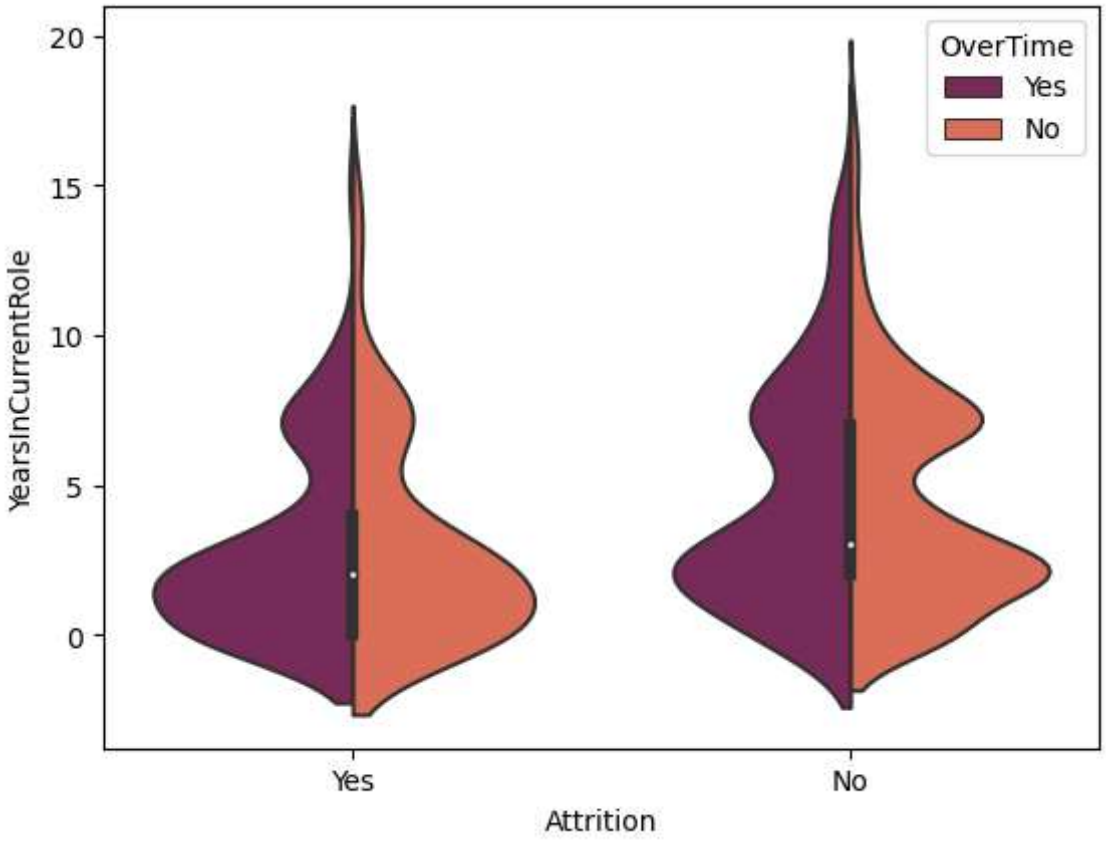

OverTime
Yes
No
StockOptionLevel
0
1
2
3
Yes
No
Attrition

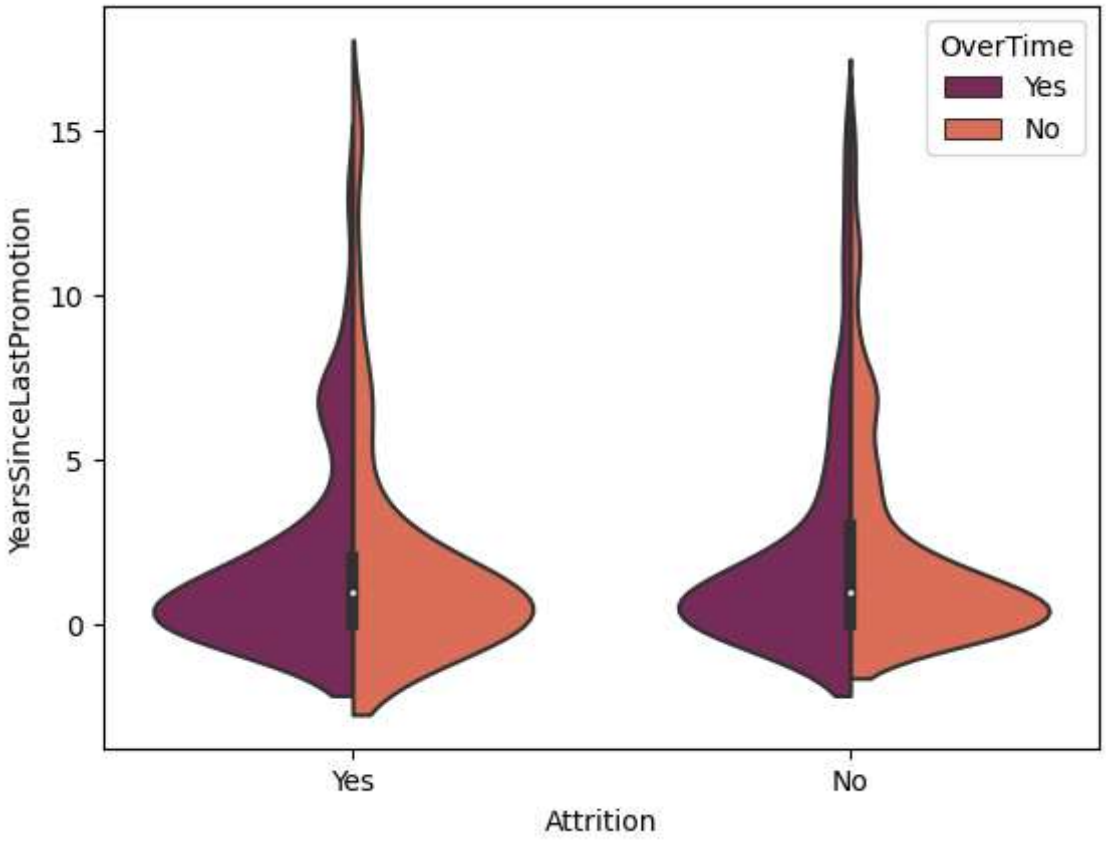

OverTime
Yes
No
TotalWorkingYears
0
10
20
30
40
Yes
No
Attrition

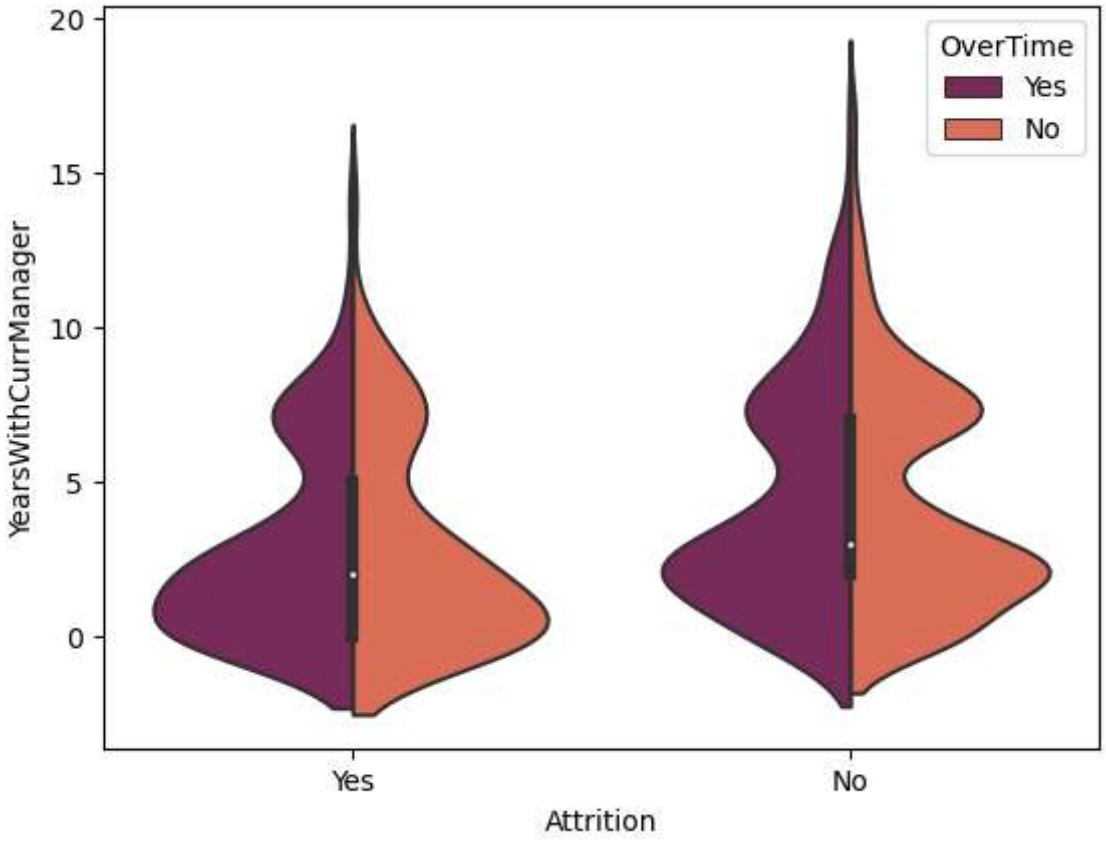

OverTime
Yes
No
TrainingTimesLastYear
7
6
5
4
3
2
1
0
−1
Yes
No
Attrition

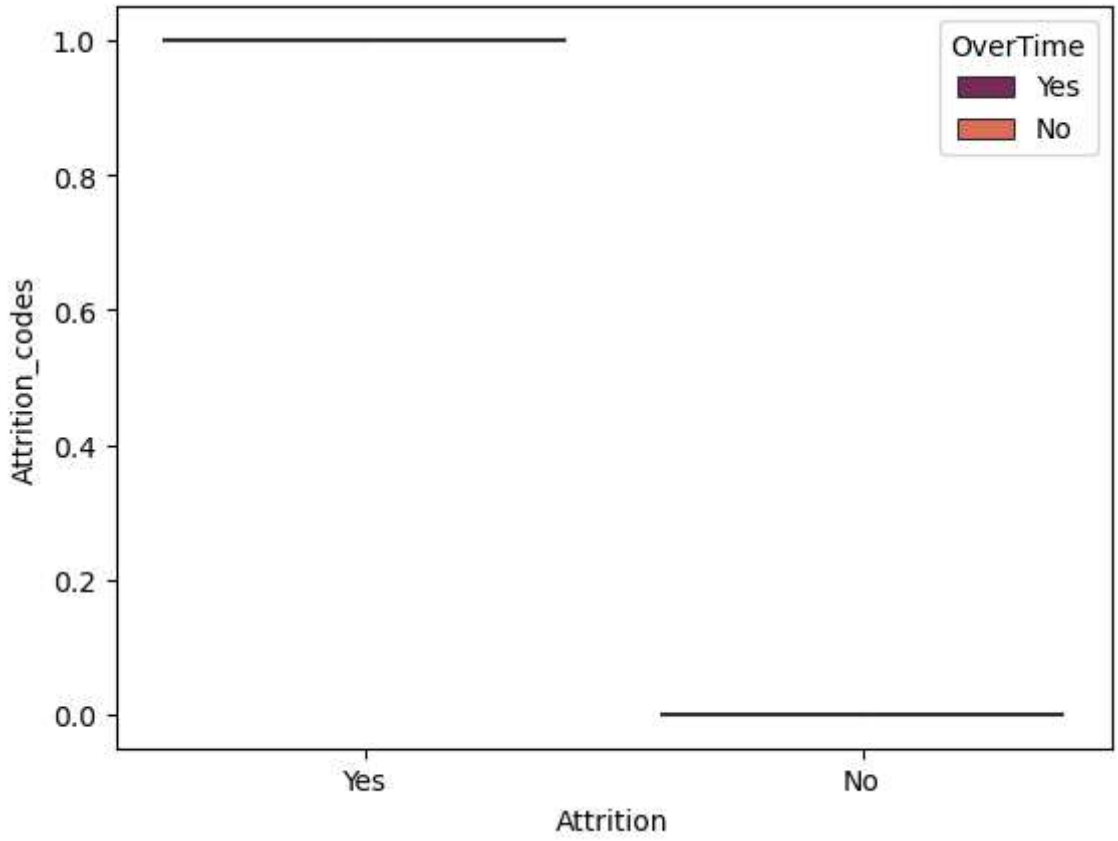

OverTime
Yes
No
WorkLifeBalance
4
3
2
1
Yes
No
Attrition

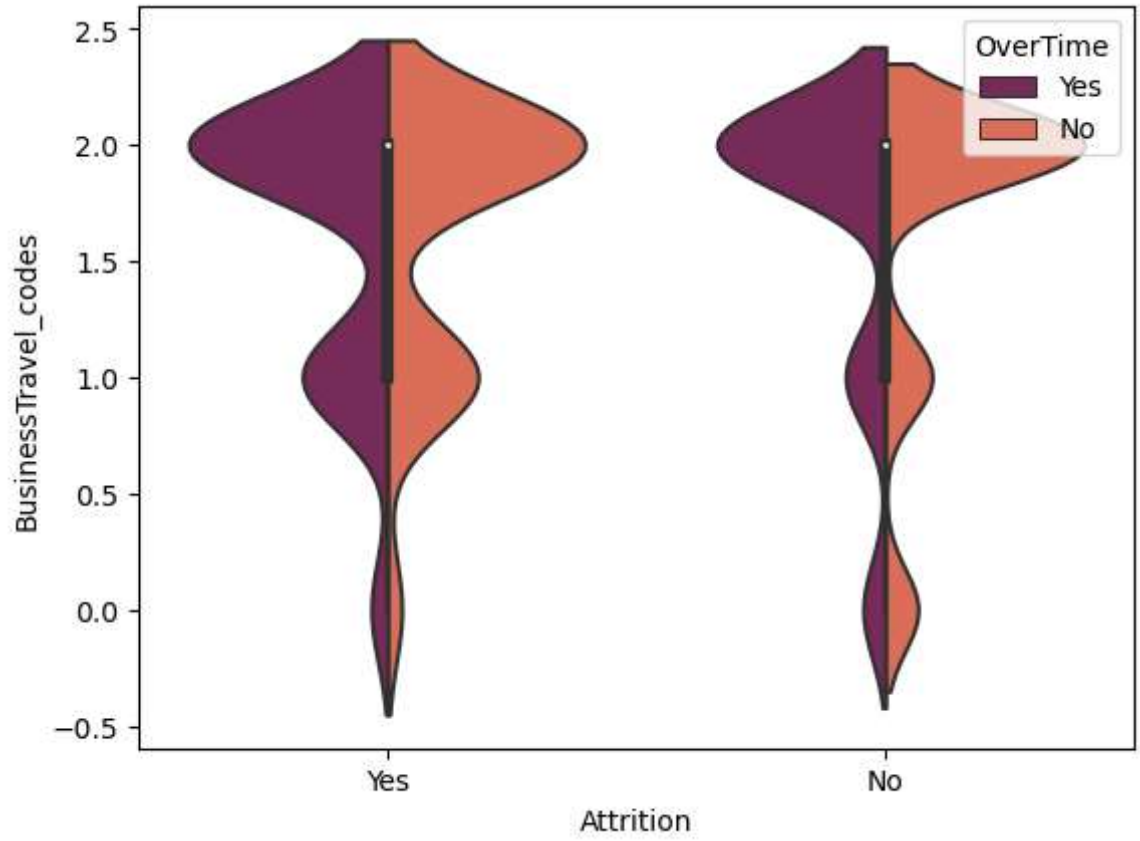

OverTime
Yes
No
YearsAtCompany
40
30
20
10
0
Yes
No
Attrition

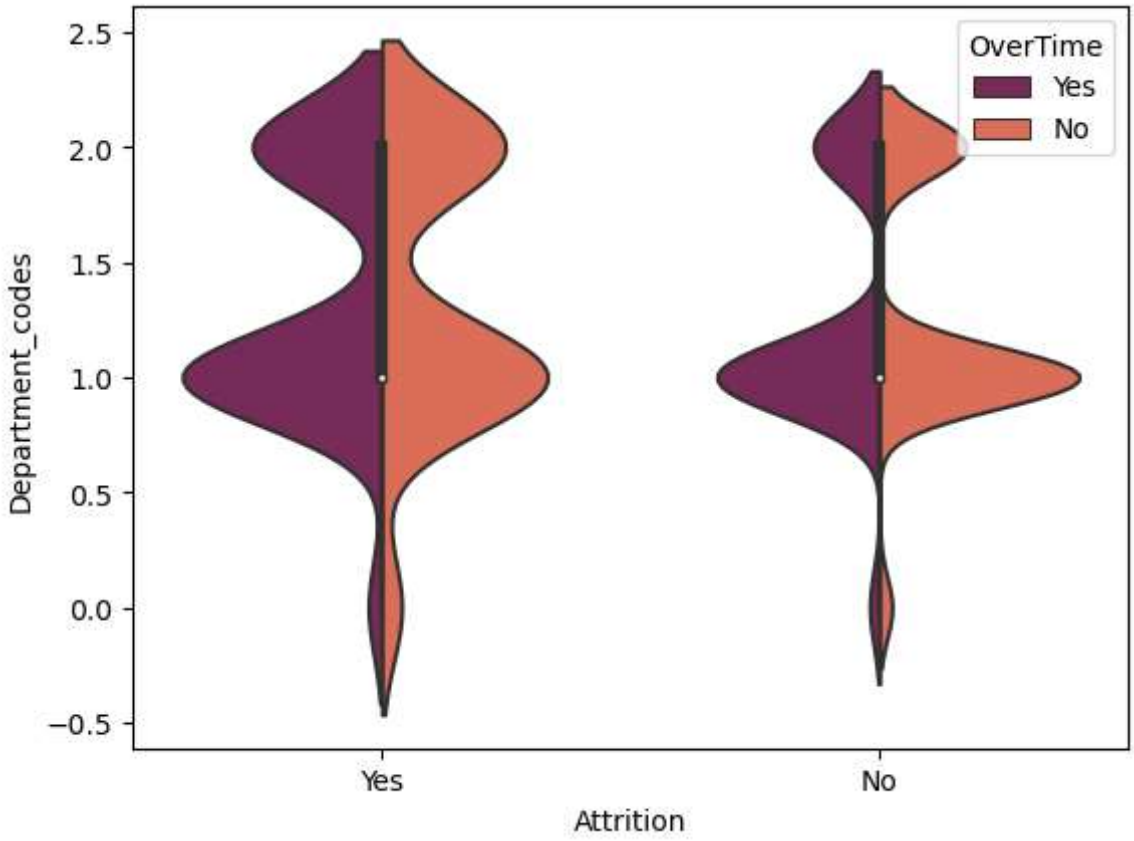

OverTime
Yes
No
YearsInCurrentRole
20
15
10
5
0
Yes
No
Attrition

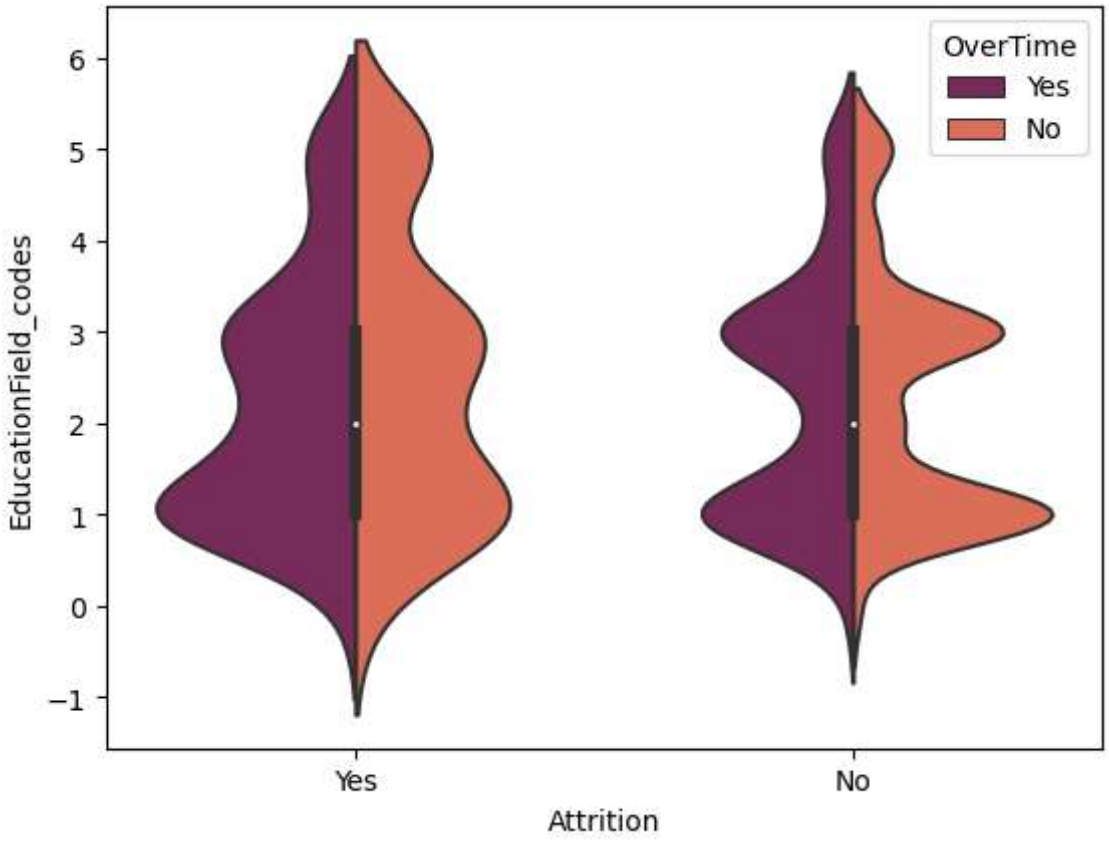

OverTime
Yes
No
YearsSinceLastPromotion
15
10
5
0
Yes
No
Attrition


OverTime
Yes
No
YearsWithCurrManager
20
15
10
5
0
Yes
No
Attrition

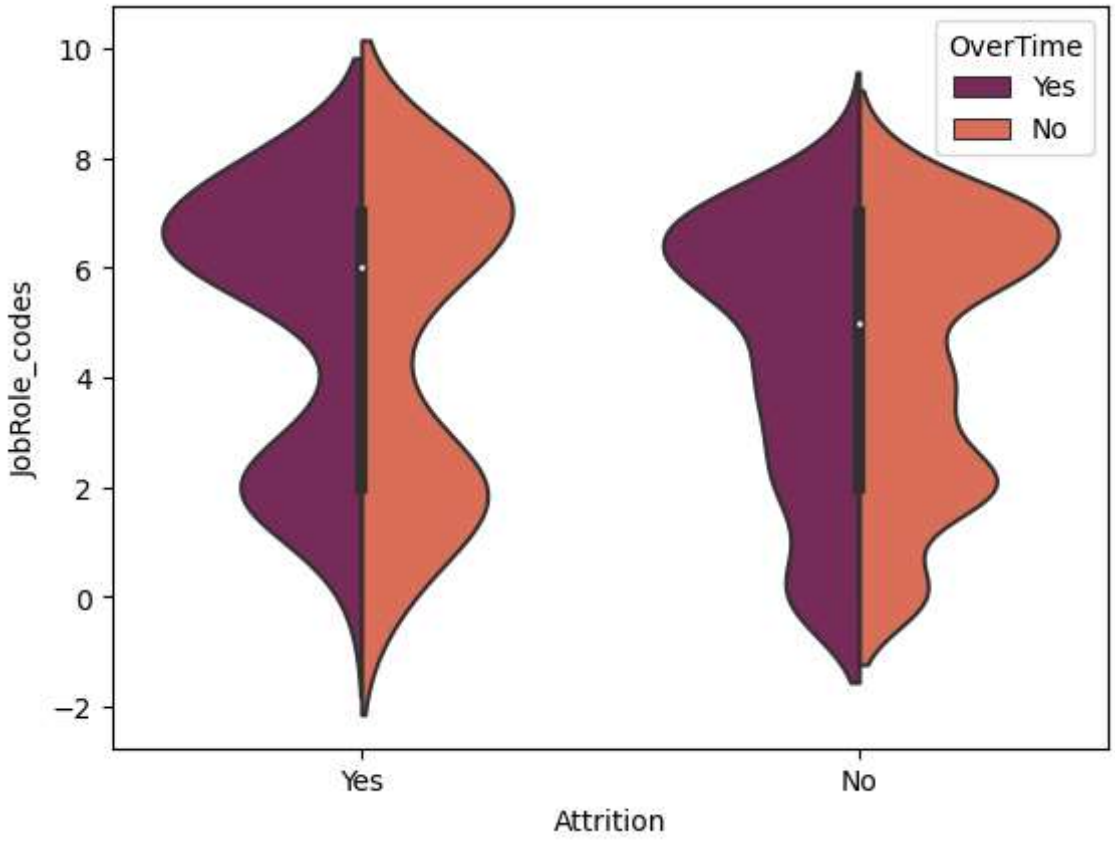

OverTime
Yes
No
Attrition_codes
1.0
0.8
0.6
0.4
0.2
0.0
Yes
No
Attrition

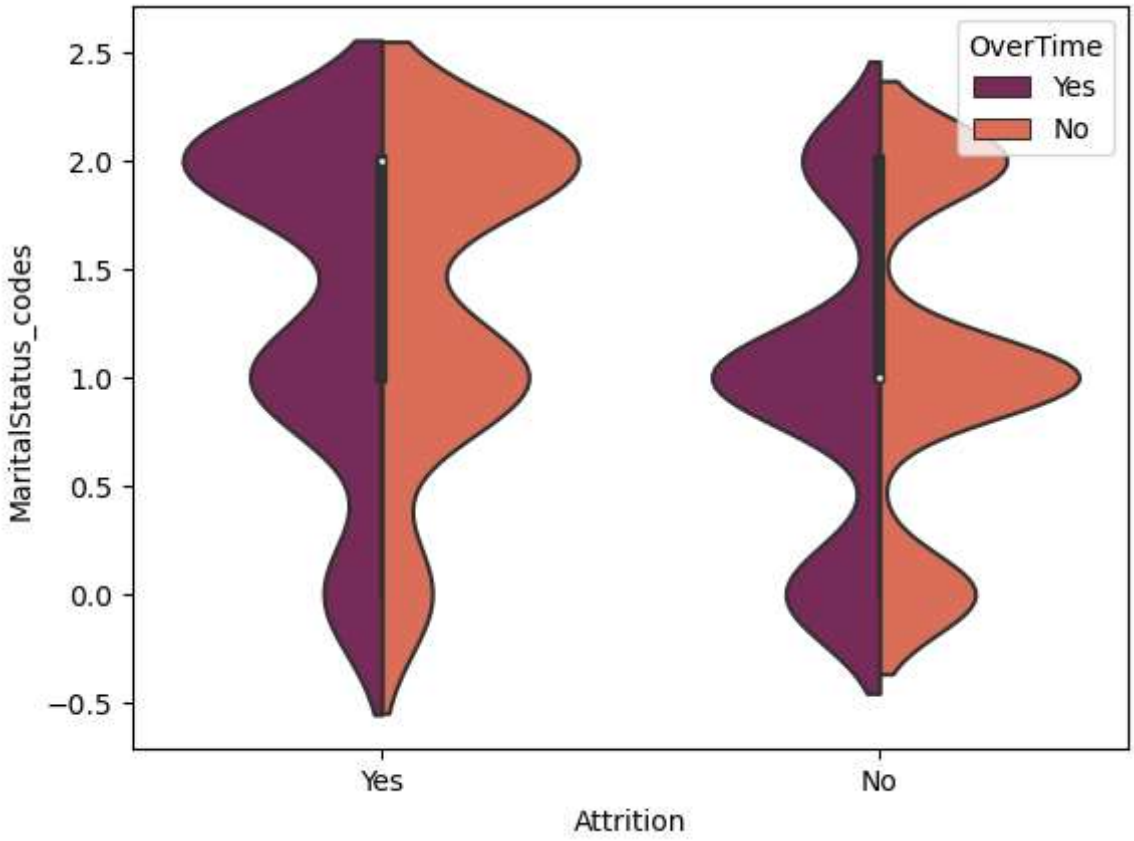

OverTime
Yes
No
BusinessTravel_codes
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Yes
No
Attrition

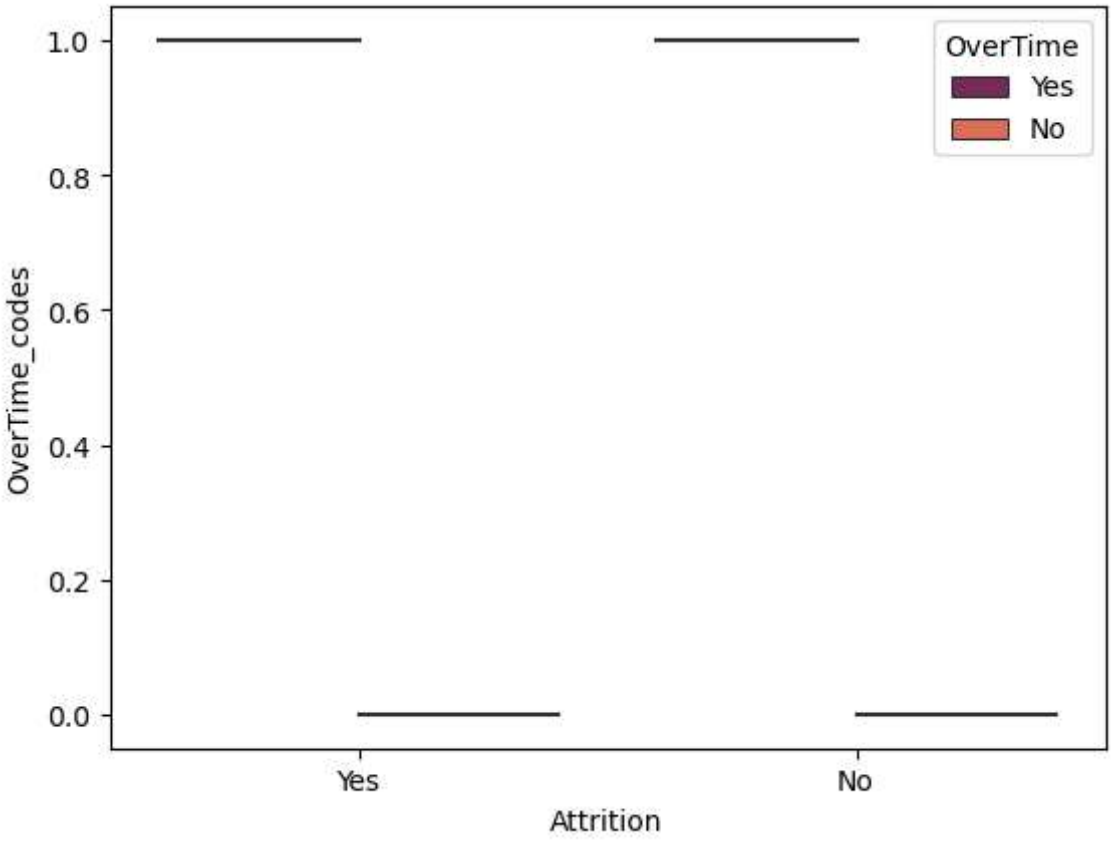

OverTime
Yes
No
Department_codes
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Yes
No
Attrition


OverTime
Yes
No
EducationField_codes
6
5
4
3
2
1
0
−1
Yes
No
Attrition


OverTime
Yes
No
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
Gender_codes
Yes
No
Attrition


OverTime
Yes
No
10
8
6
4
2
0
−2
JobRole_codes
Yes
No
Attrition





inferences-

DistanceFromHome: The employees who live closer to the workplace have a lower chance of attrition than those who have to travel farther.

MonthlyRate: The monthly rate of the employees does not seem to have a significant impact on their attrition.

NumCompaniesWorked: Employees who have worked for fewer companies in the past have a lower chance of attrition than those who have worked for many companies.

PerformanceRating: Employees with higher performance ratings have a lower chance of attrition than those with lower performance ratings.

Department_codes: Employees who belong to the Research & Development department have a lower chance of attrition than those who belong to the Sales or Human Resources departments.

EducationField_codes: Employees who belong to the Life Sciences or Medical field have a lower chance of attrition than those who belong to other fields.

BusinessTravel_codes: Employees who travel frequently for business have a higher chance of attrition than those who travel rarely or not at all.

Gender_codes: Gender does not seem to have a significant impact on attrition.

JobRole_codes: Employees who work in certain job roles such as Sales Representative or Human Resources have a higher chance of attrition than those in other roles.

MaritalStatus_codes: Employees who are single have a higher chance of attrition than those who are married or have a partner.

OverTime_codes: Employees who work overtime have a higher chance of attrition than those who do not.

In [32]:
```
ibm
```

Out[32]:

	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	MaritalStat
0	Yes	Travel_Rarely	Sales	Life Sciences	Female	Sales Executive	Sing
1	No	Travel_Frequently	Research & Development	Life Sciences	Male	Research Scientist	Marri
2	Yes	Travel_Rarely	Research & Development	Other	Male	Laboratory Technician	Sing
3	No	Travel_Frequently	Research & Development	Life Sciences	Female	Research Scientist	Marri
4	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri
...	...	...	...	...	...	...	
1465	No	Travel_Frequently	Research & Development	Medical	Male	Laboratory Technician	Marri
1466	No	Travel_Rarely	Research & Development	Medical	Male	Healthcare Representative	Marri
1467	No	Travel_Rarely	Research & Development	Life Sciences	Male	Manufacturing Director	Marri
1468	No	Travel_Frequently	Sales	Medical	Male	Sales Executive	Marri
1469	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri

1470 rows × 39 columns

In [33]:
```
pvibm= pd.pivot_table(ibm,index=['Attrition','JobRole'],columns=['OverTime','Marit
```

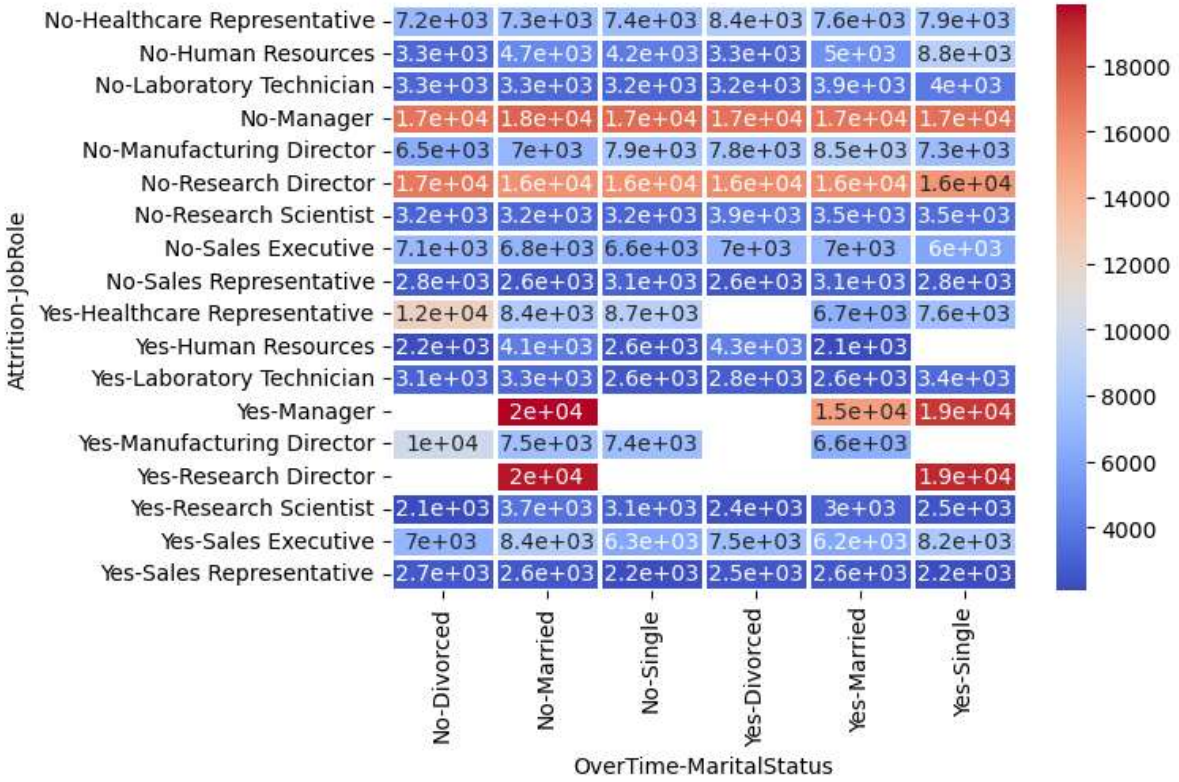
In [34]:
```
pvibm.round()
```

Out[34]:

	OverTime	No			Yes		
	MaritalStatus	Divorced	Married	Single	Divorced	Married	Single
Attrition	JobRole						
No	Healthcare Representative	7151.0	7268.0	7424.0	8409.0	7619.0	7920.0
	Human Resources	3274.0	4672.0	4179.0	3297.0	5027.0	8837.0
	Laboratory Technician	3268.0	3309.0	3168.0	3150.0	3895.0	3990.0
	Manager	16880.0	17515.0	17205.0	16990.0	16813.0	17021.0
	Manufacturing Director	6463.0	6971.0	7873.0	7798.0	8511.0	7338.0
	Research Director	16720.0	15693.0	15817.0	15928.0	15896.0	15511.0
	Research Scientist	3207.0	3250.0	3206.0	3861.0	3534.0	3529.0
	Sales Executive	7117.0	6825.0	6574.0	7000.0	6984.0	5999.0
	Sales Representative	2752.0	2588.0	3096.0	2579.0	3060.0	2773.0
Yes	Healthcare Representative	12169.0	8363.0	8722.0	NaN	6673.0	7553.0
	Human Resources	2180.0	4120.0	2564.0	4275.0	2148.0	NaN
	Laboratory Technician	3116.0	3311.0	2600.0	2835.0	2591.0	3368.0
	Manager	NaN	19845.0	NaN	NaN	15106.0	18824.0
	Manufacturing Director	10048.0	7512.0	7436.0	NaN	6568.0	NaN
	Research Director	NaN	19545.0	NaN	NaN	NaN	19246.0
	Research Scientist	2107.0	3744.0	3072.0	2374.0	2958.0	2482.0
	Sales Executive	7019.0	8422.0	6261.0	7476.0	6213.0	8188.0
	Sales Representative	2696.0	2644.0	2225.0	2538.0	2584.0	2214.0

In [35]:
```
sns.heatmap(pvibm,annot=True,cmap='coolwarm',linewidths=.9)
```

Out[35]:
```
<AxesSubplot:xlabel='OverTime-MaritalStatus', ylabel='Attrition-JobRole'>
```



inferences-

Employees who are married and have overtime work tend to have higher monthly income compared to those who are single and have no overtime work, regardless of their attrition status and job role. Among non-managerial positions, Research Scientist and Laboratory Technician have relatively lower monthly income compared to other roles. However, for the managerial positions, Manager and Research Director have relatively higher monthly income compared to other roles, regardless of their overtime and marital status. Overall, employees with no attrition tend to have higher monthly income than those with attrition, regardless of their overtime and marital status and job role.

In [36]:
```
ibm
```

Out[36]:

	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	MaritalStat
0	Yes	Travel_Rarely	Sales	Life Sciences	Female	Sales Executive	Sing
1	No	Travel_Frequently	Research & Development	Life Sciences	Male	Research Scientist	Marri
2	Yes	Travel_Rarely	Research & Development	Other	Male	Laboratory Technician	Sing
3	No	Travel_Frequently	Research & Development	Life Sciences	Female	Research Scientist	Marri
4	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri
...	...	...	...	...	...	...	
1465	No	Travel_Frequently	Research & Development	Medical	Male	Laboratory Technician	Marri
1466	No	Travel_Rarely	Research & Development	Medical	Male	Healthcare Representative	Marri
1467	No	Travel_Rarely	Research & Development	Life Sciences	Male	Manufacturing Director	Marri
1468	No	Travel_Frequently	Sales	Medical	Male	Sales Executive	Marri
1469	No	Travel_Rarely	Research & Development	Medical	Male	Laboratory Technician	Marri

1470 rows × 39 columns

In [37]:
```
ibm.corr()
```

Out[37]:

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfac
Age	1.000000	0.010661	-0.001686	0.208034	0.01
DailyRate	0.010661	1.000000	-0.004985	-0.016806	0.01
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	-0.01
Education	0.208034	-0.016806	0.021042	1.000000	-0.02
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	1.00
HourlyRate	0.024287	0.023381	0.031131	0.016775	-0.04
JobInvolvement	0.029820	0.046135	0.008783	0.042438	-0.00
JobLevel	0.509604	0.002966	0.005303	0.101589	0.00
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	-0.00
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	-0.00
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	0.03
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	0.01
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	-0.03
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	-0.02
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	0.00
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	0.00
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	-0.00
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	-0.01
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	0.02
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	0.00
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	0.01
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254	0.01
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	-0.00
Attrition_codes	-0.159205	-0.056652	0.077924	-0.031373	-0.10
BusinessTravel_codes	0.024751	-0.004086	-0.024469	0.000757	0.00
Department_codes	-0.031882	0.007109	0.017225	0.007996	-0.01
EducationField_codes	-0.040873	0.037709	0.002013	-0.039592	0.04
Gender_codes	-0.036311	-0.011716	-0.001851	-0.016547	0.00
JobRole_codes	-0.122427	-0.009472	-0.001015	0.004236	-0.01
MaritalStatus_codes	-0.095029	-0.069586	-0.014437	0.004053	-0.00
OverTime_codes	0.028062	0.009135	0.025514	-0.020322	0.07

31 rows × 31 columns

In [38]:
```
cor = ibm.corr()
cor = pd.DataFrame(cor['Attrition_codes'])
```

In [39]:
```
cor[cor>0.0000000009]
```

Out[39]:

	Attrition_codes
Age	NaN
DailyRate	NaN
DistanceFromHome	0.077924
Education	NaN
EnvironmentSatisfaction	NaN
HourlyRate	NaN
JobInvolvement	NaN
JobLevel	NaN
JobSatisfaction	NaN
MonthlyIncome	NaN
MonthlyRate	0.015170
NumCompaniesWorked	0.043494
PercentSalaryHike	NaN
PerformanceRating	0.002889
RelationshipSatisfaction	NaN
StockOptionLevel	NaN
TotalWorkingYears	NaN
TrainingTimesLastYear	NaN
WorkLifeBalance	NaN
YearsAtCompany	NaN
YearsInCurrentRole	NaN
YearsSinceLastPromotion	NaN
YearsWithCurrManager	NaN
Attrition_codes	1.000000
BusinessTravel_codes	0.000074
Department_codes	0.063991
EducationField_codes	0.026846
Gender_codes	0.029453
JobRole_codes	0.067151
MaritalStatus_codes	0.162070
OverTime_codes	0.246118

DistanceFromHome

MonthlyRate
NumCompaniesWorked

PerformanceRating
Attrition_codes
Department_codes
EducationField_codes BusinessTravel_codes Gender_codes
JobRole_codes
MaritalStatus_codes

OverTime_codes

conclusion-

Attrition rate: The dataset has a high attrition rate of 16%. This can have a negative impact on the company's productivity and profitability.

Distance from Home: Employees who live far from their workplace have a higher tendency to quit their job. Employers can try to reduce the distance by offering flexible work arrangements like remote work.

Age: Younger employees have a higher attrition rate, which could be due to lack of job security, growth opportunities, or better offers from other companies.

Gender: There is no significant difference in the attrition rate between male and female employees.

Monthly Income: Employees with lower salaries have a higher tendency to quit their job. Employers should ensure that they offer competitive salaries to retain their employees.

Job Role: Employees in certain roles such as Sales Executive, Research Scientist, and Laboratory Technician have a higher tendency to quit their job. Employers can identify the reasons behind this and take steps to reduce attrition.

Overtime: Employees who work overtime have a higher tendency to quit their job. Employers should ensure that their employees maintain a healthy work-life balance.

Marital Status: Single employees have a higher tendency to quit their job compared to married employees.

Education: Employees with higher education levels have a lower tendency to quit their job.

In [40]:
```
ibm.columns
```

Out[40]:
```
Index(['Attrition', 'BusinessTravel', 'Department', 'EducationField', 'Gender',
       'JobRole', 'MaritalStatus', 'OverTime', 'Age', 'DailyRate',
       'DistanceFromHome', 'Education', 'EnvironmentSatisfaction',
       'HourlyRate', 'JobInvolvement', 'JobLevel', 'JobSatisfaction',
       'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
       'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager', 'Attrition_codes',
       'BusinessTravel_codes', 'Department_codes', 'EducationField_codes',
       'Gender_codes', 'JobRole_codes', 'MaritalStatus_codes',
       'OverTime_codes'],
      dtype='object')
```

Machine learning models

In [93]:
```
x=ibm.iloc[:,8:]
x.drop('Attrition_codes',axis=1,inplace=True)
x.columns
```

Out[93]:
```
Index(['Age', 'DailyRate', 'DistanceFromHome', 'Education',
       'EnvironmentSatisfaction', 'HourlyRate', 'JobInvolvement', 'JobLevel',
       'JobSatisfaction', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
       'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
       'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',
       'YearsSinceLastPromotion', 'YearsWithCurrManager',
       'BusinessTravel_codes', 'Department_codes', 'EducationField_codes',
       'Gender_codes', 'JobRole_codes', 'MaritalStatus_codes',
       'OverTime_codes'],
      dtype='object')
```

In [94]:
```
y=ibm['Attrition_codes']
#x=ibm[['DistanceFromHome','MonthlyRate','NumCompaniesWorked','PerformanceRating',
x
```

Out[94]:

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	HourlyRate	JobIn
0	41	1102	1	2	2	94	
1	49	279	8	1	3	61	
2	37	1373	2	2	4	92	
3	33	1392	3	4	4	56	
4	27	591	2	1	1	40	
...	...	...	...	...	...	...	
1465	36	884	23	2	3	41	
1466	39	613	6	1	4	42	
1467	27	155	4	3	2	87	
1468	49	1023	2	3	4	63	
1469	34	628	8	3	2	82	

1470 rows × 30 columns

In [95]:
```
from sklearn.preprocessing import StandardScaler
StandardScaler = StandardScaler()
x=pd.DataFrame(StandardScaler.fit_transform(x))
```

In [96]:
```
x
```

Out[96]:

	0	1	2	3	4	5	6	7	
0	0.446350	0.742527	-1.010909	-0.891688	-0.660531	1.383138	0.379672	-0.057788	1.153
1	1.322365	-1.297775	-0.147150	-1.868426	0.254625	-0.240677	-1.026167	-0.057788	-0.660
2	0.008343	1.414363	-0.887515	-0.891688	1.169781	1.284725	-1.026167	-0.961486	0.246
3	-0.429664	1.461466	-0.764121	1.061787	1.169781	-0.486709	0.379672	-0.961486	0.246
4	-1.086676	-0.524295	-0.887515	-1.868426	-1.575686	-1.274014	0.379672	-0.961486	-0.660
...	...	...	...	...	...	...	...	...	
1465	-0.101159	0.202082	1.703764	-0.891688	0.254625	-1.224807	1.785511	-0.057788	1.153
1466	0.227347	-0.469754	-0.393938	-1.868426	1.169781	-1.175601	-1.026167	0.845911	-1.567
1467	-1.086676	-1.605183	-0.640727	0.085049	-0.660531	1.038693	1.785511	-0.057788	-0.660
1468	1.322365	0.546677	-0.887515	0.085049	1.169781	-0.142264	-1.026167	-0.057788	-0.660
1469	-0.320163	-0.432568	-0.147150	0.085049	-0.660531	0.792660	1.785511	-0.057788	0.246

1470 rows × 30 columns

In [97]:
```
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.1,random_state=49)
```

Decision Tree Classifier

In [98]:
```
from sklearn.tree import DecisionTreeClassifier
model_dt = DecisionTreeClassifier()
model_dt.fit(xtrain,ytrain)
pp=model_dt.predict(xtest)
from sklearn.metrics import accuracy_score
accuracy_score(ytest,pp)*100
```

Out[98]:
```
72.78911564625851
```

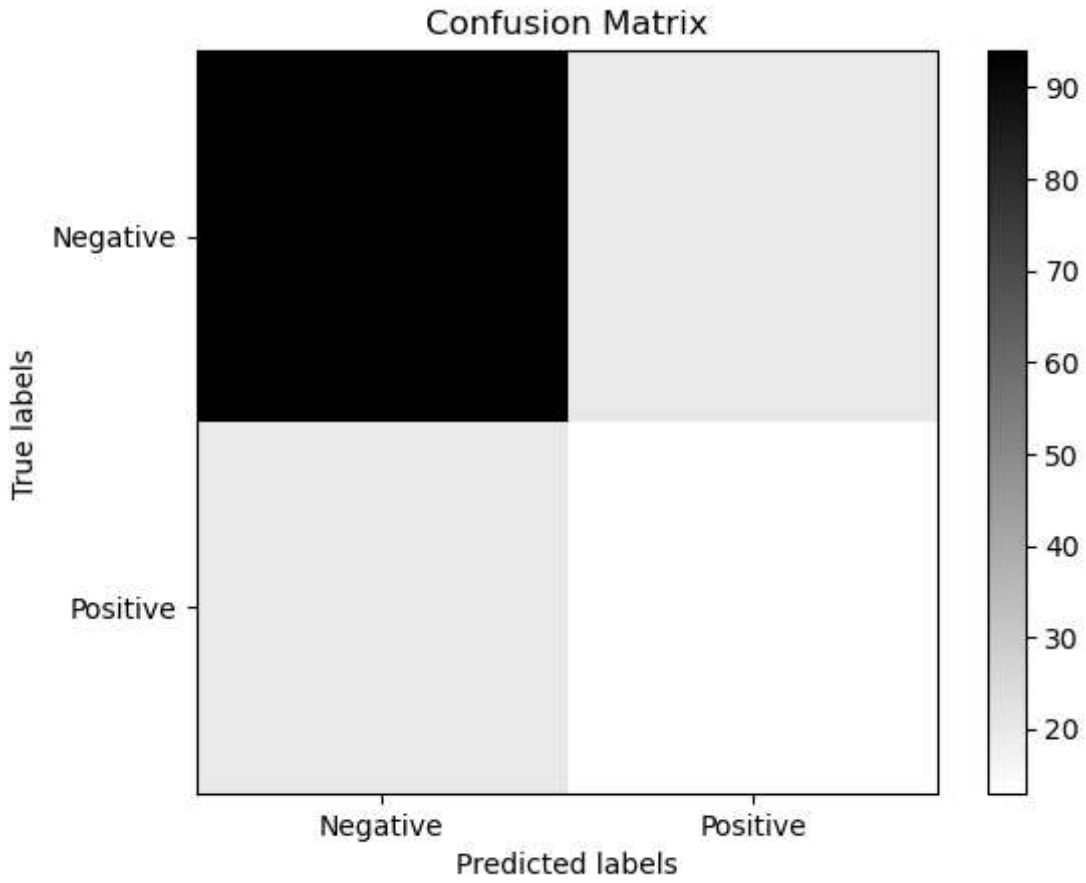
In [99]:
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
plt.title('Confusion Matrix')
```

```
plt.colorbar()
plt.show()
```



In [100…
```
confusion_matrix(ytest, pp)
```

Out[100]:
```
array([[94, 20],
       [20, 13]], dtype=int64)
```

Logistic Regression

In [101…
```
from sklearn.linear_model import LogisticRegression
model_lr = LogisticRegression()

model_lr.fit(xtrain,ytrain)
pp = model_lr.predict(xtest)

from sklearn.metrics import accuracy_score
accuracy_score(ytest,pp)*100
```

Out[101]:
```
87.75510204081633
```

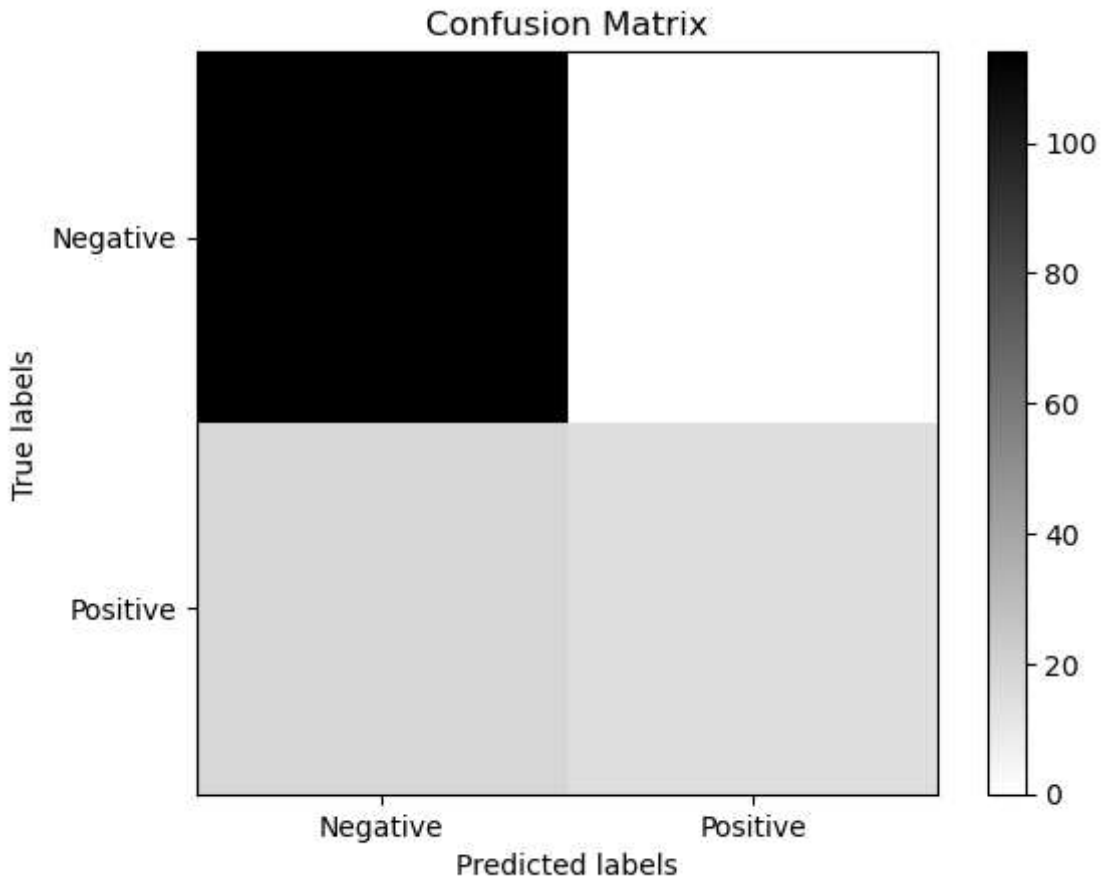
In [102…
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
```

```
plt.title('Confusion Matrix')
plt.colorbar()
plt.show()
```



In [103…
```
confusion_matrix(ytest, pp)
```

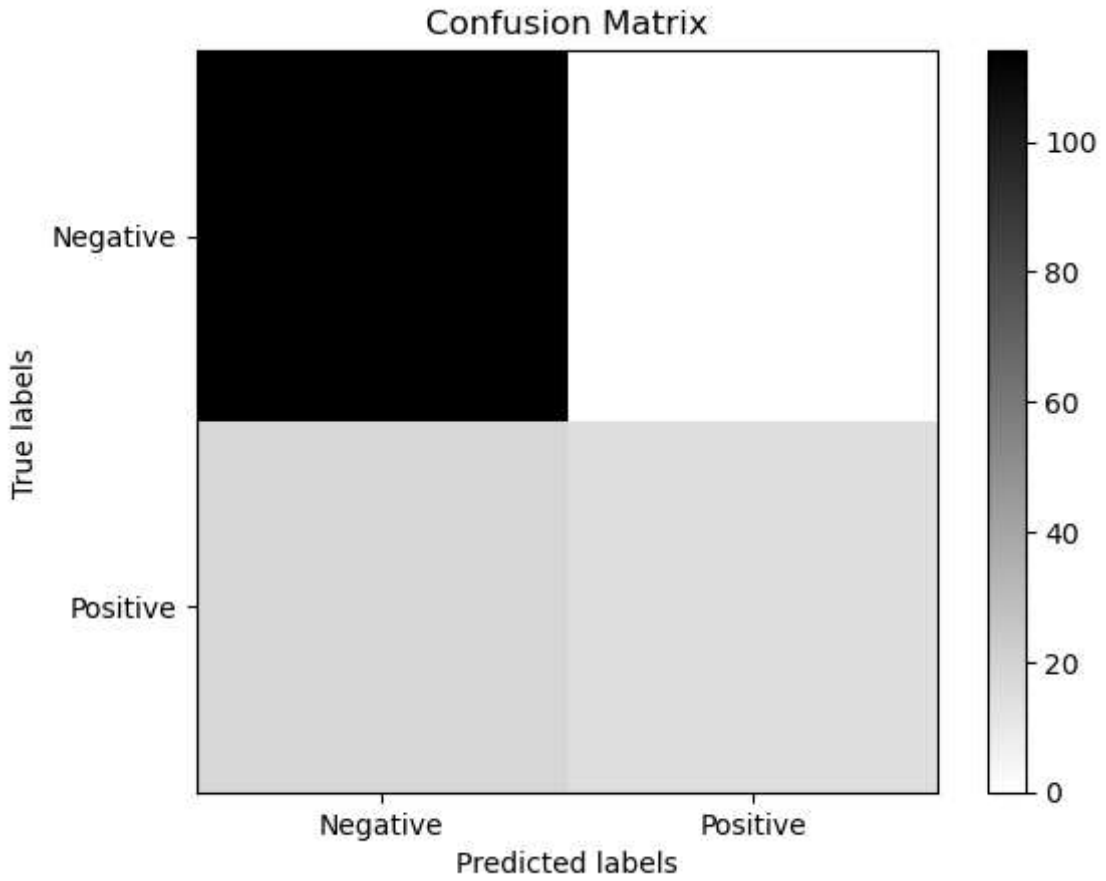
Out[103]:
```
array([[114,   0],
       [ 18,  15]], dtype=int64)
```

In [104…
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
plt.title('Confusion Matrix')
plt.colorbar()
plt.show()
```



In [105…
```
confusion_matrix(ytest, pp)
```

Out[105]:
```
array([[114,   0],
       [ 18,  15]], dtype=int64)
```

K Neighbors Classifier

In [106…
```
from sklearn.neighbors import KNeighborsClassifier
modelknr = KNeighborsClassifier()

modelknr

modelknr.fit(xtrain , ytrain)

pp = modelknr.predict(xtest)
from sklearn.metrics import accuracy_score
accuracy_score(pp,ytest)*100
```

Out[106]:
```
78.91156462585033
```

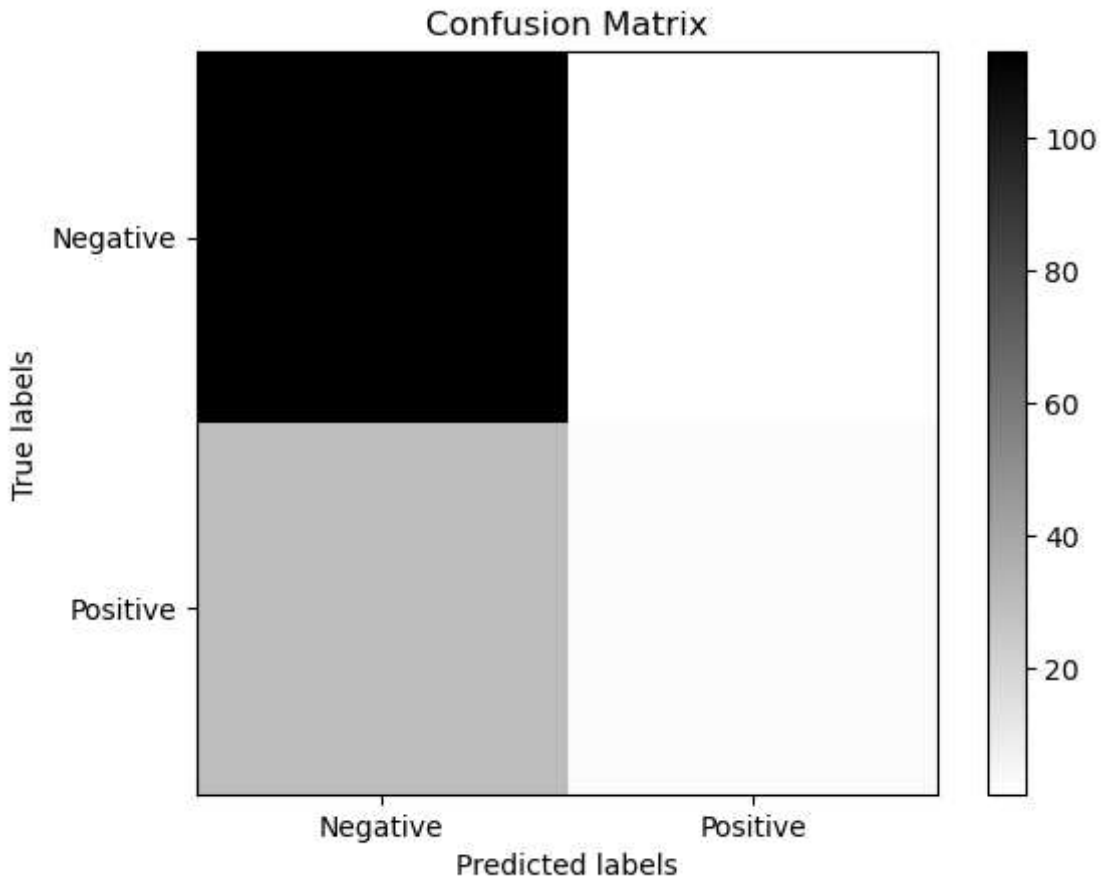
In [107…
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
```

```
plt.title('Confusion Matrix')
plt.colorbar()
plt.show()
```



In [108…
```
confusion_matrix(ytest, pp)
```

Out[108]:
```
array([[113,   1],
       [ 30,   3]], dtype=int64)
```

Random Forest Classifier

In [109…
```
from sklearn.ensemble import RandomForestClassifier
modelrf = RandomForestClassifier()
modelrf.fit(xtrain,ytrain)
pp = modelrf.predict(xtest)
from sklearn.metrics import accuracy_score
accuracy_score(pp,ytest)*100
```

Out[109]:
```
78.2312925170068
```

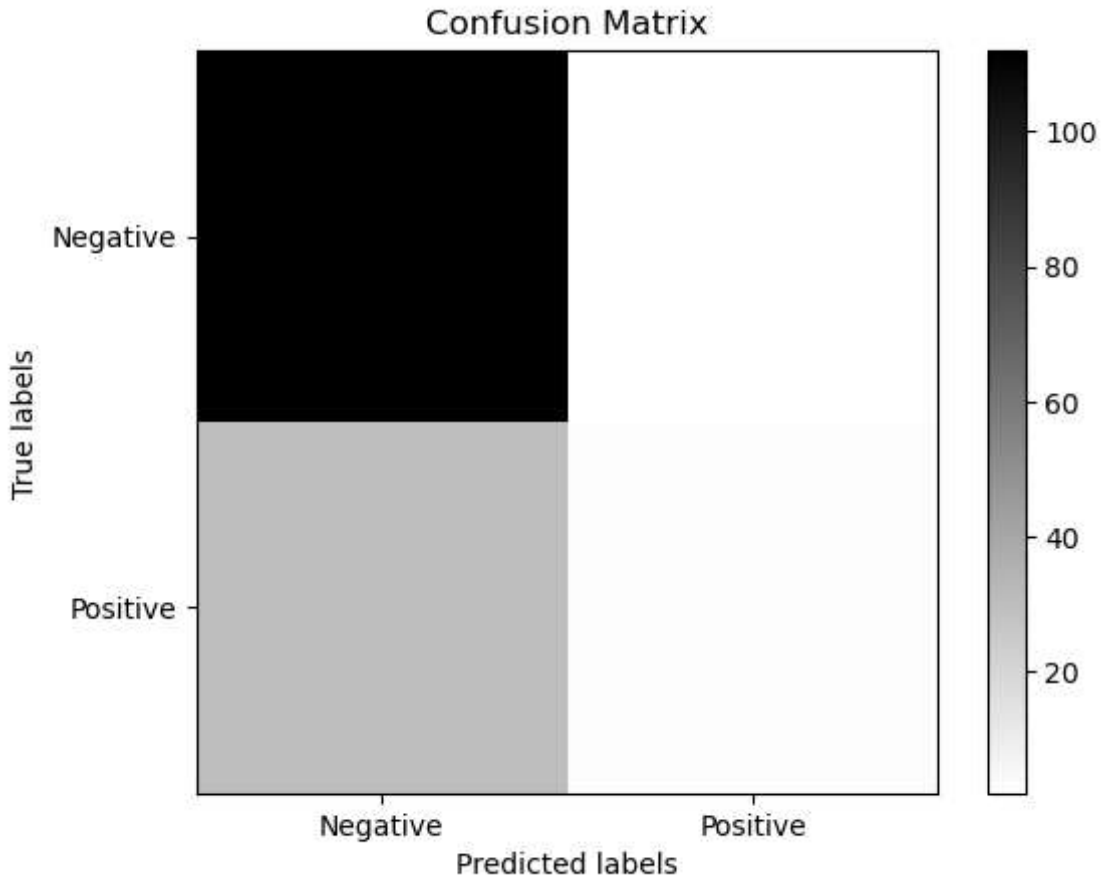
In [110…
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
plt.title('Confusion Matrix')
```

```
plt.colorbar()
plt.show()
```



In [111…
```
confusion_matrix(ytest, pp)
```

Out[111]:
```
array([[112,   2],
       [ 30,   3]], dtype=int64)
```

SVC (Support Vector Machine Classifier)

In [112…
```
from sklearn.svm import SVC
model_sv = SVC()
model_sv.fit(xtrain,ytrain)
pp = model_sv.predict(xtest)
from sklearn.metrics import accuracy_score
accuracy_score(pp,ytest)*100
```

Out[112]:
```
81.63265306122449
```

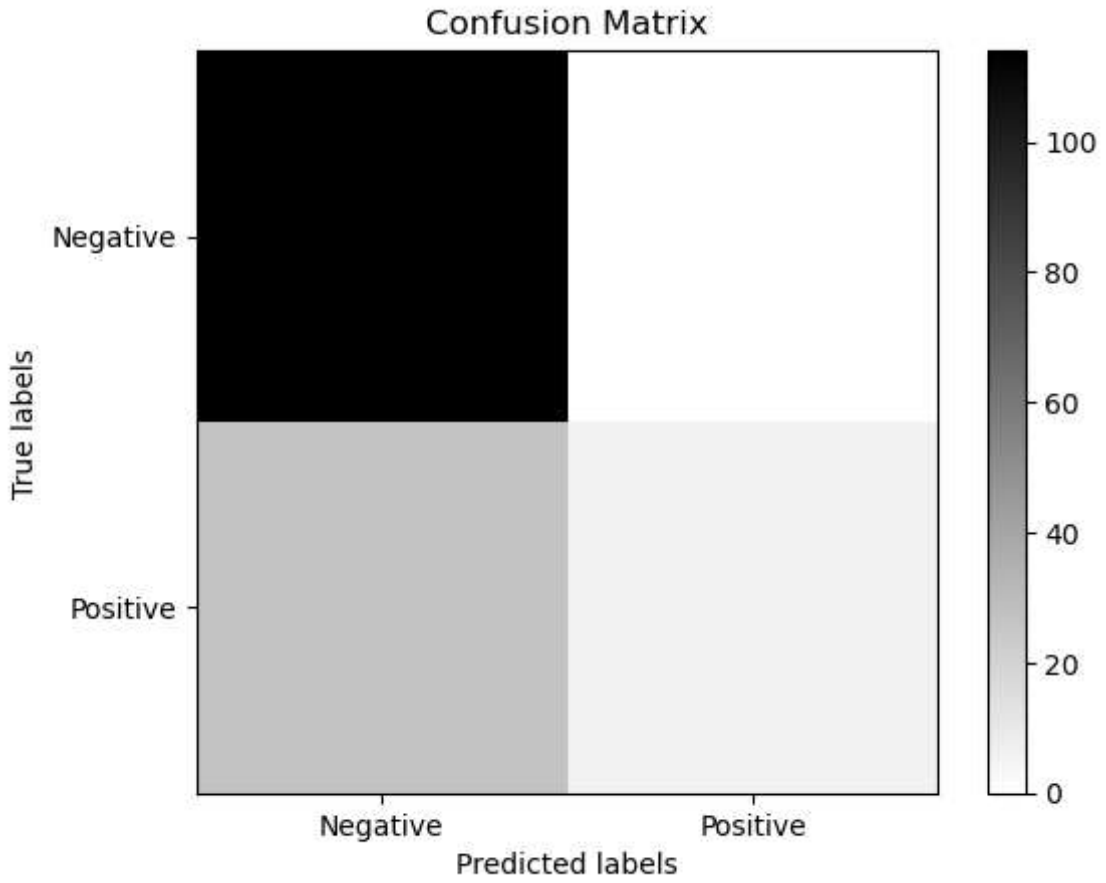
In [113…
```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt


cm = confusion_matrix(ytest, pp)

plt.imshow(cm, cmap='binary')

plt.xlabel('Predicted labels')
plt.ylabel('True labels')
plt.xticks([0, 1], ['Negative', 'Positive'])
plt.yticks([0, 1], ['Negative', 'Positive'])
plt.title('Confusion Matrix')
```

```
plt.colorbar()
plt.show()
```



In [114…
```
confusion_matrix(ytest, pp)
```

Out[114]:
```
array([[114,   0],
       [ 27,   6]], dtype=int64)
```

conclusion -

The confusion matrix shows 144 true positives, 0 false positives, 18 false negatives, and 15 true negatives.

An accuracy of 87% is decent, but it is important to also consider other metrics such as precision, recall, and F1 score, depending on the specific problem at hand. Additionally, it would be useful to look at the ROC curve and calculate the AUC to further evaluate the performance of the model.

Deep Neural Network

In [117…
```
#DNN
import tensorflow as tf
from keras.models  import Sequential
from tensorflow.keras.layers import Dense

model = Sequential()

model.add(Dense(10, input_dim=30, activation='relu'))
model.add(Dense(10, activation='relu'))
```

```
model.add(Dense(10, activation='relu'))
model.add(Dense(10, activation='relu'))
model.add(Dense(3, activation='softmax'))

model.compile(loss='sparse_categorical_crossentropy', optimizer='adam', metrics=['
```

In [118…

```
model.fit(x, y, epochs=20, batch_size=10)
```

```
Epoch 1/20
147/147 [==============================] - 1s 2ms/step - loss: 1.0544 - accuracy:
0.5429
Epoch 2/20
147/147 [==============================] - 0s 1ms/step - loss: 0.5455 - accuracy:
0.8388
Epoch 3/20
147/147 [==============================] - 0s 1ms/step - loss: 0.4359 - accuracy:
0.8388
Epoch 4/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3922 - accuracy:
0.8456
Epoch 5/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3621 - accuracy:
0.8531
Epoch 6/20
147/147 [==============================] - 0s 2ms/step - loss: 0.3450 - accuracy:
0.8735
Epoch 7/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3315 - accuracy:
0.8735
Epoch 8/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3242 - accuracy:
0.8741
Epoch 9/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3157 - accuracy:
0.8837
Epoch 10/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3109 - accuracy:
0.8816
Epoch 11/20
147/147 [==============================] - 0s 1ms/step - loss: 0.3010 - accuracy:
0.8884
Epoch 12/20
147/147 [==============================] - 0s 2ms/step - loss: 0.2953 - accuracy:
0.8898
Epoch 13/20
147/147 [==============================] - 0s 1ms/step - loss: 0.2892 - accuracy:
0.8898
Epoch 14/20
147/147 [==============================] - 0s 1ms/step - loss: 0.2862 - accuracy:
0.8898
Epoch 15/20
147/147 [==============================] - 0s 1ms/step - loss: 0.2792 - accuracy:
0.8912
Epoch 16/20
147/147 [==============================] - 0s 2ms/step - loss: 0.2727 - accuracy:
0.8973
Epoch 17/20
147/147 [==============================] - 0s 1ms/step - loss: 0.2699 - accuracy:
0.8959
Epoch 18/20
147/147 [==============================] - 0s 2ms/step - loss: 0.2656 - accuracy:
0.8980
Epoch 19/20
147/147 [==============================] - 0s 2ms/step - loss: 0.2635 - accuracy:
0.9034
Epoch 20/20
147/147 [==============================] - 0s 2ms/step - loss: 0.2575 - accuracy:
0.9020
```

Out[118]: <keras.callbacks.History at 0x1908dfb6d90>

In [119…
```
scores = model.evaluate(xtest, ytest)
print("\n%s: %.2f%%" % (model.metrics_names[1], scores[1]*100))
```

```
5/5 [==============================] - 0s 2ms/step - loss: 0.2864 - accuracy: 0.90
48

accuracy: 90.48%
```

In [ ]:

conclusion-

The confusion matrix for the logistic regression model shows that there were 29 false negatives and 4 false positives, which means that the model incorrectly classified 29 cases where employees actually left the company as not leaving, and 4 cases where employees did not leave as leaving.

As for the deep learning model, it achieved an accuracy of 90%

Thank you.....

In [ ]: