

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('D:\SPN PERS\DATA SCIENCE\CSV FILE\WA_Fn-UseC_-HR-
Employee-Attrition.csv')
```

```
df.head(3)
```

	Age	Attrition	BusinessTravel	DailyRate	Department
0	41	Yes	Travel_Rarely	1102	Sales
1	49	No	Travel_Frequently	279	Research & Development
2	37	Yes	Travel_Rarely	1373	Research & Development

EmployeeNumber	DistanceFromHome	Education	EducationField	EmployeeCount
0	1	2	Life Sciences	1
1				
1	8	1	Life Sciences	1
2				
2	2	2	Other	1
4				

	RelationshipSatisfaction	StandardHours	StockOptionLevel
0	1	80	0
1	4	80	1
2	2	80	0

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

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

```
[3 rows x 35 columns]
```

```
df.columns
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate',
      'Department',
      'DistanceFromHome', 'Education', 'EducationField',
      'EmployeeCount',
      'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender',
      'HourlyRate',
      'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
      'MaritalStatus', 'MonthlyIncome', 'MonthlyRate',
      'NumCompaniesWorked',
      'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
      'RelationshipSatisfaction', 'StandardHours',
      'StockOptionLevel',
      'TotalWorkingYears', 'TrainingTimesLastYear',
      'WorkLifeBalance',
      'YearsAtCompany', 'YearsInCurrentRole',
      'YearsSinceLastPromotion',
      'YearsWithCurrManager'],
      dtype='object')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1470 entries, 0 to 1469
```

```
Data columns (total 35 columns):
```

#	Column	Non-Null Count	Dtype
0	Age	1470 non-null	int64
1	Attrition	1470 non-null	object
2	BusinessTravel	1470 non-null	object
3	DailyRate	1470 non-null	int64
4	Department	1470 non-null	object
5	DistanceFromHome	1470 non-null	int64
6	Education	1470 non-null	int64
7	EducationField	1470 non-null	object
8	EmployeeCount	1470 non-null	int64
9	EmployeeNumber	1470 non-null	int64
10	EnvironmentSatisfaction	1470 non-null	int64
11	Gender	1470 non-null	object
12	HourlyRate	1470 non-null	int64
13	JobInvolvement	1470 non-null	int64
14	JobLevel	1470 non-null	int64
15	JobRole	1470 non-null	object
16	JobSatisfaction	1470 non-null	int64
17	MaritalStatus	1470 non-null	object
18	MonthlyIncome	1470 non-null	int64
19	MonthlyRate	1470 non-null	int64
20	NumCompaniesWorked	1470 non-null	int64
21	Over18	1470 non-null	object
22	OverTime	1470 non-null	object
23	PercentSalaryHike	1470 non-null	int64

24	PerformanceRating	1470	non-null	int64
25	RelationshipSatisfaction	1470	non-null	int64
26	StandardHours	1470	non-null	int64
27	StockOptionLevel	1470	non-null	int64
28	TotalWorkingYears	1470	non-null	int64
29	TrainingTimesLastYear	1470	non-null	int64
30	WorkLifeBalance	1470	non-null	int64
31	YearsAtCompany	1470	non-null	int64
32	YearsInCurrentRole	1470	non-null	int64
33	YearsSinceLastPromotion	1470	non-null	int64
34	YearsWithCurrManager	1470	non-null	int64

dtypes: int64(26), object(9)

memory usage: 402.1+ KB

df.describe()

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

	EmployeeNumber	EnvironmentSatisfaction	HourlyRate
JobInvolvement \			
count	1470.000000	1470.000000	1470.000000
1470.000000			
mean	1024.865306	2.721769	65.891156
2.729932			
std	602.024335	1.093082	20.329428
0.711561			
min	1.000000	1.000000	30.000000
1.000000			
25%	491.250000	2.000000	48.000000
2.000000			
50%	1020.500000	3.000000	66.000000
3.000000			
75%	1555.750000	4.000000	83.750000

3.000000

max 2068.000000 4.000000 100.000000

4.000000

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

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

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

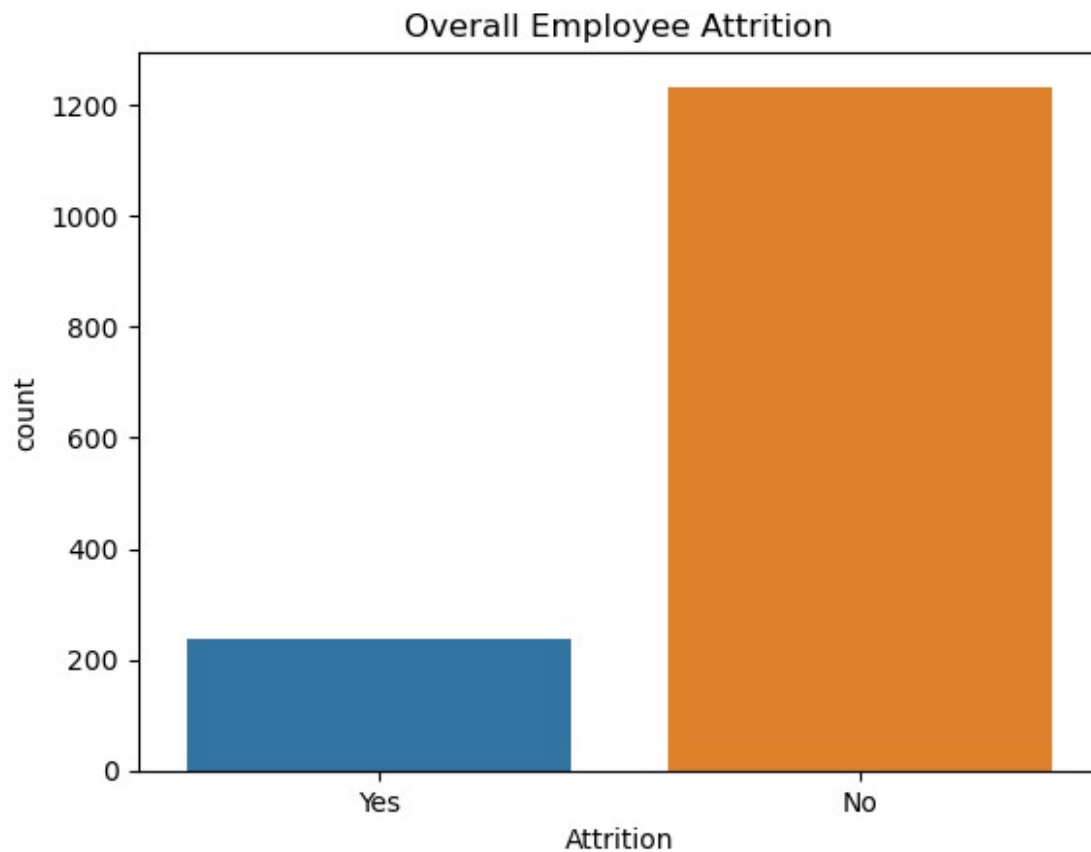
	YearsSinceLastPromotion	YearsWithCurrManager
count	1470.000000	1470.000000
mean	2.187755	4.123129
std	3.222430	3.568136
min	0.000000	0.000000
25%	0.000000	2.000000
50%	1.000000	3.000000
75%	3.000000	7.000000
max	15.000000	17.000000

[8 rows x 26 columns]

Recommended Visualizations in Jupyter Notebook

```
sns.countplot(x='Attrition', data=df)
plt.title('Overall Employee Attrition')
```

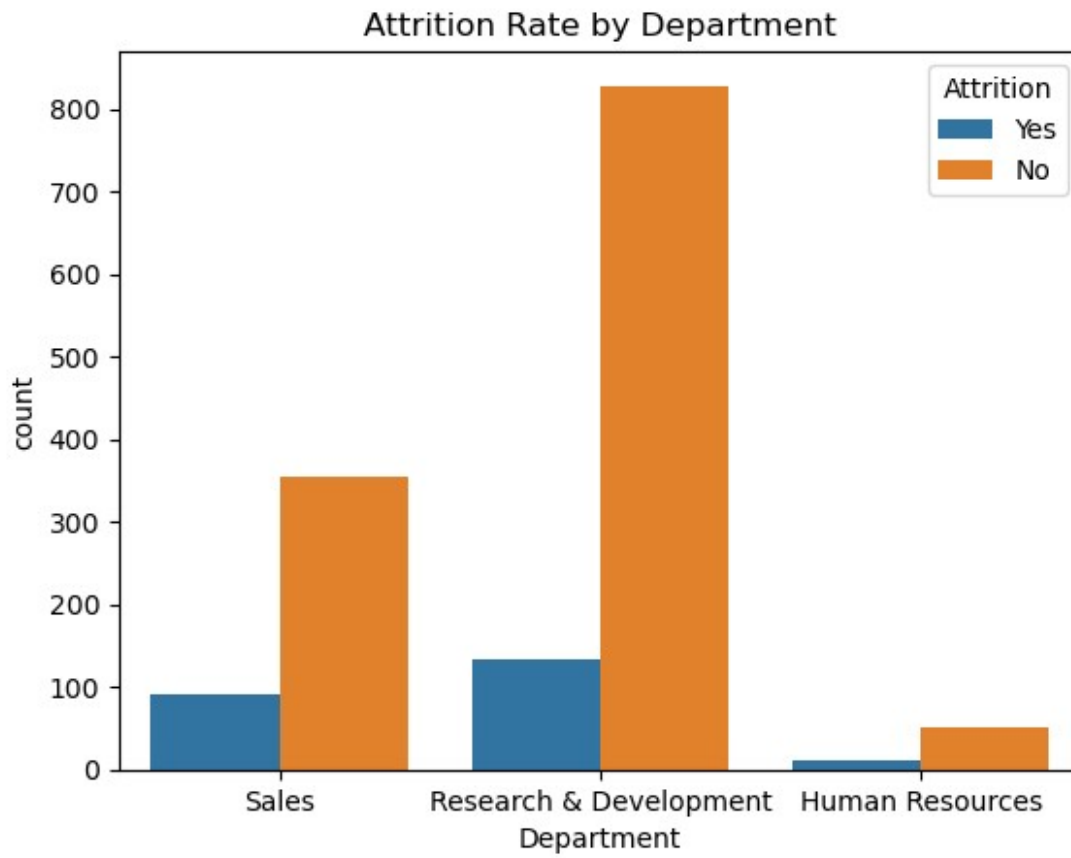
```
Text(0.5, 1.0, 'Overall Employee Attrition')
```



```
#2. Attrition by Department
```

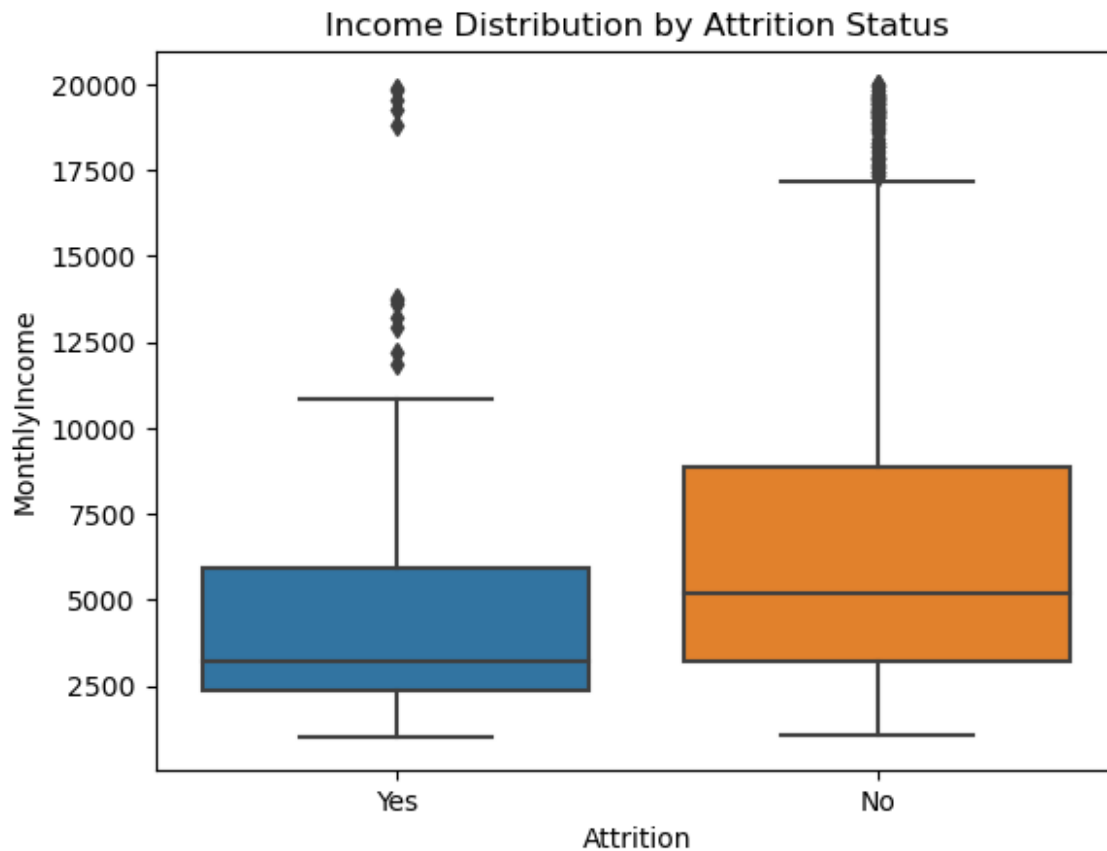
```
sns.countplot(x='Department', hue='Attrition', data=df)  
plt.title('Attrition Rate by Department')
```

```
Text(0.5, 1.0, 'Attrition Rate by Department')
```



#3. Monthly Income Distribution by Attrition

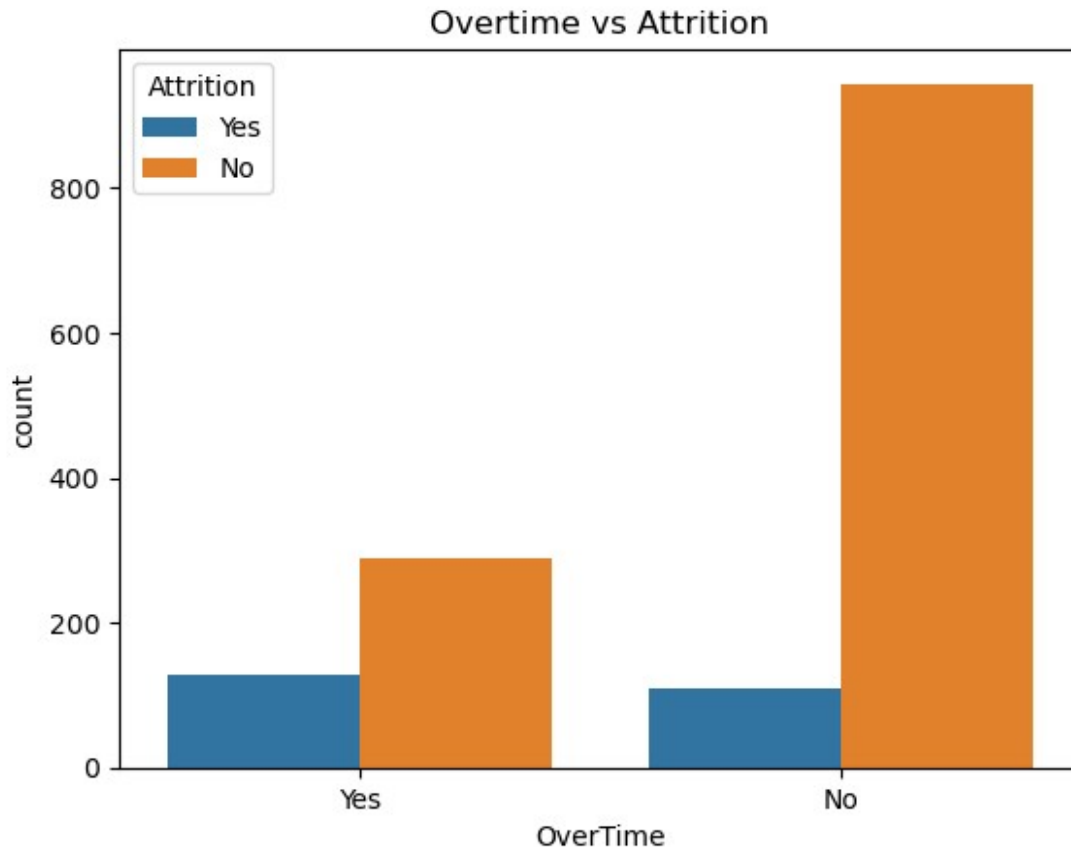
```
sns.boxplot(x='Attrition', y='MonthlyIncome', data=df)
plt.title('Income Distribution by Attrition Status')
Text(0.5, 1.0, 'Income Distribution by Attrition Status')
```



#4. *Attrition vs OverTime*

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

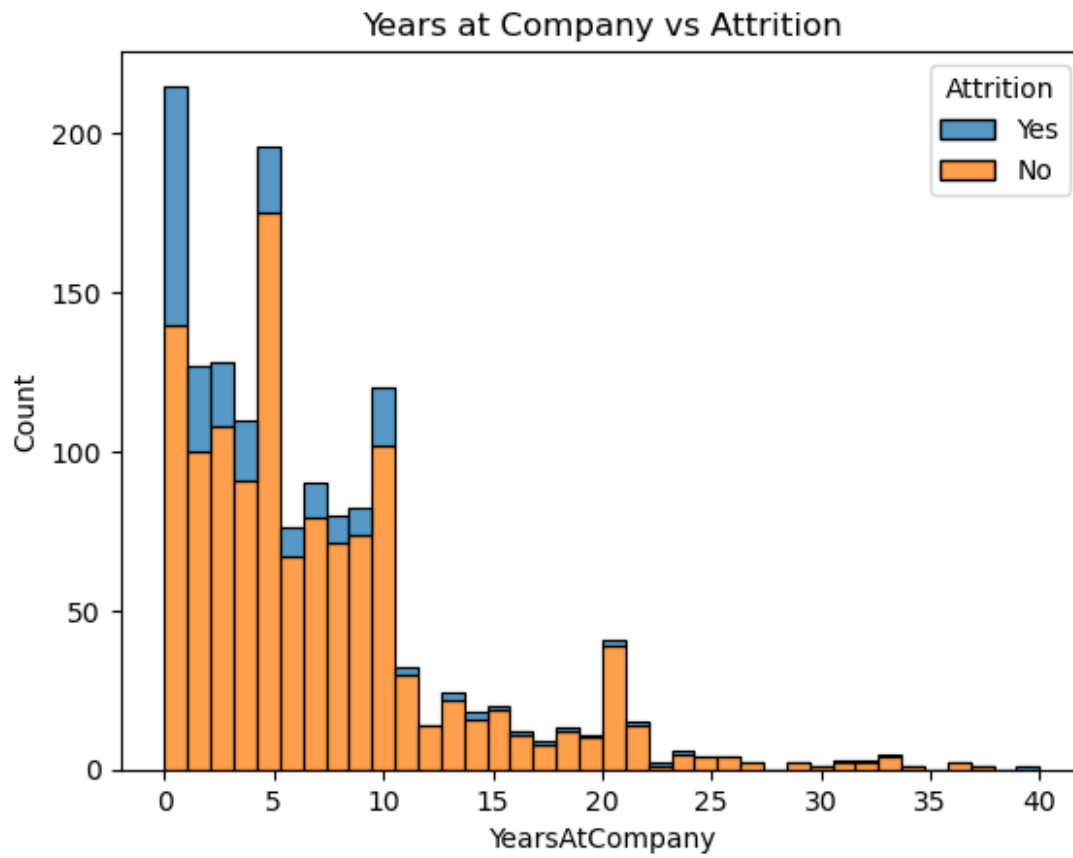
```
Text(0.5, 1.0, 'Overtime vs Attrition')
```



#5. Years at Company vs Attrition (Histogram)

```
sns.histplot(data=df, x='YearsAtCompany', hue='Attrition',  
multiple='stack')  
plt.title('Years at Company vs Attrition')
```

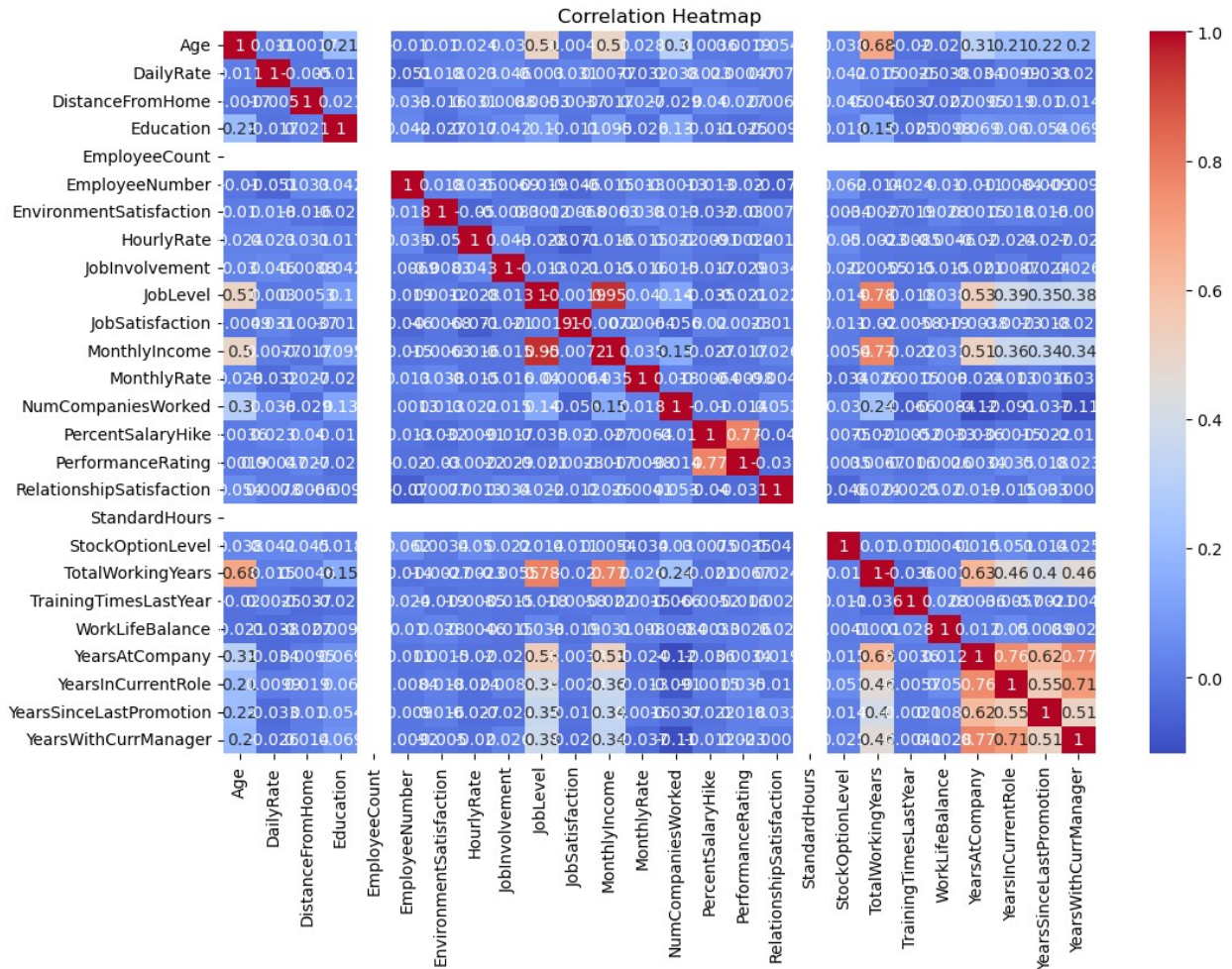
```
Text(0.5, 1.0, 'Years at Company vs Attrition')
```

#6. Correlation Heatmap

```
plt.figure(figsize=(12,8))  
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')  
plt.title('Correlation Heatmap')
```

```
Text(0.5, 1.0, 'Correlation Heatmap')
```



#7. Job Role vs Job Satisfaction

```
plt.figure(figsize=(12,6))
sns.boxplot(x='JobRole', y='JobSatisfaction', data=df)
plt.xticks(rotation=45)
plt.title('Job Satisfaction by Role')
```

```
Text(0.5, 1.0, 'Job Satisfaction by Role')
```

