DATA ANALYSIS PROJECT ON H.R DATA-SET

Importing Libraries

```
In [53]:  import pandas as pd
          import matplotlib.pyplot as plt
          %matplotlib inline
          import seaborn as sns
          import warnings
          warnings.filterwarnings('ignore')
```

Loading the dataset

```
In [38]:  df = pd.read_csv(r'HR Data.csv')
          pd.set_option('display.max_columns', None)
          df
```

Out[38]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Emplo Nur
0	Yes	Travel_Rarely	35 - 44	Ex-Employees	Sales	Life Sciences	STAFF-1	
1	No	Travel_Frequently	45 - 54	Current Employees	R&D	Life Sciences	STAFF-2	
2	Yes	Travel_Rarely	35 - 44	Ex-Employees	R&D	Other	STAFF-4	
3	No	Travel_Frequently	25 - 34	Current Employees	R&D	Life Sciences	STAFF-5	
4	No	Travel_Rarely	25 - 34	Current Employees	R&D	Medical	STAFF-7	
...	...	...	...	...	...	...	...	
1465	Yes	Non-Travel	25 - 34	Ex-Employees	R&D	Technical Degree	STAFF-1905	
1466	Yes	Travel_Frequently	25 - 34	Ex-Employees	R&D	Life Sciences	STAFF-1868	
1467	Yes	Travel_Frequently	35 - 44	Ex-Employees	Sales	Other	STAFF-1667	
1468	Yes	Travel_Rarely	Under 25	Ex-Employees	R&D	Life Sciences	STAFF-1878	
1469	Yes	Travel_Rarely	Under 25	Ex-Employees	Sales	Life Sciences	STAFF-1702	

1470 rows × 39 columns

INTRODUCTION OF PROJECT

TARGET COLUMN == ATTRITION

-ATTRITION :-Attrition is the departure of employees from the organization for any reason or in simple words we can say that employees who want to leave the company.

-ATTRITION "YES" :- Employees wants to leave the company.

-ATTRITION "NO" :- Employees don't want to leave.

PROBLEM STATEMENT

This project aims to uncover the factors contributing to employee attrition. Through analysis of various dataset features, we seek to understand their impact on attrition rates within the organization.

Exploratory Data Analysis and Data Cleaning

In [110]:
```python
# Visualizing top 5 rows
df.head()
```

Out[110]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Employee Number
0	Yes	Travel_Rarely	35 - 44	Ex-Employees	Sales	Life Sciences	STAFF-1	
1	No	Travel_Frequently	45 - 54	Current Employees	R&D	Life Sciences	STAFF-2	
2	Yes	Travel_Rarely	35 - 44	Ex-Employees	R&D	Other	STAFF-4	
3	No	Travel_Frequently	25 - 34	Current Employees	R&D	Life Sciences	STAFF-5	
4	No	Travel_Rarely	25 - 34	Current Employees	R&D	Medical	STAFF-7	

In [111]:
```python
# Visualizing bottom 5 rows
df.tail()
```

Out[111]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Emplo Nur
1465	Yes	Non-Travel	25 - 34	Ex-Employees	R&D	Technical Degree	STAFF-1905	
1466	Yes	Travel_Frequently	25 - 34	Ex-Employees	R&D	Life Sciences	STAFF-1868	
1467	Yes	Travel_Frequently	35 - 44	Ex-Employees	Sales	Other	STAFF-1667	
1468	Yes	Travel_Rarely	Under 25	Ex-Employees	R&D	Life Sciences	STAFF-1878	
1469	Yes	Travel_Rarely	Under 25	Ex-Employees	Sales	Life Sciences	STAFF-1702	

In [112]:
```python
df.shape #shows number of rows and column
```

Out[112]: (1470, 39)

In [113]:
```python
df.columns #shows all column name in dataframe
```

Out[113]:
```
Index(['Attrition', 'Business Travel', 'CF_age band', 'CF_attrition labe
l',
       'Department', 'Education Field', 'emp no', 'Employee Number', 'Gend
er',
       'Job Role', 'Marital Status', 'Over Time', 'Over18',
       'Training Times Last Year', 'Age', 'CF_current Employee', 'Daily Ra
te',
       'Distance From Home', 'Education', 'Employee Count',
       'Environment Satisfaction', 'Hourly Rate', 'Job Involvement',
       'Job Level', 'Job Satisfaction', 'Monthly Income', 'Monthly Rate',
       'Num Companies Worked', 'Percent Salary Hike', 'Performance Ratin
g',
       'Relationship Satisfaction', 'Standard Hours', 'Stock Option Leve
l',
       'Total Working Years', 'Work Life Balance', 'Years At Company',
       'Years In Current Role', 'Years Since Last Promotion',
       'Years With Curr Manager'],
      dtype='object')
```

In [114]: `df.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   Business Travel             1470 non-null    object
 2   CF_age band                 1470 non-null    object
 3   CF_attrition label          1470 non-null    object
 4   Department                  1470 non-null    object
 5   Education Field             1470 non-null    object
 6   emp no                      1470 non-null    object
 7   Employee Number             1470 non-null    int64
 8   Gender                      1470 non-null    object
 9   Job Role                    1470 non-null    object
 10  Marital Status              1470 non-null    object
 11  Over Time                   1470 non-null    object
 12  Over18                      1470 non-null    object
 13  Training Times Last Year    1470 non-null    int64
 14  Age                         1470 non-null    int64
 15  CF_current Employee         1470 non-null    int64
 16  Daily Rate                  1470 non-null    int64
 17  Distance From Home          1470 non-null    int64
 18  Education                   1470 non-null    object
 19  Employee Count              1470 non-null    int64
 20  Environment Satisfaction    1470 non-null    int64
 21  Hourly Rate                 1470 non-null    int64
 22  Job Involvement             1470 non-null    int64
 23  Job Level                   1470 non-null    int64
 24  Job Satisfaction            1470 non-null    int64
 25  Monthly Income              1470 non-null    int64
 26  Monthly Rate                1470 non-null    int64
 27  Num Companies Worked        1470 non-null    int64
 28  Percent Salary Hike         1470 non-null    int64
 29  Performance Rating          1470 non-null    int64
 30  Relationship Satisfaction   1470 non-null    int64
 31  Standard Hours              1470 non-null    int64
 32  Stock Option Level          1470 non-null    int64
 33  Total Working Years         1470 non-null    int64
 34  Work Life Balance           1470 non-null    int64
 35  Years At Company            1470 non-null    int64
 36  Years In Current Role       1470 non-null    int64
 37  Years Since Last Promotion  1470 non-null    int64
 38  Years With Curr Manager     1470 non-null    int64
dtypes: int64(26), object(13)
memory usage: 448.0+ KB
```

```
In [115]: for col in df.describe(include='object').columns: # check all categorical co
              print(col)
              print(df[col].unique())
              print('-'*50)
```

```
Attrition
['Yes' 'No']
--------------------------------------------------
Business Travel
['Travel_Rarely' 'Travel_Frequently' 'Non-Travel']
--------------------------------------------------
CF_age band
['35 - 44' '45 - 54' '25 - 34' 'Over 55' 'Under 25']
--------------------------------------------------
CF_attrition label
['Ex-Employees' 'Current Employees']
--------------------------------------------------
Department
['Sales' 'R&D' 'HR']
--------------------------------------------------
Education Field
['Life Sciences' 'Other' 'Medical' 'Marketing' 'Technical Degree'
 'Human Resources']
--------------------------------------------------
emp no
['STAFF-1' 'STAFF-2' 'STAFF-4' ... 'STAFF-1667' 'STAFF-1878' 'STAFF-1702']
--------------------------------------------------
Gender
['Female' 'Male']
--------------------------------------------------
Job Role
['Sales Executive' 'Research Scientist' 'Laboratory Technician'
 'Manufacturing Director' 'Healthcare Representative' 'Manager'
 'Sales Representative' 'Research Director' 'Human Resources']
--------------------------------------------------
Marital Status
['Single' 'Married' 'Divorced']
--------------------------------------------------
Over Time
['Yes' 'No']
--------------------------------------------------
Over18
['Y']
--------------------------------------------------
Education
['Associates Degree' 'High School' "Master's Degree" "Bachelor's Degree"
 'Doctoral Degree']
--------------------------------------------------
```

In [116]: `pd.isnull(df)` *# checking null values*

Out[116]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Employee Number	Ge
0	False	False	False	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	
1465	False	False	False	False	False	False	False	False	F
1466	False	False	False	False	False	False	False	False	F
1467	False	False	False	False	False	False	False	False	F
1468	False	False	False	False	False	False	False	False	F
1469	False	False	False	False	False	False	False	False	F

1470 rows × 39 columns

In [117]:
```python
pd.isnull(df).sum()
```

Out[117]:
```
Attrition                      0
Business Travel                0
CF_age band                    0
CF_attrition label             0
Department                     0
Education Field                0
emp no                         0
Employee Number                0
Gender                         0
Job Role                       0
Marital Status                 0
Over Time                      0
Over18                         0
Training Times Last Year       0
Age                            0
CF_current Employee            0
Daily Rate                     0
Distance From Home             0
Education                      0
Employee Count                 0
Environment Satisfaction       0
Hourly Rate                    0
Job Involvement                0
Job Level                      0
Job Satisfaction               0
Monthly Income                 0
Monthly Rate                   0
Num Companies Worked           0
Percent Salary Hike            0
Performance Rating             0
Relationship Satisfaction      0
Standard Hours                 0
Stock Option Level             0
Total Working Years            0
Work Life Balance              0
Years At Company               0
Years In Current Role          0
Years Since Last Promotion     0
Years With Curr Manager        0
dtype: int64
```

Shows no null value present in the dataframe

In [118]:
```python
# Use the describe() method to get summary statistics
df.describe()
```

Out[118]:

	Employee Number	Training Times Last Year	Age	CF_current Employee	Daily Rate	Distance From Home	Emplo Co
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	147
mean	1024.865306	2.799320	36.923810	0.838776	802.485714	9.192517	
std	602.024335	1.289271	9.135373	0.367863	403.509100	8.106864	
min	1.000000	0.000000	18.000000	0.000000	102.000000	1.000000	
25%	491.250000	2.000000	30.000000	1.000000	465.000000	2.000000	
50%	1020.500000	3.000000	36.000000	1.000000	802.000000	7.000000	
75%	1555.750000	3.000000	43.000000	1.000000	1157.000000	14.000000	
max	2068.000000	6.000000	60.000000	1.000000	1499.000000	29.000000	

In [119]:
```python
df.describe(include='object')
```

Out[119]:

	Attrition	Business Travel	CF_age band	CF_attrition label	Department	Education Field	emp no	Gender
count	1470	1470	1470	1470	1470	1470	1470	1470
unique	2	3	5	2	3	6	1470	2
top	No	Travel_Rarely	25 - 34	Current Employees	R&D	Life Sciences	STAFF-1	Male
freq	1233	1043	554	1233	961	606	1	882

Data Analysis and Visulaization

ANALYSIS ON CATEGORICAL COLUMN WRT TARGET COLUMN(ATTRITION)

In [120]:
```python
ax=sns.countplot(data=df, x='Attrition')
plt.show()
```
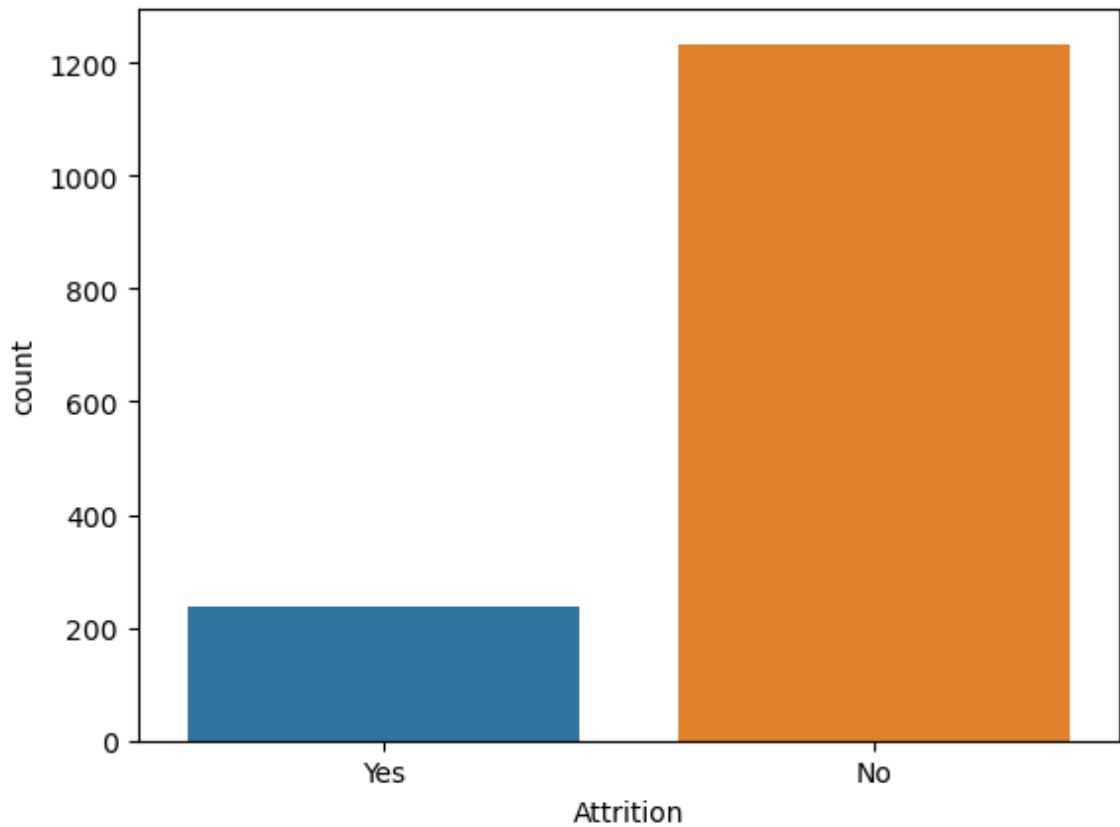

Observation:- Data of attrition say that it has more number of "No" values and less number of "Yes"

1. IMPACT OF BUSINESS TRAVEL ON ATTRITION

In [121]:
```python
ax=sns.countplot(data=df, x='Business Travel', hue='Attrition')
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```
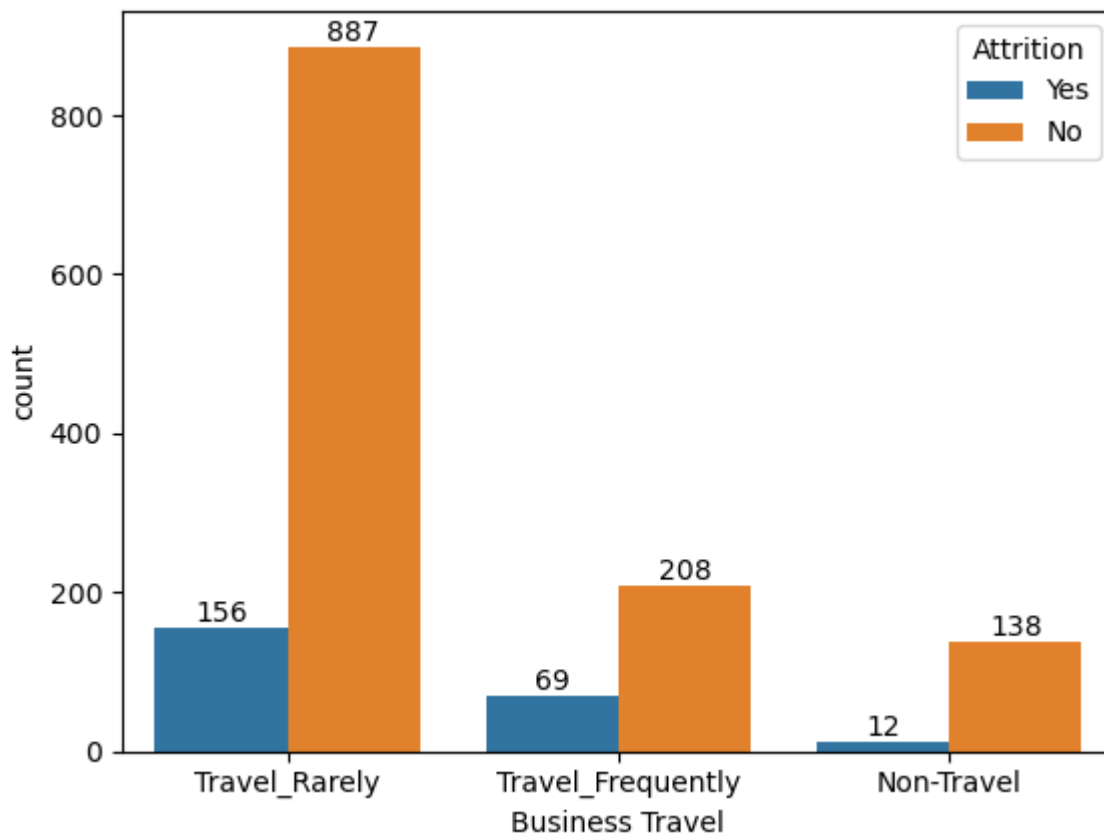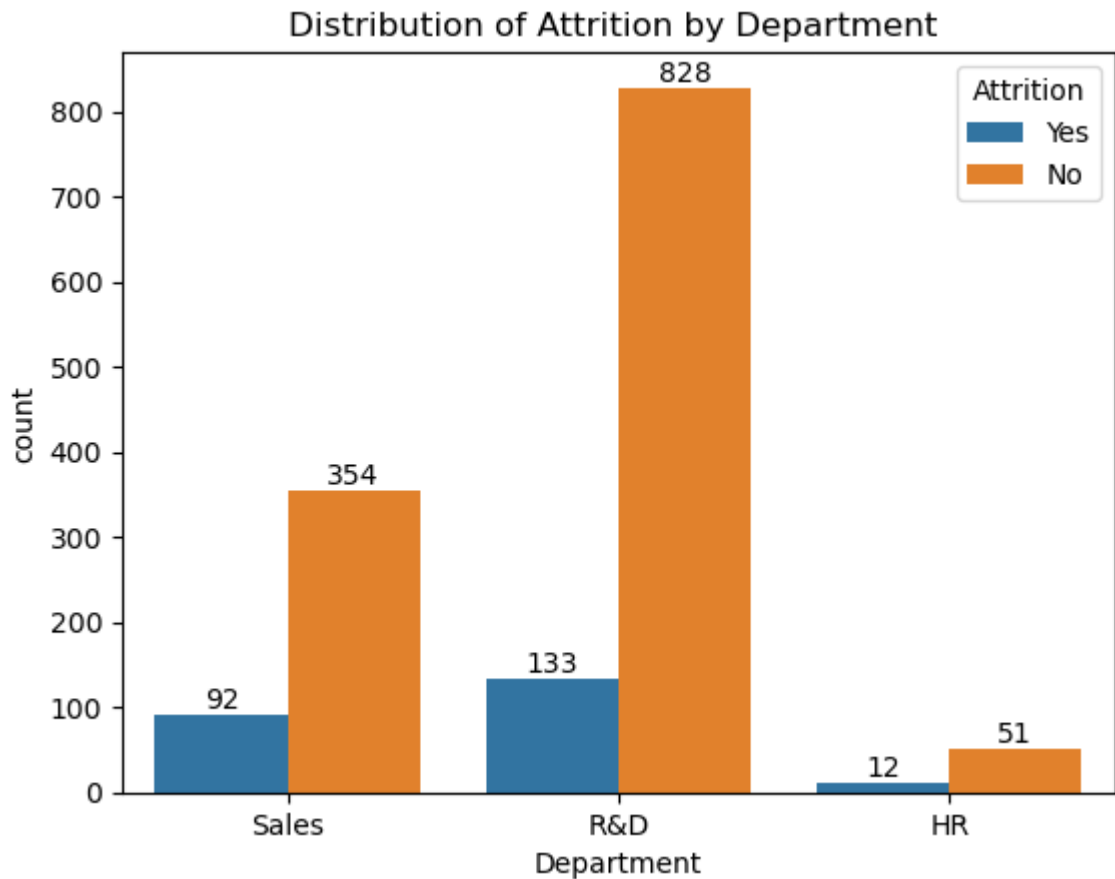

Observation:-
-Graph tells us that company has more count or more number of employes
whos travels rarely it means travel rate of company is less
-There are more employees which travel rarely and are not satisfied with
their job.
-Non traveller have least count as well as least attrition.

2. IMPACT OF DEPARTMENT ON ATTRITION

In [54]:
```python
ax2=sns.countplot(data=df, x='Department', hue='Attrition' )
plt.title('Distribution of Attrition by Department ')
for bars in ax2.containers:
    ax2.bar_label(bars)
plt.show()
```


observation:-
- Research and Development (R&D) department have more number attrition (133 employees) as compared to other 2 department
- After R & D department second highest attrition(92 employees) sales department.
- HR department have least attrition with just 12 employees

3. IMPACT OF EDUCATION FIELD ON ATTRITION

In [73]:
```python
ax3=sns.countplot(data=df, x='Attrition', hue='Education Field')
for bars in ax3.containers:
    ax3.bar_label(bars)
plt.show()
```
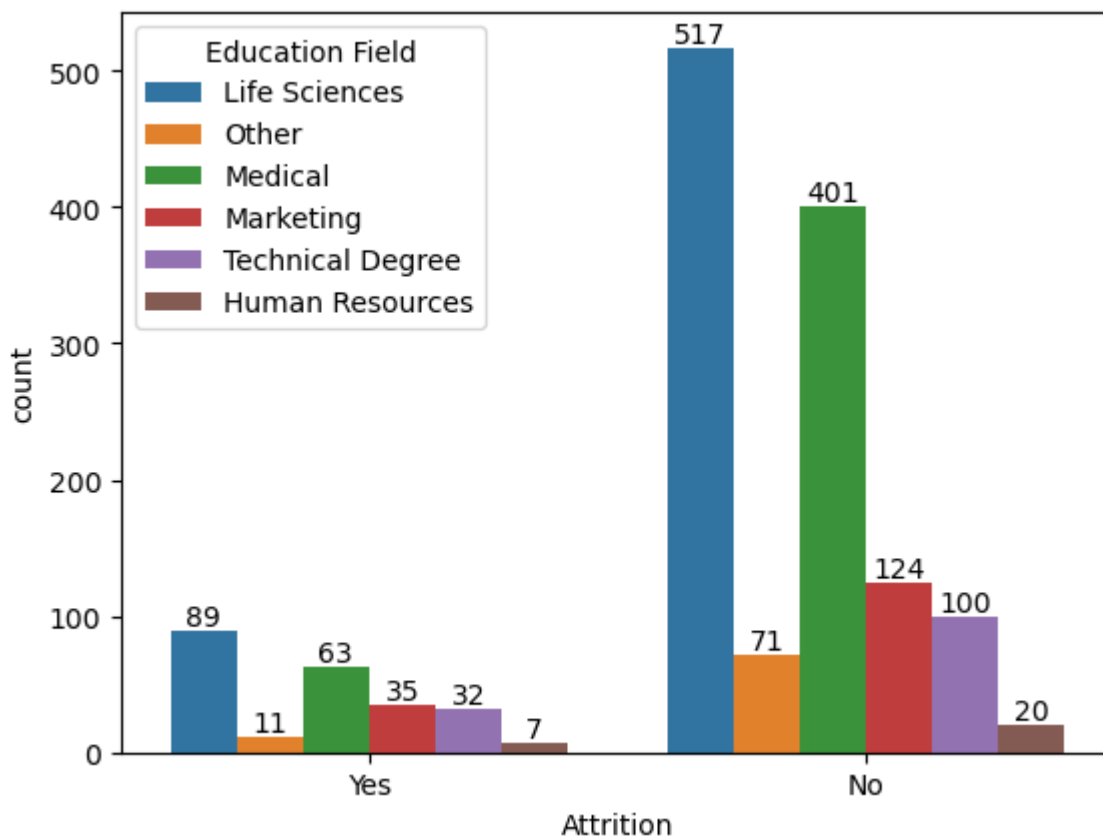


```
Observation:-
- In this graph shows the life science and medical education fileds are
more employees as compared to other
- Nearly 89 number of employees are there who are from life science
education background will leave the company anf followed by
medical education employees
- As we conclude from analysis of department and attrition, here also HR
education background employees have leat attrition
```

4. IMPACT OF GENDER ON ATTRITION

In [83]:
```python
ax4=sns.countplot(data=df, x='Attrition', hue='Gender')
for bars in ax4.containers:
    ax4.bar_label(bars)
plt.show()
```
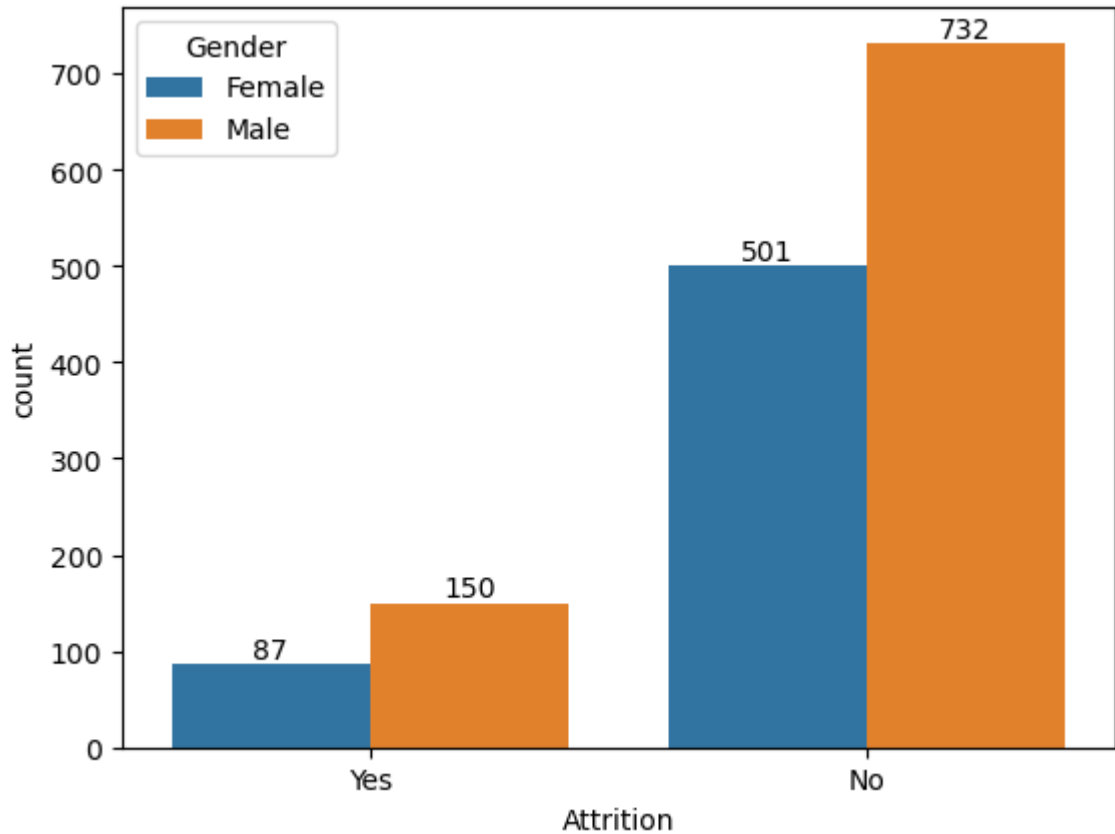


```
observation:-
- Male employee are more as compared to female
- Males are more likely to quit the job rather than females
```

5. IMPACT OF OVERTIME ON ATTRITION

In [72]:
```python
ax5=sns.countplot(data=df,x='Over Time', hue='Attrition')
for bars in ax5.containers:
    ax5.bar_label(bars)
plt.show()
```
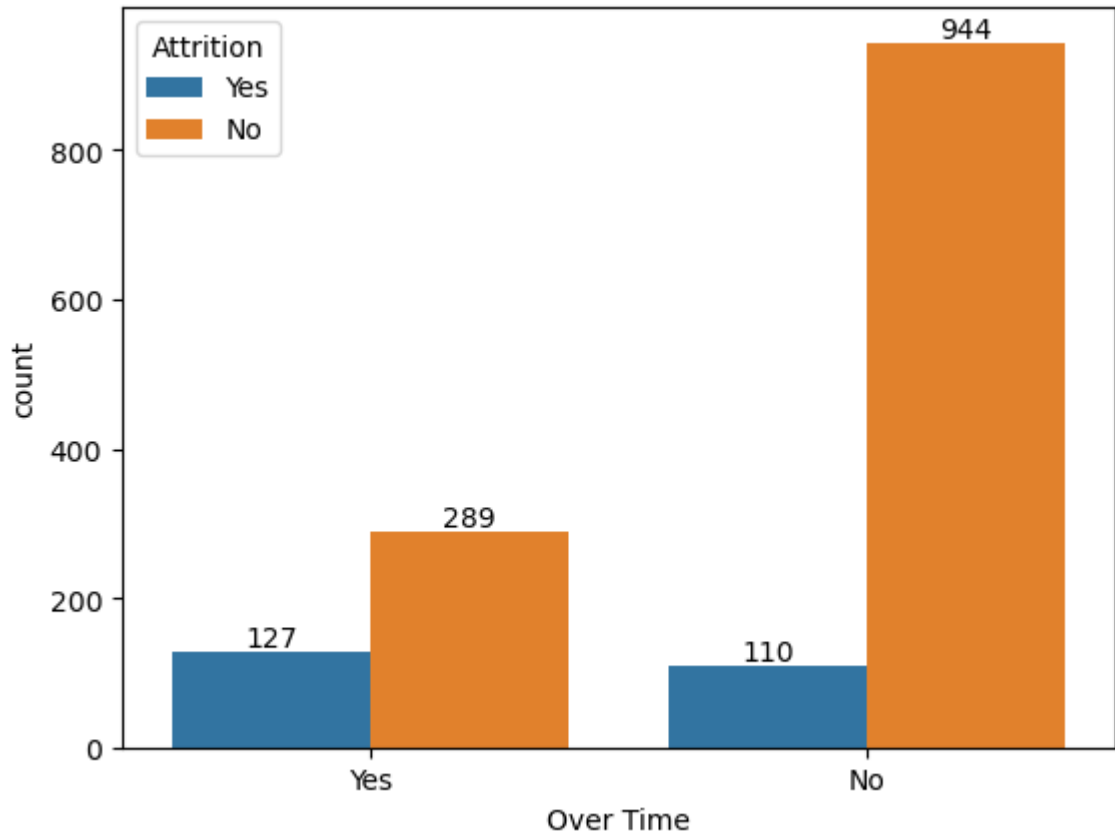

observation:-
- As far as 'Attrition yes' there is a minor difference between the employees who are doing overtime and those who are not doing overtime.
- so we can say that the overtime feature does not much affect attrition.
- But we can conclude that most of employees are not doing overtime.

6. IMPACT OF JOB ROLE ON ATTRITION

In [82]:
```python
plt.figure(figsize=(15,6))
ax1=sns.countplot(data=df, x='Job Role', hue='Attrition')
plt.title('Distribution of Attrition by Job Role')
plt.xlabel('Job Role')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.legend(title='Attrition')
for bars in ax1.containers:
    ax1.bar_label(bars)
plt.show()
```
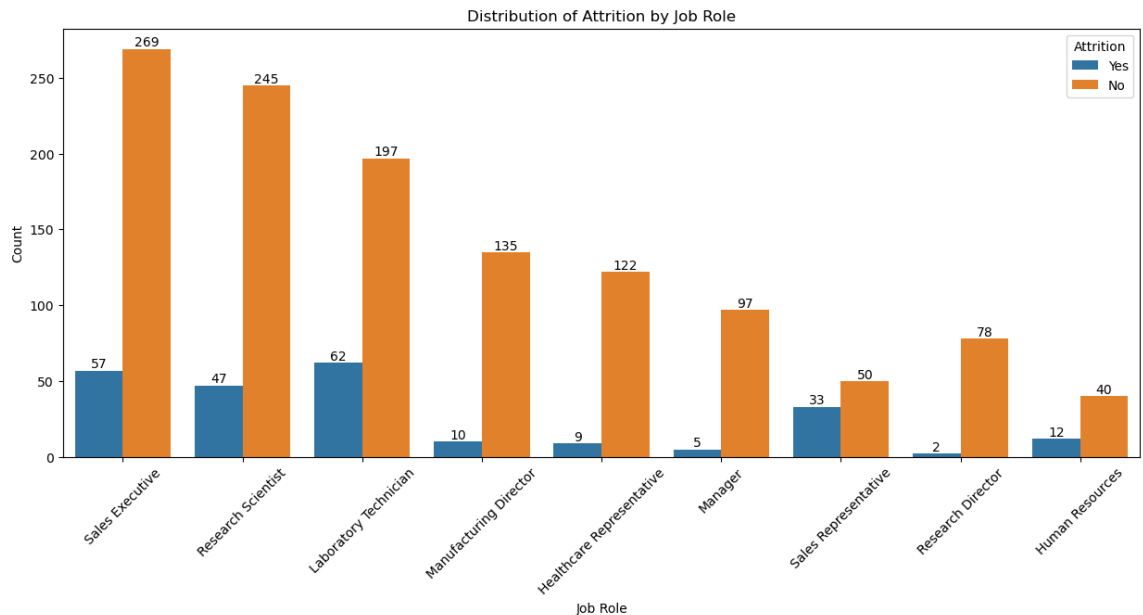


```
observation:-
- In this graph shows laboratory technician(62 employees) , sales
executive(57 employees) and research scientist(47 employees) are top 3 job
roles in which employees have their 'Attrition yes'
-Apart from this it can also seen that there are more number of employees
in sales executive job role
- There are less number of Research director(2 employees) who leave
company
```

ANALYSIS ON CONTINUOS DATA W.R.T TARGET COLUMN

1. IMPACT OF AGE ON ATTRITION

In [91]:
```python
sns.histplot(data=df, x='Age', hue='Attrition')
plt.xticks(range(15,70,5))
plt.show()
```
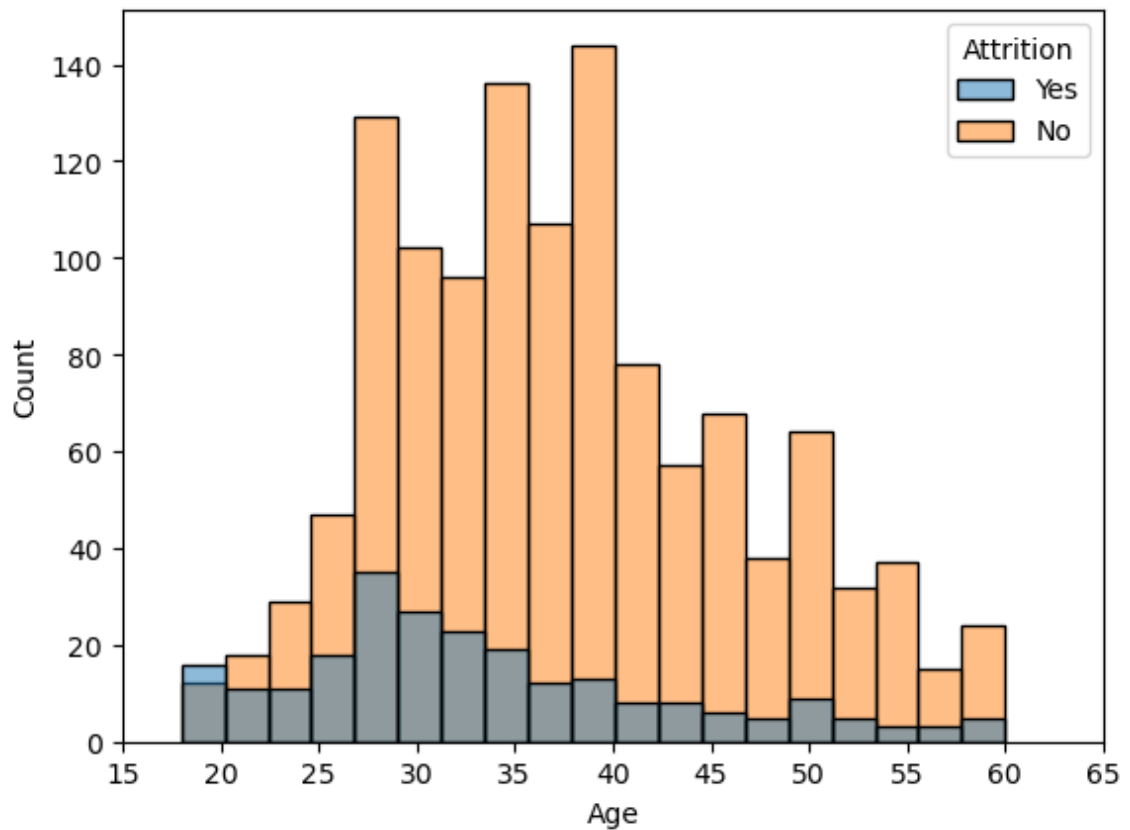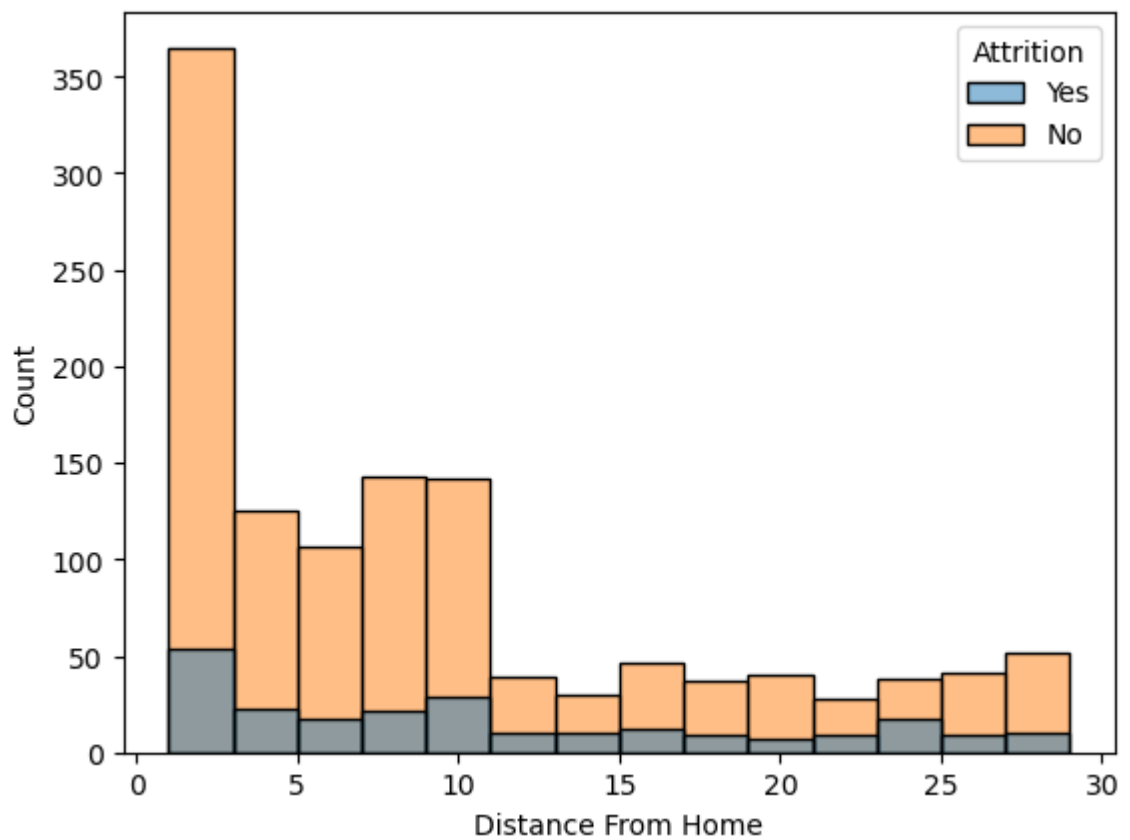


```
observation:-
- In the histplot shows employess in range of 20 to 35 are more likely to
leave the job.
- After 40 age, the distribution tells us that 'Higher the age lesser will
be attrition'
```

2. IMPACT OF DISTANCE FROM HOME AND ATTRITION

In [93]:
```python
sns.histplot(data=df, x='Distance From Home', hue='Attrition')
plt.show()
```
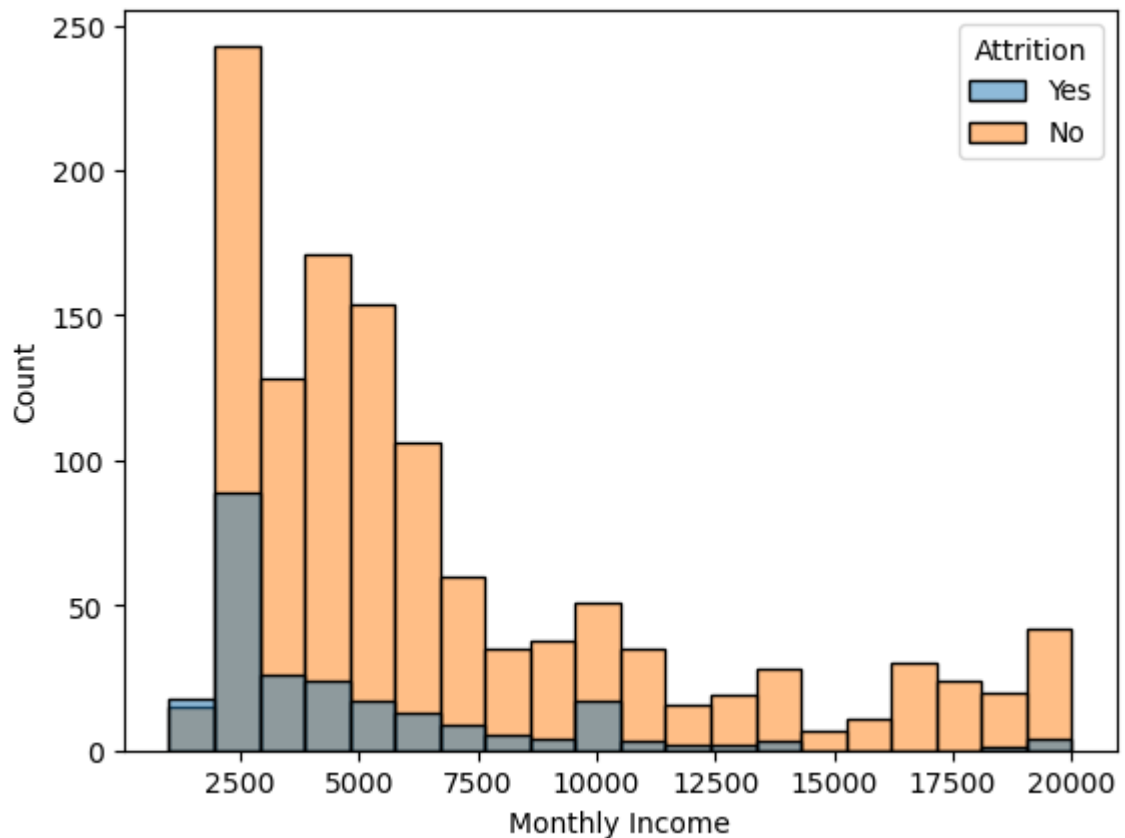

Observation:-
- In this histplot shows 0 to 10 km range more number of employees are working.
- In 0 to 5 km range distance employees are more likely to leave the job

3. HOW MONTHLY INCOME GIVES TRENDS W.R.T ATTRITION

In [94]:
```python
sns.histplot(data=df, x='Monthly Income', hue='Attrition')
plt.show
```

Out[94]: `<function matplotlib.pyplot.show(close=None, block=None)>`


```
Observation:-
- In this histplot shows employees who have income 2500 are more likely to
qquit their job, because 2500 is the least range of income .
- That means Lesser the income attrition is High and Higher the monthly
income less attrition
```

4. IMPACT OF NUMBER OF COMPANIES WORKED ON ATTRITION

In [95]:
```python
sns.histplot(data=df, x='Num Companies Worked', hue='Attrition')
plt.show()
```
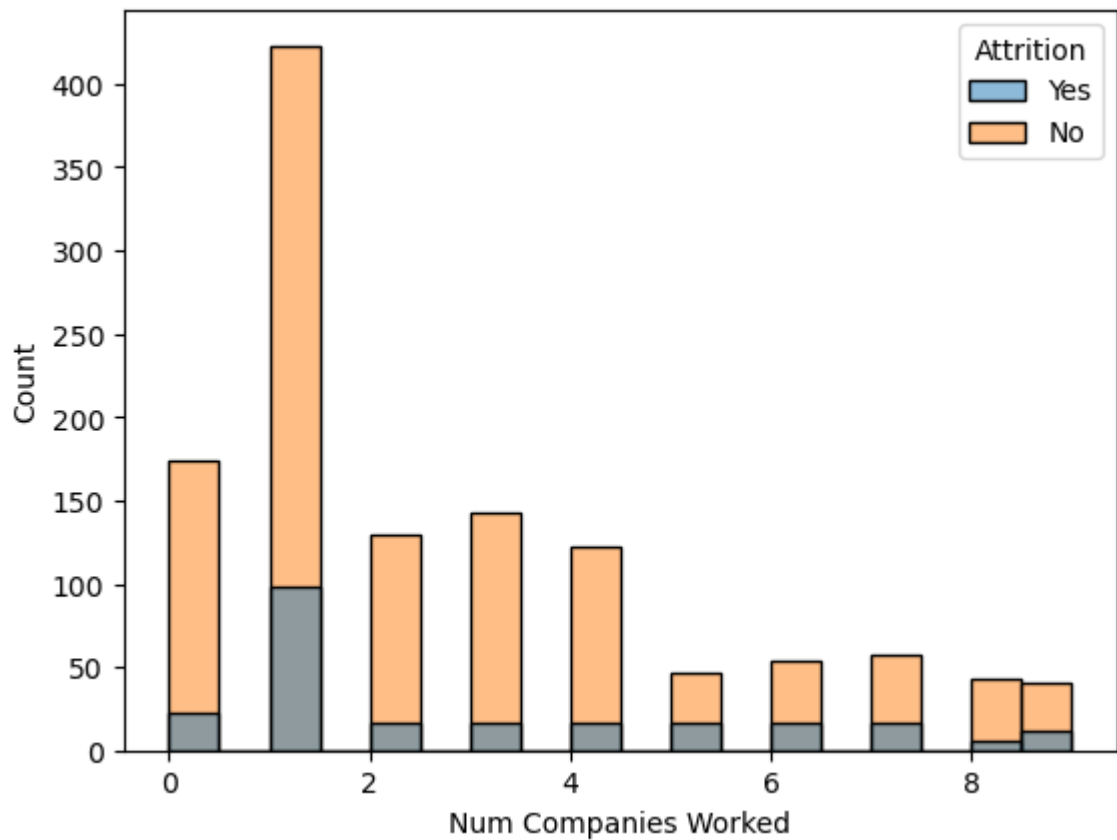

In [ ]:
```
observation:-
- In this hisplot shows number of past company workde 1 are more attrition |
- Thats way it is not majorly affect Attrition
```
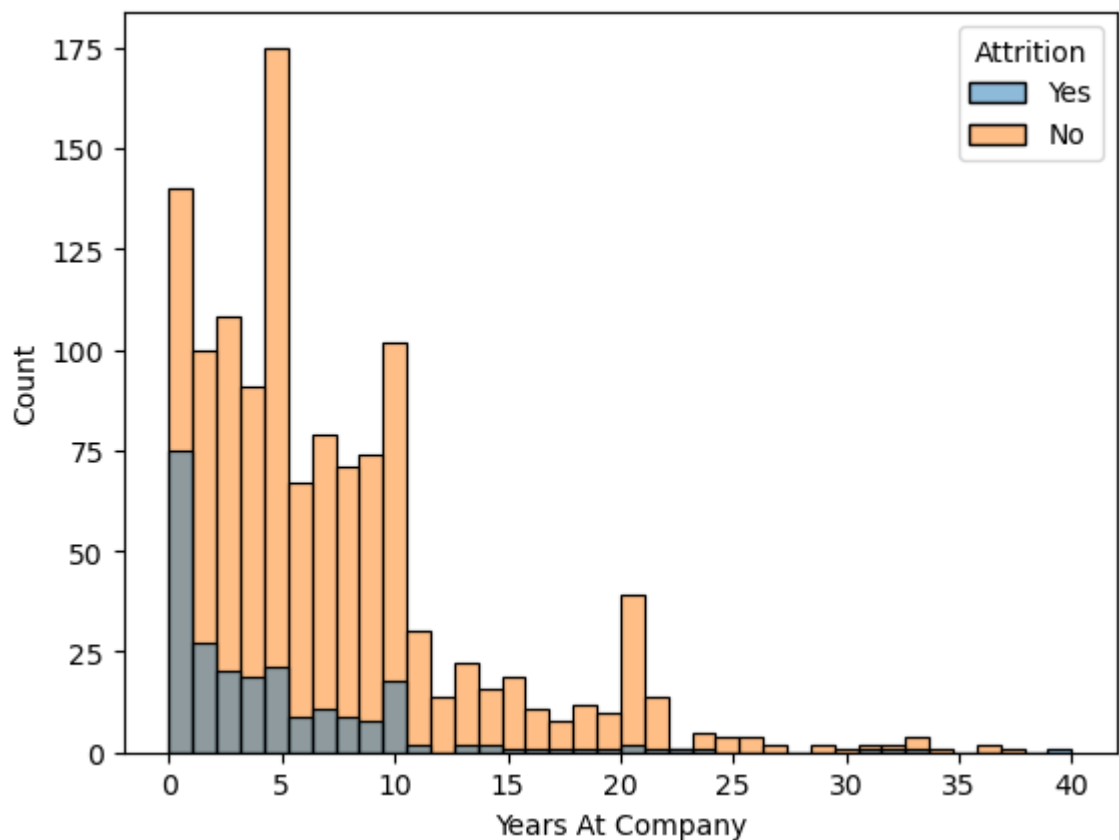
5. IMPACT OF SALARY HIKE ON ATTRITION

In [96]:
```python
sns.histplot(data=df, x='Percent Salary Hike', hue='Attrition')
plt.show()
```
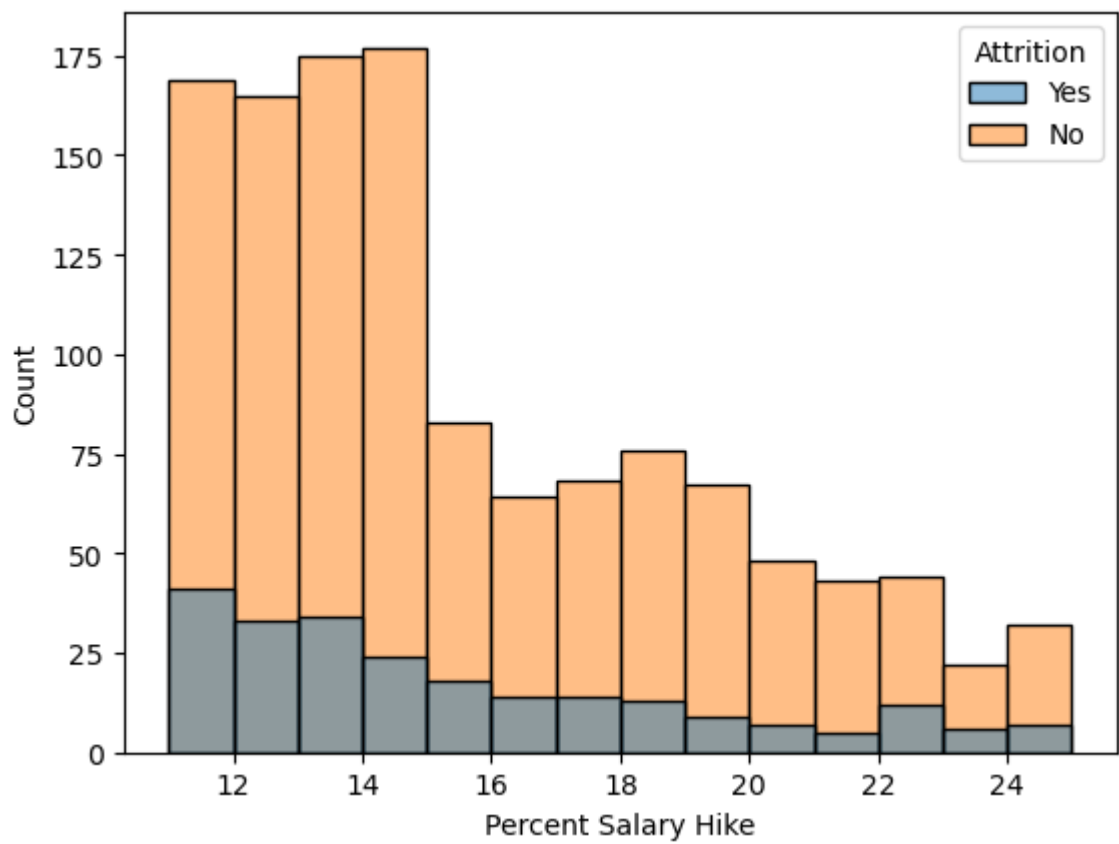

In [ ]:
```
observation:-
- In the Histplot shows Higher the percentage salary hike ,lesser will be tl
```

6. IMPACT OF YEARS AT THE COMPANY ON ATTRITION

In [97]:
```python
sns.histplot(data=df, x='Years At Company', hue='Attrition')
plt.show()
```



```
Observation:-
- Among freshers (employees with 0 years of experience), more than half of
them have left the company ("attrition yes").
- Employees who have been with the company for 0 to 10 years are also
likely to leave their jobs
```

ANALYSIS ON DISCRETE DATA W.R.T TARGET COLUMN

1. IMPACT OF JOB SATISFACTION ON ATTRITION

In [122]:
```python
ax10=sns.countplot(data=df, x='Job Satisfaction', hue='Attrition')
for bars in ax10.containers:
    ax10.bar_label(bars)
plt.show()
```
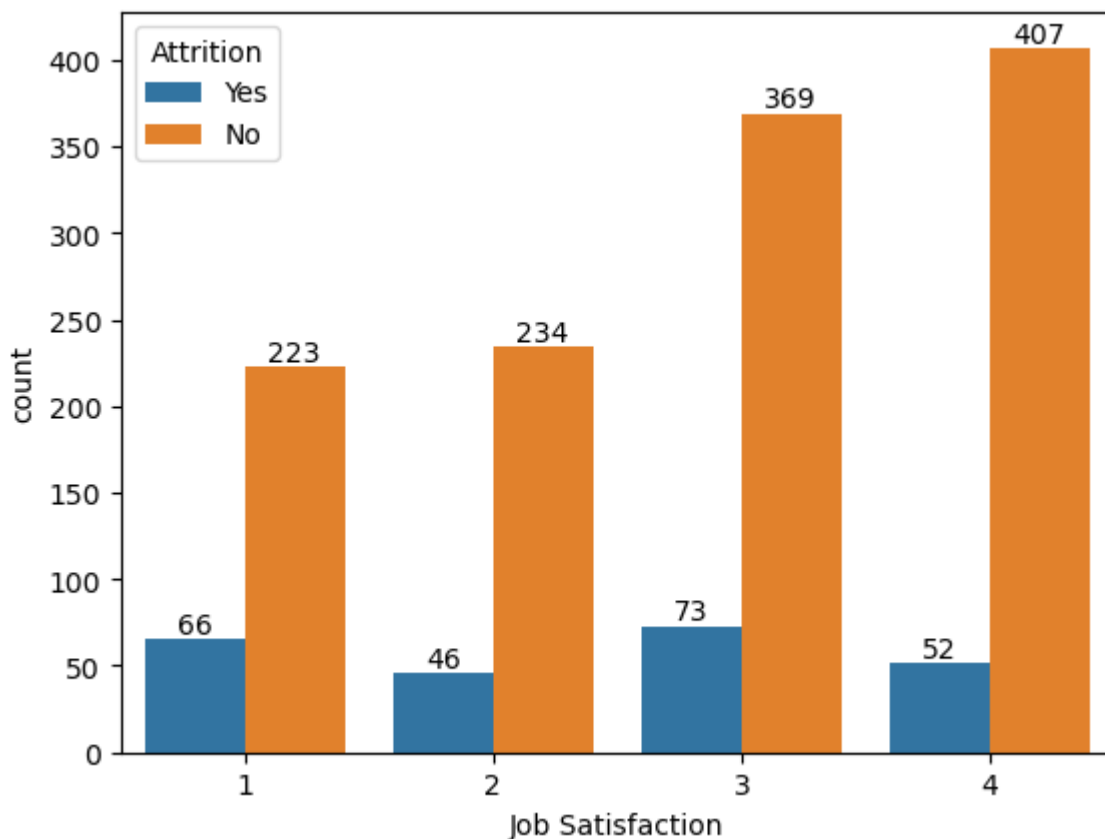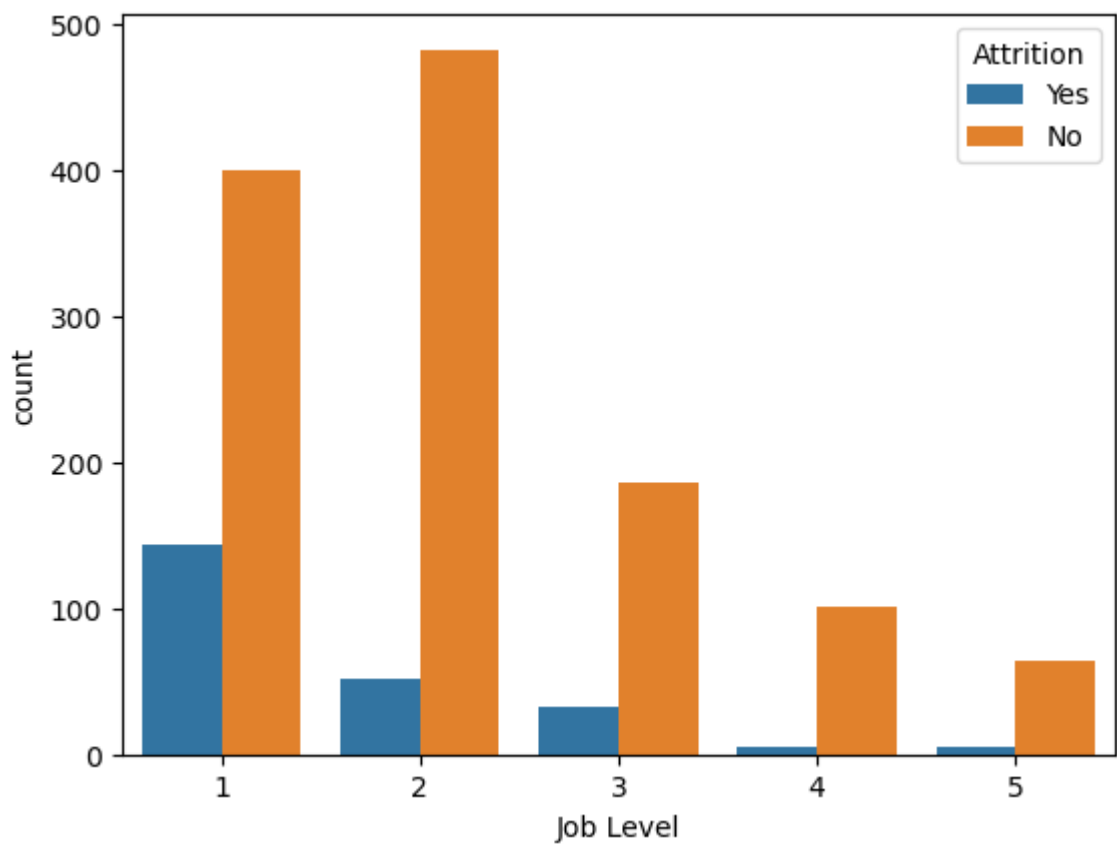


```
observation:-
- In this graph shows job satisfaction rating3 and 1 'attration yes' is
more
- Increase the rate of job satisfaction gives me to increase 'attrition
no'
```
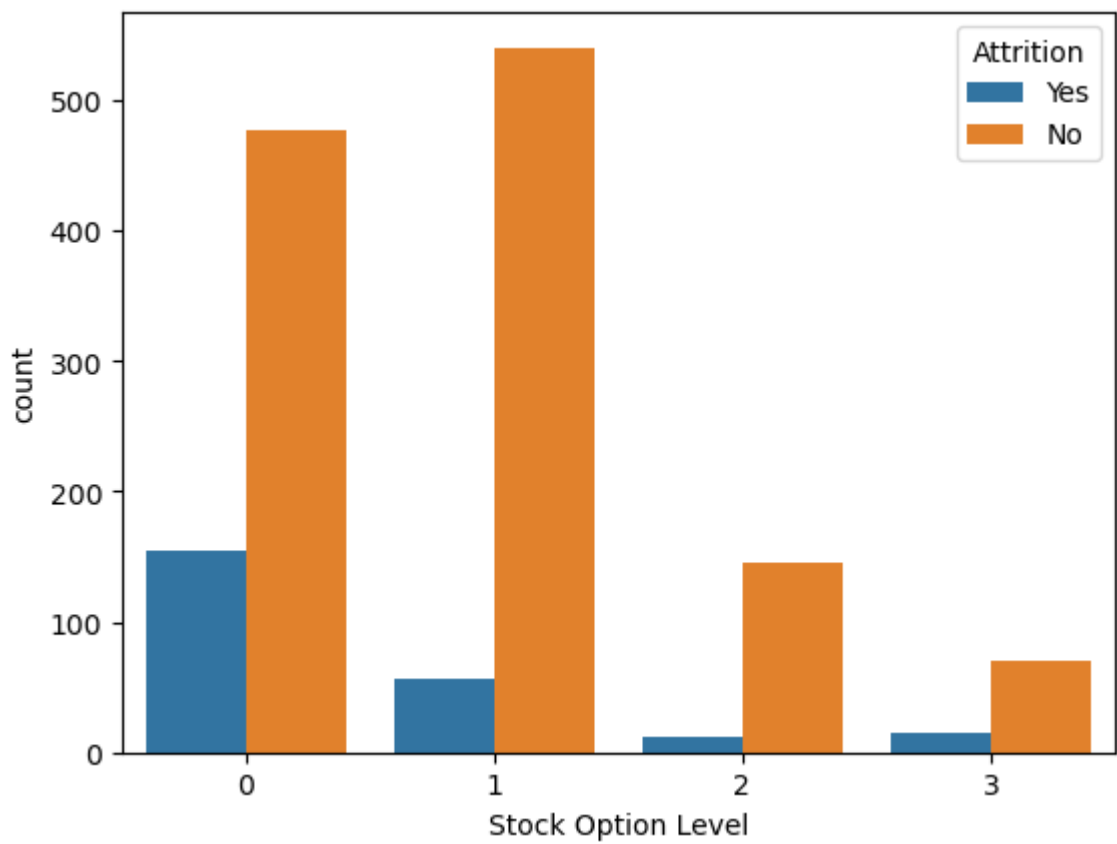
2. IMPACT OF JOB LEVEL ON ATTRITION

In [123]:
```python
sns.countplot(data=df, x='Job Level', hue='Attrition')
plt.show()
```


observation:- Increase in job level, Decrease in chance of leaving the company for Employees
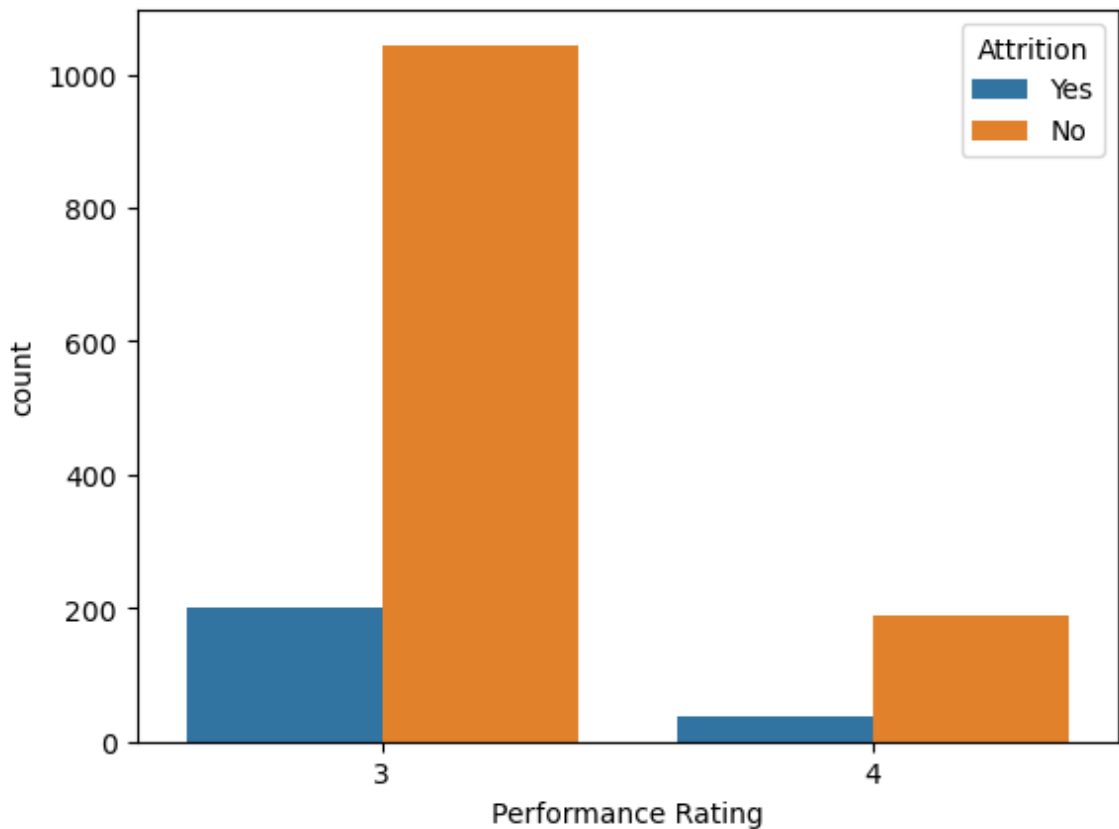
3. IMPACT OF STOCK OPTION LEVEL ON ATTRITION

In [124]:
```python
sns.countplot(data=df, x='Stock Option Level', hue='Attrition')
plt.show()
```


observation:- For the employees who's not having stock option are likely to quit

4. IMPACT OF PERFORMANCE RATING ON ATTRITION

In [125]:
```python
sns.countplot(data=df, x='Performance Rating', hue='Attrition')
plt.show()
```



```
observation:-
- In this countplot shows performance rating range lies on 3 -4
- However performance rating of employee 3 have more likely to quit or we
can say that company wants to fire the employees.
```

IDENTITY HOTSPOTS

This are top 10 factors which are affecting attrition "Yes"(Employees wants to leave the company)

- Impact of years at the company on attrition (Major Factors of affrecting attrition (employees want to leave the company))
- Impact of Monthly Income on attrition (Major Factors of affecting attrition (employees want to leave the company)
- Impact of salary hike on attrition (Major Factors of affecting attrition (employees want to leave the company)
- Impact of age on attrition
- Impact of performance rating on attrition
- Impact of stock level option on attrition
- Impact of job satisfaction on attrition
- Impact of job role on attrition
- Impact of Department on attrition

- Impact of Business Travel on attrition

Explore Relationships

Relationship 2 major factors of attrition monthly income and years at company

In [130]:
```
sns.scatterplot(data=df, x='Monthly Income', y='Years At Company', hue='Att
plt.show()
```
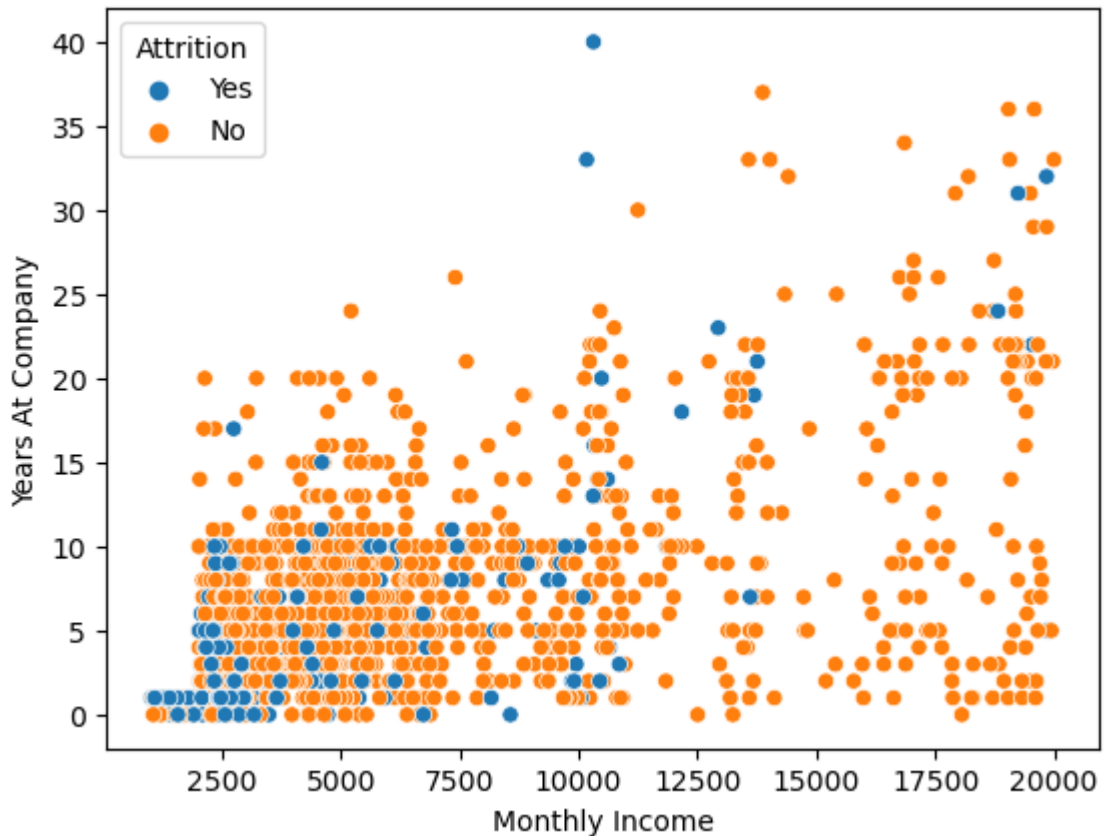
observation:- In this relationship in the scatterplot suggests that
fresher whos experience 0-5 yeas and monthly income 2500 are more likely
to leave jobs.

Conclusion and Recommendation:

- Invest in Employee Development: Implement comprehensive training and development programs to nurture employees' skills and facilitate career advancement opportunities.
- Enhance Compensation and Benefits: Review and adjust compensation packages to remain competitive in the market and ensure employees feel valued for their contributions.
- Feedback Mechanisms: Establish regular feedback mechanisms to solicit input from fresher employees regarding their experiences, challenges, and suggestions for improvement.
- Career Path Development: Create clear and transparent career paths for fresher employees, outlining opportunities for growth, advancement, and skill development.

- Salary Adjustment: Consider adjusting the salary structure for fresher employees to ensure it aligns with industry standards and reflects the value of their contributions. This can help improve employee satisfaction and reduce turnover rates among this demographic.

In [ ]:

- Salary Adjustment: Consider adjusting the salary structure for fresher employees to ensure it aligns with industry standards and reflects the value of their contributions. This can help improve employee satisfaction and reduce turnover rates among this demographic.