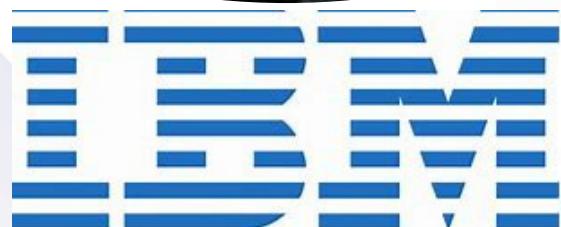


IBM HR Analytics Employee Attrition & Performance

PRESENTED BY **Sana Khan**
Data Science intern at Unified Mentor

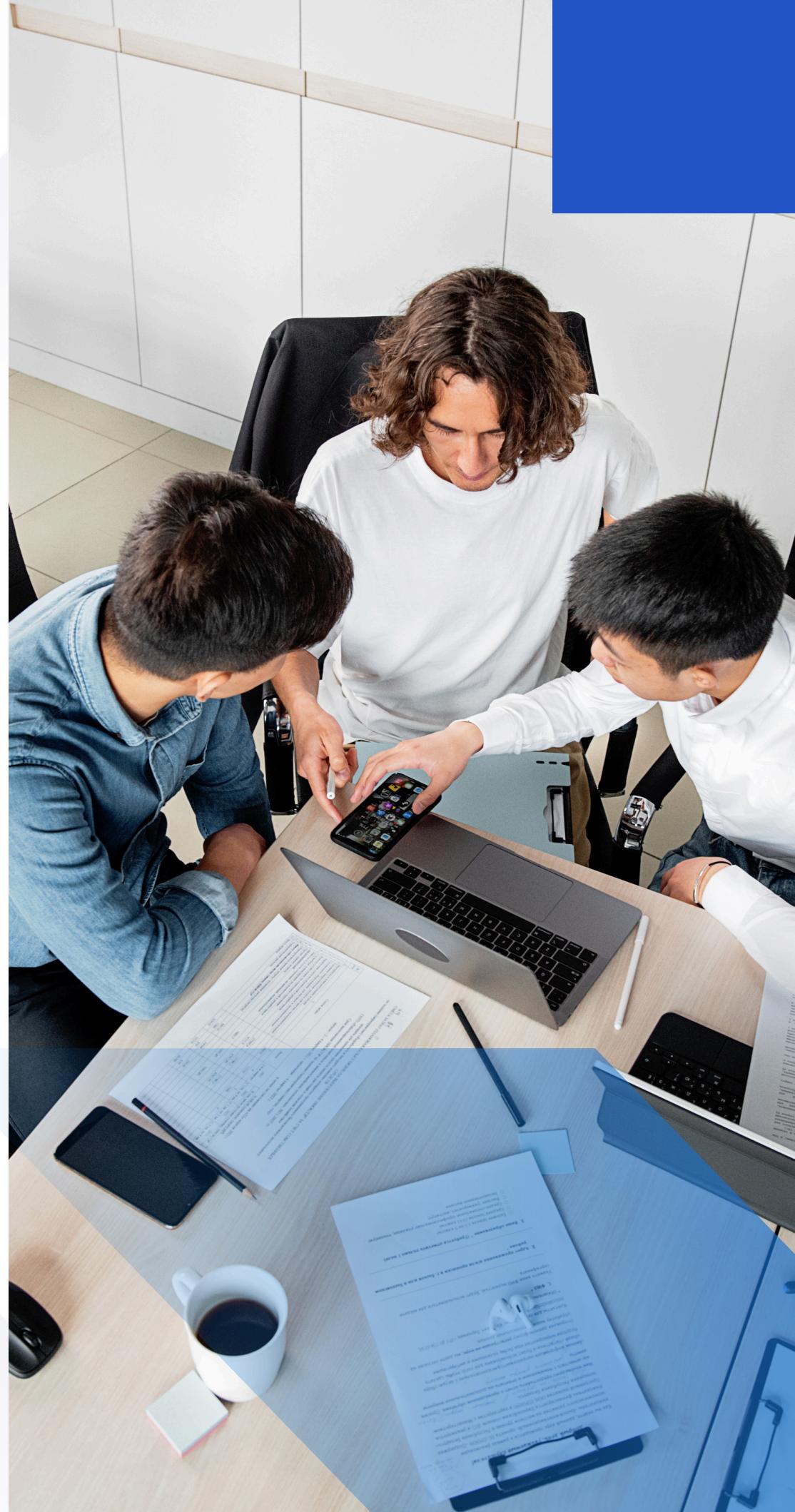




About The Company

IBM (International Business Machines Corporation) is a global technology and consulting company headquartered in Armonk, New York. Founded in 1911, IBM is a leader in innovation, offering solutions in cloud computing, AI, quantum computing, and data analytics across industries

 www.ibm.com 



Project Objectives

- Understand employee attrition patterns.
- Explore demographics affecting attrition.
- Analyze job satisfaction, overtime, salary, and other factors.
- Build predictive models for attrition forecasting.

Project Aim

To analyze employee attrition data to uncover factors influencing employee turnover and provide actionable insights to improve employee retention.

Problem Statements

What is the overall employee attrition rate?

Which job roles, age groups, and departments have higher attrition?

Does overtime influence attrition rates?

How does monthly income relate to attrition?

What other factors are strong indicators of employee turnover?

Tools & Libraries Used

- Python
- pandas
- numpy
- matplotlib
- seaborn
- scikit-learn
- Jupyter Notebook

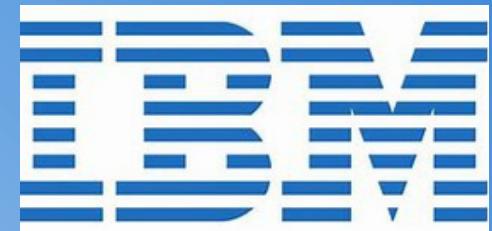


Dataset Overview

- Dataset mentor
- 1470 employee records.
- 35 features (Age, Gender, Education, Department, Monthly Income, Overtime, etc.).
- Target Variable: Attrition (Yes/No)



Data Cleaning



```
[2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
[3]: df = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition (1).csv')
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80

5 rows × 35 columns

Activate Windows
Go to Settings to activate Windows.

	action	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
	1	80	0	8	0	1	6	4	0	5
	4	80	1	10	3	3	10	7	1	7
	2	80	0	7	3	3	0	0	0	0
	3	80	0	8	3	3	8	7	3	0
	4	80	1	6	3	3	2	2	2	2

```
In [4]: print(df.shape)
```

```
[1]: print(df.shape)
print(df.info())
print(df.describe())
print(df.isnull().sum())
print(df.duplicated().sum())
```

(1470, 35)
<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

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

None

Age DailyRate DistanceFromHome Education

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement	JobLevel	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000	1470.000000
mean	36.923810	802.485714	9.192517	2.912925	1.0	2.761224	7.008163	4.229252	2.063946	...	2.712245	80.0	2.187755	4.123129	2.063946
std	9.135373	403.509100	8.106864	1.024165	0.0	0.706476	6.126525	3.623137	1.106940	...	1.081209	0.0	3.222430	3.568136	1.106940
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	0.000000	0.000000	1.000000	...	1.000000	80.0	0.000000	0.000000	1.000000
25%	30.000000	465.000000	2.000000	2.000000	1.0	1.000000	10.000000	3.000000	2.000000	...	2.000000	80.0	1.000000	2.000000	2.000000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1.000000	15.000000	3.000000	3.000000	...	7.000000	80.0	3.000000	6.000000	3.000000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1.000000	40.000000	6.000000	4.000000	...	14.000000	80.0	40.000000	6.000000	14.000000
max	60.000000	1499.000000	29.000000	5.000000	1.0	4.000000	40.000000	18.000000	4.000000	...	100.000000	80.0	40.000000	18.000000	100.000000

	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]

Data Cleaning

[8 rows x 26 columns]

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

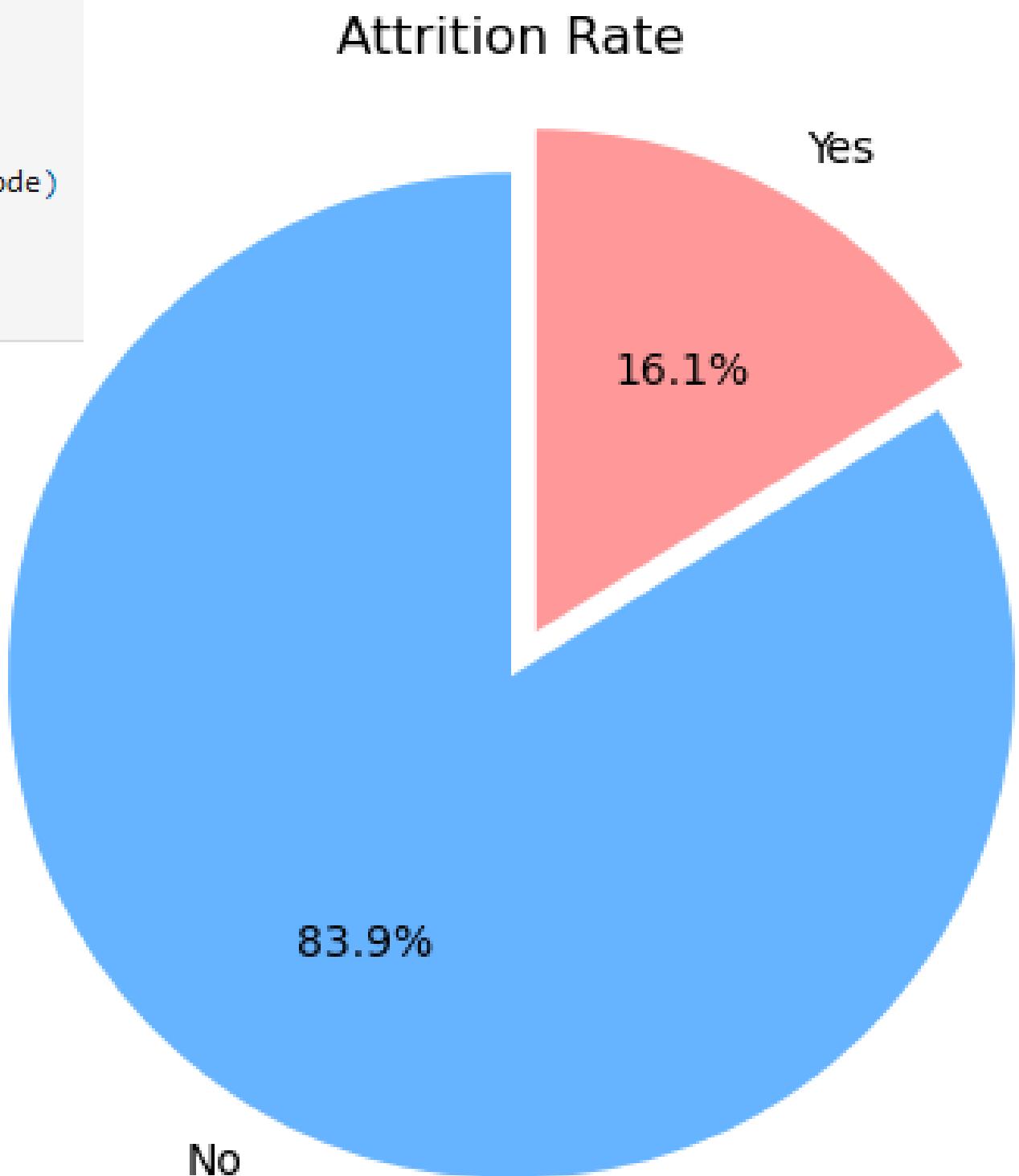
HourlyRate	0
JobInvolvement	0
JobLevel	0
JobRole	0
JobSatisfaction	0
MaritalStatus	0
MonthlyIncome	0
MonthlyRate	0
NumCompaniesWorked	0
Over18	0
Overtime	0
PercentSalaryHike	0
PerformanceRating	0
RelationshipSatisfaction	0
StandardHours	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	0
YearsInCurrentRole	0
YearsSinceLastPromotion	0
YearsWithCurrManager	0
dtype: int64	
0	

- No missing values detected.
- No duplicate entries found.
- Converted categorical data using label encoding.
- Prepared data for visualization and modeling.

Attrition Rate Pie Chart

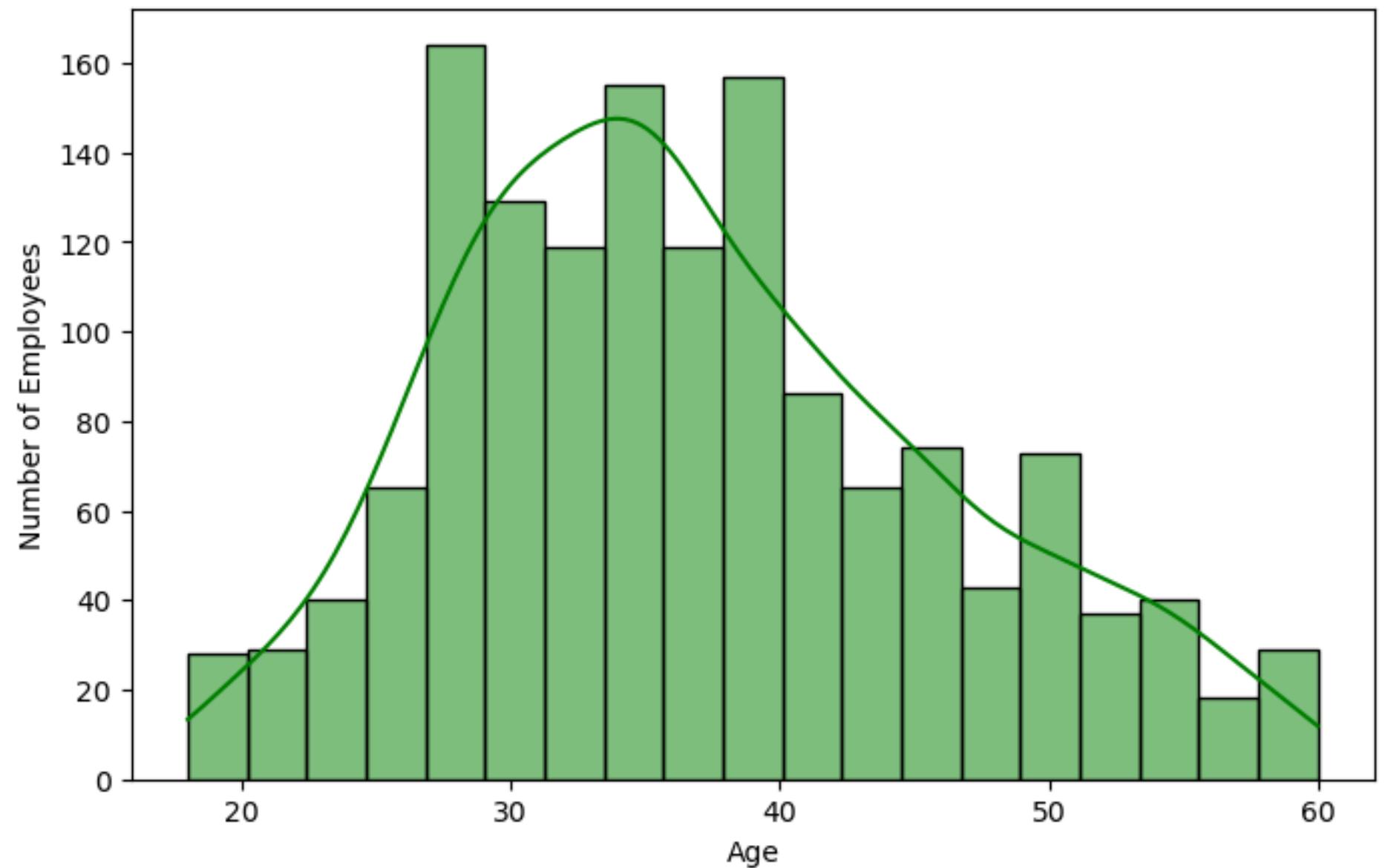
```
5]: labels = df['Attrition'].value_counts().index  
sizes = df['Attrition'].value_counts().values  
  
colors = ['#66b3ff', '#ff9999']  
explode = (0.1, 0)  
  
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=90, explode=explode)  
plt.title('Attrition Rate')  
plt.axis('equal')  
plt.show()
```

~16.1% of
employees left
the company.



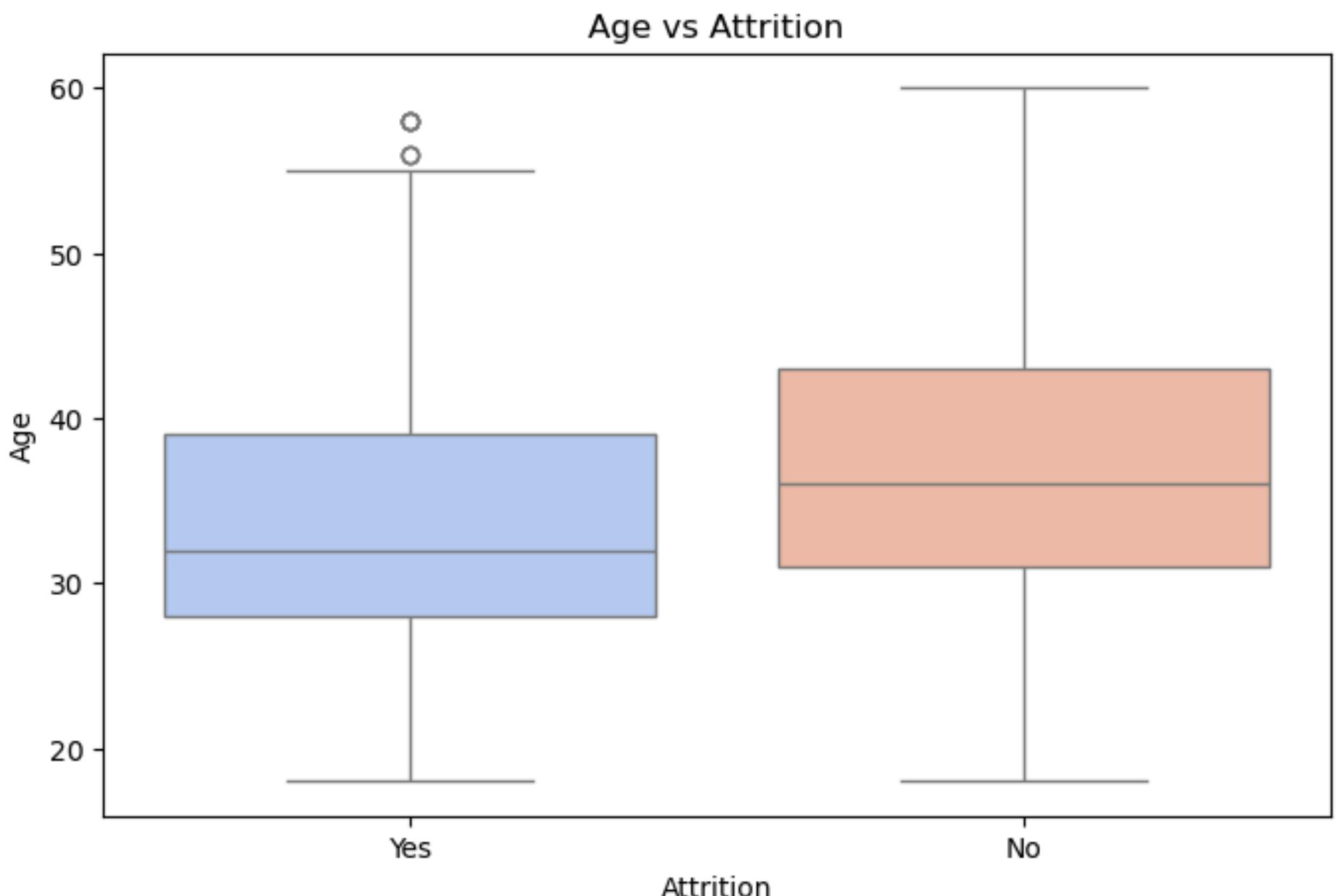
Age Distribution

Age Distribution



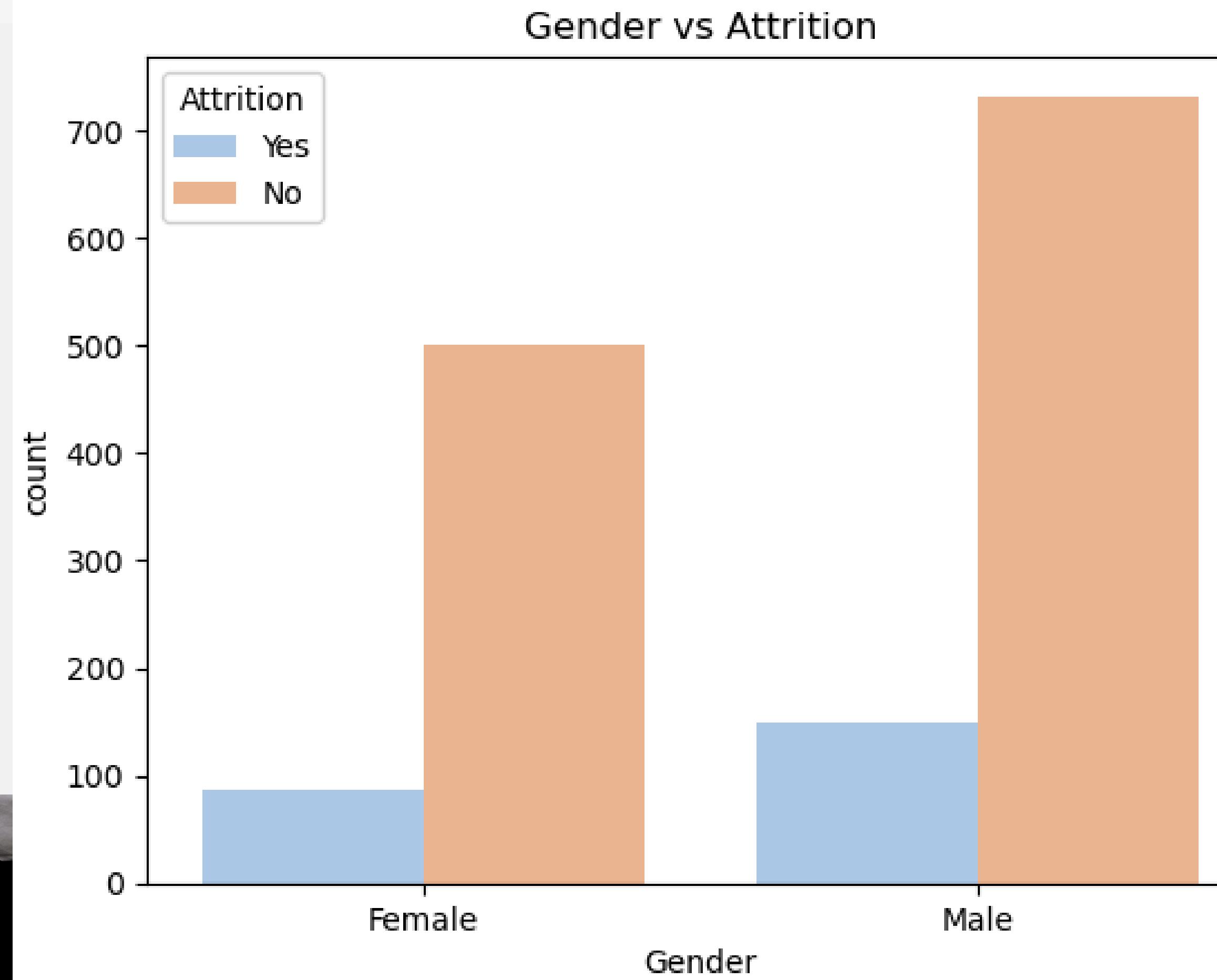
```
[]: plt.figure(figsize=(8,5))
sns.histplot(df['Age'], kde=True, color='green')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Number of Employees')
plt.show()
```

```
7]: plt.figure(figsize=(8,5))
sns.boxplot(x='Attrition', y='Age', data=df, palette='coolwarm')
plt.title('Age vs Attrition')
plt.show()
```



```
[8]: sns.countplot(x='Gender', hue='Attrition', data=df, palette='pastel')
plt.title('Gender vs Attrition')
plt.show()
```

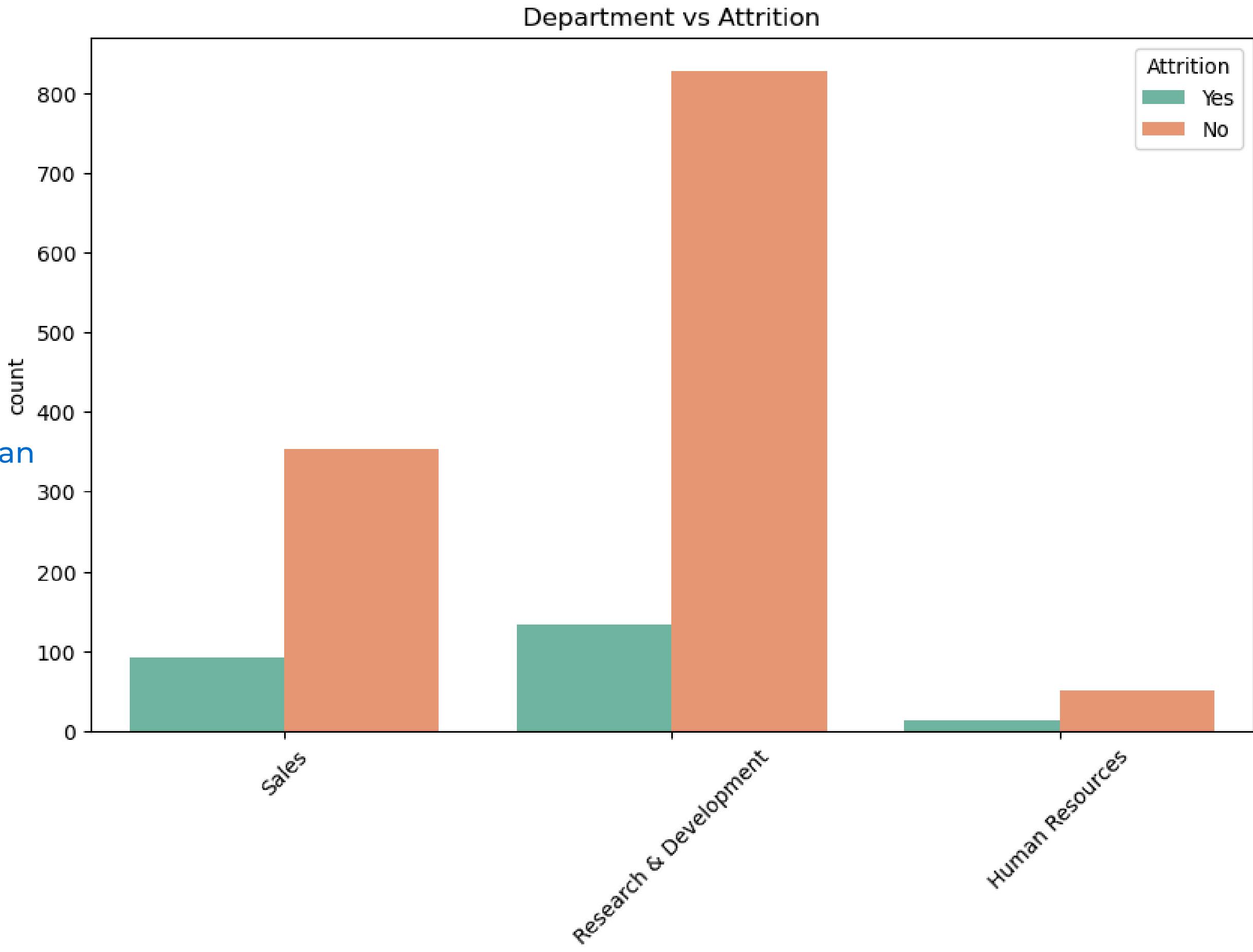
Gender vs Attrition



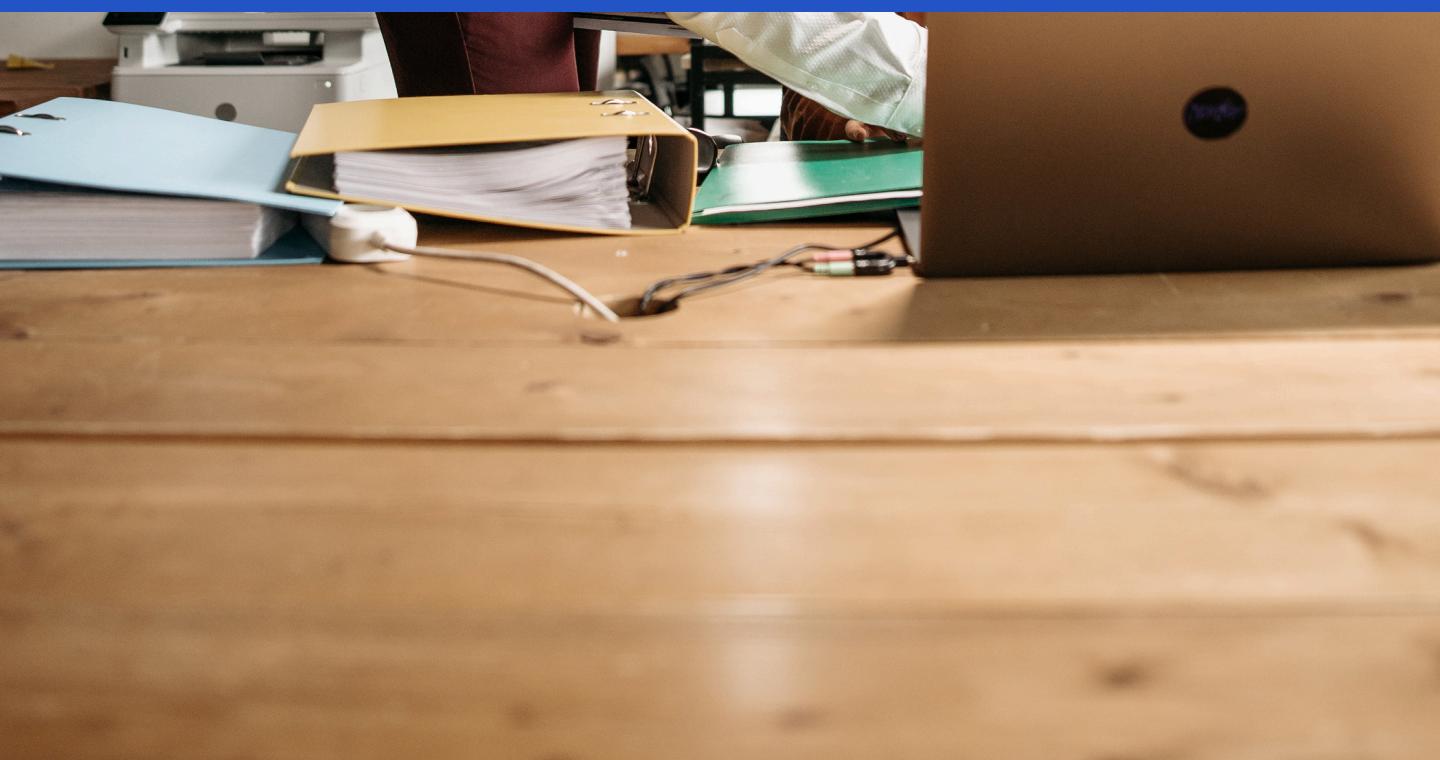
Department-wise Attrition

```
: plt.figure(figsize=(10,6))
sns.countplot(x='Department', hue='Attrition', data=df, palette='Set2')
plt.title('Department vs Attrition')
plt.xticks(rotation=45)
plt.show()
```

- Highest attrition from Sales and Human Resources departments.

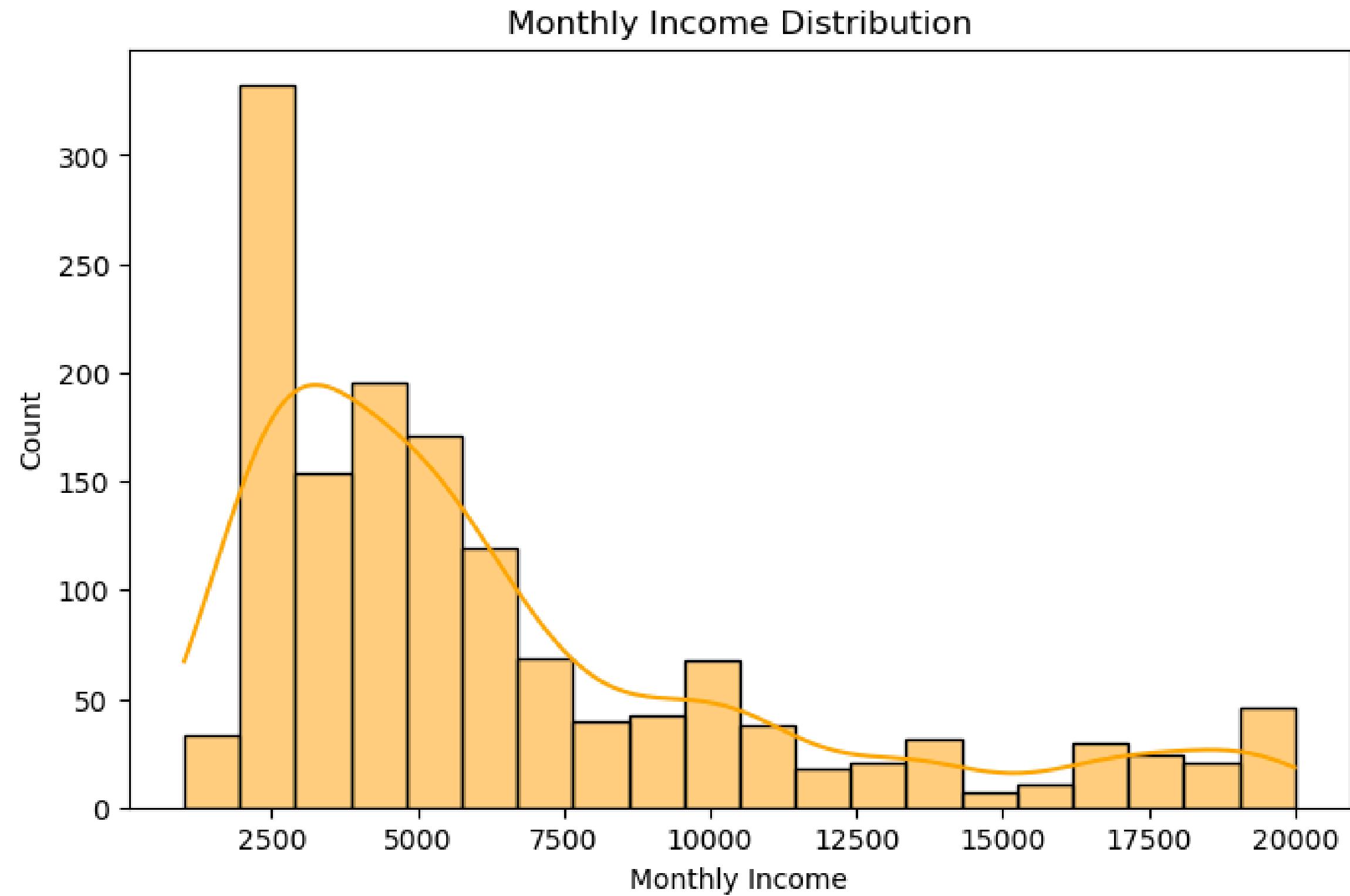


Monthly Income Distribution

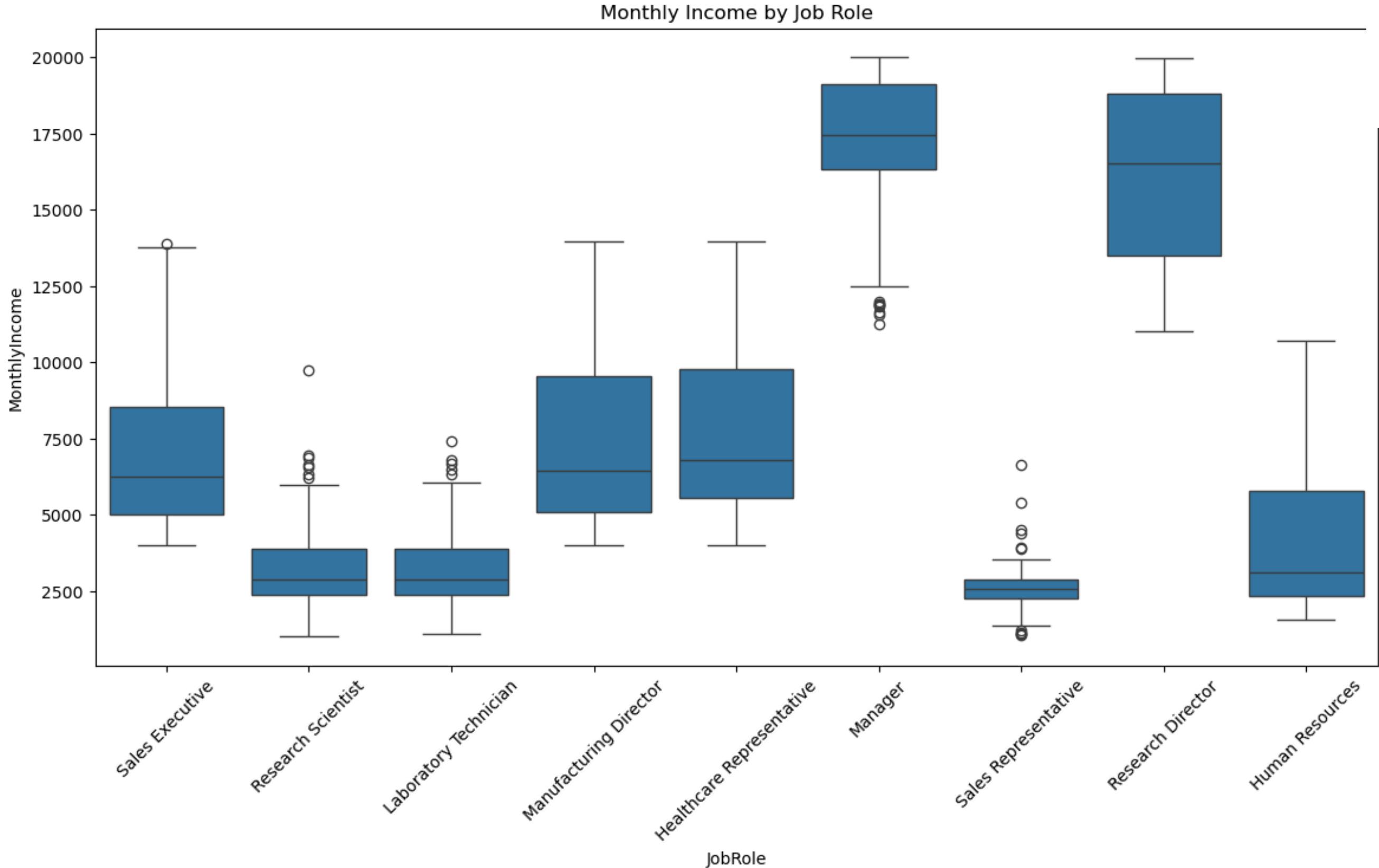


```
In[1]:  
plt.figure(figsize=(8,5))  
sns.histplot(df['MonthlyIncome'], kde=True, color='orange')  
plt.title('Monthly Income Distribution')  
plt.xlabel('Monthly Income')  
plt.ylabel('Count')  
plt.show()
```

Wide income spread; lower income groups show higher attrition.



Monthly Income vs Job Role (Boxplot)

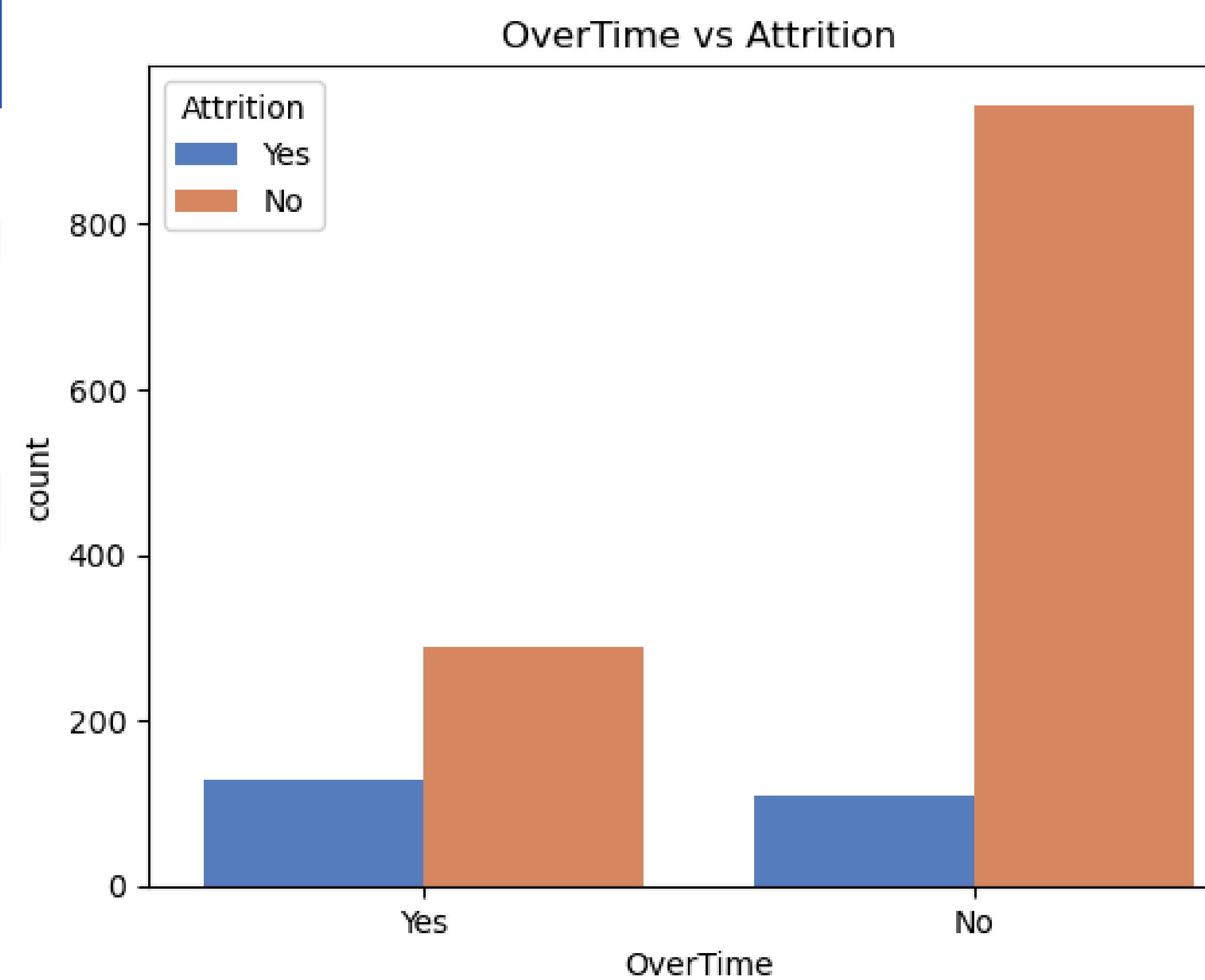


```
: plt.figure(figsize=(14,7))
sns.boxplot(x='JobRole', y='MonthlyIncome', data=df)
plt.xticks(rotation=45)
plt.title('Monthly Income by Job Role')
plt.show()
```

- Managers and Research Directors have higher incomes.
- Sales Executives and Laboratory Technicians show lower incomes.

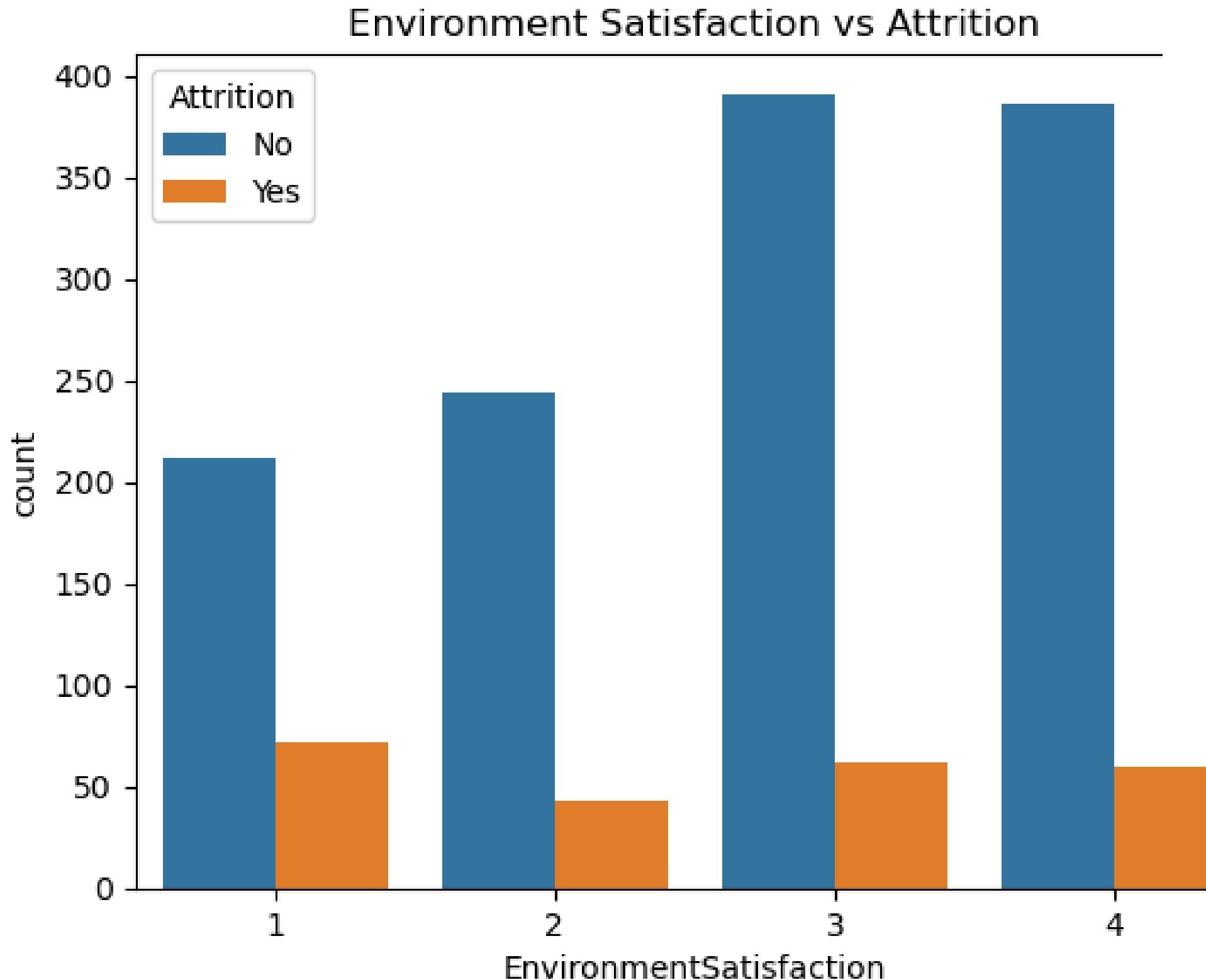
Overtime vs Attrition

```
[12]: sns.countplot(x='OverTime', hue='Attrition', data=df, palette='muted')
plt.title('OverTime vs Attrition')
plt.show()
```



- Employees working overtime are more likely to leave.

Environment Satisfaction Vs Attrition



```
3]: sns.countplot(x='EnvironmentSatisfaction', hue='Attrition', data=df)  
plt.title('Environment Satisfaction vs Attrition')  
plt.show()
```

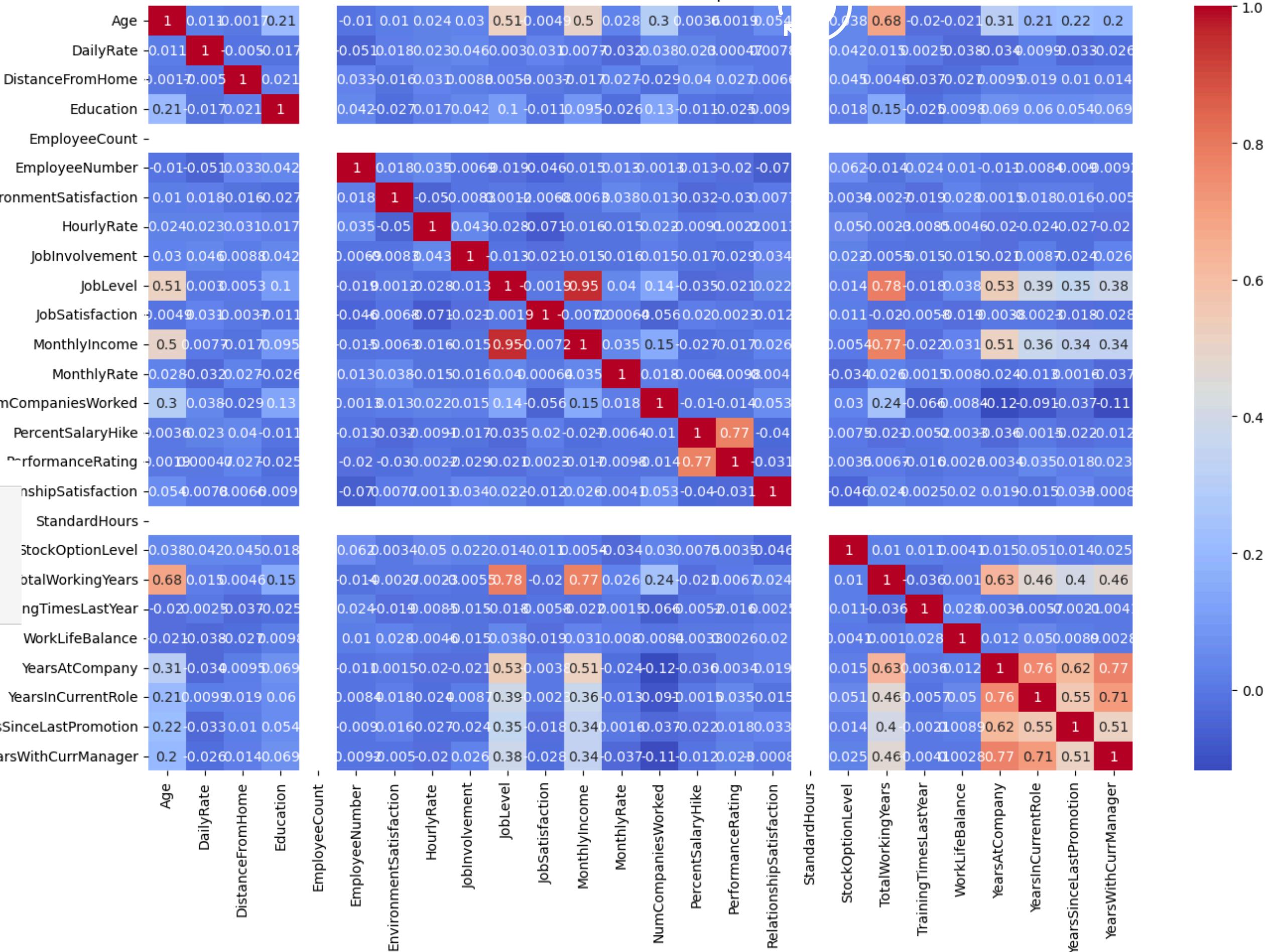
- Employees with low satisfaction levels are more prone to leave.



Corelation Heatmap

Overtime, MonthlyIncome, JobSatisfaction have strong correlations with Attrition.

```
: plt.figure(figsize=(16,10))
corr = df.select_dtypes(include=['int64', 'float64']).corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```





Key Insights

- Overtime, lower salaries, and low satisfaction increase attrition.
- Young employees and certain job roles experience higher turnover.
- Work-life balance is critical to retain employees.





Conclusion & Learnings

- Identified key drivers of attrition.
- Importance of improving job satisfaction and compensation.
- Hands-on experience with data cleaning, visualization, and machine learning.
- Gained practical knowledge of real-world HR analytics.



Thank You!



sanaekhan25@gmail.com

