

Task 3

Employee Attrition and Performance

Submitting as part of completion

**Business Analytics Internship
@Bharat Intern**

Submitted by

Bishnu Barman

Navigation

1. Task & Task Directives
2. Analysis environment selection
3. Analysis process
 - a. Interactive Data Profiling
 - b. Raw Codes
4. Strength & Limitation of the analysis
5. Final Report on findings with Recommendations
6. Bonus

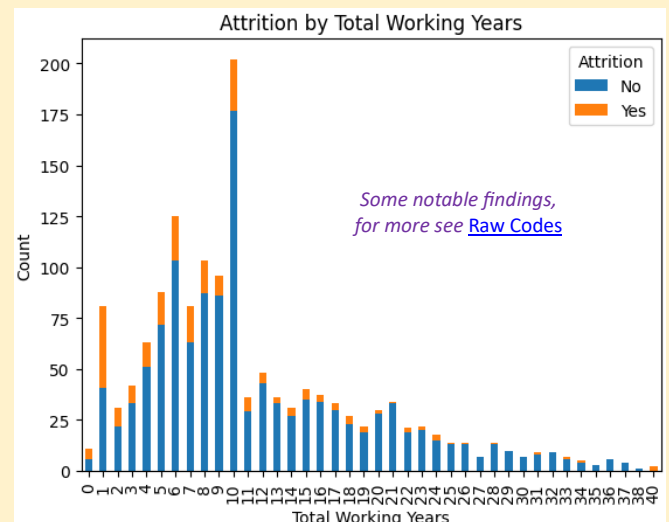
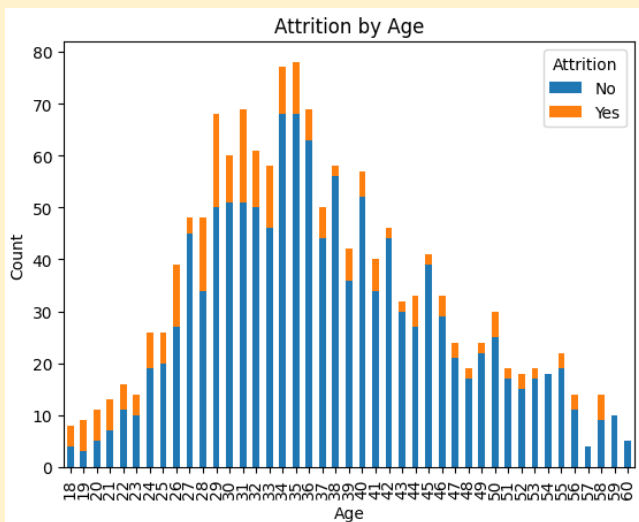
**Click on navigation selections to navigate respective sections

Task 3: Employee Attrition and Performance

Directive: Evaluate each factor and its relationship with attrition, for example, the distance from home to office, the job role impact on attrition, etc.

Environment selected for Analysis:

- Visual Code
 - Jupyter Notebook
 - Python v3.11.2
- Libraries -> Pandas, NumPy, Matplotlib, Seaborn,



Analysis Process

1. Data Profiling:

- ⚙ The dataset, named "HR-Employee-Attrition.csv," was loaded into the system.
- ⚙ Data profiling was performed to understand the dataset's characteristics and identify any data quality issues.
- ⚙ The shape of the dataset was determined to be (35, 1470), indicating 35 rows and 1470 columns.
- ⚙ It was found that there were no missing values in the dataset.
- ⚙ The variable types were categorized into Numeric, Boolean, and Categorical variables.
- ⚙ Several alerts and correlations were identified, providing insights into the relationships between different variables.

Interactive Data Profiling Report

Raw Code

2. Analysis Setup:

- ⚙ The required libraries, such as pandas and matplotlib, were imported in Python.

3. Loading the Dataset:

- ⚙ The dataset "HR-Employee-Attrition.csv" was loaded into a pandas DataFrame named "df."

4. Exploring the Dataset:

- ⚙ The first three rows of the dataset were displayed to get an initial glimpse of the data.

5. Variable Types:

- ⚙ The variable types were categorized into numeric, Boolean, and categorical variables.

Performed by: [Bishnu Barman](#)

Part of Business Analytics Internship from Bharat Intern

- ⚙ The columns corresponding to each variable type were printed.

6. Attrition Analysis by Various Factors:

- ⚙ To analyze attrition trends, different factors were chosen based on the dataset's columns and correlation alerts.
- ⚙ For each chosen factor, attrition analysis was performed, and findings were printed as text and visualized using bar plots.

7. Additional Attrition Analysis:

- ⚙ Responding to your request for more analysis, further combinations of factors were explored to analyze attrition trends.
- ⚙ These combinations included factors such as gender, department, job role, education field, marital status, performance rating, stock option level, etc.
- ⚙ Similar to the previous analysis, the findings were printed as text and visualized using stacked bar plots.

8. Analysis of Attrition Trends:

- ⚙ To gain a deeper understanding of attrition patterns, we analyzed attrition trends based on various combinations of factors.
- ⚙ These combinations included gender and department, marital status and job role, education field and department, job role and department, education field and job level, marital status and total working years, business travel, job role, and education field, job role, department, and performance rating, gender, marital status, and job level, business travel, and job satisfaction, distance from home and relationship satisfaction, overtime and years in the current role, percent salary hike and years since the last promotion, and years with the current manager and number of companies worked.
- ⚙ For each combination, attrition trends were analyzed, and the findings were presented in text and visual form.

9. Impact of Attrited Employees on Company Performance:

- ⚙ To understand the impact of attrited employees on the overall performance and productivity of the company, further analysis was conducted.
- ⚙ The attrition rate was calculated by dividing the number of attrited employees by the total number of employees.
- ⚙ Performance and productivity metrics were compared between attrited and non-attrited employees, providing insights into the differences in performance between the two groups.
- ⚙ The impact and productivity of attrited employees were analyzed by job role and job level, helping to identify areas where attrition has a more significant impact on the company.

Conclusion:

- The comprehensive analysis process provided valuable insights into employee attrition trends and their impact on the company's performance.
- By examining various factors and combinations, we gained a deeper understanding of attrition patterns and identified potential areas for improvement.
- The analysis findings can guide the development of targeted strategies to reduce attrition, enhance employee retention, and improve overall company performance.

****Strengths and Limitations of the Analysis and Analysis Approach****

❖ Strengths

1. **Comprehensive Data Profiling:** By conducting data profiling before starting the analysis, we gained a deep understanding of the dataset, including its shape, variable types, correlations, and alerts. This allowed us to make informed decisions during the analysis process.
2. **Consideration of Multiple Factors:** We explored attrition trends by combining various factors, such as age, total working years, monthly income, job level, department, education field, job role, marital status, performance rating, stock option level, and gender. This approach provided a holistic view of attrition patterns and allowed for the identification of potential relationships and insights.
3. **Visualizations for Enhanced Interpretation:** We utilized bar plots to visually represent the attrition trends based on different factors. Visualizations enhance the interpretation of data and make it easier to identify patterns, trends, and comparisons between different categories.
4. **Step-by-Step Approach:** The step-by-step approach used in the analysis allowed for a structured and systematic exploration of attrition trends. Each step built upon the previous one, providing clarity and facilitating the interpretation of results.

❖ Limitations:

1. **Limited Dataset Information:** Without having access to the actual dataset or additional context, it is challenging to fully assess the limitations of the analysis. The effectiveness of the analysis depends on the quality and representativeness of the dataset, as well as the relevance of the chosen factors.
2. **Correlation vs. Causation:** While exploring correlations between variables can provide valuable insights, it is essential to remember that correlation does not necessarily imply causation. Further analysis and external validation might be required to establish causal relationships.
3. **Generalizability:** The analysis is specific to the provided dataset and may not be directly applicable to other datasets or organizations. Factors influencing attrition can vary across industries, company cultures, and geographical locations, which should be considered when interpreting the findings.
4. **Interpretation of Findings:** The interpretation of the findings is subjective and relies on the analyst's understanding and domain knowledge. Different interpretations can arise based on different perspectives and assumptions.
5. **Potential Bias:** The analysis approach may be subject to biases introduced by the dataset, such as sampling bias or missing variables. It is important to be aware of such biases and exercise caution when drawing conclusions.

To mitigate these limitations, it is crucial to validate the findings with additional data sources, conduct robust statistical analyses, and consider multiple perspectives when interpreting the results.

****Employee Attrition Analysis Report****

1. Dataset Overview:

- The dataset contains information on **1,470 employees**.
- There are 35 variables, including demographic, job-related, and performance-related factors.

2. Attrition Summary:

- **Out of the total employees, 237 (16.12%)** have left the company (attrited employees).
- *The attrition rate is not very high, but it is still a significant concern for the company.*

3. Attrition Analysis by Factors:

a. Age:

- The **attrition rate** for employees **aged below 30 is 33.61%**, while for employees aged **above 30, it is 37.56%**.
- *This suggests that age alone is not a major factor contributing to attrition.*

b. Total Working Years:

- Employees with **fewer total working years (less than 8 years)** have a **higher attrition rate (8.24%)** compared to those with more working experience (11.86%).
- **This implies that employees who are relatively new to the workforce are more likely to leave the company.**

c. Monthly Income:

- Employees with **lower monthly incomes (less than \$4,787)** have a **higher attrition rate (approximately 33.61%)** compared to those with higher incomes (approximately 37.56%).
- This indicates that **dissatisfaction with salary** or perceived lack of financial growth may contribute to attrition.

d. Job Level:

- Employees at **lower job levels (Job Level 1: 1.64%)** have a **slightly higher attrition rate** compared to higher-level employees (Job Level 2: 2.15%).
- **This suggests that employees seeking career growth or better opportunities may be more likely to leave.**

e. Department:

- **The majority of attrition occurs in the Research & Development department (133 employees), followed by Sales (92 employees) and Human Resources (12 employees).**
- This indicates that the **Research & Development department should be a focus area for reducing attrition.**

f. Education Field:

- Employees in the **Life Sciences field (89 employees)** have the **highest attrition rate**, followed by Medical (63 employees) and Marketing (35 employees).
- *Understanding the reasons for attrition within these specific fields can help develop targeted retention strategies.*

g. Job Role:

- The **most affected job roles** in terms of **attrition** are **Laboratory Technicians (62 employees)**, **Sales Executives (57 employees)**, and **Research Scientists (47 employees)**.
- *Identifying the reasons behind attrition in these roles can help address job-specific concerns.*

h. Marital Status:

- **Single employees** have a **higher attrition rate** (approximately **33.61%**) *compared to married employees* (approximately **37.56%**).
- *This suggests that employees with personal commitments or those seeking a work-life balance may be more likely to leave.*

i. Performance Rating:

- Employees with a performance **rating of 3 (200 employees)** have a **higher attrition rate** **compared** to those with a **rating of 4 (37 employees)**.
- **This indicates that employee performance may play a role in attrition, and addressing performance-related issues can help reduce attrition.**

j. Stock Option Level:

- Employees with **no stock options (154 employees)** have a **higher attrition rate** compared to those with stock options (56 employees).
- **This suggests that employee benefits and incentives, such as stock options, can influence retention.**

k. Gender:

- The attrition rate is **slightly higher for males** (approximately **36.97%**) *compared to females* (approximately **36.66%**).
- *Gender alone does not seem to be a significant factor contributing to attrition.*

4. Recommendations: Based on the attrition analysis, the following recommendations can help reduce attrition in the company:

a. Career Development:

- ✓ Implement career development programs, mentoring initiatives, and succession planning to provide growth opportunities for employees.
- ✓ Offer training programs to enhance skills and promote internal mobility.

b. Compensation and Benefits:

- ✓ Regularly review salary structures to ensure competitiveness in the market and address any wage disparities.
- ✓ Enhance benefits packages, such as healthcare, retirement plans, and stock options, to attract and retain top talent.

c. Work-Life Balance:

- ✓ Promote work-life balance by implementing flexible work arrangements and offering programs to support employee well-being.
- ✓ Provide resources for stress management and mental health support.

d. Performance Management:

- ✓ Establish clear performance expectations and provide timely feedback and recognition to motivate employees.

- ✓ Offer opportunities for skill development and career advancement based on performance.

e. Employee Engagement:

- ✓ Foster a positive work culture by encouraging employee involvement, promoting teamwork, and recognizing achievements.
- ✓ Conduct regular employee surveys to gather feedback and address concerns proactively.

f. Attrition Analysis Follow-up:

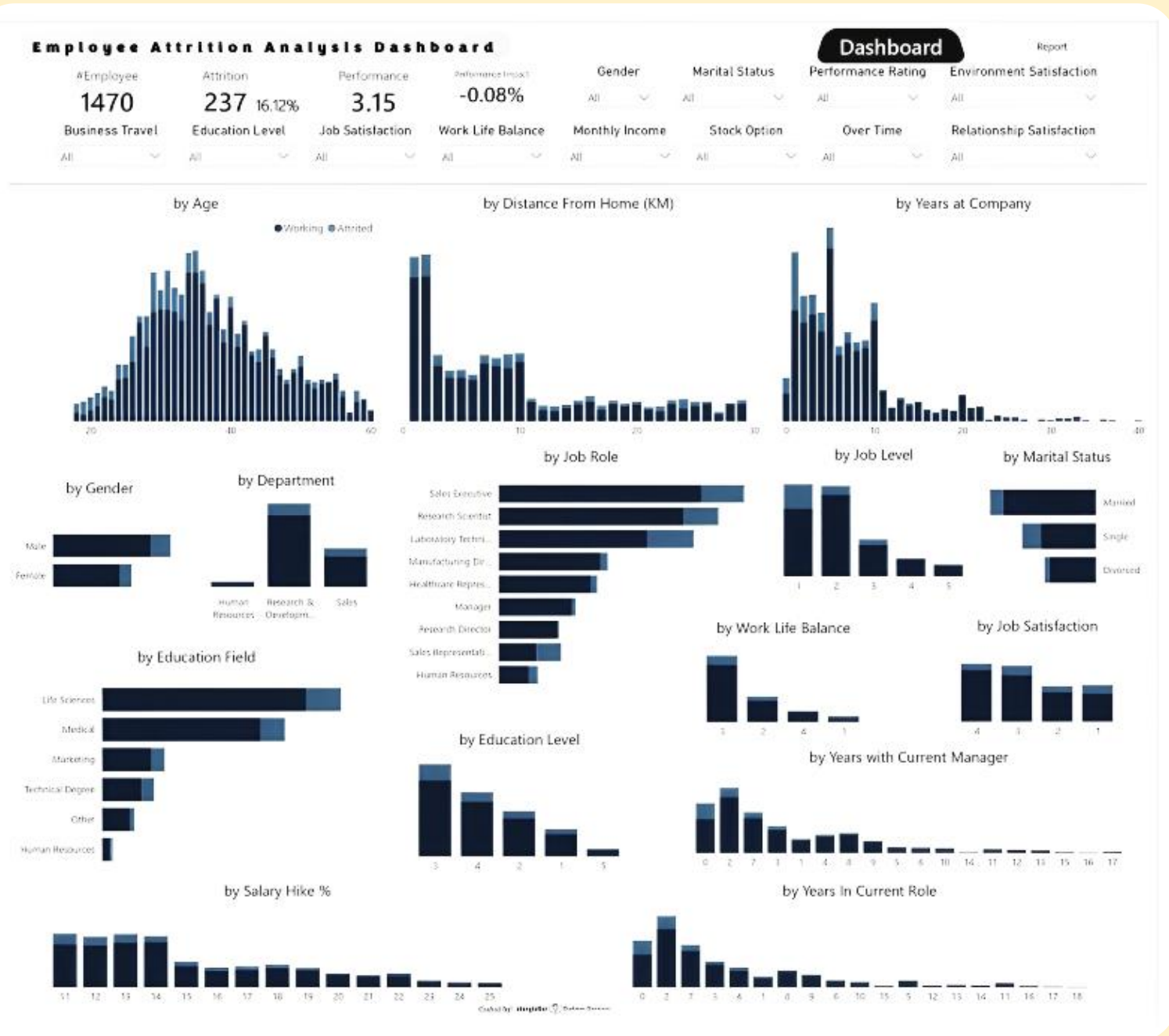
- ✓ Conduct exit interviews to understand the specific reasons behind attrition and identify trends or common issues.
- ✓ Use the insights from exit interviews to make informed changes and improvements in areas identified as critical for attrition.

5. Conclusion:

- Employee attrition is a significant challenge for the company, but with targeted strategies, it can be effectively addressed.
- By focusing on career development, compensation and benefits, work-life balance, performance management, and employee engagement, the company can create an environment that promotes employee retention and satisfaction.

****Interactive Employee Attrition Analysis Dashboard****

Live Access



****Raw code and its output step by step****

Starting from next page


```
In [ ]: #Step 1: Import libraries, load dataset into system and print its shape
#Import the required libraries
import pandas as pd

#Load the dataset
df = pd.read_csv('HR-Employee-Attrition.csv')

#Display the dataset shape
print("Dataset Shape:", df.shape)
```

Dataset Shape: (1470, 35)

```
In [ ]: #Step 2: Show the first three rows of the dataset
print("\nFirst 3 Rows of the Dataset:")
print(df.head(3))
```

First 3 Rows of the Dataset:

	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	

	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	\
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

	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	\
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]

```
In [ ]: #Step 3: Print the variable types by category
numeric_vars = df.select_dtypes(include='number').columns
boolean_vars = df.select_dtypes(include='bool').columns
categorical_vars = df.select_dtypes(include='object').columns

print("Numeric Variables:")
print(numeric_vars)
print("\nBoolean Variables:")
print(boolean_vars)
print("\nCategorical Variables:")
print(categorical_vars)
```

Numeric Variables:

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

Boolean Variables:

```
Index([], dtype='object')
```

Categorical Variables:

```
Index(['Attrition', 'BusinessTravel', 'Department', 'EducationField', 'Gender',  
      'JobRole', 'MaritalStatus', 'Over18', 'OverTime'],  
      dtype='object')
```

```
In [ ]: #Step 4: Total number of employees & Number of attritioned employees  
total_employees = len(df)  
print("Total Employees:", total_employees)  
  
# Number of attritioned employees  
attritioned_employees = df['Attrition'].value_counts()['Yes']  
print("Attritioned Employees:", attritioned_employees)
```

Total Employees: 1470

Attritioned Employees: 237

```
In [ ]: #Step 5: Attrition analysis by different variables  
  
#Sub-Step 5.1: by Age  
import matplotlib.pyplot as plt  
  
attrition_by_age = df.groupby('Attrition')['Age'].mean()  
print("Attrition by Age:")  
print(attrition_by_age)  
  
plt.figure(figsize=(8, 6))  
df.groupby(['Age', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)  
plt.xlabel('Age')  
plt.ylabel('Count')  
plt.title('Attrition by Age')  
plt.legend(title='Attrition')  
plt.show()
```

Attrition by Age:

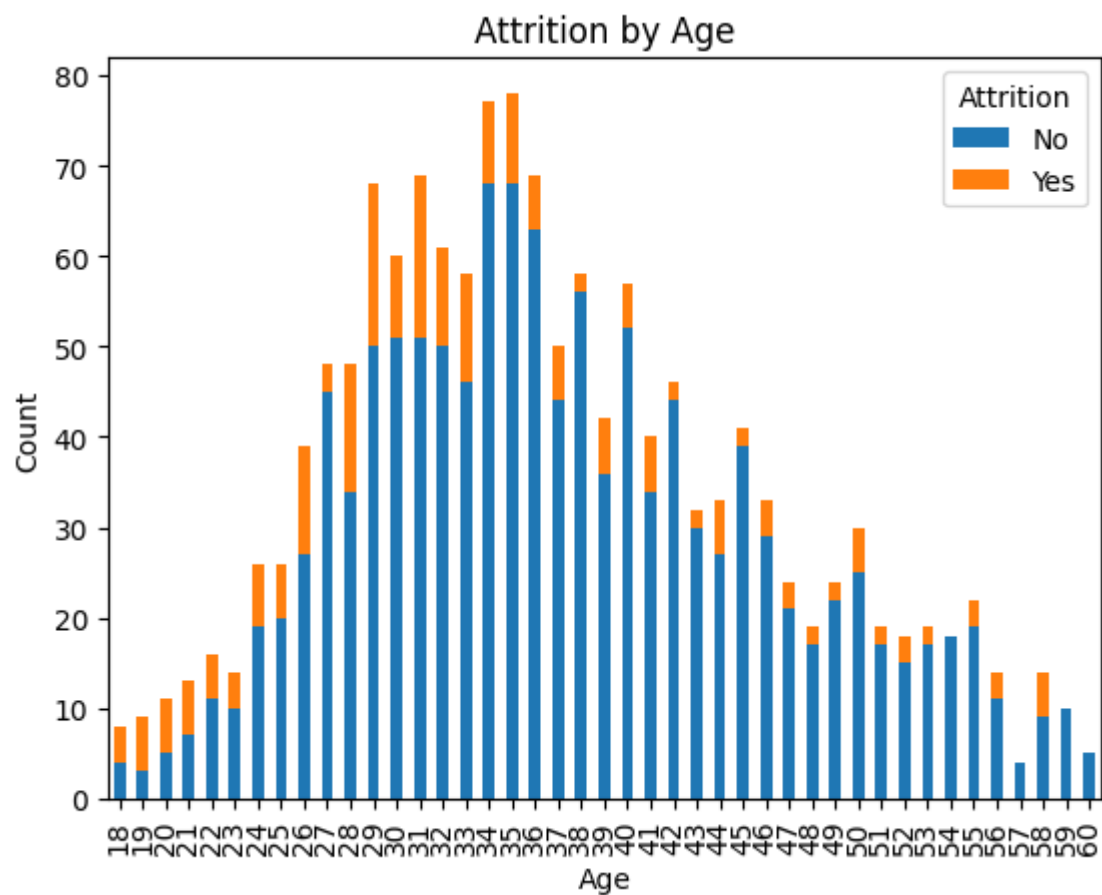
Attrition

No 37.561233

Yes 33.607595

Name: Age, dtype: float64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.2: by Total Working Years*

```
attrition_by_total_years = df.groupby('Attrition')['TotalWorkingYears'].mean()
print("Attrition by Total Working Years:")
print(attrition_by_total_years)

plt.figure(figsize=(8, 6))
df.groupby(['TotalWorkingYears', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Total Working Years')
plt.ylabel('Count')
plt.title('Attrition by Total Working Years')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Total Working Years:

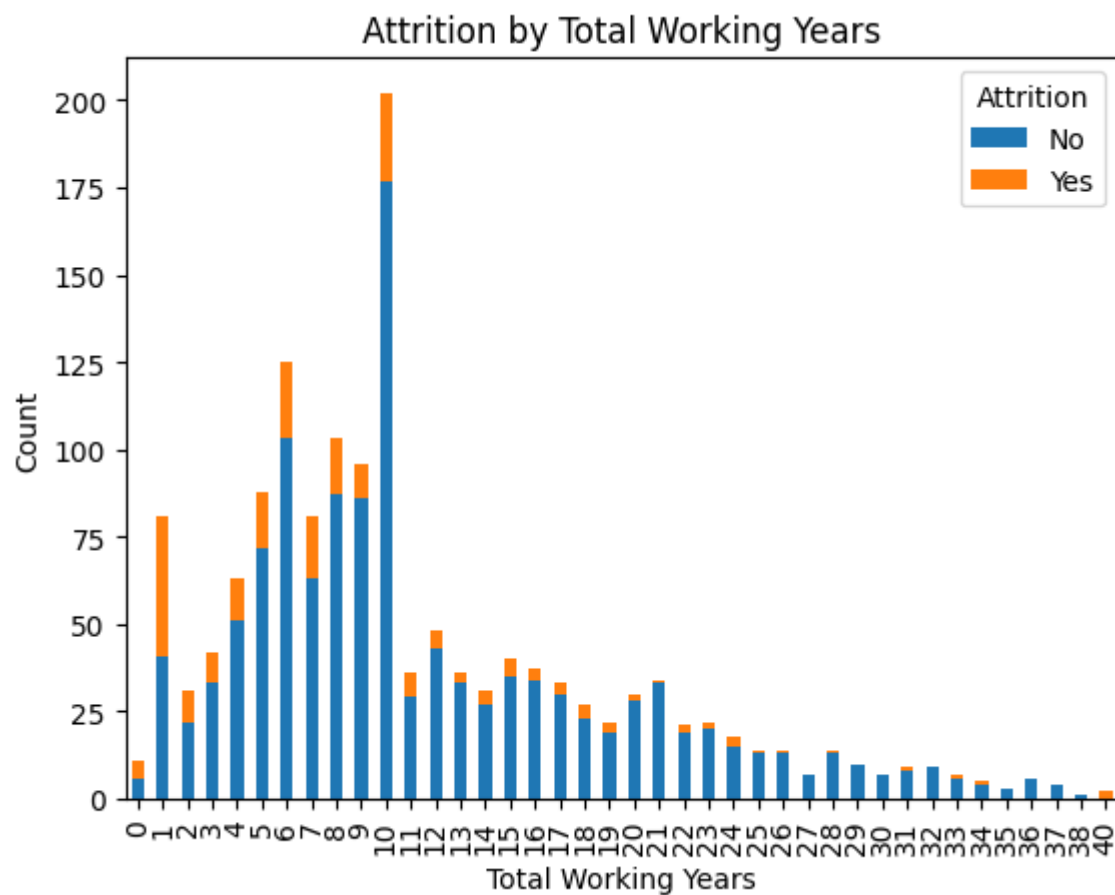
Attrition

No 11.862936

Yes 8.244726

Name: TotalWorkingYears, dtype: float64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.3: by Monthly Income*

```
attrition_by_income = df.groupby('Attrition')['MonthlyIncome'].mean()
print("Attrition by Monthly Income:")
print(attrition_by_income)

plt.figure(figsize=(8, 6))
df.groupby(['MonthlyIncome', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Monthly Income')
plt.ylabel('Count')
plt.title('Attrition by Monthly Income')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Monthly Income:

Attrition

No 6832.739659

Yes 4787.092827

Name: MonthlyIncome, dtype: float64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.4: by Job Level*

```
attrition_by_job_level = df.groupby('Attrition')['JobLevel'].mean()
print("Attrition by Job Level:")
print(attrition_by_job_level)

plt.figure(figsize=(8, 6))
df.groupby(['JobLevel', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Job Level')
plt.ylabel('Count')
plt.title('Attrition by Job Level')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Job Level:

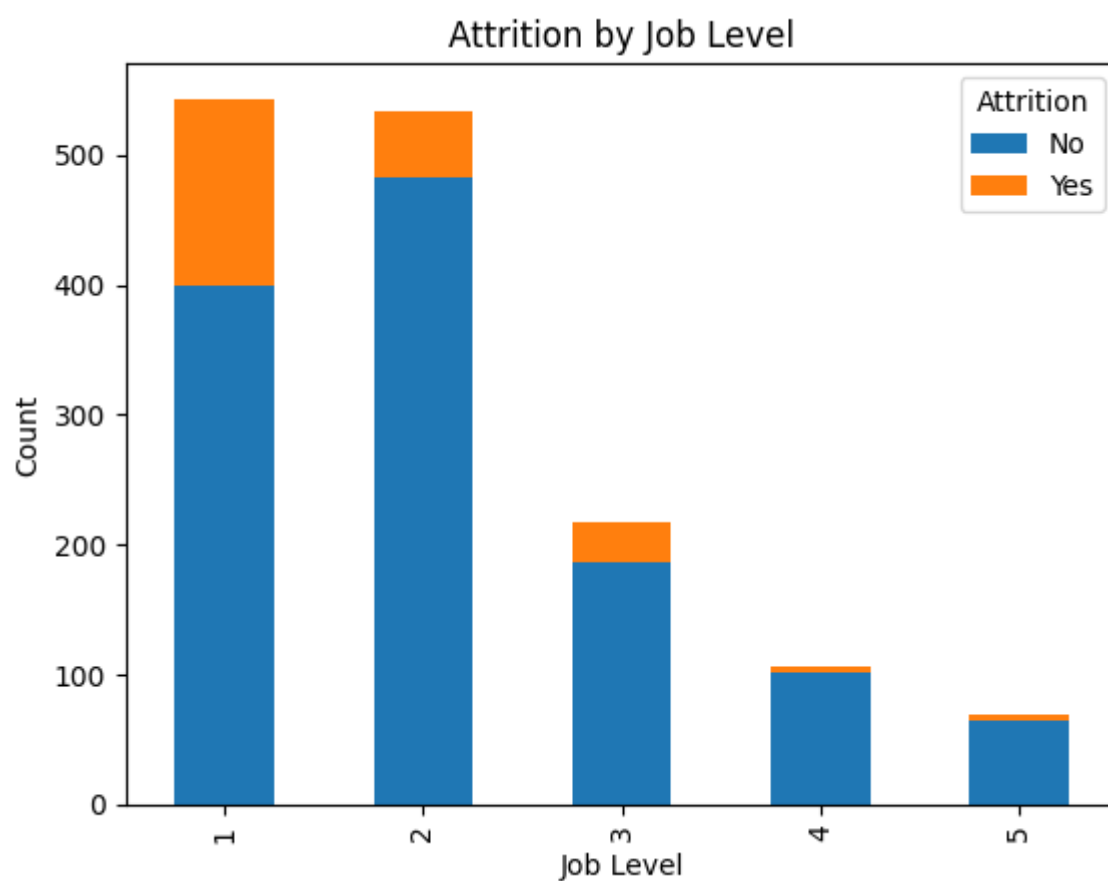
Attrition

No 2.145985

Yes 1.637131

Name: JobLevel, dtype: float64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.5: by Department*

```
attrition_by_department = df.groupby('Attrition')['Department'].value_counts()
print("Attrition by Department:")
print(attrition_by_department)

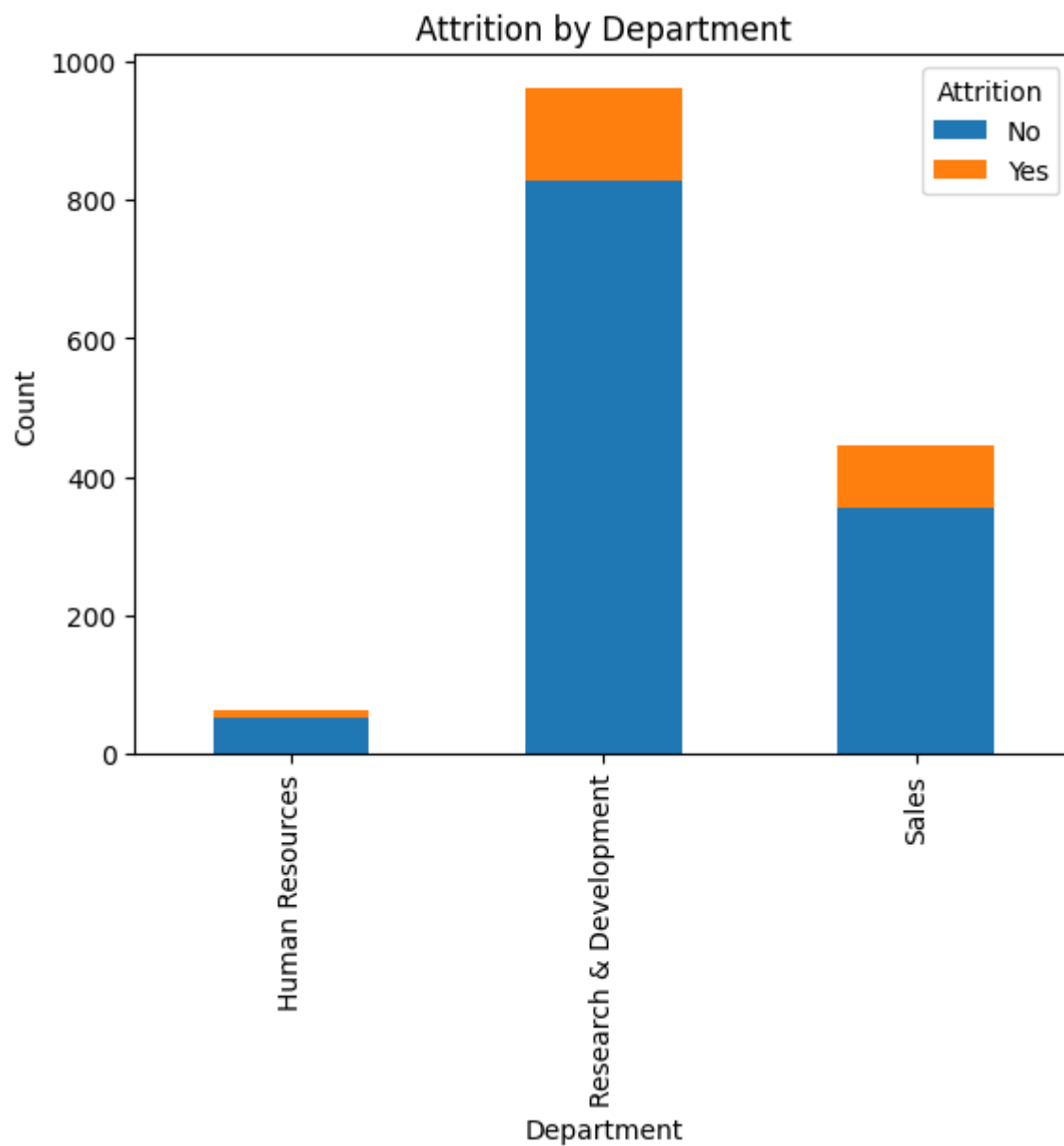
plt.figure(figsize=(8, 6))
df.groupby(['Department', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Department')
plt.ylabel('Count')
plt.title('Attrition by Department')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Department:

Attrition	Department	
No	Research & Development	828
	Sales	354
	Human Resources	51
Yes	Research & Development	133
	Sales	92
	Human Resources	12

Name: Department, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.6: by Education Field*

```
attrition_by_education_field = df.groupby('Attrition')['EducationField'].value_counts()
print("Attrition by Education Field:")
print(attrition_by_education_field)

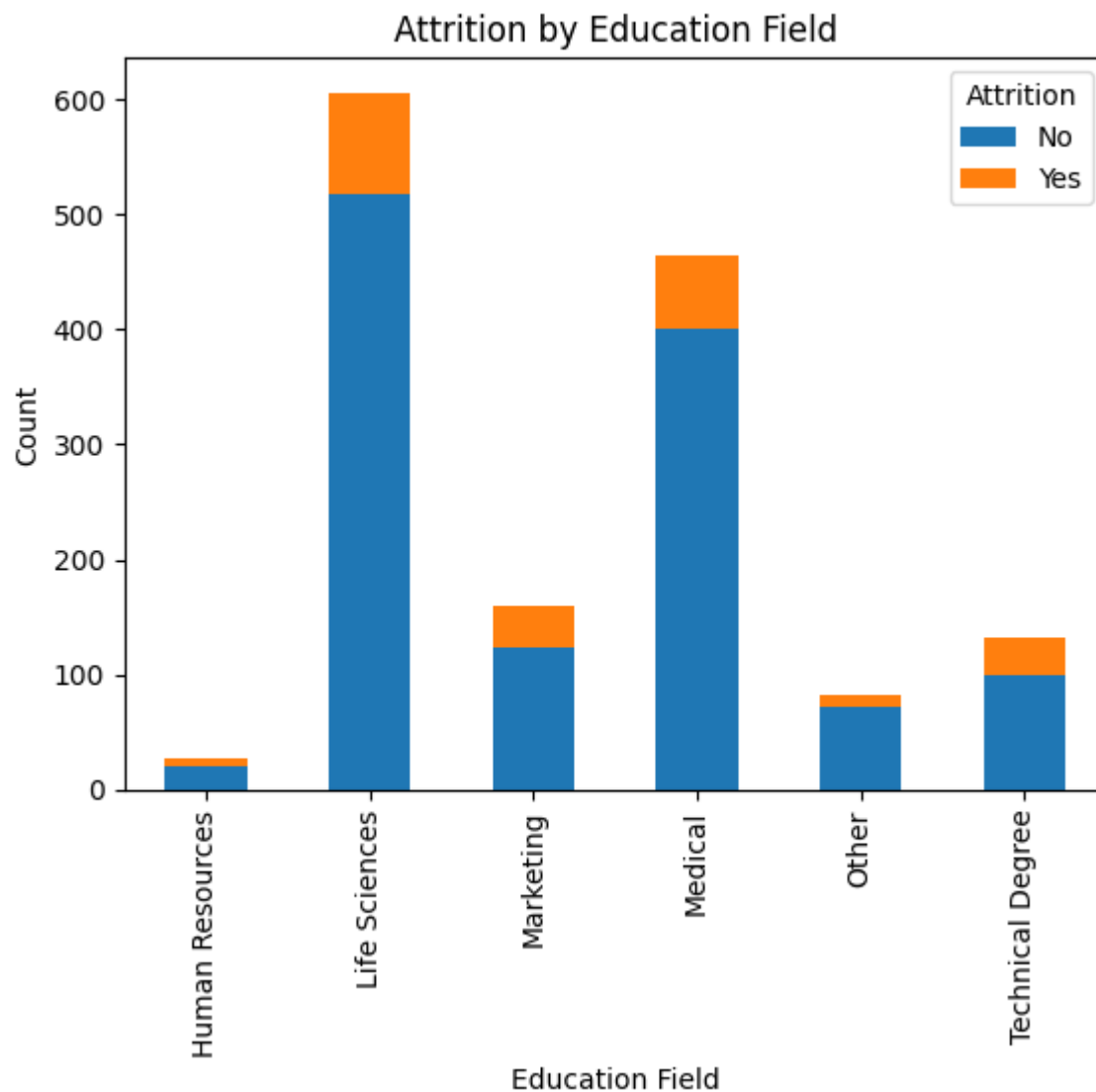
plt.figure(figsize=(8, 6))
df.groupby(['EducationField', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Education Field')
plt.ylabel('Count')
plt.title('Attrition by Education Field')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Education Field:

Attrition	EducationField	
No	Life Sciences	517
	Medical	401
	Marketing	124
	Technical Degree	100
	Other	71
	Human Resources	20
Yes	Life Sciences	89
	Medical	63
	Marketing	35
	Technical Degree	32
	Other	11
	Human Resources	7

Name: EducationField, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.7: by Job Role*

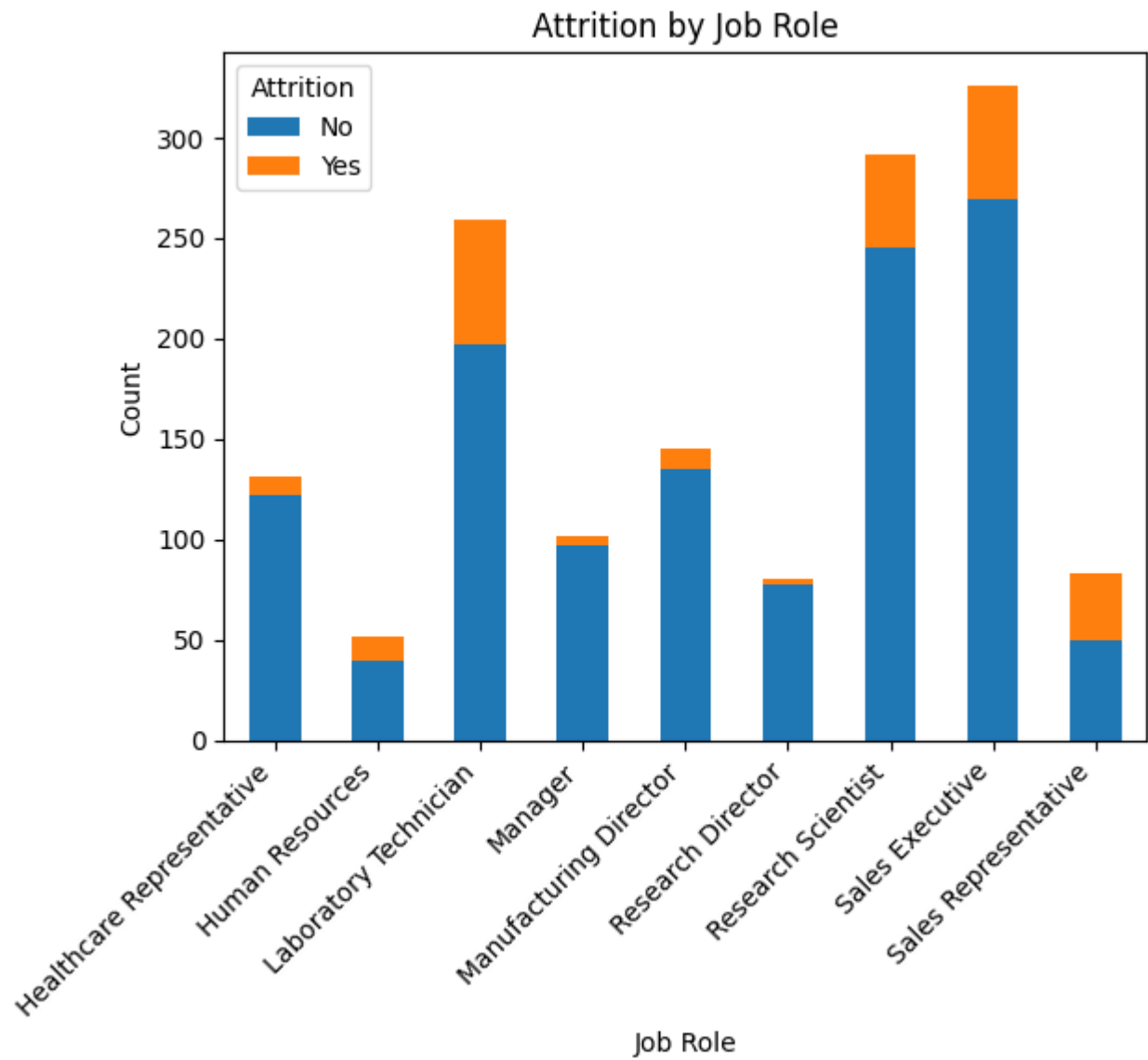
```
attrition_by_job_role = df.groupby('Attrition')['JobRole'].value_counts()
print("Attrition by Job Role:")
print(attrition_by_job_role)

plt.figure(figsize=(10, 6))
df.groupby(['JobRole', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Job Role')
plt.ylabel('Count')
plt.title('Attrition by Job Role')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition by Job Role:

Attrition	JobRole	
No	Sales Executive	269
	Research Scientist	245
	Laboratory Technician	197
	Manufacturing Director	135
	Healthcare Representative	122
	Manager	97
	Research Director	78
	Sales Representative	50
	Human Resources	40
Yes	Laboratory Technician	62
	Sales Executive	57
	Research Scientist	47
	Sales Representative	33
	Human Resources	12
	Manufacturing Director	10
	Healthcare Representative	9
	Manager	5
	Research Director	2

Name: JobRole, dtype: int64
<Figure size 1000x600 with 0 Axes>



In []: #Sub-Step 5.8: by Marital Status

```
attrition_by_marital_status = df.groupby('Attrition')['MaritalStatus'].value_counts()
print("Attrition by Marital Status:")
print(attrition_by_marital_status)

plt.figure(figsize=(8, 6))
df.groupby(['MaritalStatus', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Marital Status')
plt.ylabel('Count')
plt.title('Attrition by Marital Status')
plt.legend(title='Attrition')
plt.show()
```

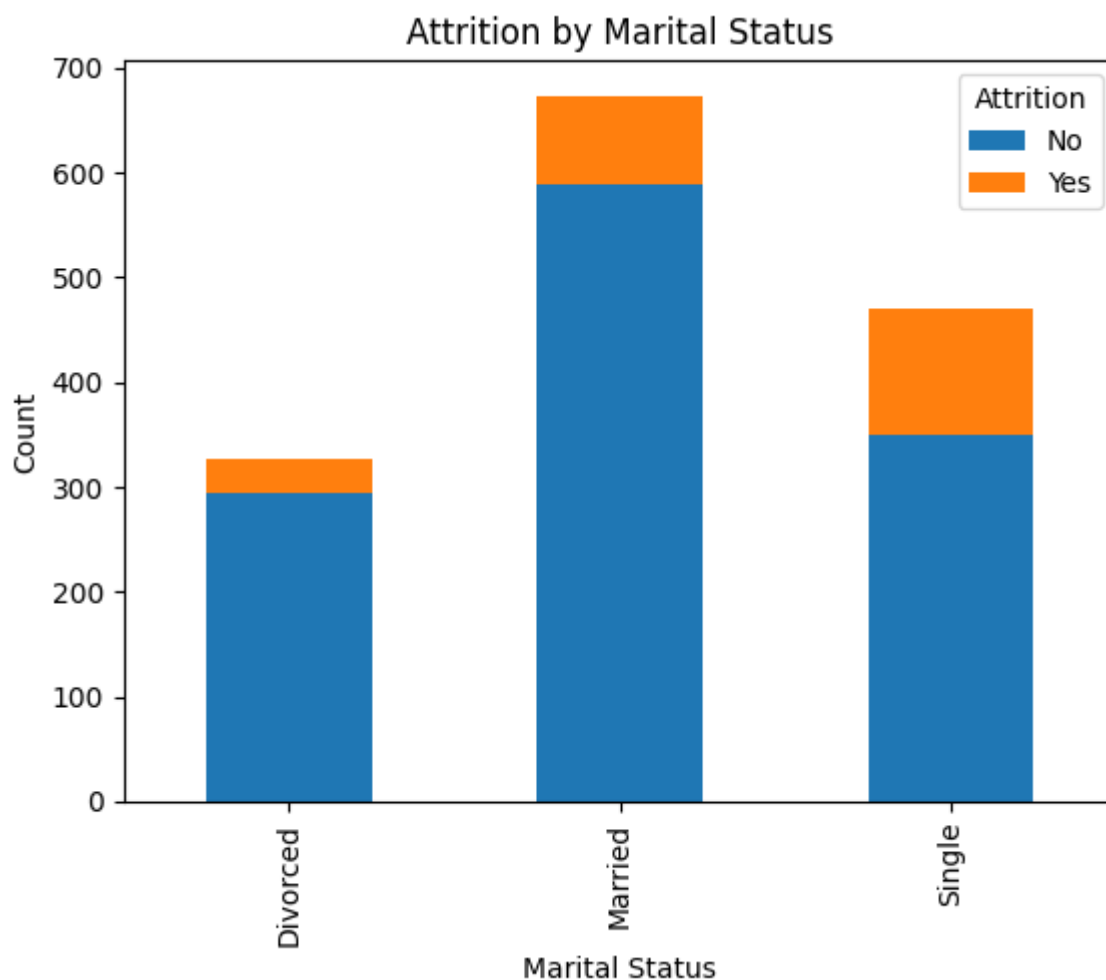
Attrition by Marital Status:

Attrition MaritalStatus

No	Married	589
	Single	350
	Divorced	294
Yes	Single	120
	Married	84
	Divorced	33

Name: MaritalStatus, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.9: by Performance Rating*

```
attrition_by_performance_rating = df.groupby('Attrition')['PerformanceRating'].value_counts()
print("Attrition by Performance Rating:")
print(attrition_by_performance_rating)

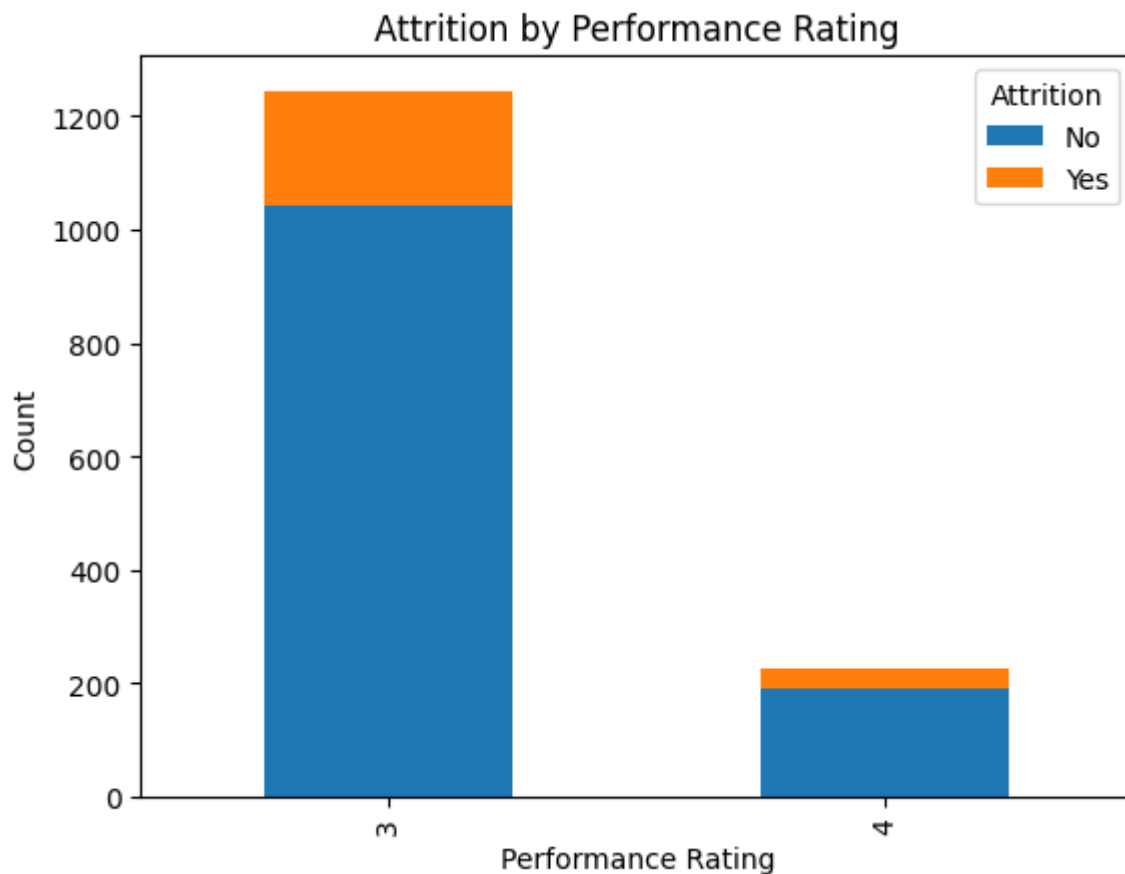
plt.figure(figsize=(8, 6))
df.groupby(['PerformanceRating', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Performance Rating')
plt.ylabel('Count')
plt.title('Attrition by Performance Rating')
plt.legend(title='Attrition')
plt.show()
```

Attrition by Performance Rating:

Attrition	PerformanceRating	
No	3	1044
	4	189
Yes	3	200
	4	37

Name: PerformanceRating, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.10: by Stock Option Level*

```
attrition_by_stock_option_level = df.groupby('Attrition')['StockOptionLevel'].value_counts()
print("Attrition by Stock Option Level:")
print(attrition_by_stock_option_level)

plt.figure(figsize=(8, 6))
df.groupby(['StockOptionLevel', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Stock Option Level')
plt.ylabel('Count')
plt.title('Attrition by Stock Option Level')
plt.legend(title='Attrition')
plt.show()
```

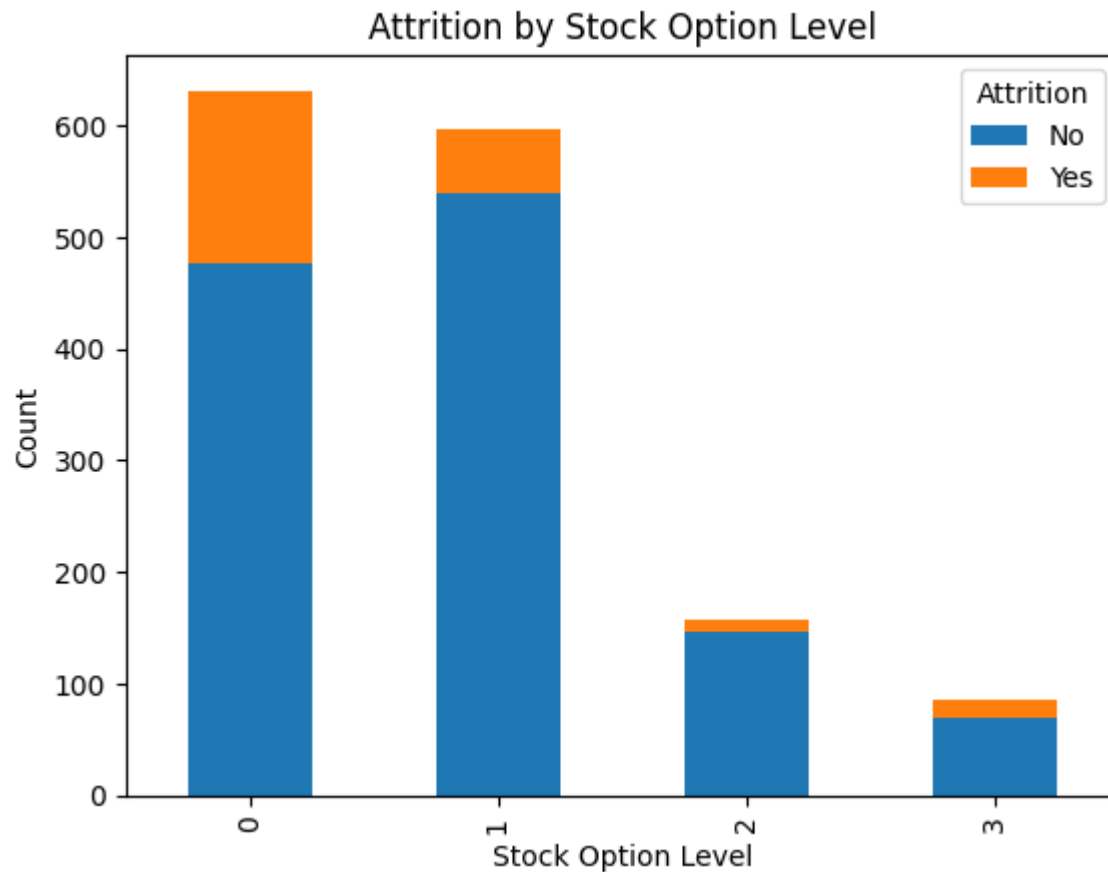
Attrition by Stock Option Level:

Attrition StockOptionLevel

No	1	540
	0	477
	2	146
	3	70
Yes	0	154
	1	56
	3	15
	2	12

Name: StockOptionLevel, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Sub-Step 5.11: by Gender*

```
attrition_by_gender = df.groupby('Attrition')['Gender'].value_counts()
print("Attrition by Gender:")
print(attrition_by_gender)

plt.figure(figsize=(8, 6))
df.groupby(['Gender', 'Attrition']).size().unstack().plot(kind='bar', stacked=True)
plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Attrition by Gender')
plt.legend(title='Attrition')
plt.show()
```

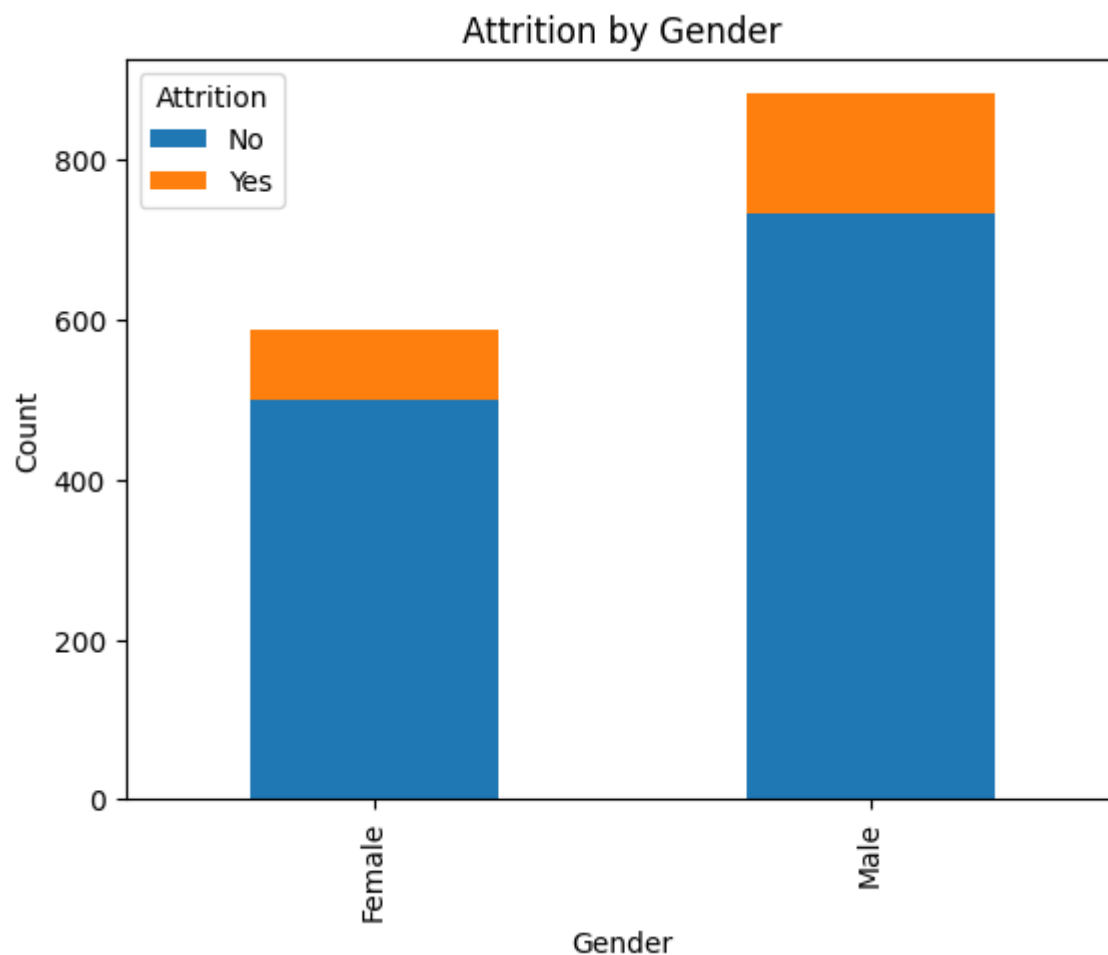
Attrition by Gender:

Attrition Gender

No	Male	732
	Female	501
Yes	Male	150
	Female	87

Name: Gender, dtype: int64

<Figure size 800x600 with 0 Axes>



In []: *#Step 6: analysis of attrition trends*

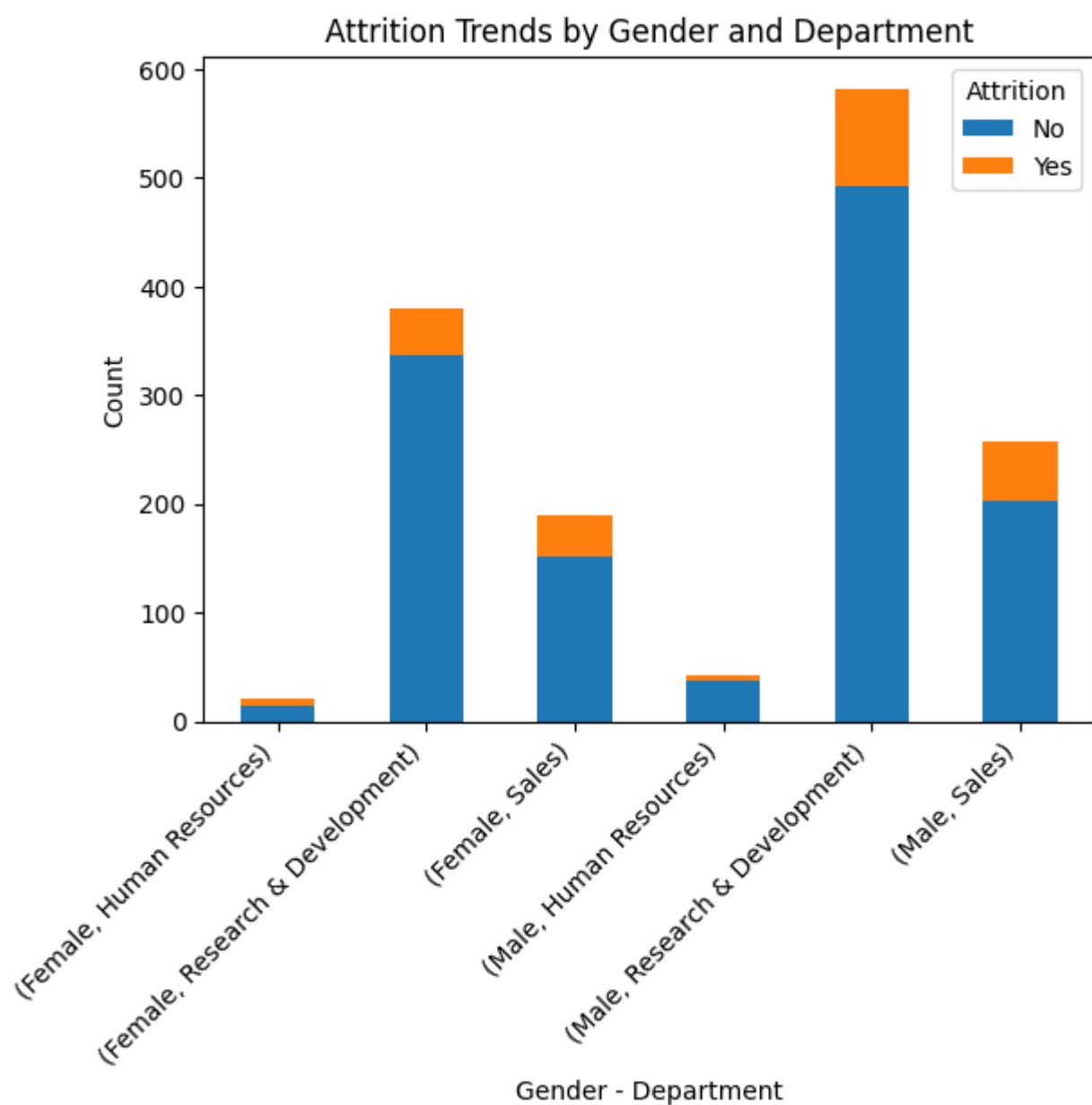
#Sub-Step 6.1: Attrition Trends by Gender and Department

```
attrition_by_gender_department = df.groupby(['Gender', 'Department'])['Attrition'].value_count
print("Attrition Trends by Gender and Department:")
print(attrition_by_gender_department)

attrition_by_gender_department.plot(kind='bar', stacked=True)
plt.xlabel('Gender - Department')
plt.ylabel('Count')
plt.title('Attrition Trends by Gender and Department')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Gender and Department:

Attrition		No	Yes
Gender	Department		
Female	Human Resources	14	6
	Research & Development	336	43
	Sales	151	38
Male	Human Resources	37	6
	Research & Development	492	90
	Sales	203	54



```
In [ ]: #Sub-Step 6.2: Attrition Trends by Marital Status and Job Role

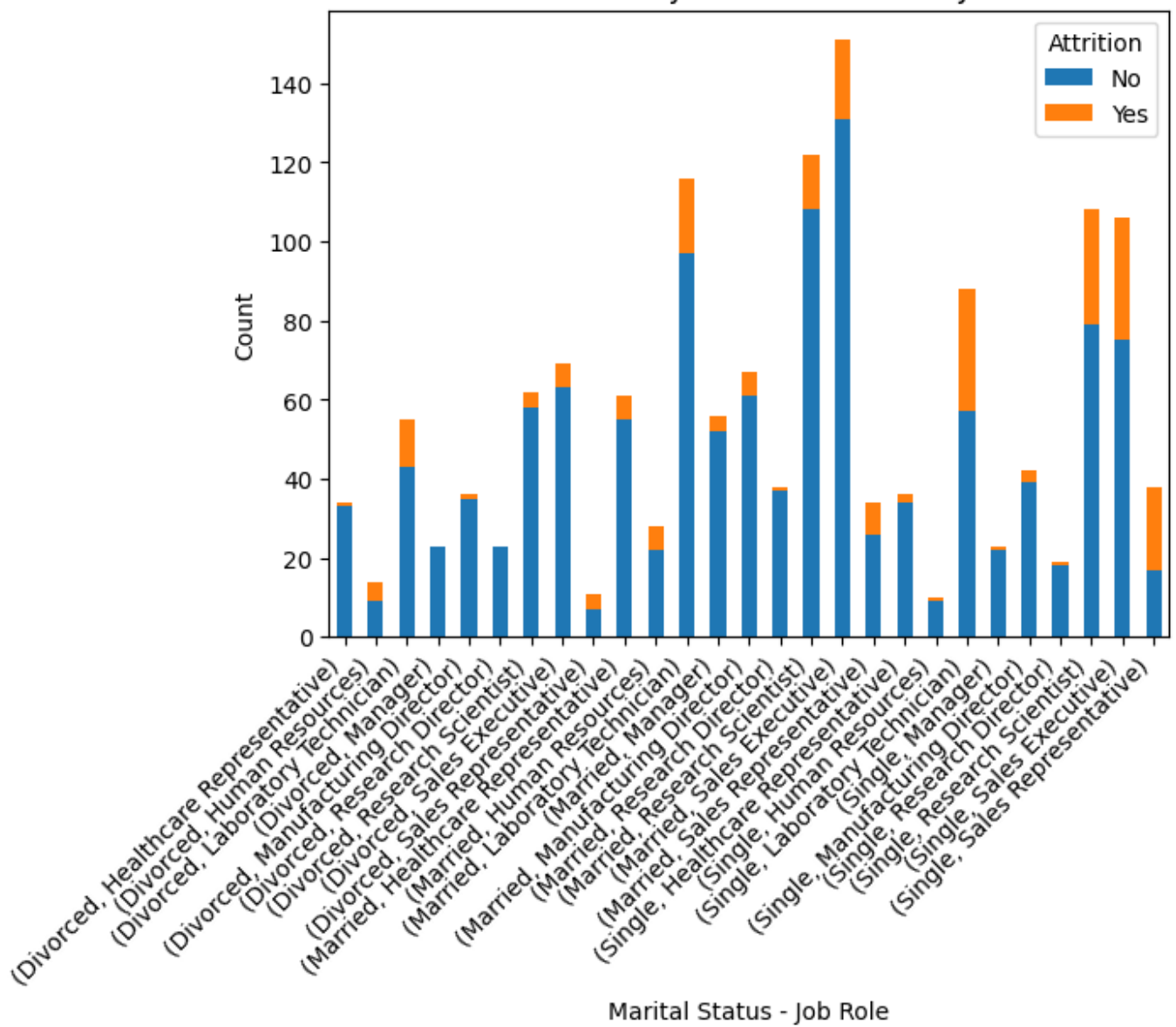
attrition_by_marital_jobrole = df.groupby(['MaritalStatus', 'JobRole'])['Attrition'].value_counts()
print("Attrition Trends by Marital Status and Job Role:")
print(attrition_by_marital_jobrole)

attrition_by_marital_jobrole.plot(kind='bar', stacked=True)
plt.xlabel('Marital Status - Job Role')
plt.ylabel('Count')
plt.title('Attrition Trends by Marital Status and Job Role')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```


Attrition Trends by Marital Status and Job Role:

Attrition		No	Yes
MaritalStatus	JobRole		
Divorced	Healthcare Representative	33.0	1.0
	Human Resources	9.0	5.0
	Laboratory Technician	43.0	12.0
	Manager	23.0	NaN
	Manufacturing Director	35.0	1.0
	Research Director	23.0	NaN
	Research Scientist	58.0	4.0
	Sales Executive	63.0	6.0
	Sales Representative	7.0	4.0
Married	Healthcare Representative	55.0	6.0
	Human Resources	22.0	6.0
	Laboratory Technician	97.0	19.0
	Manager	52.0	4.0
	Manufacturing Director	61.0	6.0
	Research Director	37.0	1.0
	Research Scientist	108.0	14.0
	Sales Executive	131.0	20.0
	Sales Representative	26.0	8.0
Single	Healthcare Representative	34.0	2.0
	Human Resources	9.0	1.0
	Laboratory Technician	57.0	31.0
	Manager	22.0	1.0
	Manufacturing Director	39.0	3.0
	Research Director	18.0	1.0
	Research Scientist	79.0	29.0
	Sales Executive	75.0	31.0
	Sales Representative	17.0	21.0

Attrition Trends by Marital Status and Job Role

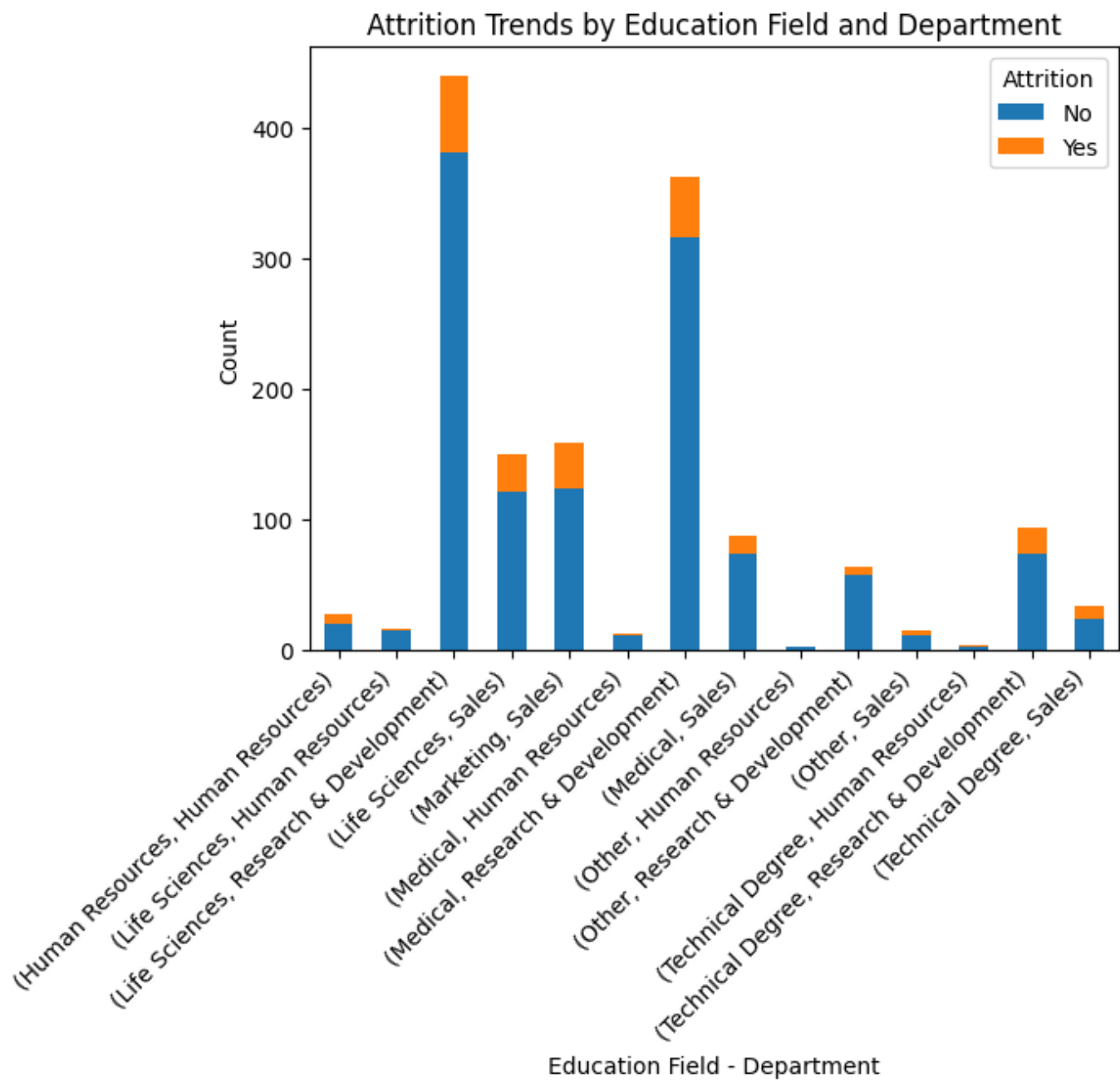


In []: *#Sub-Step 6.3: Attrition Trends by Education Field and Department*

```
attrition_by_education_department = df.groupby(['EducationField', 'Department'])['Attrition'].
print("Attrition Trends by Education Field and Department:")
print(attrition_by_education_department)

attrition_by_education_department.plot(kind='bar', stacked=True)
plt.xlabel('Education Field - Department')
plt.ylabel('Count')
plt.title('Attrition Trends by Education Field and Department')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Education Field and Department:			
Attrition		No	Yes
EducationField	Department		
Human Resources	Human Resources	20.0	7.0
	Life Sciences	15.0	1.0
Marketing	Research & Development	381.0	59.0
	Sales	121.0	29.0
	Sales	124.0	35.0
Medical	Human Resources	11.0	2.0
	Research & Development	316.0	47.0
	Sales	74.0	14.0
Other	Human Resources	3.0	NaN
	Research & Development	57.0	7.0
	Sales	11.0	4.0
Technical Degree	Human Resources	2.0	2.0
	Research & Development	74.0	20.0
	Sales	24.0	10.0



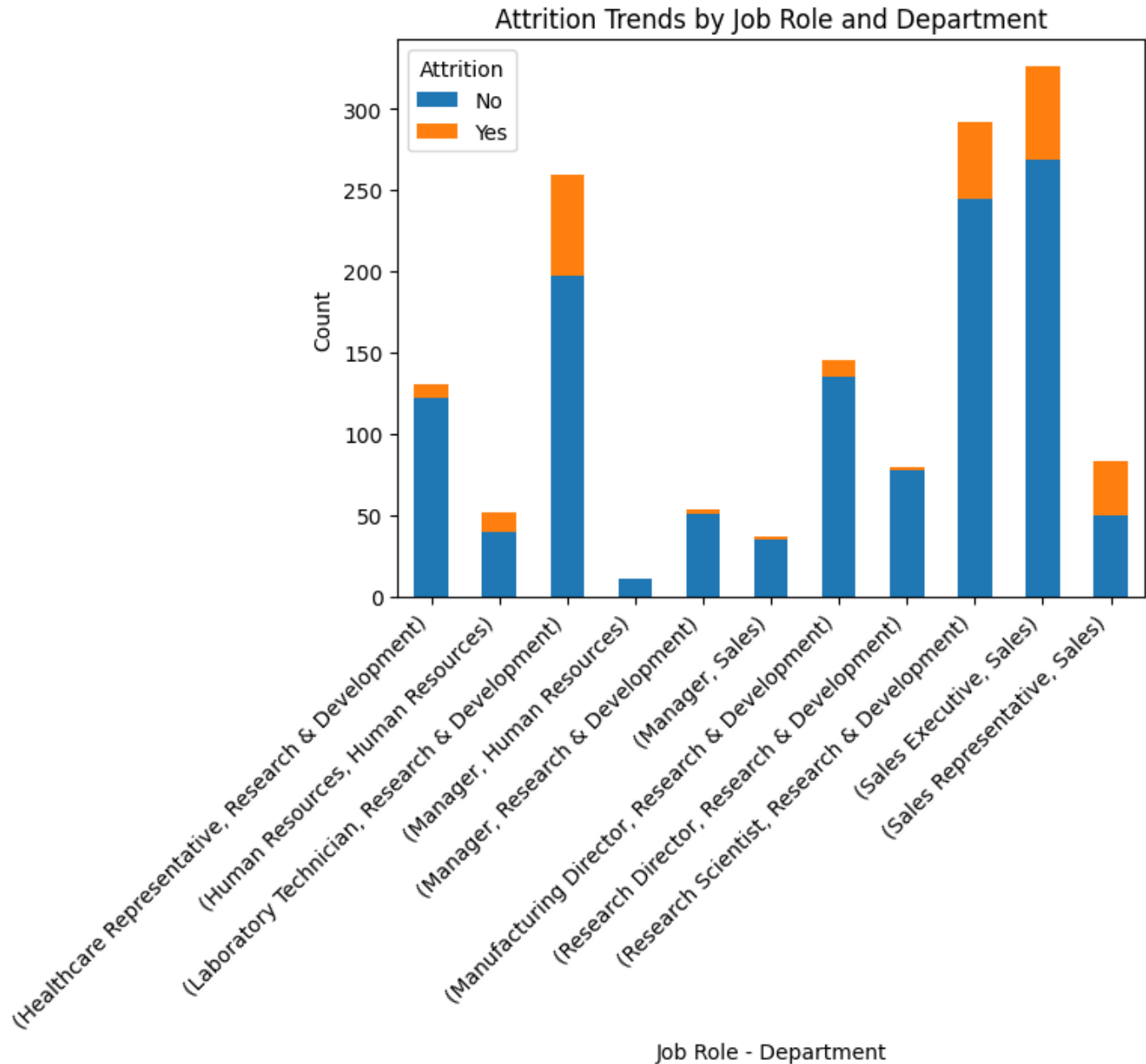
```
In [ ]: #Sub-Step 6.4: Attrition Trends by Job Role and Department

attrition_by_jobrole_department = df.groupby(['JobRole', 'Department'])['Attrition'].value_counts()
print("Attrition Trends by Job Role and Department:")
print(attrition_by_jobrole_department)

attrition_by_jobrole_department.plot(kind='bar', stacked=True)
plt.xlabel('Job Role - Department')
plt.ylabel('Count')
plt.title('Attrition Trends by Job Role and Department')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Job Role and Department:

Attrition		No	Yes
JobRole	Department		
Healthcare Representative	Research & Development	122.0	9.0
Human Resources	Human Resources	40.0	12.0
Laboratory Technician	Research & Development	197.0	62.0
Manager	Human Resources	11.0	NaN
	Research & Development	51.0	3.0
	Sales	35.0	2.0
Manufacturing Director	Research & Development	135.0	10.0
Research Director	Research & Development	78.0	2.0
Research Scientist	Research & Development	245.0	47.0
Sales Executive	Sales	269.0	57.0
Sales Representative	Sales	50.0	33.0



In []: *#Sub-Step 6.5: Attrition Trends by Education Field and Job Level*

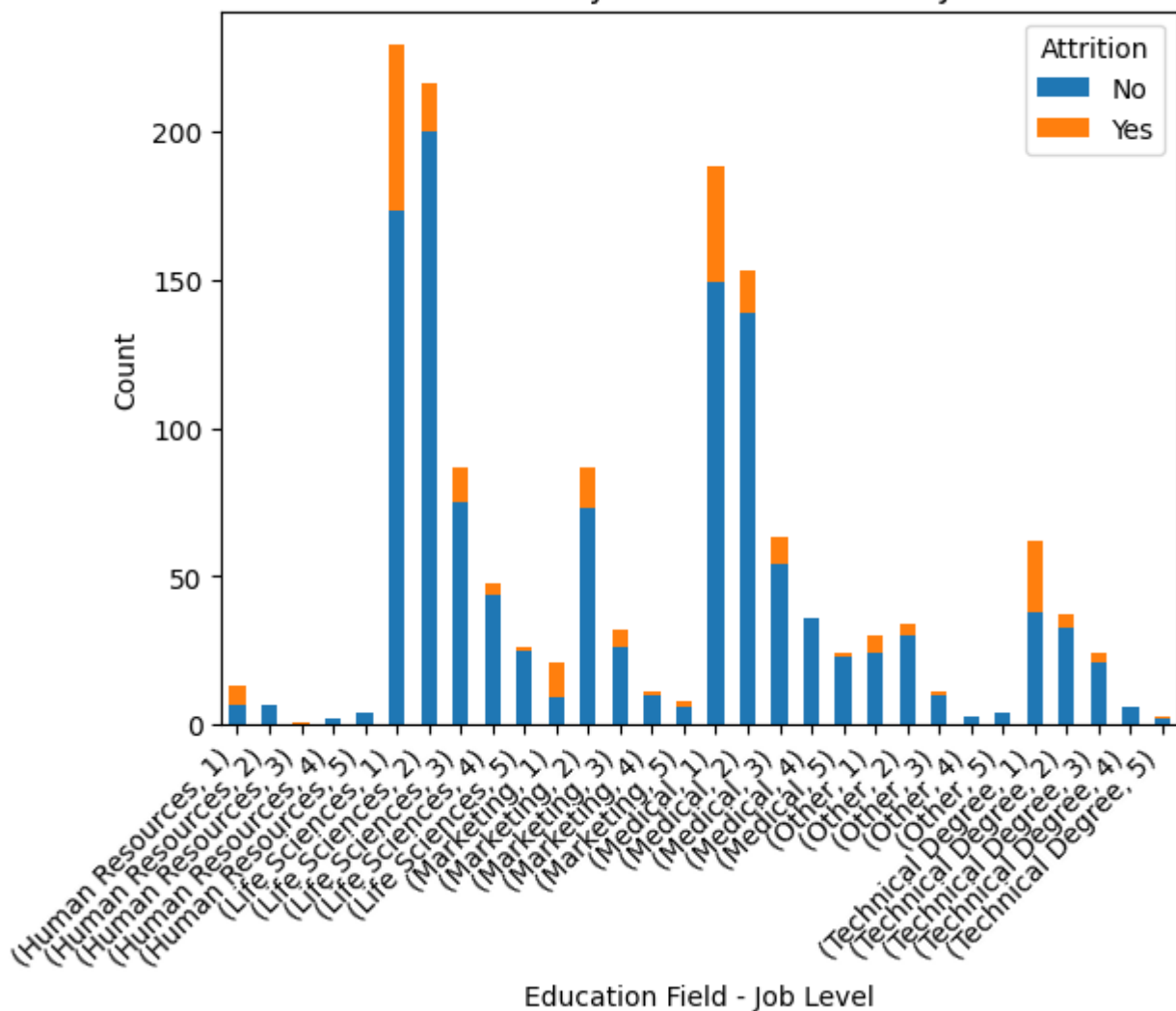
```
attrition_by_education_joblevel = df.groupby(['EducationField', 'JobLevel'])['Attrition'].value_counts()
print("Attrition Trends by Education Field and Job Level:")
print(attrition_by_education_joblevel)

attrition_by_education_joblevel.plot(kind='bar', stacked=True)
plt.xlabel('Education Field - Job Level')
plt.ylabel('Count')
plt.title('Attrition Trends by Education Field and Job Level')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Education Field and Job Level:

Attrition		No	Yes
EducationField	JobLevel		
Human Resources	1	7.0	6.0
	2	7.0	NaN
	3	NaN	1.0
	4	2.0	NaN
	5	4.0	NaN
Life Sciences	1	173.0	56.0
	2	200.0	16.0
	3	75.0	12.0
	4	44.0	4.0
	5	25.0	1.0
Marketing	1	9.0	12.0
	2	73.0	14.0
	3	26.0	6.0
	4	10.0	1.0
	5	6.0	2.0
Medical	1	149.0	39.0
	2	139.0	14.0
	3	54.0	9.0
	4	36.0	NaN
	5	23.0	1.0
Other	1	24.0	6.0
	2	30.0	4.0
	3	10.0	1.0
	4	3.0	NaN
	5	4.0	NaN
Technical Degree	1	38.0	24.0
	2	33.0	4.0
	3	21.0	3.0
	4	6.0	NaN
	5	2.0	1.0

Attrition Trends by Education Field and Job Level



In []: *#Sub-Step 6.6: Attrition Trends by Marital Status and Total Working Years*

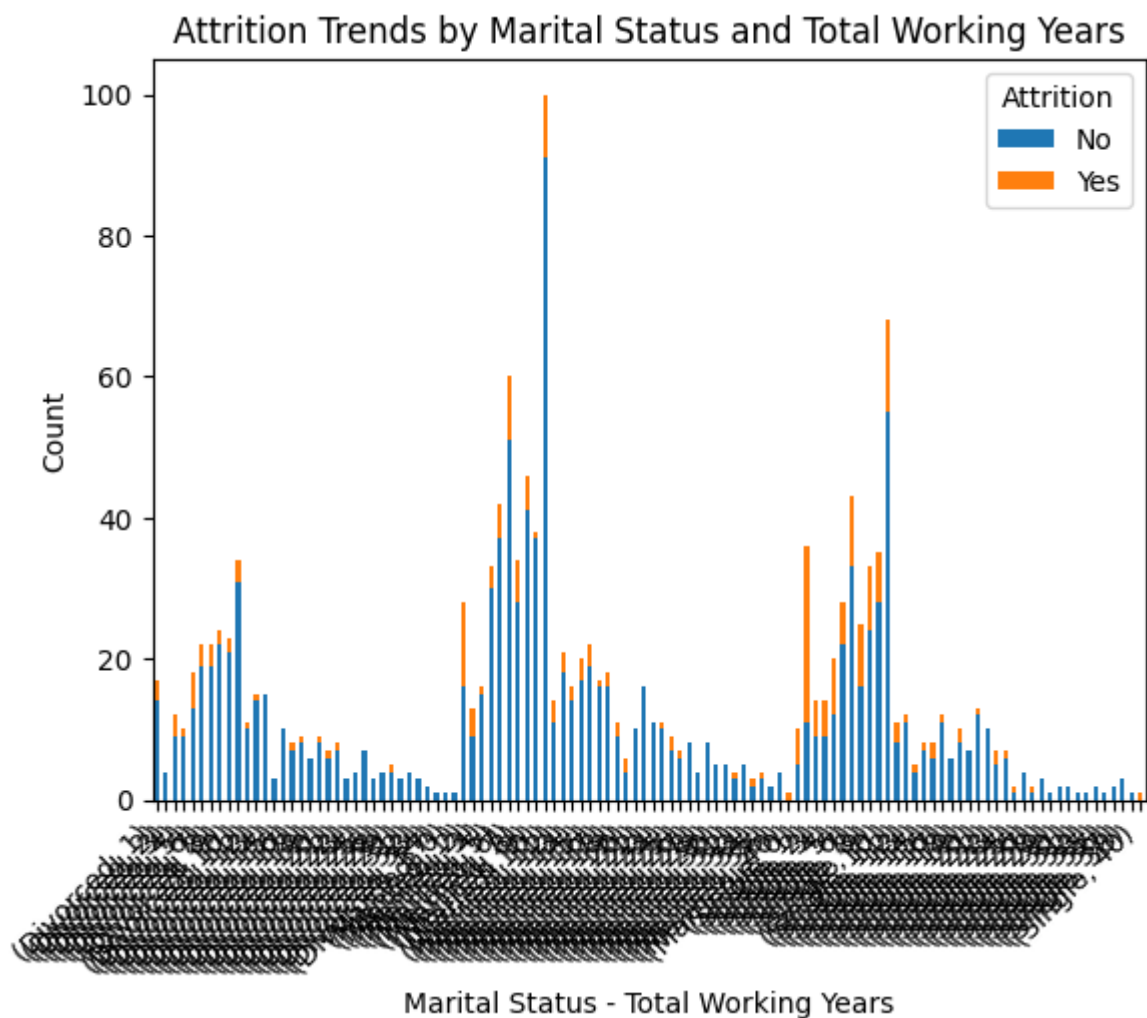
```
attrition_by_marital_totalyears = df.groupby(['MaritalStatus', 'TotalWorkingYears'])['Attrition']
print("Attrition Trends by Marital Status and Total Working Years:")
print(attrition_by_marital_totalyears)

attrition_by_marital_totalyears.plot(kind='bar', stacked=True)
plt.xlabel('Marital Status - Total Working Years')
plt.ylabel('Count')
plt.title('Attrition Trends by Marital Status and Total Working Years')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Marital Status and Total Working Years:

Attrition		No		Yes	
MaritalStatus	TotalWorkingYears				
Divorced	1	14.0	3.0		
	2	4.0	NaN		
	3	9.0	3.0		
	4	9.0	1.0		
	5	13.0	5.0		
...					
Single	35	1.0	NaN		
	36	2.0	NaN		
	37	3.0	NaN		
	38	1.0	NaN		
	40	NaN	1.0		

[110 rows x 2 columns]



```
In [ ]: #Sub-Step 6.7: Attrition Trends by Business Travel, Job Role, and Education Field

attrition_by_travel_jobrole_education = df.groupby(['BusinessTravel', 'JobRole', 'EducationField'])
print("Attrition Trends by Business Travel, Job Role, and Education Field:")
print(attrition_by_travel_jobrole_education)

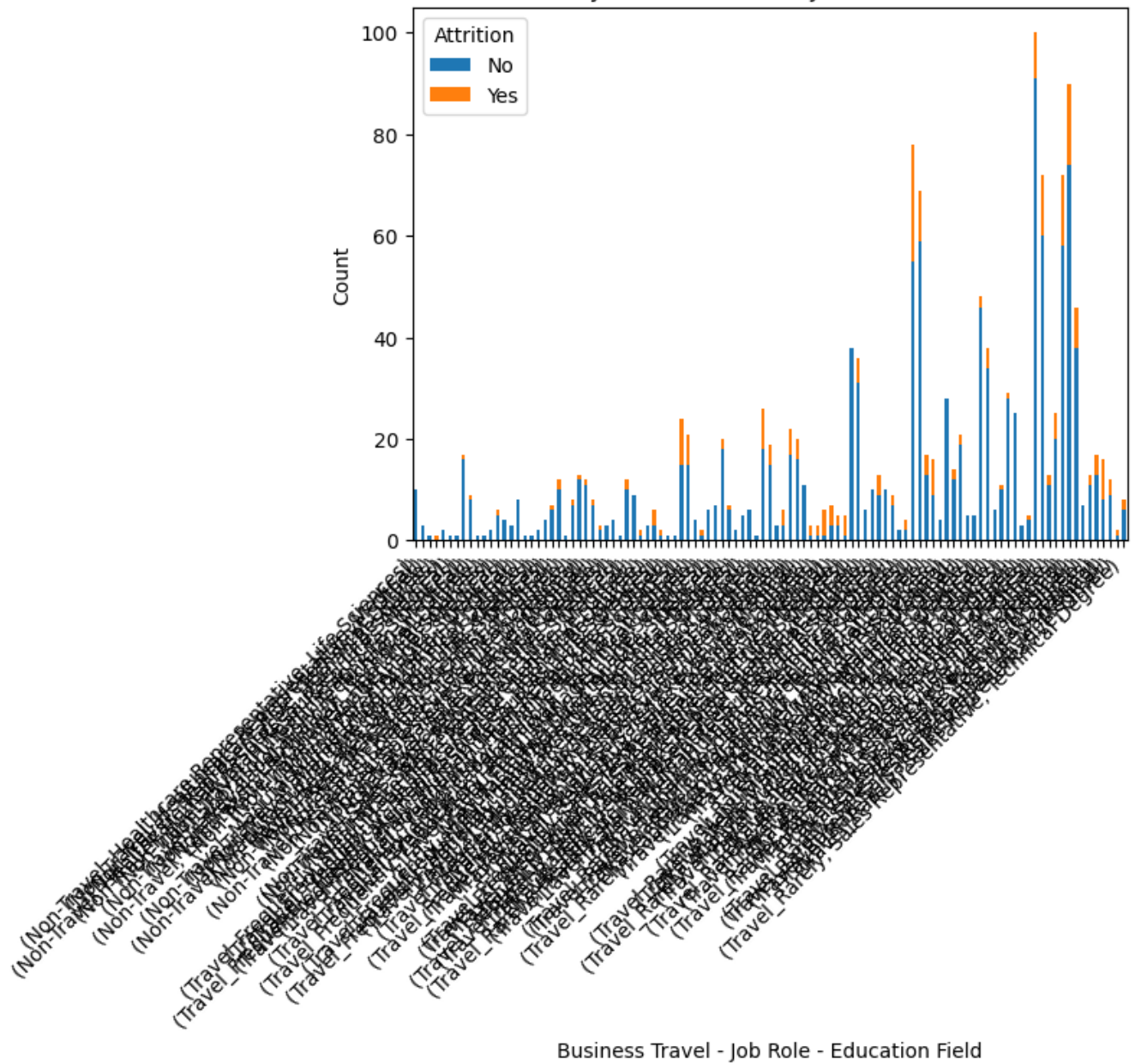
attrition_by_travel_jobrole_education.plot(kind='bar', stacked=True)
plt.xlabel('Business Travel - Job Role - Education Field')
plt.ylabel('Count')
plt.title('Attrition Trends by Business Travel, Job Role, and Education Field')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Business Travel, Job Role, and Education Field:

Attrition				No	Yes
BusinessTravel	JobRole		EducationField		
Non-Travel	Healthcare Representative	Life Sciences	10.0	NaN	
		Medical	3.0	NaN	
		Other	1.0	NaN	
		Technical Degree	NaN	1.0	
		Human Resources	2.0	NaN	
...			
Travel_Rarely	Sales Representative	Life Sciences	13.0	4.0	
		Marketing	8.0	8.0	
		Medical	9.0	3.0	
		Other	1.0	1.0	
		Technical Degree	6.0	2.0	

[105 rows x 2 columns]

Attrition Trends by Business Travel, Job Role, and Education Field



Business Travel - Job Role - Education Field

```
In [ ]: #Sub-Step 6.8: Attrition Trends by Job Role, Department, and Performance Rating

attrition_by_jobrole_department_performance = df.groupby(['JobRole', 'Department', 'Performance Rating'])
print("Attrition Trends by Job Role, Department, and Performance Rating:")
print(attrition_by_jobrole_department_performance)

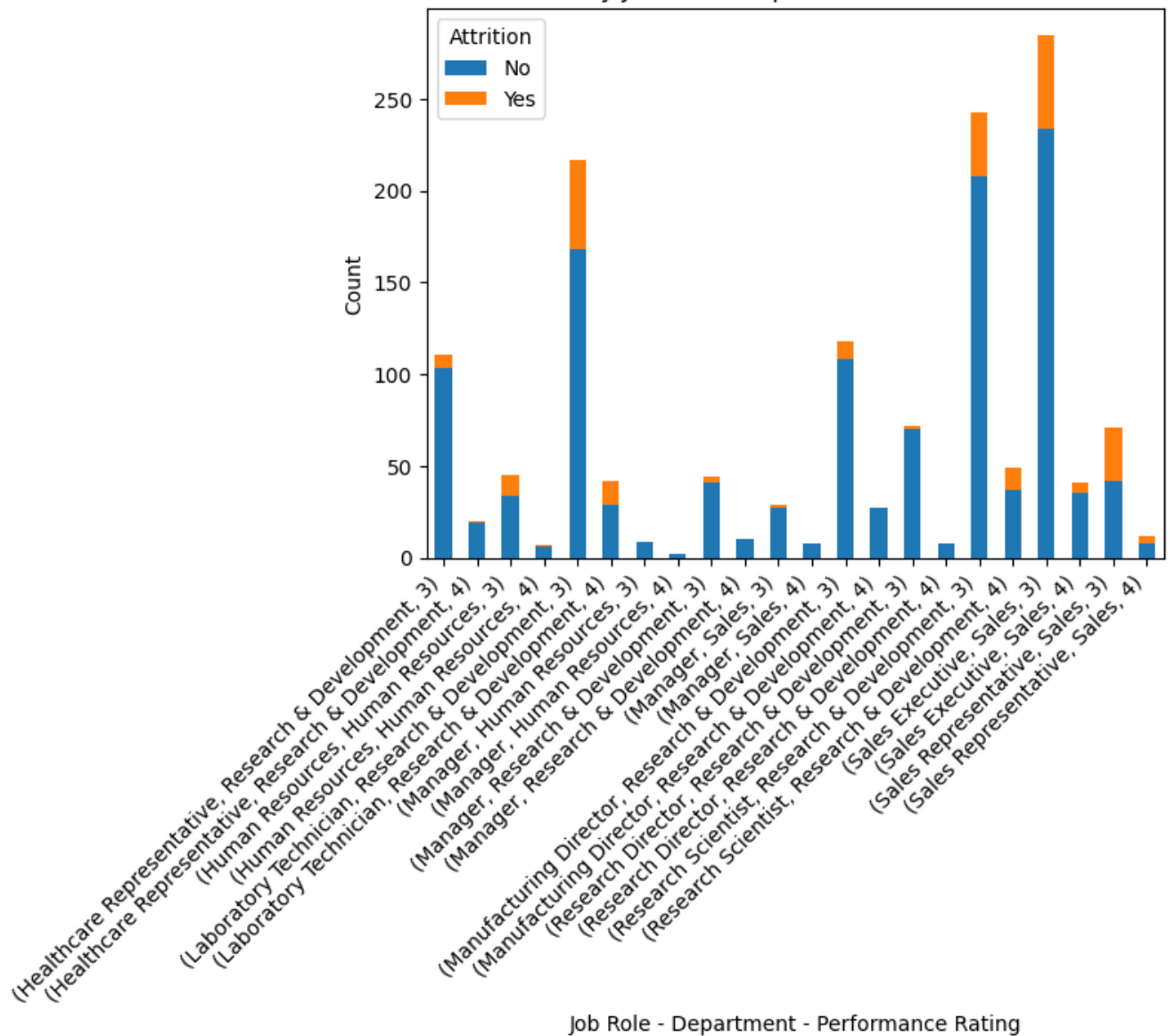
attrition_by_jobrole_department_performance.plot(kind='bar', stacked=True)
plt.xlabel('Job Role - Department - Performance Rating')
plt.ylabel('Count')
plt.title('Attrition Trends by Job Role, Department, and Performance Rating')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Job Role, Department, and Performance Rating:

Attrition			No \
JobRole	Department	PerformanceRating	
Healthcare Representative	Research & Development	3	103.0
		4	19.0
Human Resources	Human Resources	3	34.0
		4	6.0
Laboratory Technician	Research & Development	3	168.0
		4	29.0
Manager	Human Resources	3	9.0
		4	2.0
	Research & Development	3	41.0
		4	10.0
	Sales	3	27.0
		4	8.0
	Research & Development	3	108.0
		4	27.0
Research Director	Research & Development	3	70.0
		4	8.0
Research Scientist	Research & Development	3	208.0
		4	37.0
Sales Executive	Sales	3	234.0
		4	35.0
Sales Representative	Sales	3	42.0
		4	8.0

Attrition			Yes
JobRole	Department	PerformanceRating	
Healthcare Representative	Research & Development	3	8.0
		4	1.0
Human Resources	Human Resources	3	11.0
		4	1.0
Laboratory Technician	Research & Development	3	49.0
		4	13.0
Manager	Human Resources	3	NaN
		4	NaN
	Research & Development	3	3.0
		4	NaN
	Sales	3	2.0
		4	NaN
	Research & Development	3	10.0
		4	NaN
Research Director	Research & Development	3	2.0
		4	NaN
Research Scientist	Research & Development	3	35.0
		4	12.0
Sales Executive	Sales	3	51.0
		4	6.0
Sales Representative	Sales	3	29.0
		4	4.0

Attrition Trends by Job Role, Department, and Performance Rating



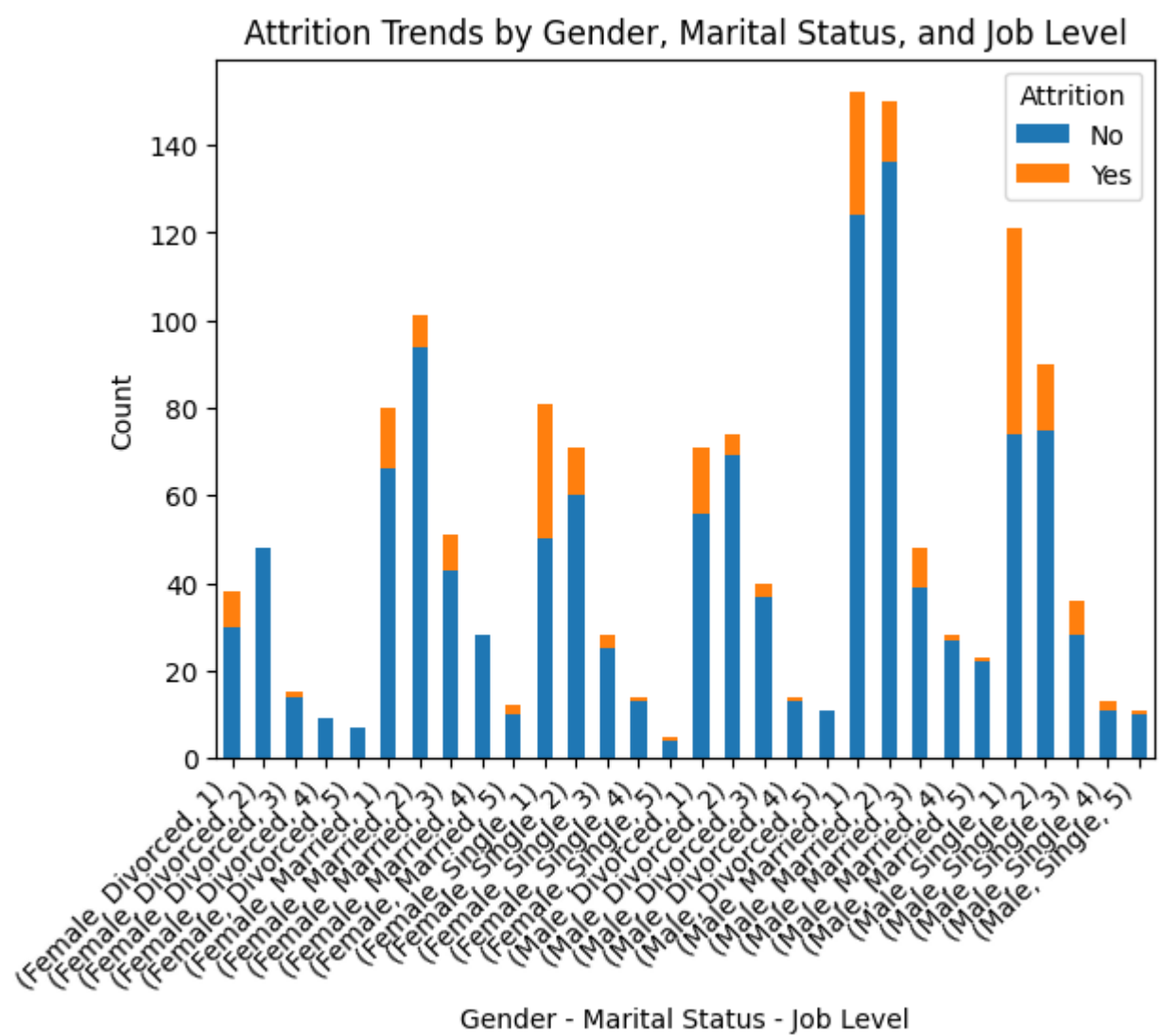
In []: *#Sub-Step 6.9: Attrition Trends by Gender, Marital Status, and Job Level*

```
attrition_by_gender_marital_joblevel = df.groupby(['Gender', 'MaritalStatus', 'JobLevel'])['Attrition']
print("Attrition Trends by Gender, Marital Status, and Job Level:")
print(attrition_by_gender_marital_joblevel)

attrition_by_gender_marital_joblevel.plot(kind='bar', stacked=True)
plt.xlabel('Gender - Marital Status - Job Level')
plt.ylabel('Count')
plt.title('Attrition Trends by Gender, Marital Status, and Job Level')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Gender, Marital Status, and Job Level:

Attrition		No	Yes
Gender	MaritalStatus	JobLevel	
Female	Divorced	1	30.0
		2	48.0
		3	14.0
		4	9.0
		5	7.0
	Married	1	66.0
		2	94.0
		3	43.0
		4	28.0
		5	10.0
	Single	1	50.0
		2	60.0
		3	25.0
		4	13.0
		5	4.0
Male	Divorced	1	56.0
		2	69.0
		3	37.0
		4	13.0
		5	11.0
	Married	1	124.0
		2	136.0
		3	39.0
		4	27.0
		5	22.0
	Single	1	74.0
		2	75.0
		3	28.0
		4	11.0
		5	10.0



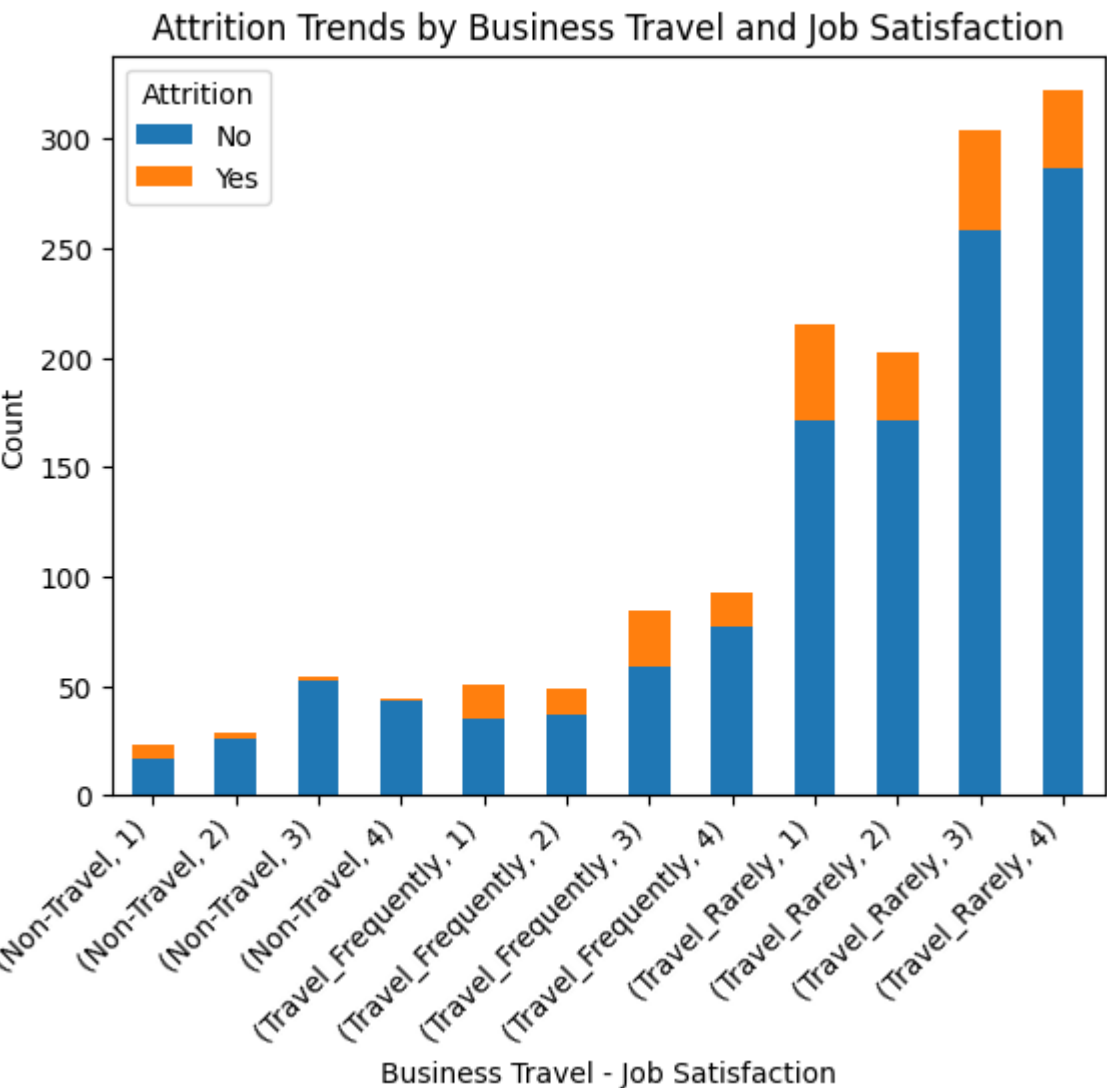
In []: #Sub-Step 6.10: Attrition Trends by Business Travel and Job Satisfaction

```
attrition_by_travel_jobsatisfaction = df.groupby(['BusinessTravel', 'JobSatisfaction'])['Attrition'].count()
print("Attrition Trends by Business Travel and Job Satisfaction:")
print(attrition_by_travel_jobsatisfaction)

attrition_by_travel_jobsatisfaction.plot(kind='bar', stacked=True)
plt.xlabel('Business Travel - Job Satisfaction')
plt.ylabel('Count')
plt.title('Attrition Trends by Business Travel and Job Satisfaction')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Business Travel and Job Satisfaction:

Attrition		No	Yes
BusinessTravel	JobSatisfaction		
Non-Travel	1	17	6
	2	26	3
	3	52	2
	4	43	1
Travel_Frequently	1	35	16
	2	37	12
	3	59	25
	4	77	16
Travel_Rarely	1	171	44
	2	171	31
	3	258	46
	4	287	35



In []: #Sub-Step 6.11: Attrition Trends by Distance from Home and Relationship Satisfaction

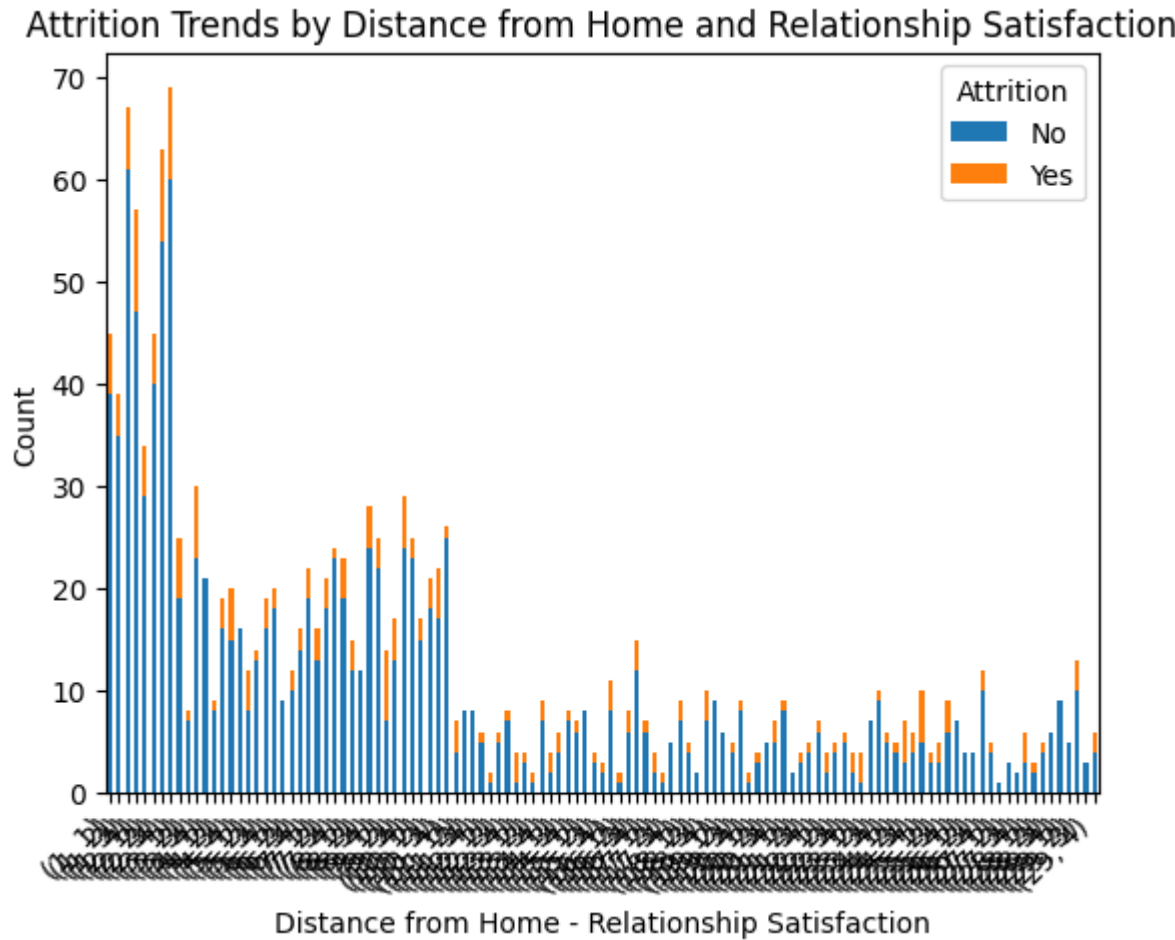
```
attrition_by_distance_relationship = df.groupby(['DistanceFromHome', 'RelationshipSatisfaction'])
print("Attrition Trends by Distance from Home and Relationship Satisfaction:")
print(attrition_by_distance_relationship)

attrition_by_distance_relationship.plot(kind='bar', stacked=True)
plt.xlabel('Distance from Home - Relationship Satisfaction')
plt.ylabel('Count')
plt.title('Attrition Trends by Distance from Home and Relationship Satisfaction')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Distance from Home and Relationship Satisfaction:

Attrition		No	Yes
DistanceFromHome	RelationshipSatisfaction		
1	1	39.0	6.0
	2	35.0	4.0
	3	61.0	6.0
	4	47.0	10.0
2	1	29.0	5.0
...	
28	4	9.0	NaN
29	1	5.0	NaN
	2	10.0	3.0
	3	3.0	NaN
	4	4.0	2.0

[115 rows x 2 columns]



In []: *#Sub-Step 6.12: Attrition Trends by OverTime and Years in Current Role*

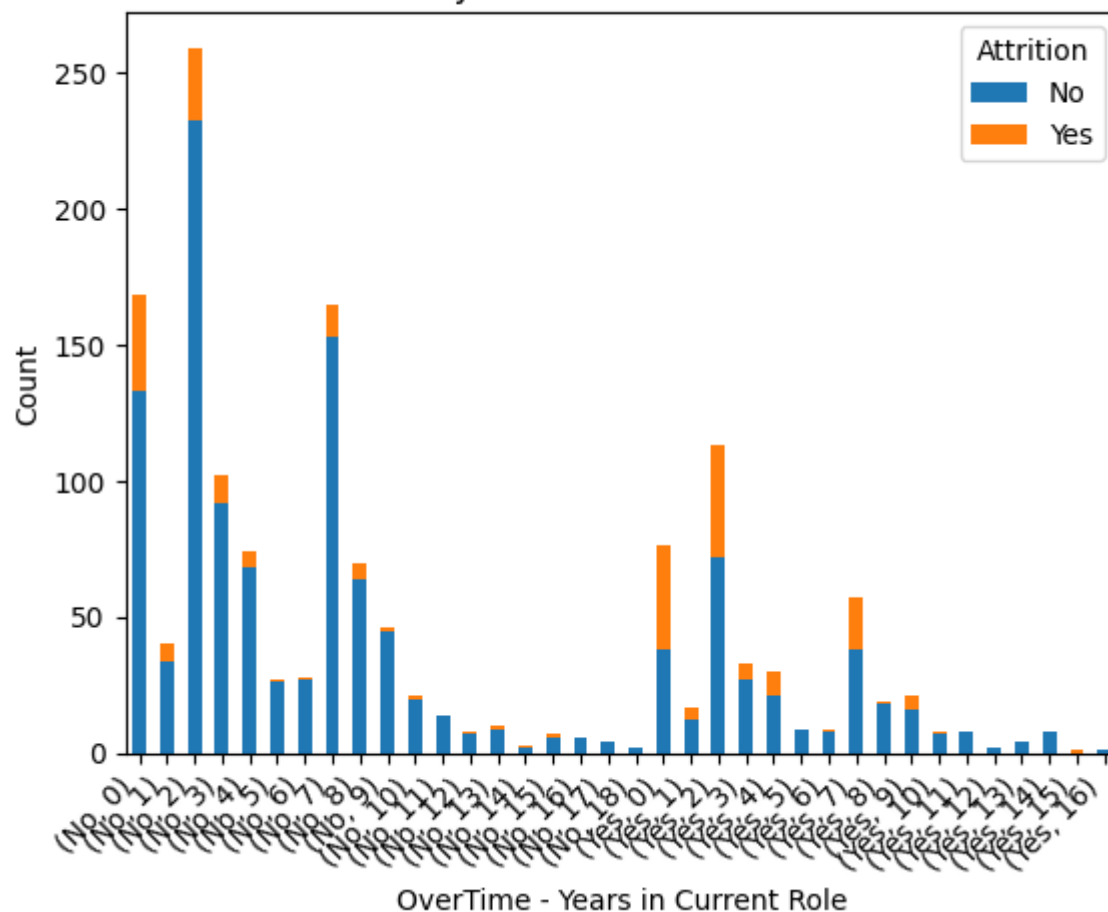
```
attrition_by_overtime_yearsrole = df.groupby(['OverTime', 'YearsInCurrentRole'])['Attrition'].
print("Attrition Trends by OverTime and Years in Current Role:")
print(attrition_by_overtime_yearsrole)

attrition_by_overtime_yearsrole.plot(kind='bar', stacked=True)
plt.xlabel('OverTime - Years in Current Role')
plt.ylabel('Count')
plt.title('Attrition Trends by OverTime and Years in Current Role')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by OverTime and Years in Current Role:

Attrition		No	Yes
OverTime	YearsInCurrentRole		
No	0	133.0	35.0
	1	34.0	6.0
	2	232.0	27.0
	3	92.0	10.0
	4	68.0	6.0
	5	26.0	1.0
	6	27.0	1.0
	7	153.0	12.0
	8	64.0	6.0
	9	45.0	1.0
	10	20.0	1.0
	11	14.0	NaN
	12	7.0	1.0
	13	9.0	1.0
	14	2.0	1.0
	15	6.0	1.0
	16	6.0	NaN
	17	4.0	NaN
Yes	18	2.0	NaN
	0	38.0	38.0
	1	12.0	5.0
	2	72.0	41.0
	3	27.0	6.0
	4	21.0	9.0
	5	9.0	NaN
	6	8.0	1.0
	7	38.0	19.0
	8	18.0	1.0
	9	16.0	5.0
	10	7.0	1.0
	11	8.0	NaN
	12	2.0	NaN
	13	4.0	NaN
	14	8.0	NaN
	15	NaN	1.0
	16	1.0	NaN

Attrition Trends by OverTime and Years in Current Role



```
In [ ]: #Sub-Step 6.13: Attrition Trends by Percent Salary Hike and Years Since Last Promotion

attrition_by_hike_promotion = df.groupby(['PercentSalaryHike', 'YearsSinceLastPromotion'])['Attrition']
print("Attrition Trends by Percent Salary Hike and Years Since Last Promotion:")
print(attrition_by_hike_promotion)

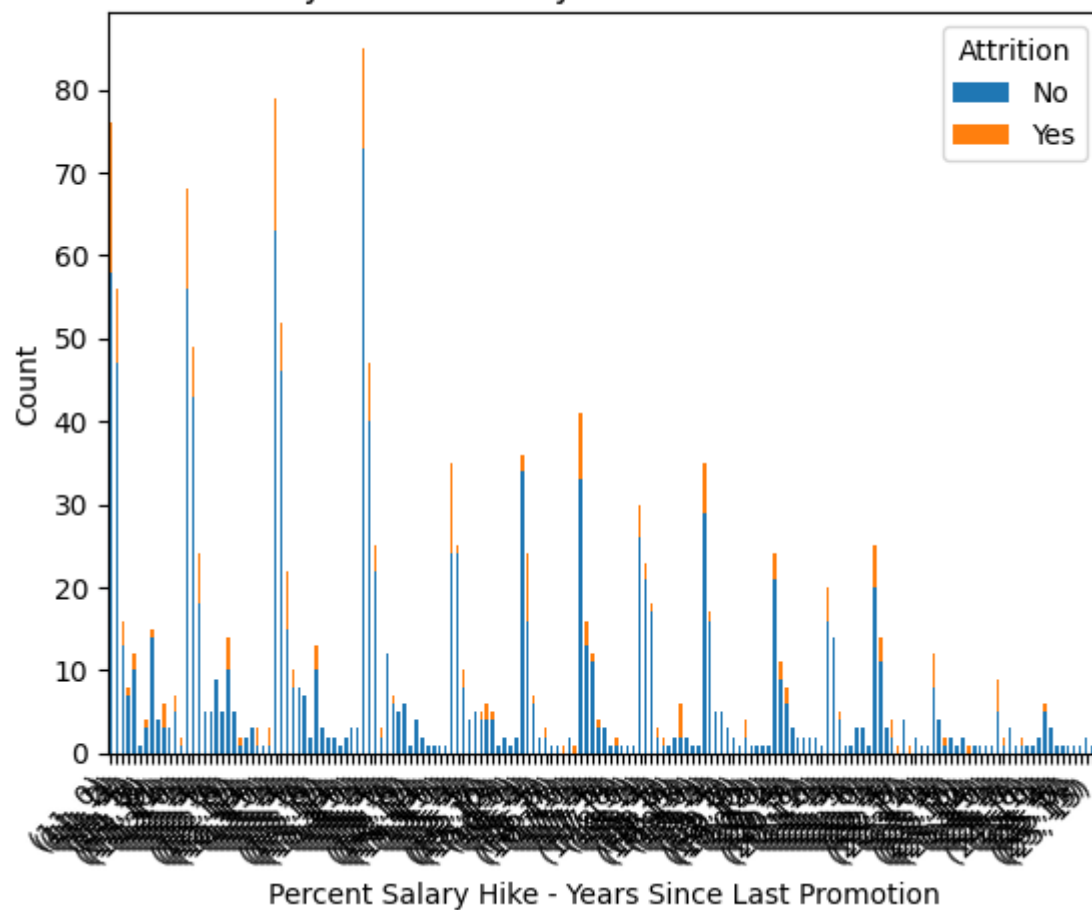
attrition_by_hike_promotion.plot(kind='bar', stacked=True)
plt.xlabel('Percent Salary Hike - Years Since Last Promotion')
plt.ylabel('Count')
plt.title('Attrition Trends by Percent Salary Hike and Years Since Last Promotion')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Percent Salary Hike and Years Since Last Promotion:

Attrition		No	Yes
PercentSalaryHike	YearsSinceLastPromotion		
11	0	58.0	18.0
	1	47.0	9.0
	2	13.0	3.0
	3	7.0	1.0
	4	10.0	2.0
...
25	6	1.0	NaN
	7	1.0	NaN
	11	2.0	NaN
	13	1.0	NaN
	15	1.0	NaN

[169 rows x 2 columns]

Attrition Trends by Percent Salary Hike and Years Since Last Promotion



```
In [ ]: #Sub-Step 6.14: Attrition Trends by Years with Current Manager and Num Companies Worked

attrition_by_manager_companies = df.groupby(['YearsWithCurrManager', 'NumCompaniesWorked'])['Attrition'].count()
print("Attrition Trends by Years with Current Manager and Num Companies Worked:")
print(attrition_by_manager_companies)

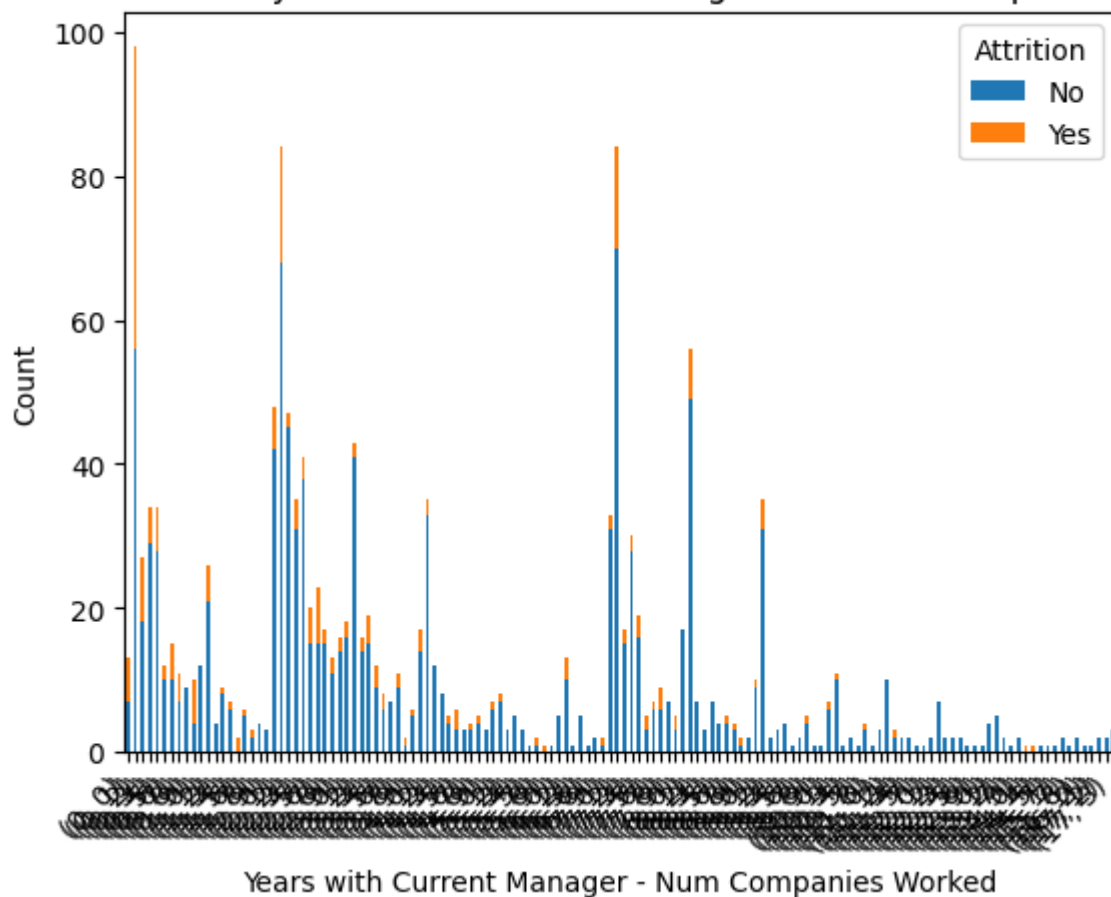
attrition_by_manager_companies.plot(kind='bar', stacked=True)
plt.xlabel('Years with Current Manager - Num Companies Worked')
plt.ylabel('Count')
plt.title('Attrition Trends by Years with Current Manager and Num Companies Worked')
plt.legend(title='Attrition')
plt.xticks(rotation=45, ha='right')
plt.show()
```

Attrition Trends by Years with Current Manager and Num Companies Worked:

Attrition		No	Yes
YearsWithCurrManager	NumCompaniesWorked		
0	0	7.0	6.0
	1	56.0	42.0
	2	18.0	9.0
	3	29.0	5.0
	4	28.0	6.0
...
16	1	1.0	NaN
	2	1.0	NaN
17	1	2.0	NaN
	2	2.0	NaN
	3	3.0	NaN

[136 rows x 2 columns]

Attrition Trends by Years with Current Manager and Num Companies Worked



In []: *#Step 7: analysis the impact of attrited employees on the overall performance and productivity*

#Sub-Step 7.1: Calculate the attrition rate

```
attrition_rate = df['Attrition'].value_counts(normalize=True) * 100
print("Attrition Rate:")
print(attrition_rate)
```

Attrition Rate:

No 83.877551

Yes 16.122449

Name: Attrition, dtype: float64

In []: *#Sub-Step 7.2: Analyze performance and productivity metrics for attrited and non-attrited empl*

```
attrited_employees = df[df['Attrition'] == 'Yes']
non_attrited_employees = df[df['Attrition'] == 'No']

attrited_performance = attrited_employees['PerformanceRating'].mean()
non_attrited_performance = non_attrited_employees['PerformanceRating'].mean()

attrited_productivity = attrited_employees['TotalWorkingYears'].mean()
non_attrited_productivity = non_attrited_employees['TotalWorkingYears'].mean()

print("Performance Rating - Attrited Employees: {:.2f}".format(attrited_performance))
print("Performance Rating - Non-Attrited Employees: {:.2f}".format(non_attrited_performance))
print("Total Working Years - Attrited Employees: {:.2f}".format(attrited_productivity))
print("Total Working Years - Non-Attrited Employees: {:.2f}".format(non_attrited_productivity))
```

Performance Rating - Attrited Employees: 3.16

Performance Rating - Non-Attrited Employees: 3.15

Total Working Years - Attrited Employees: 8.24

Total Working Years - Non-Attrited Employees: 11.86

In []: *#Sub-Step 7.3: Analyze impact and productivity by department*

```
department_impact = df.groupby('Department')['Attrition'].value_counts(normalize=True) * 100
department_productivity = df.groupby('Department')['TotalWorkingYears'].mean()

print("Attrition Impact by Department:")
print(department_impact)
print("Average Total Working Years by Department:")
print(department_productivity)
```

Attrition Impact by Department:

Department	Attrition	
Human Resources	No	80.952381
	Yes	19.047619
Research & Development	No	86.160250
	Yes	13.839750
Sales	No	79.372197
	Yes	20.627803

Name: Attrition, dtype: float64

Average Total Working Years by Department:

Department	
Human Resources	11.555556
Research & Development	11.342352
Sales	11.105381

Name: TotalWorkingYears, dtype: float64

In []: *#Sub-Step 7.4: Analyze impact and productivity by job role*

```
jobrole_impact = df.groupby('JobRole')['Attrition'].value_counts(normalize=True) * 100
jobrole_productivity = df.groupby('JobRole')['TotalWorkingYears'].mean()

print("Attrition Impact by Job Role:")
print(jobrole_impact)
print("Average Total Working Years by Job Role:")
print(jobrole_productivity)
```

Attrition Impact by Job Role:		
JobRole	Attrition	
Healthcare Representative	No	93.129771
	Yes	6.870229
Human Resources	No	76.923077
	Yes	23.076923
Laboratory Technician	No	76.061776
	Yes	23.938224
Manager	No	95.098039
	Yes	4.901961
Manufacturing Director	No	93.103448
	Yes	6.896552
Research Director	No	97.500000
	Yes	2.500000
Research Scientist	No	83.904110
	Yes	16.095890
Sales Executive	No	82.515337
	Yes	17.484663
Sales Representative	No	60.240964
	Yes	39.759036

Name: Attrition, dtype: float64

Average Total Working Years by Job Role:

JobRole	
Healthcare Representative	14.068702
Human Resources	8.173077
Laboratory Technician	7.656371
Manager	24.549020
Manufacturing Director	12.786207
Research Director	21.400000
Research Scientist	7.715753
Sales Executive	11.101227
Sales Representative	4.674699

Name: TotalWorkingYears, dtype: float64

In []: *#Sub-Step 7.5: Analyze impact and productivity by job Level*

```

joblevel_impact = df.groupby('JobLevel')['Attrition'].value_counts(normalize=True) * 100
joblevel_productivity = df.groupby('JobLevel')['TotalWorkingYears'].mean()

print("Attrition Impact by Job Level:")
print(joblevel_impact)
print("Average Total Working Years by Job Level:")
print(joblevel_productivity)

```

Attrition Impact by Job Level:

JobLevel	Attrition	
1	No	73.664825
	Yes	26.335175
2	No	90.262172
	Yes	9.737828
3	No	85.321101
	Yes	14.678899
4	No	95.283019
	Yes	4.716981
5	No	92.753623
	Yes	7.246377

Name: Attrition, dtype: float64

Average Total Working Years by Job Level:

JobLevel	
1	5.891344
2	10.423221
3	15.119266
4	25.471698
5	26.376812

Name: TotalWorkingYears, dtype: float64