# IBM Employee Attrition and Performance Analysis Using Pandas

Submitted By:

JISA VARGHESE

# TABLE OF CONTENTS

# INTRODUCTION

Employee attrition is a critical concern for organizations striving to retain top talent and maintain operational efficiency. High turnover rates can lead to increased costs, loss of institutional knowledge, and disruption in workplace culture. Analyzing attrition trends and performance data can help organizations proactively address the underlying factors contributing to employee turnover.

This report leverages **Pandas and Seaborn library** to perform an in-depth analysis of the **IBM HR Analytics Employee Attrition & Performance** dataset. Pandas provides robust tools for data manipulation, allowing for the identification of key patterns, trends, and relationships within the dataset.

Through this report, we aim to uncover actionable insights that can help HR teams implement strategies to reduce attrition, improve employee satisfaction, and optimize workforce performance.

# OBJECTIVES

The primary objectives of this Pandas-based analysis are:

1. **Understand Employee Attrition Patterns**
    - Analyze attrition trends by age, gender, marital status, and department.
    - Identify which groups are most vulnerable to attrition.
2. **Evaluate Income and Performance Relationships**
    - Explore the correlation between monthly income, performance ratings, and attrition.
    - Assess whether financial incentives could help reduce turnover.
3. **Analyze the Impact of Tenure and Distance**
    - Investigate how factors like years with the company, years under the current manager, and commuting distance influence attrition.
4. **Provide Actionable Insights**
    - Use findings to propose strategic interventions to improve retention and enhance employee satisfaction. This analysis aims to empower HR teams with data-driven recommendations to address workforce challenges effectively.

# DATASET OVERVIEW

The dataset consists of **1,470** records and **35** features, including demographic details, professional characteristics, satisfaction levels, and attrition status.

## ATTRIBUTES:

- **Age**: Age of the employee in years.
- **Attrition**: Whether the employee left the company (Yes/No).
- **BusinessTravel**: Frequency of business travel (e.g., Rarely, Often).
- **DailyRate**: Daily income of the employee in USD.
- **Department**: Department of the employee (e.g., Sales, R&D).
- **DistanceFromHome**: Distance from home to the workplace (in miles).
- **Education**: Education level (1: Below College, 5: Doctorate).
- **EducationField**: Field of education (e.g., Life Sciences, Technical Degree, Marketing).
- **EmployeeCount**: Always 1
- **EmployeeNumber**: Unique identifier for each employee (not useful for analysis).
- **EnvironmentSatisfaction**: Employee's satisfaction with the work environment (1-4).

- **Gender**: Gender of the employee (Male/Female).
- **HourlyRate**: Hourly rate of the employee's pay.
- **JobInvolvement**: Employee's involvement level in their job (1-4).
- **JobLevel:** Level of the job (1-5).
- **JobRole**: Designation of the employee (e.g., Manager, Analyst).
- **JobSatisfaction**: Employee's job satisfaction (1: Low, 4: High).
- **MaritalStatus**: Marital status of the employee (Single/Married/Divorced).
- **MonthlyIncome**: Monthly salary in USD.
- **MonthlyRate**: Monthly rate of the employee's pay.
- **NumCompaniesWorked**: Number of companies the employee has worked for.
- **Over18**: whether an employee is above the age of 18(yes/no).
- **OverTime**: Whether the employee works overtime (Yes/No).
- **PercentSalaryHike**: Percentage increase in salary.
- **PerformanceRating**: Performance rating (1: Low, 4: Excellent).
- **RelationshipSatisfaction**: Satisfaction level with workplace relationships (1-4).
- **StandardHours**: Standard hours (always 80, irrelevant for analysis).
- **StockOptionLevel**: Level of stock options granted to the employee (0-3).

- **TotalWorkingYears**: Total years of professional experience.
- **TrainingTimesLastYear**: Number of training sessions attended last year.
- **WorkLifeBalance**: Work-life balance rating (1-4).
- **YearsAtCompany**: Total years spent at the company.
- **YearsInCurrentRole**: Number of years in the current role.
- **YearsSinceLastPromotion**: Years since the last promotion.
- **YearsWithCurrManager:** Years working with the current manager.

# EDA STEP BY STEP PROCESS

1. Understanding the Dataset

2. Data Cleaning

3. Data type Conversion

4. Feature Engineering

5.Outlier detection and removal

6. Data Analysis

   - Univariate analysis
   - Bivariate analysis
   - Multivariate analysis

7. Visualisation

8. Insights Extraction

9. Conclusion and Recommendations

# DATA PREPROCESSING

## 1.DATA UNDERSTANDING

Exploring the structure and characteristics of the dataset to identify key features for analysis.

```
df.shape()
df.info()
df.describe()
df.dtypes
df.head()
df.tail()
```

## 2.HANDLING MISSING AND DUPLICATE VALUES:

No missing or duplicate values were found in the dataset after checking for null entries.

```
df.isnull().sum()
df.duplicated().sum()
```

## 3.DATA TYPE CONVERSION:

Convert categorical variables such as **Gender**, **BusinessTravel**, **Attrition**, etc., to appropriate categorical types for analysis.

```
categorical_columns = ['BusinessTravel', 'Department', 'Gender', 'JobRole',
'MaritalStatus', 'OverTime', 'Attrition']
df[categorical_columns] = df[categorical_columns].apply(lambda x:
x.astype('category'))
```

## 4.FEATURE ENGINEERING:

Created new features, such as **AgeGroup**, **RecentlyPromoted**, **IncomeLevel**, based on domain knowledge.

```
def age_category(age):
    if age < 30:
        return 'Young'
    elif age < 50:
        return 'Adult'
    else:
        return 'Senior'
df['Age_Category'] = df['Age'].apply(age_category)

def income_category(income):
    if income < 4000:
        return 'Low'
    elif income < 8000:
        return 'Medium'
    elif income < 12000:
        return 'High'
    else:
      return 'Very High'
df['Income_Category'] = df['MonthlyIncome'].apply(income_category)
```

```
df['RecentlyPromoted'] = df['YearsSinceLastPromotion'].apply(lambda x: 'Yes' if x == 0 else 'No')
```

## 5.COLUMNS NOT USEFUL FOR ANALYSIS:

- **EmployeeCount:** All values are 1. This column does not provide any meaningful variation for analysis**.**
- **StandardHours:** All values are 80. This column is uniform and not useful for identifying trends or insights.
- **EmployeeNumber:** This is likely an identifier and not meaningful for analysis or modeling.
- **Over18:** Only contains the value 'Yes' for all rows

```
columns_to_drop = ['EmployeeCount', 'StandardHours', 'EmployeeNumber', 'Over18']
df = df.drop(columns=columns_to_drop, axis=1)
```

# 6.OUTLIER DETECTION USING IQR METHOD AND BOXPLOT

Outlier detection was performed on numerical columns using the Box plot method. Outliers that were deemed errors were removed, while genuine extreme values were retained for analysis**.**

```
def detect_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = data[(data[column] < lower_bound) | (data[column] > upper_bound)]
    print(f"Outliers detected in '{column}':")
    print(outliers)
    return outliers
numerical_columns = df.select_dtypes(include=['int64', 'float64']).columns
for col in numerical_columns:
    print(f"\nAnalyzing column: {col}")
    detect_outliers_iqr(df, col)
numerical_columns = df.select_dtypes(include=['float64', 'int64']).columns
for column in numerical_columns:
    plt.figure(figsize=(8, 5))
    sns.boxplot(data=df, x=column)
    plt.title(f'Box Plot for {column}')
    plt.show()
```

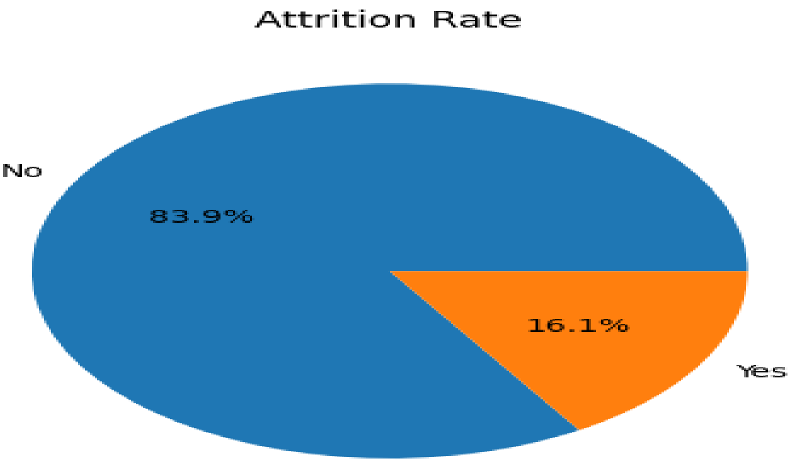
# DATA ANALYSIS

## 1.OVERALL ATTRITION RATE

The attrition rate is **16%** of total employees , indicating that approximately 1 in 6 employees have left the organization.

```
attrition_counts = df['Attrition'].value_counts()
print(attrition_counts)
attrition_percentages = df['Attrition'].value_counts(normalize=True) * 100
print(attrition_percentages)
attrition_counts.plot(kind='pie',autopct='%1.1f%%')
plt.title('Attrition Rate')
plt.ylabel('')
plt.show()
```

```
Attrition              Attrition   Percentage
No          1233             83.877551
Yes         237            16.122449
```



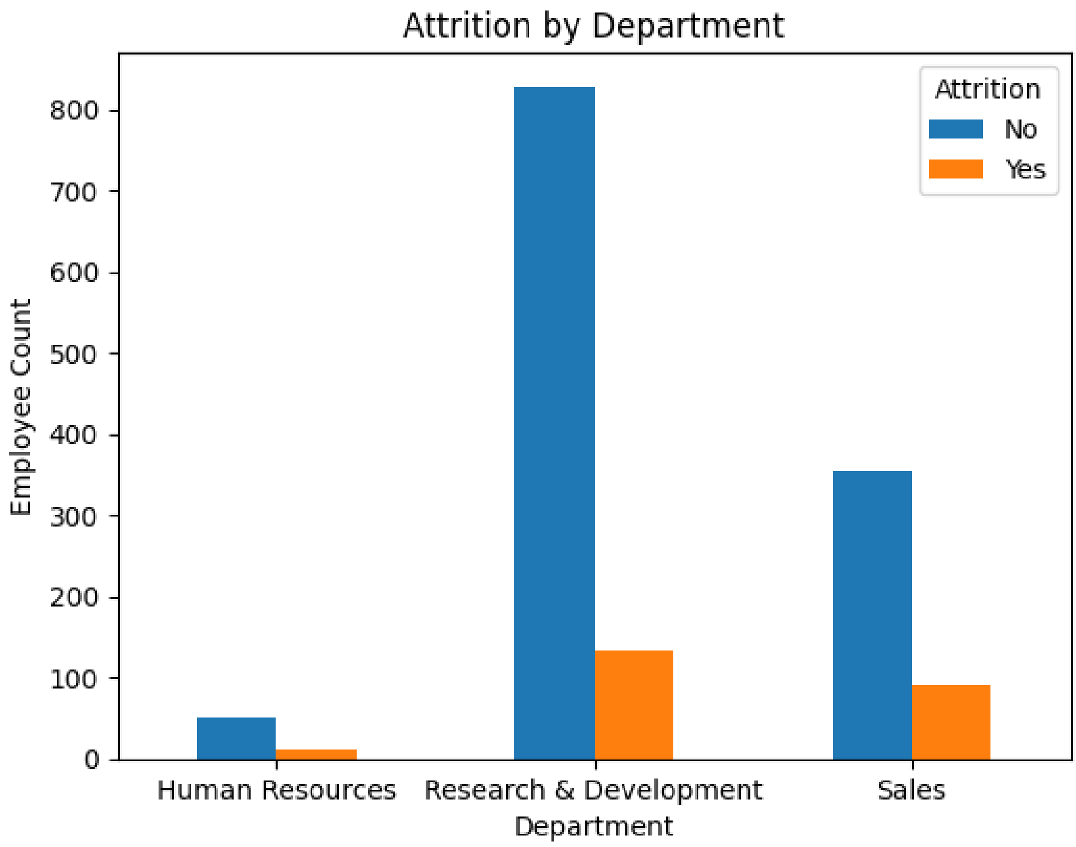
## 2. ATTRITION BY DEPARTMENT:

Highest attrition in **Sales (20%)** and **HR departments** (19%)

The high attrition rate in Sales and HR departments is likely due to high job stress, performance pressure in Sales, and emotional burnout or limited career growth opportunities in HR.

```
department_attrition =
df.groupby('Department')['Attrition'].value_counts().unstack()
print(department_attrition)
department_attrition_percent =
df.groupby('Department')['Attrition'].value_counts(normalize=True).unstack()
* 100
print(department_attrition_percent)
department_attrition.plot(kind='bar')
plt.title('Attrition by Department')
plt.ylabel('Employee Count')
plt.xlabel('Department')
plt.xticks(rotation=0)
plt.show()
```

| *Attrition* | *No* | *Yes* | *Attritionpercenatage* |
|---|---|---|---|
| *Department* | | | |
| *Human Resources* | *51* | *12* | *19.04* |
| *Research & Development* | *828* | *133* | *13.83* |
| *Sales* | *354* | *92* | *20.62* |

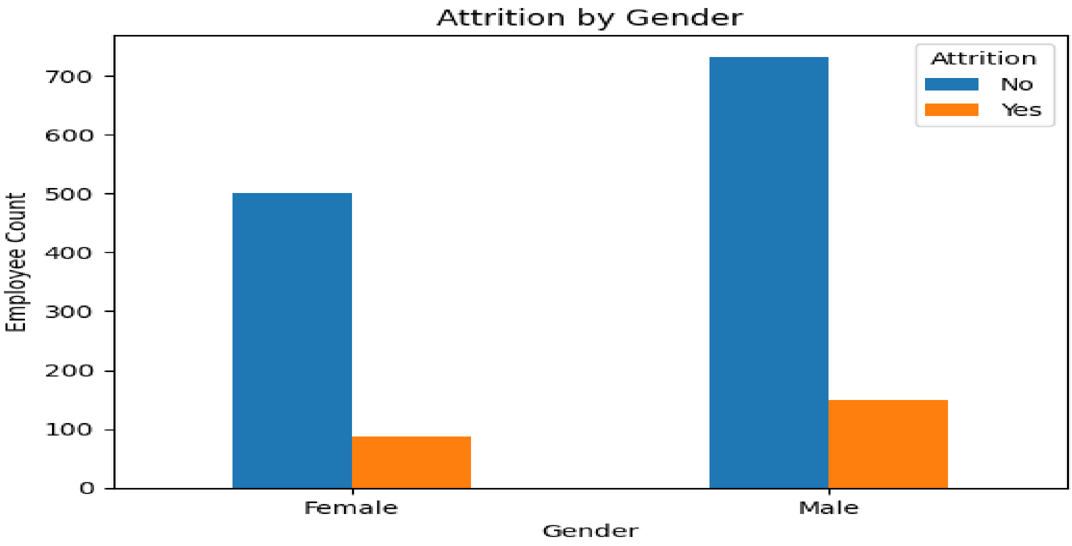


# 3.ATTRITION BY GENDER

**Males (17%)** have a slightly higher attrition rate than **females (15%).** This may be due to greater pursuit of external career opportunities or dissatisfaction with current job roles and compensation.

```
gender_attrition = df.groupby('Gender')['Attrition'].value_counts().unstack()
print(gender_attrition)
gender_attrition_percent =
df.groupby('Gender')['Attrition'].value_counts(normalize=True).unstack()*100
print(gender_attrition_percent)
gender_attrition.plot(kind='bar')
plt.title('Attrition by Gender')
plt.ylabel('Employee Count')
plt.xticks(rotation=0)
plt.show()
```

```
Attrition    No    Yes     Attrition percentage
Gender
Female      501    87          14.79
Male        732    150         17.00
```

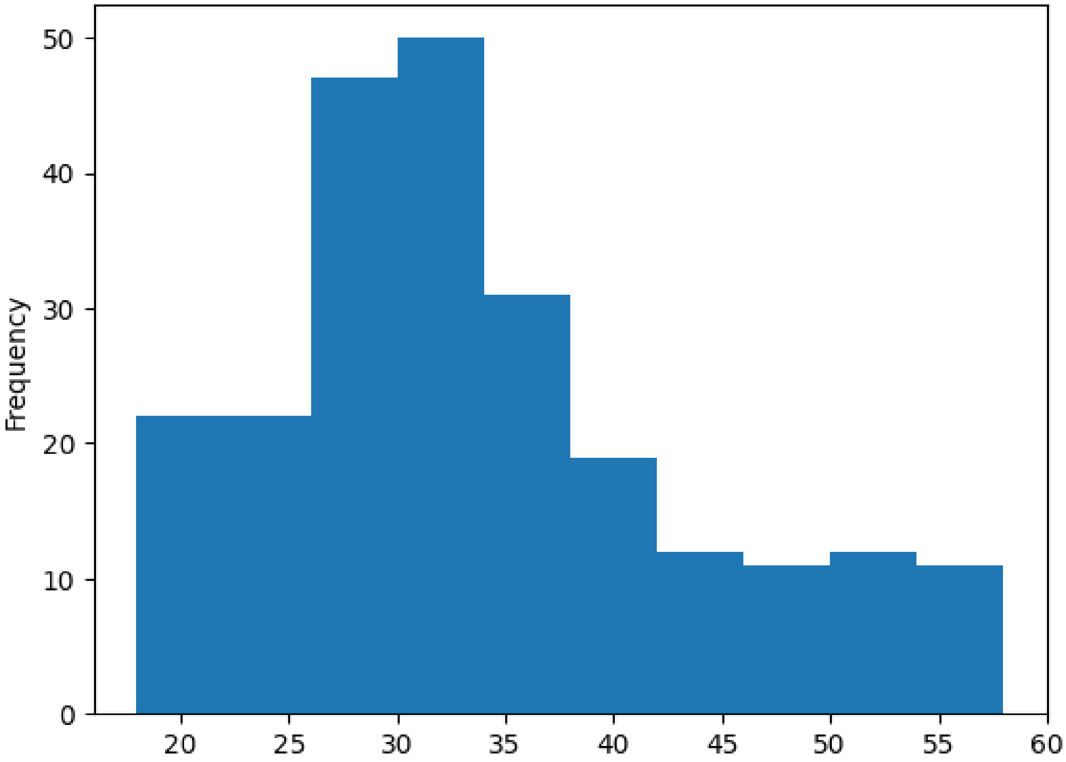


# 4.ATTRITION BY AGE GROUP:

Employees in the **20–30 age group (28%)** have the highest attrition rate, while those aged **40 and above (12.6%)** show significantly lower attrition rates.

This is due to career exploration, better job opportunities, and a desire for rapid growth, while older employees value stability and job security.

*df[df['Attrition'] == 'Yes']['Age'].plot(kind='hist')*

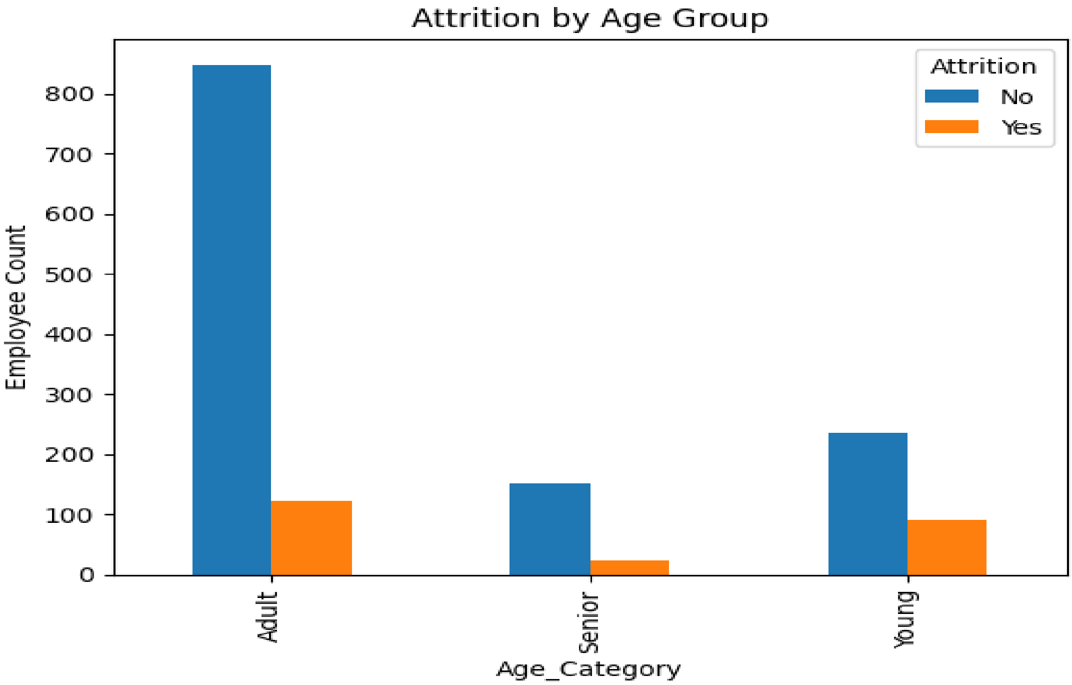


*age_attrition = df.groupby('Age_Category')['Attrition'].value_counts().unstack()*

*print(age_attrition)*

```
age_attrition_percentage= df.groupby('Age_Category')['Attrition'].value_counts(normalize=True).unstack()*100

print(age_attrition_percentage)

age_attrition.plot(kind='bar')

plt.title('Attrition by Age Group')

plt.ylabel('Employee Count')

plt.show()
```

```
Attrition          No   Yes            Attrition percentage

Age_Category


Adult             848   123                 12.66

Senior             150    23                13.29

Young             235      91               27.91
```



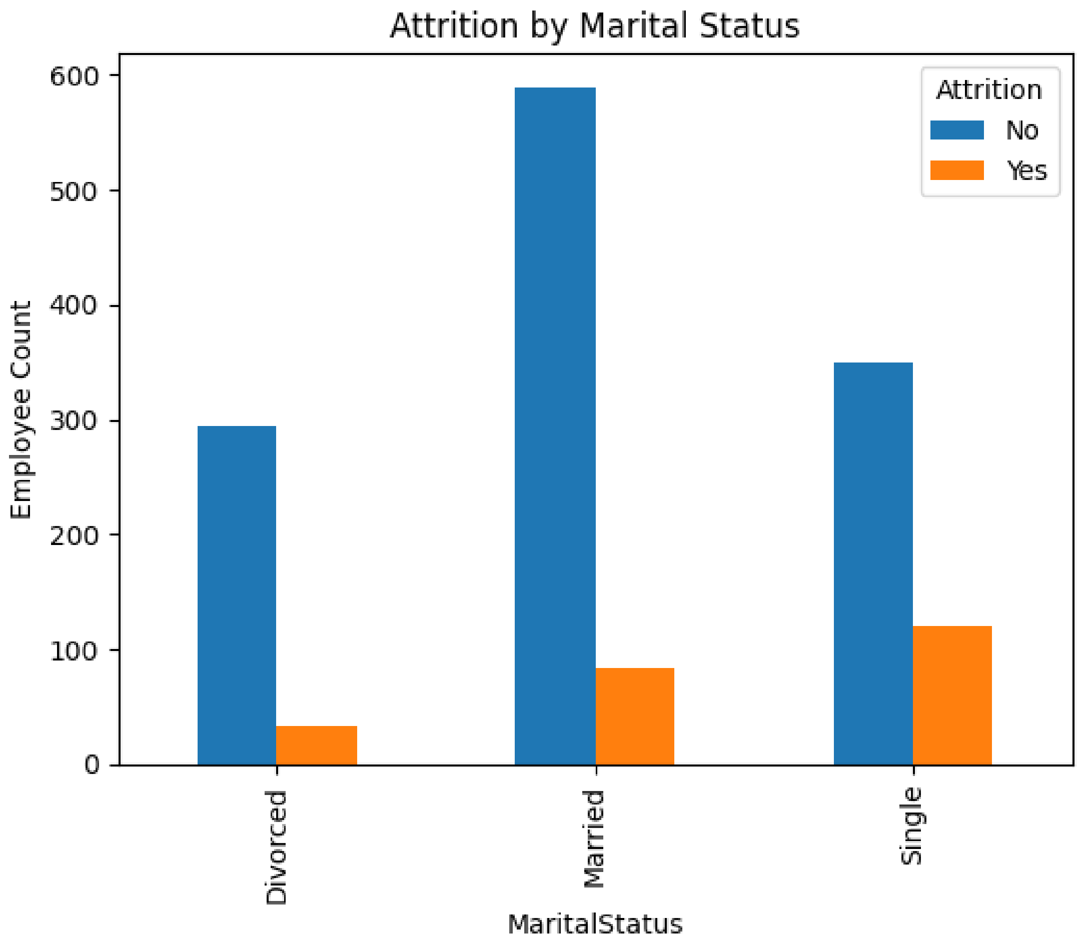
## 5.ATTRITION BY MARITAL STATUS

**Single** employees have a higher attrition rate (**25.5%)**.

This may be due to greater flexibility and fewer personal commitments, allowing them to pursue new opportunities more freely.

```
marital_attrition = df.groupby('MaritalStatus')['Attrition'].value_counts().unstack()
print(marital_attrition)
marital_attrition_percent= df.groupby('MaritalStatus')['Attrition'].value_counts(normalize=True).unstack()*100
print(marital_attrition_percent)
marital_attrition.plot(kind='bar')
plt.title('Attrition by Marital Status')
plt.ylabel('Employee Count')
plt.show()
```

```
Attrition          No    Yes        Attrition percentage
MaritalStatus


Divorced          294     33             10.09
Married           589     84             12.48
Single            350    120             25.53
```


Attrition by Marital Status
Employee Count
0
100
200
300
400
500
600
Attrition
No
Yes
Divorced
Married
Single
MaritalStatus

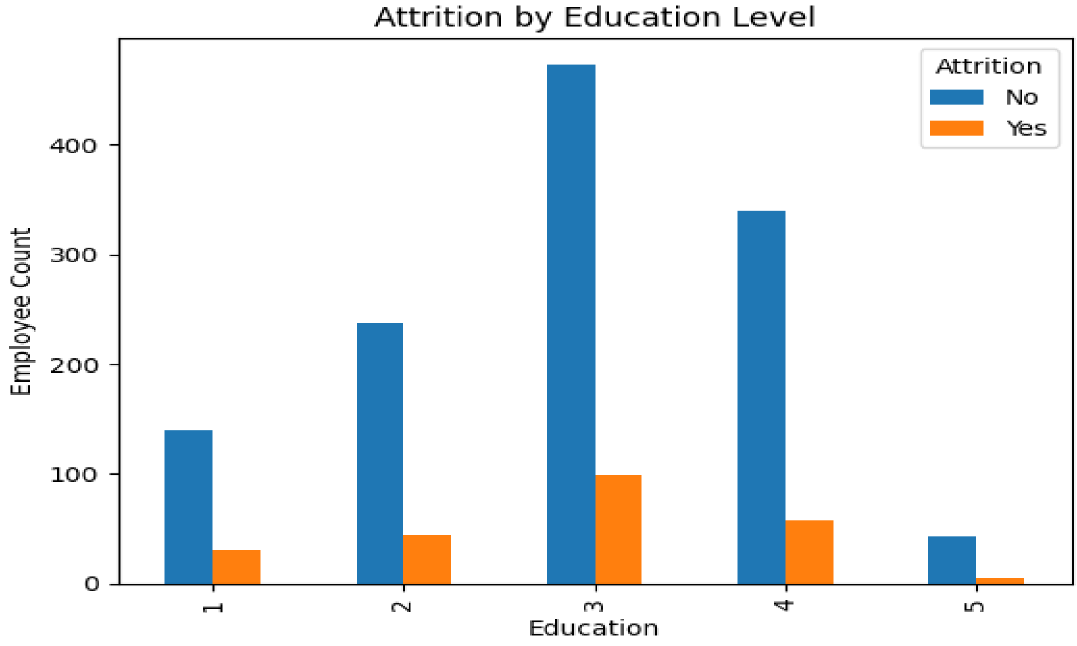
# 6.ATTRITION BY EDUCATION LEVEL

As **education level increases**, **attrition percentage decreases**.

This indicates that employees with **higher education** might be more satisfied with their roles or have better **career growth opportunities** within the organization.

```
education_attrition = df.groupby('Education')['Attrition'].value_counts().unstack()
print(education_attrition)
education_attrition_percent= df.groupby('Education')['Attrition'].value_counts(normalize=True).unstack()*100
print(education_attrition_percent)
education_attrition.plot(kind='bar')
plt.title('Attrition by Education Level')
plt.ylabel('Employee Count')
plt.show()
```

```
Attrition      No     Yes             Attrition percentage
Education
1               139    31                18.23
2               238    44               15.60
3              473    99                17.30
4              340    58               14.57
5                43    5                10.41
```

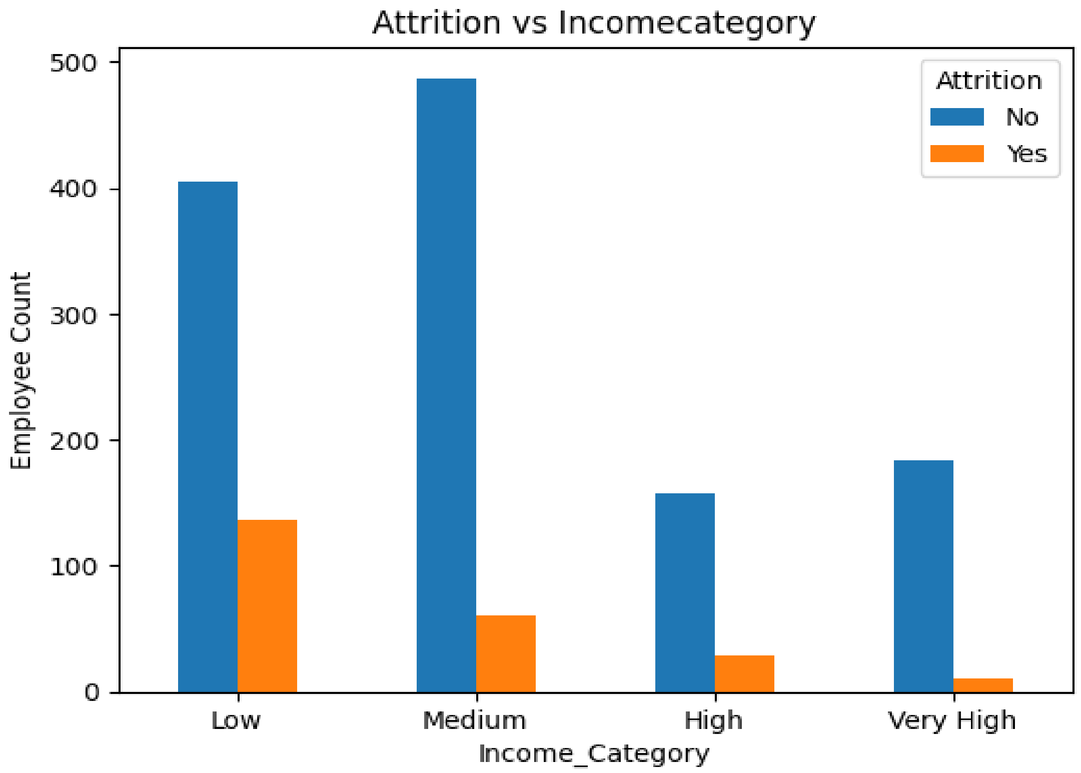


# 7. ATTRITION VS. MONTHLY INCOME

**Lower-income employees** (earning below **$4,000**) have a significantly higher attrition rate, while those earning above **$8,000** are less likely to leave, suggesting dissatisfaction with financial compensation as a key factor.

```
income_attrition = df.groupby('Income_Category')['Attrition'].value_counts().unstack()
print(income_attrition)
income_attrition_percentage= df.groupby('Income_Category')['Attrition'].value_counts(normalize=True).unstack()*100
print(income_attrition_percentage)
income_attrition = income_attrition.reindex(['Low', 'Medium', 'High', 'Very High'])
income_attrition.plot(kind='bar')
plt.title('Attrition vs Incomecategory')
plt.ylabel('Employee Count')
plt.xticks(rotation=0)
plt.show()
```

```
Attrition          No    Yes        Attrition percentage
Income_Category
High              157    29            15.59
Low               405   137            25.27
Medium            487    60            10.96
Very High         184    11             5.64
```

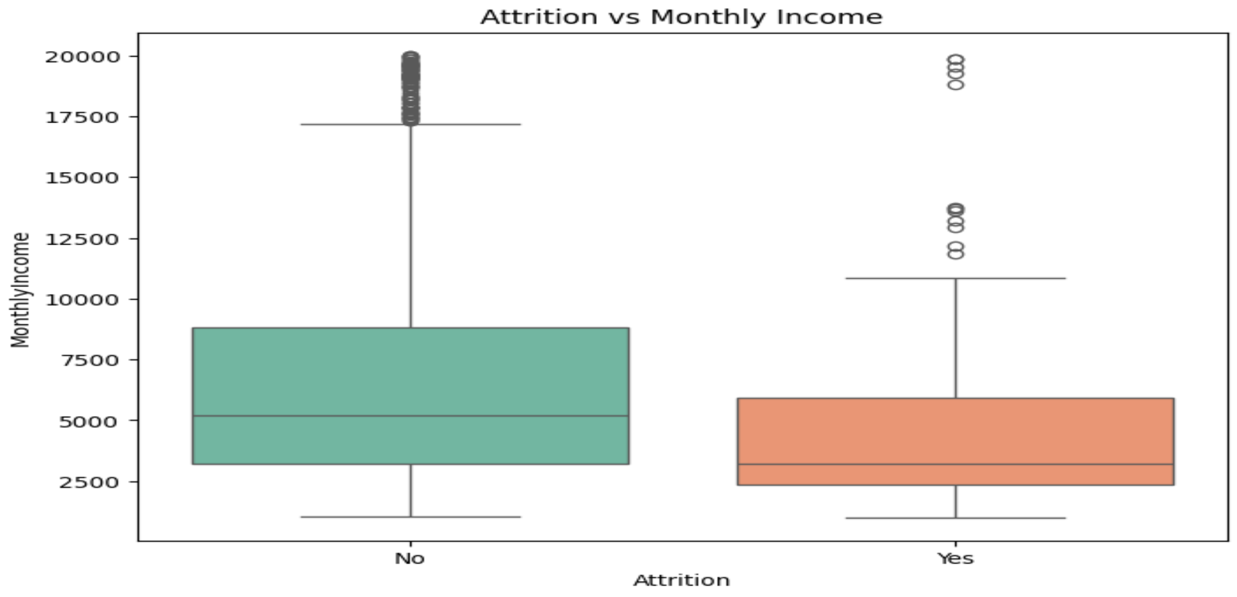


*plt.figure(figsize=(8, 6))*

*sns.boxplot(x='Attrition', y='MonthlyIncome', data=df, palette='Set2')*

*plt.title('Attrition vs Monthly Income')*

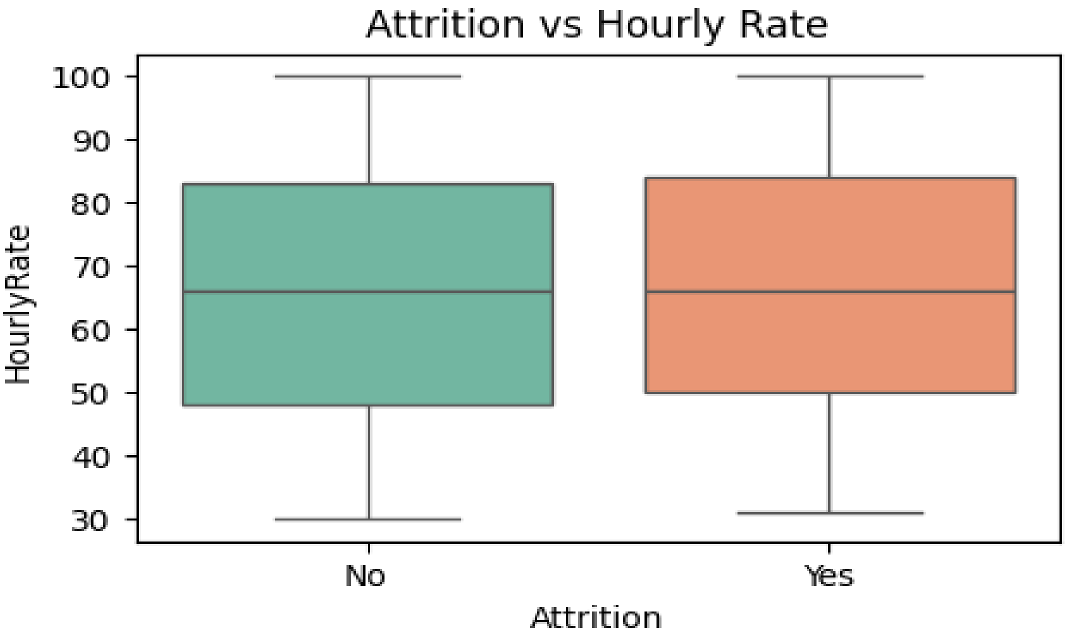
*plt.show()*



# 8. ATTRITION VS. HOURLY RATE

Very slight difference is observed.

```
plt.figure(figsize=(5, 3))
sns.boxplot(x='Attrition', y='HourlyRate', data=df, palette='Set2')
plt.title('Attrition vs Hourly Rate')
plt.show()
```



# 9.ATTRITION VS. JOB SATISFACTION

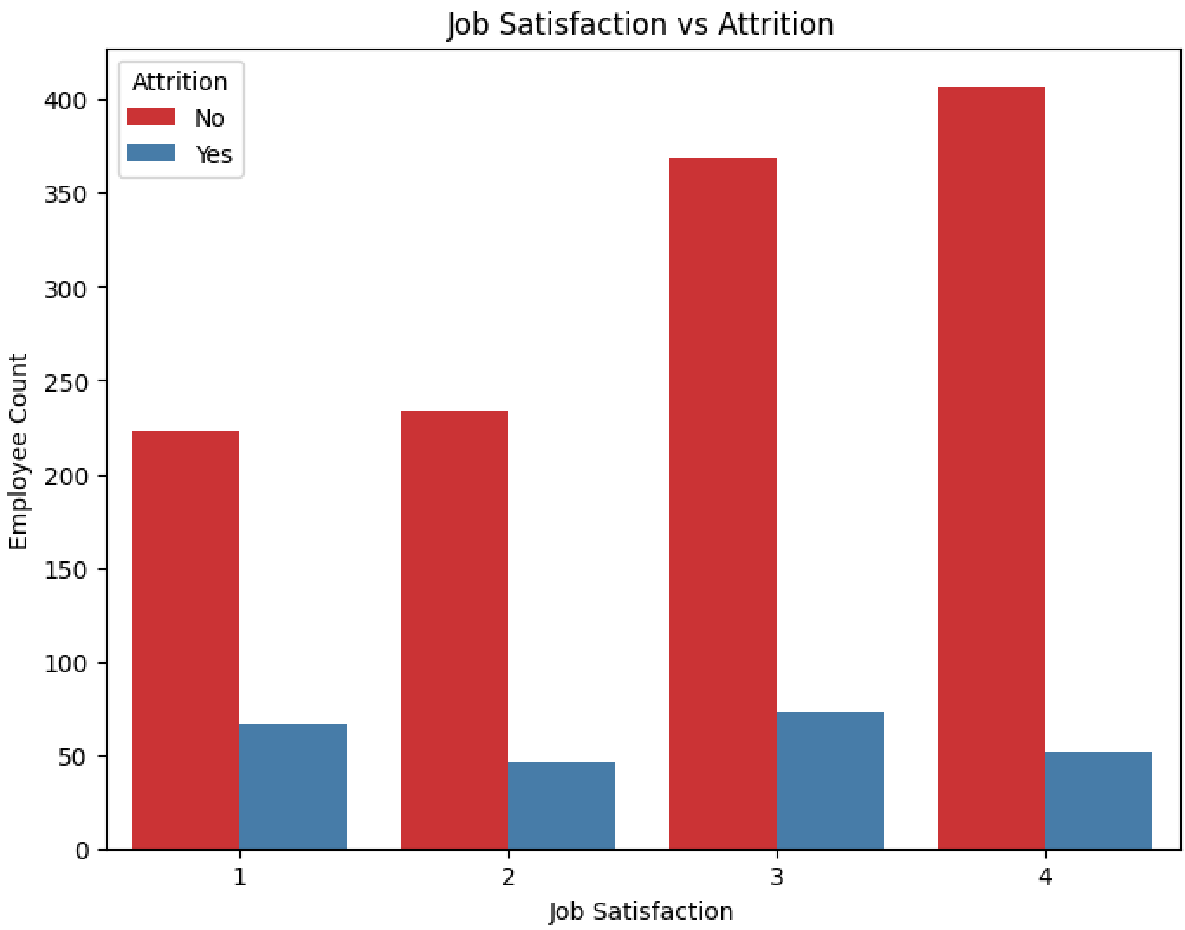
Employees with **low job satisfaction (1 or 2)** are far more likely to leave compared to those with high satisfaction (**3 or 4**).

**Attrition Rate for Low Job Satisfaction**: Around **22-23%** for satisfaction level 1, indicating extreme dissatisfaction.

**Attrition Rate for High Job Satisfaction**: Drops to **11-12%** for satisfaction level 4, showing that engaged and satisfied employees are more likely to stay.

```
satisfaction_counts = df.groupby('JobSatisfaction')['Attrition'].value_counts().unstack()
print(satisfaction_counts)
satisfaction_counts_percentage = df.groupby('JobSatisfaction')['Attrition'].value_counts(normalize=True).unstack() * 100
print(satisfaction_counts_percentage)
plt.figure(figsize=(8, 6))
sns.countplot(x='JobSatisfaction', hue='Attrition', data=df, palette='Set1')
plt.title('Job Satisfaction vs Attrition')
plt.xlabel('Job Satisfaction')
plt.ylabel('Employee Count')
plt.show()
```

```
Attrition           No    Yes           Attrition percentage
JobSatisfaction
1                   223    66               22.83
2                   234    46               16.42
3                   369    73              16.51
4                  407     52              11.32
```

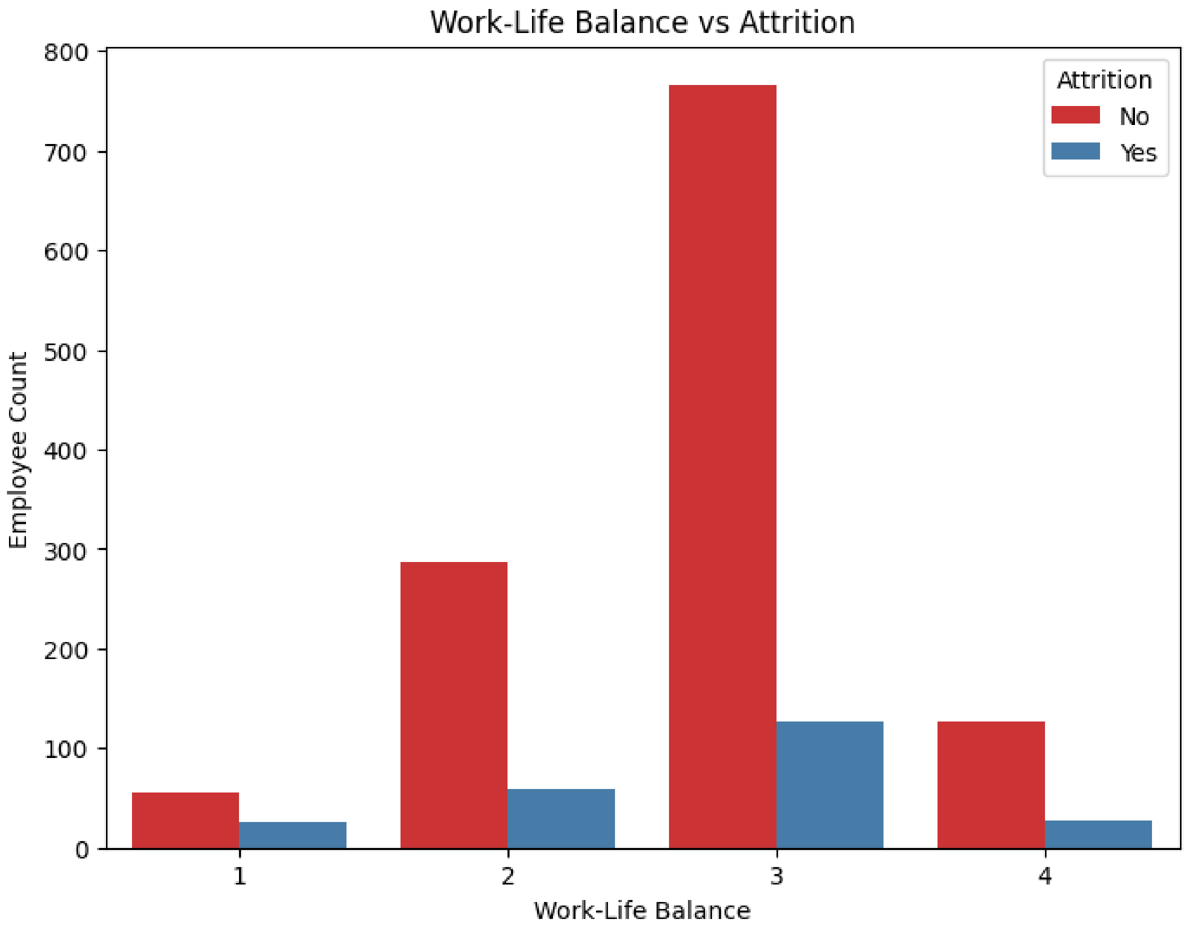


## 10.ATTRITION VS. WORK-LIFE BALANCE

Employees with **poor work-life balance (rating of 1**) are more likely to leave compared to those with higher ratings (**3–4**). This may be due to stress, burnout, and dissatisfaction with the inability to manage both personal and professional responsibilities.

```
Worklife_counts =
df.groupby('WorkLifeBalance')['Attrition'].value_counts().unstack()
print(Worklife_counts)
plt.figure(figsize=(8, 6))
sns.countplot(x='WorkLifeBalance', hue='Attrition', data=df, palette='Set1')
plt.title('Work-Life Balance vs Attrition')
plt.xlabel('Work-Life Balance')
plt.ylabel('Employee Count')
plt.show()
```

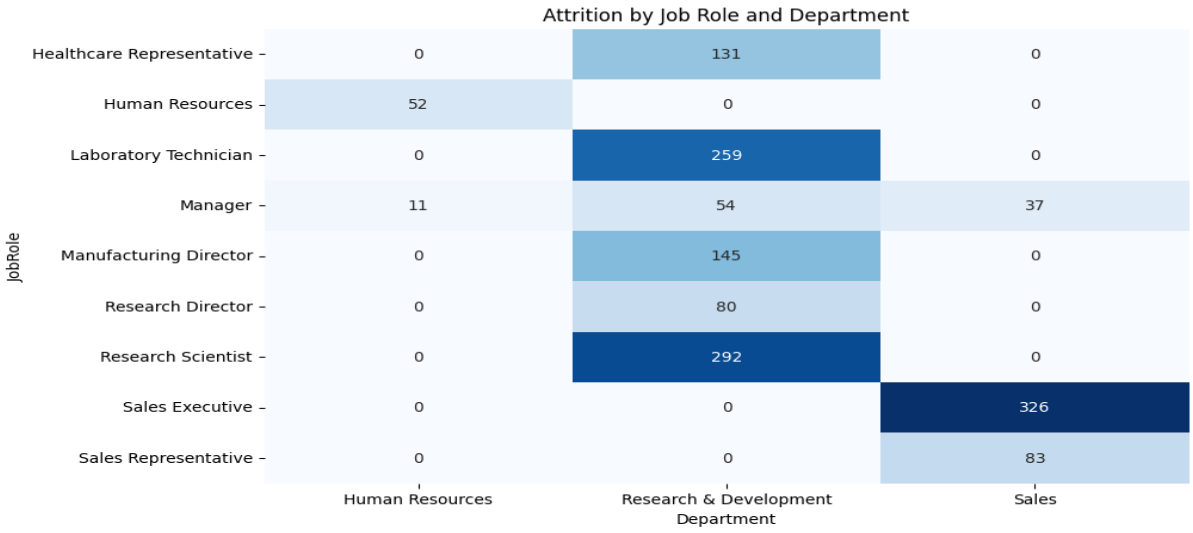
```
Attrition            No    Yes         Attrition percentage
WorkLifeBalance
1                    55     25              31.25
2                    286    58             16.86
3                    766   127             14.221
4                    126    27             17.64
```


Work-Life Balance vs Attrition
Employee Count
0
100
200
300
400
500
600
700
800
Work-Life Balance
1
2
3
4
Attrition
No
Yes

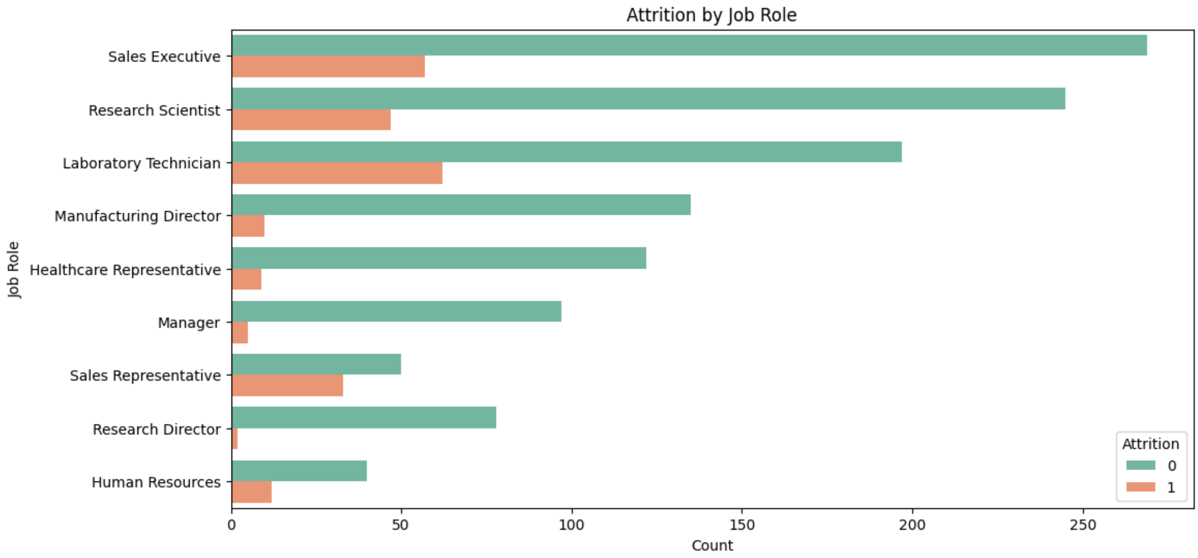
# 11. ATTRITION BY JOB ROLE AND DEPARTMENT:

**Sales Executives** in the **Sales Department** have the highest attrition percentage (**24%).** Employees in **Research scientist** and **Laboratory Technician** roles have similar high attrition rates (**23.08%** and **23.94%** respectively).

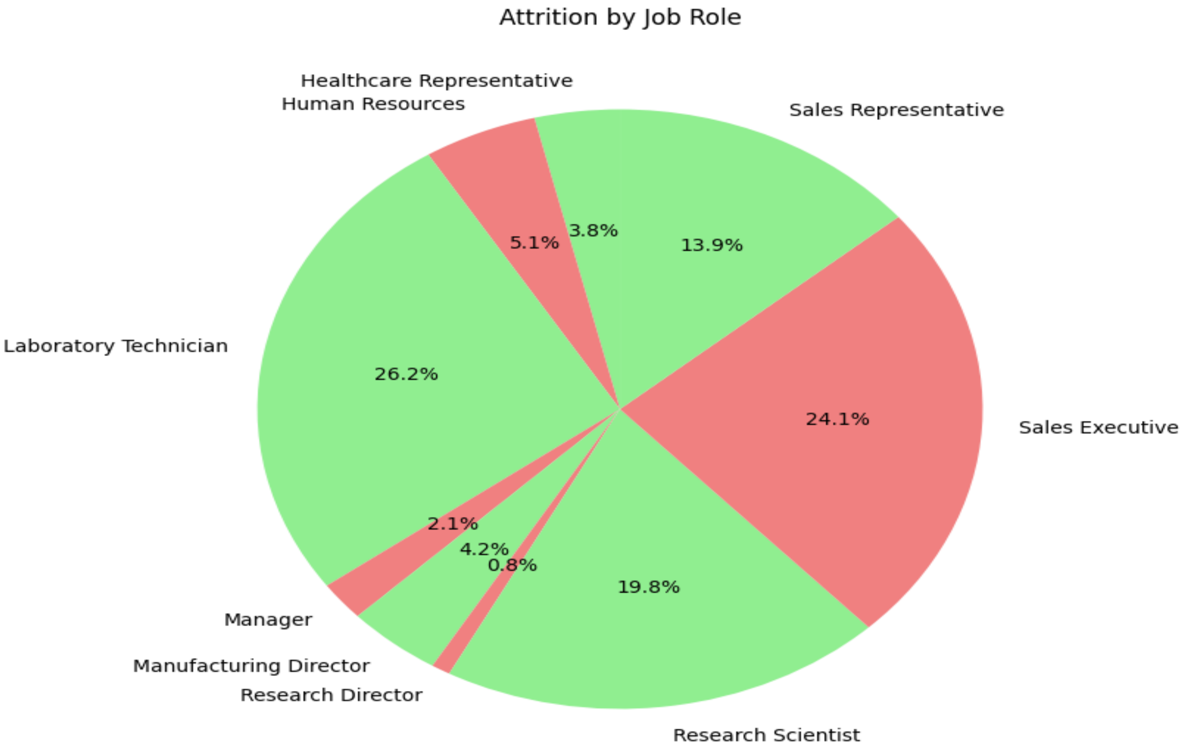
```
attrition_jobrole_dept = pd.crosstab(df['JobRole'], df['Department'],
values=df['Attrition'], aggfunc='count')

plt.figure(figsize=(10, 6))

sns.heatmap(attrition_jobrole_dept, annot=True, fmt='d', cmap='Blues',
cbar=False)

plt.title('Attrition by Job Role and Department')

plt.show()
```



```
plt.figure(figsize=(12, 6))

sns.countplot(y='JobRole', hue='Attrition', data=df, palette='Set2',
order=df['JobRole'].value_counts().index)

plt.title('Attrition by Job Role')

plt.xlabel('Count')

plt.ylabel('Job Role')

plt.show()
```



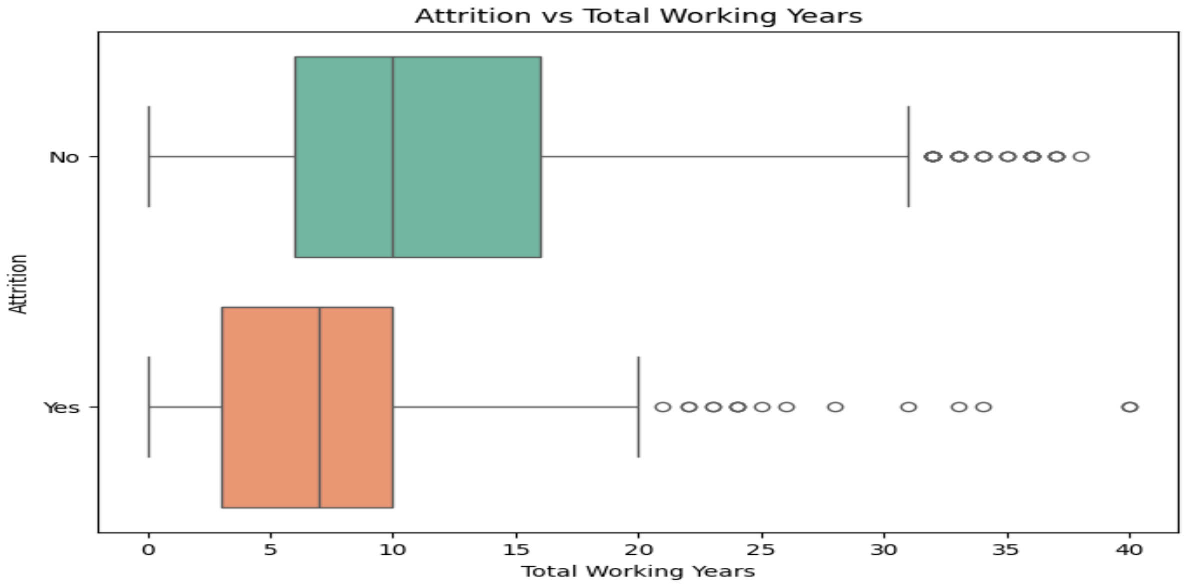
```
plt.figure(figsize=(8, 8))
job_role_attrition.plot(kind='pie', autopct='%1.1f%%', startangle=90,
colors=['lightgreen', 'lightcoral'])
plt.title('Attrition by Job Role')
plt.ylabel('')
plt.show()
```



## 12. ATTRITION VS. TOTAL WORKING YEARS:

Employees with fewer years at the company **(<2 years**) are more likely to leave. This is due to a lack of long-term commitment, unmet career expectations, and better opportunities for growth elsewhere.

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='TotalWorkingYears', y='Attrition', data=df, hue='Attrition', palette='Set2')
plt.title('Attrition vs Total Working Years')
plt.xlabel('Total Working Years')
plt.ylabel('Attrition')
plt.show()
```



## 13.ATTRITION VS. BUSINESS TRAVEL

The workers who **travel a lot** are more likely to quit than other employees (**25%**). This may be due to work-life balance challenges, increased stress, and the desire for more stability or local opportunities.

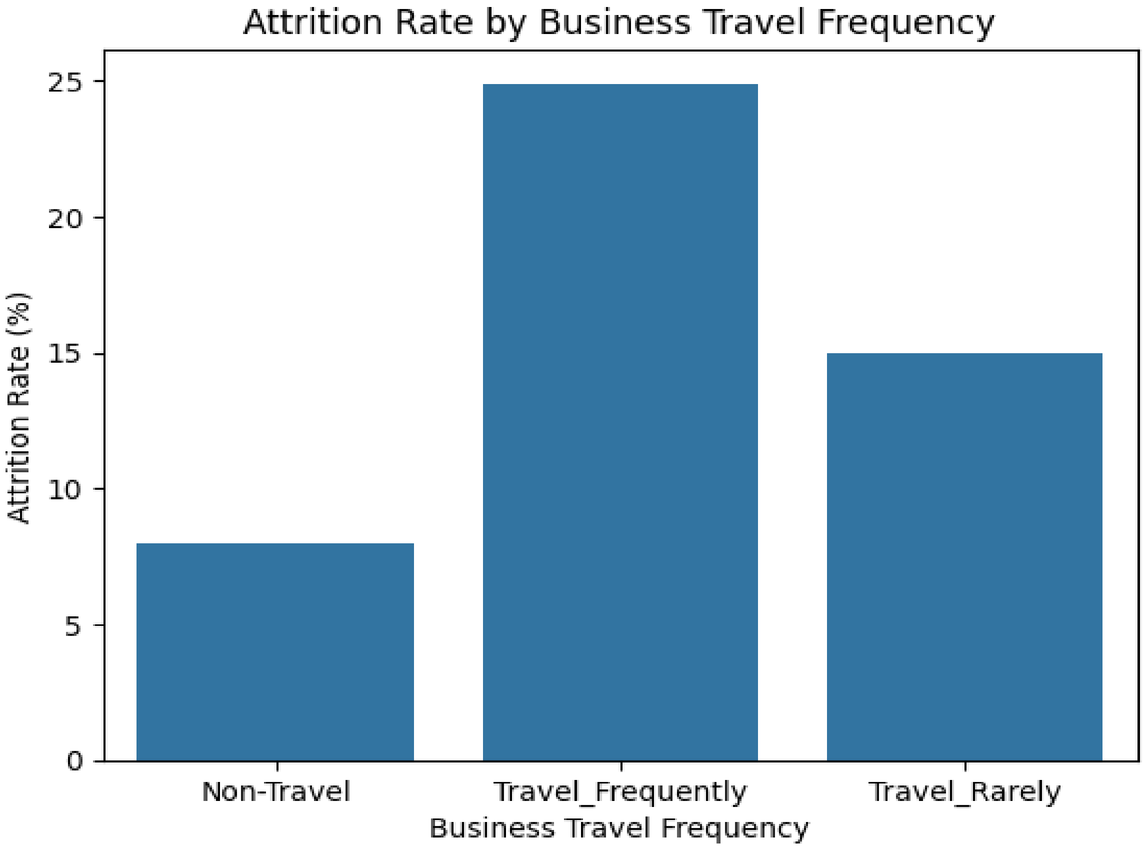
Frequent business travel often leads to **work-life balance challenges**, **increased stress**, and a desire for **stability** or **local opportunities**.

Employees who travel a lot may experience burnout from the pressures of constant travel, which can impact both personal and professional life. This could contribute to higher turnover as employees seek roles that offer more stability or fewer travel requirements.

Despite the potential for career advancement, the stress associated with frequent travel seems to outweigh these benefits for some employees.

```
travel_attrition = df.groupby(['BusinessTravel', 'Attrition']).size().unstack()
travel_attrition['Attrition_Rate'] = travel_attrition['Yes'] /
(travel_attrition['Yes'] + travel_attrition['No']) * 100
print(travel_attrition)
sns.barplot(x=travel_attrition.index, y=travel_attrition['Attrition_Rate'])
plt.title("Attrition Rate by Business Travel Frequency")
plt.ylabel("Attrition Rate (%)")
plt.xlabel("Business Travel Frequency")
plt.show()
```

```
Attrition              No    Yes     Attrition_Rate
BusinessTravel
Non-Travel            138    12       8.000000
Travel_Frequently     208    69      24.909747
Travel_Rarely         887   156      14.956855
```



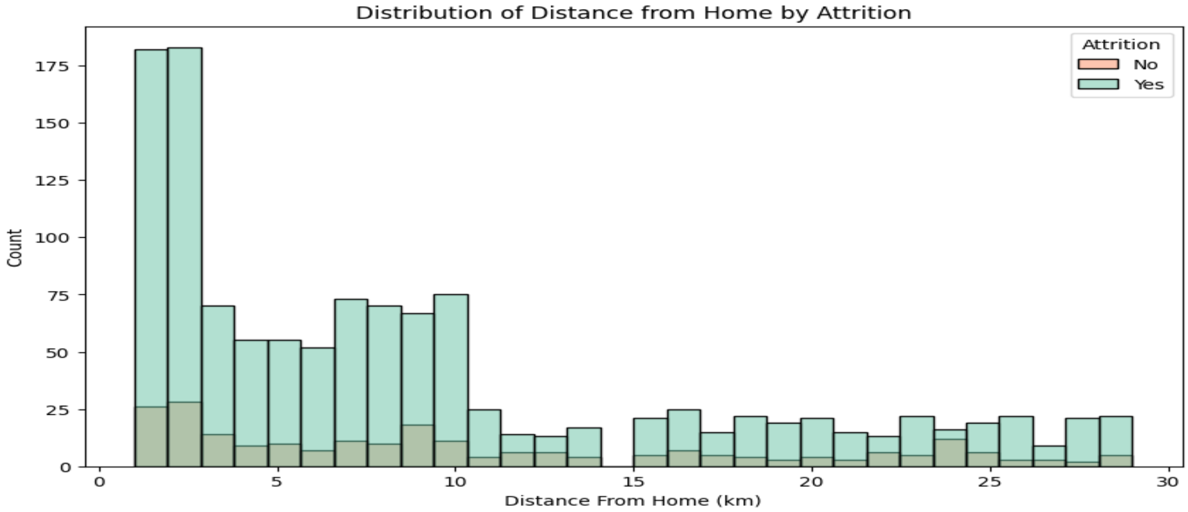
# 14.ATTRITION VS. DISTANCE FROM HOME

Employees living within **10–15 km of the workplace** have the lowest attrition rates, while those commuting more than **20 km and <5km** show significantly higher attrition.

**Low attrition for moderate commute distances (10-15 km)** may indicate a balance between convenience and job satisfaction.

Employees with **very short (under 5 km)** or **very long commutes (over 20 km)** experience higher attrition, potentially due to factors such as **commute fatigue**, **work-life balance issues**, or **lower job satisfaction**.

```
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x='DistanceFromHome', hue='Attrition', bins=30,
kde=False, palette='Set2')
plt.title("Distribution of Distance from Home by Attrition")
plt.xlabel("Distance From Home (km)")
plt.ylabel("Count")
plt.legend(title='Attrition', labels=['No', 'Yes'])
plt.show()
```



# 15.ATTRITION VS. OVER TIME

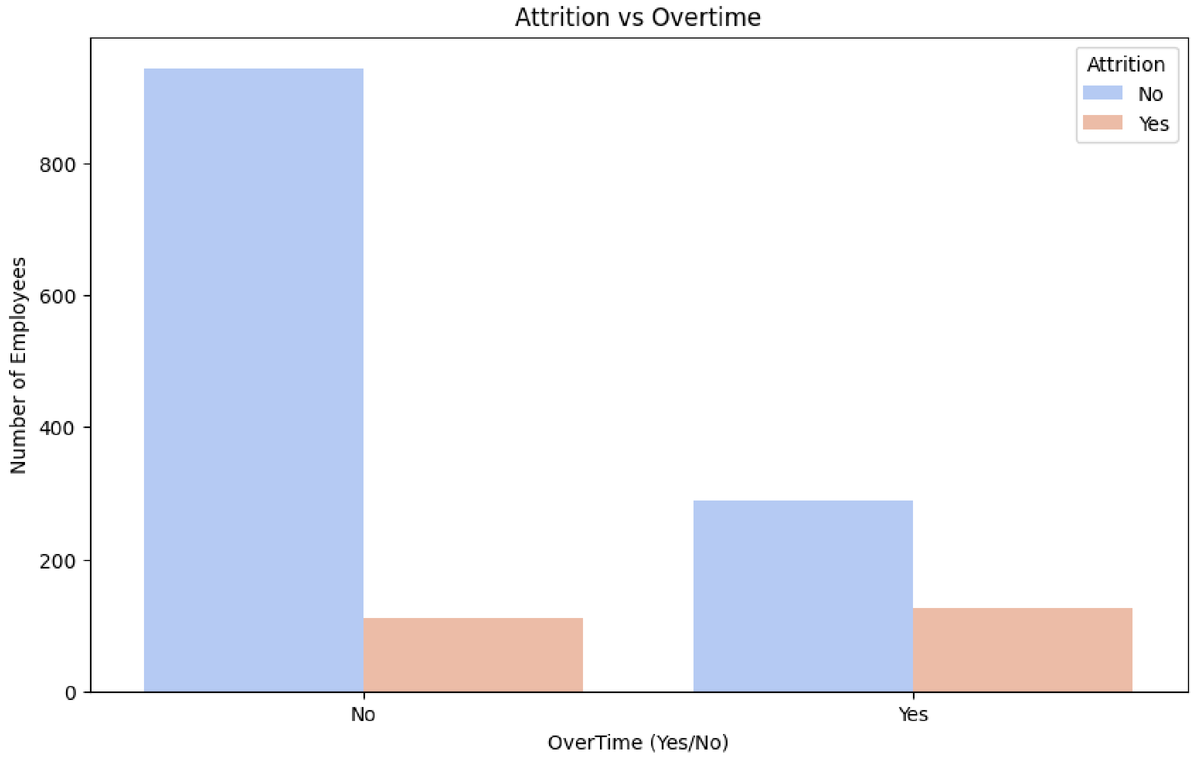
Employees who work overtime are **2.5 times** more likely to leave compared to those who don’t. This indicates that over time plays a significant role in attrition rates.

Employees who regularly work overtime may experience **burnout**, **work-life imbalance**, and **increased stress**, which can ultimately lead them to seek more stable or manageable roles elsewhere.

The findings emphasize that long working hours, while sometimes necessary for productivity, can have negative effects on employee well-being and job satisfaction, driving them to leave the company.

```
overtime_attrition = df.groupby(['OverTime', 'Attrition']).size().unstack()
overtime_attrition_percentage =
overtime_attrition.div(overtime_attrition.sum(axis=1), axis=0) * 100
print("Counts of Attrition based on Overtime:\n", overtime_attrition)
print("\nPercentage of Attrition based on Overtime:\n",
overtime_attrition_percentage)
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='OverTime', hue='Attrition', palette='coolwarm')
plt.title("Attrition vs Overtime")
plt.xlabel("OverTime (Yes/No)")
plt.ylabel("Number of Employees")
plt.show()
```

```
Counts of Attrition based on Overtime:
 Attrition         No    Yes
OverTime
No               944    110
Yes              289     127

Percentage of Attrition based on Overtime:
 Attrition          No              Yes
OverTime
No            89.563567       10.436433
Yes           69.471154       30.528846
```

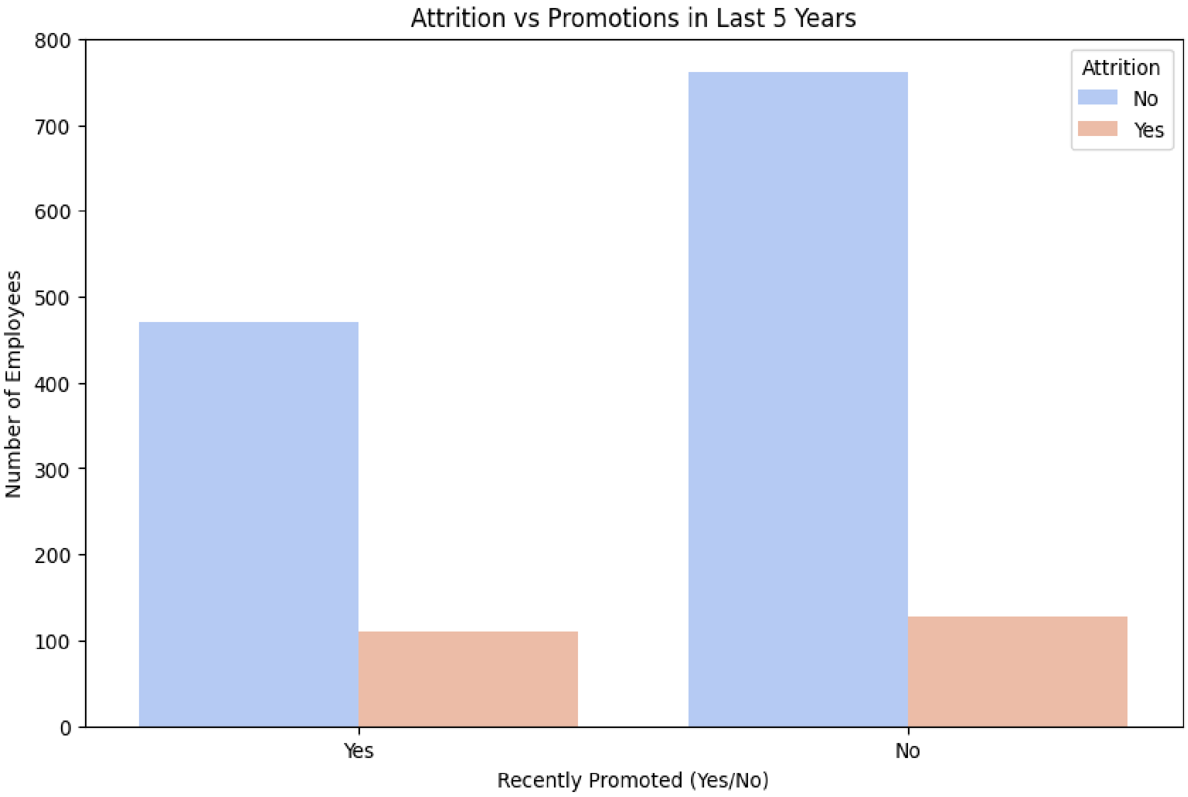


# 16.ATTRITION VS. PROMOTION

Employees **who have not been promoted in the last 5 years** show significantly higher attrition rates. This may be due to stagnation, lack of career growth, and frustration with limited advancement opportunities.

Promotion is a key factor in employee **motivation** and **loyalty**, and its absence can create frustration, leading to higher turnover. This highlights the importance of a well-structured **promotion system** and clear career development pathways.

```
promotion_attrition = df.groupby(['YearsSinceLastPromotion', 'Attrition']).size().unstack()

promotion_attrition_percentage = promotion_attrition.div(promotion_attrition.sum(axis=1), axis=0) * 100

print("Counts of Attrition based on Promotions:\n", promotion_attrition)

print("\nPercentage of Attrition based on Promotions:\n", promotion_attrition_percentage)

df['RecentlyPromoted'] = df['YearsSinceLastPromotion'].apply(lambda x: 'Yes' if x == 0 else 'No')

plt.figure(figsize=(10, 6))

sns.countplot(data=df, x='RecentlyPromoted', hue='Attrition', palette='coolwarm')

plt.title("Attrition vs Promotions in Last 5 Years")

plt.xlabel("Recently Promoted (Yes/No)")

plt.ylabel("Number of Employees")

plt.show()
```


Attrition vs Promotions in Last 5 Years
Number of Employees
0
100
200
300
400
500
600
700
800
Yes
No
Recently Promoted (Yes/No)
Attrition
No
Yes

## 17. ATTRITION VS. JOB INVOLVEMENT

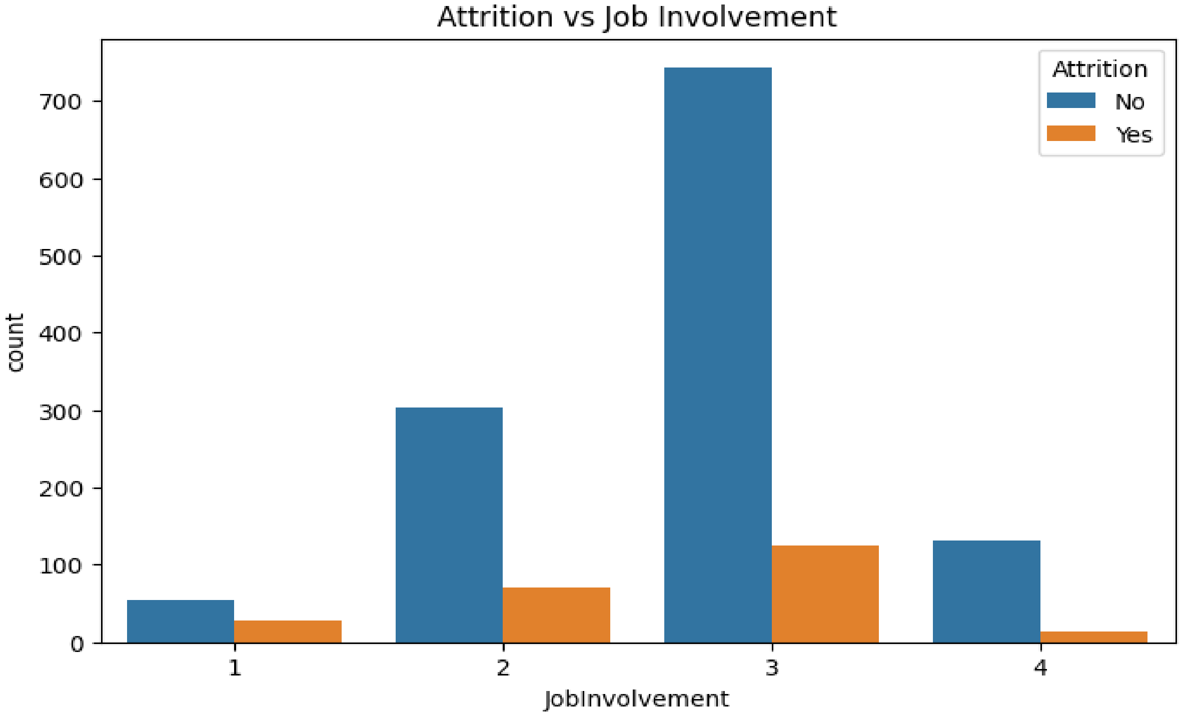
Employees with **low job involvement** are more likely to leave (**33.75%**).

**Low job involvement** suggests a lack of engagement or emotional investment in work, which often leads to **disengagement** and **higher turnover**. Employees who are not actively involved in their work may feel disconnected, leading to dissatisfaction and eventually seeking opportunities elsewhere.

On the other hand, **higher job involvement** typically correlates with **stronger engagement**, **job satisfaction**, and **commitment to the organization**, which helps improve retention rates. Engaged employees are more likely to stay because they feel valued and motivated by their roles.

```
job_involvement_attrition = 
df.groupby('JobInvolvement')['Attrition'].value_counts().unstack()

print(job_involvement_attrition)

job_involvement_attrition_percentage = 
df.groupby('JobInvolvement')['Attrition'].value_counts(normalize=True).unstac
k() * 100

print(job_involvement_attrition_percentage)

plt.figure(figsize=(8,5))

sns.countplot(data=df, x='JobInvolvement', hue='Attrition')

plt.title('Attrition vs Job Involvement')

plt.show()
```

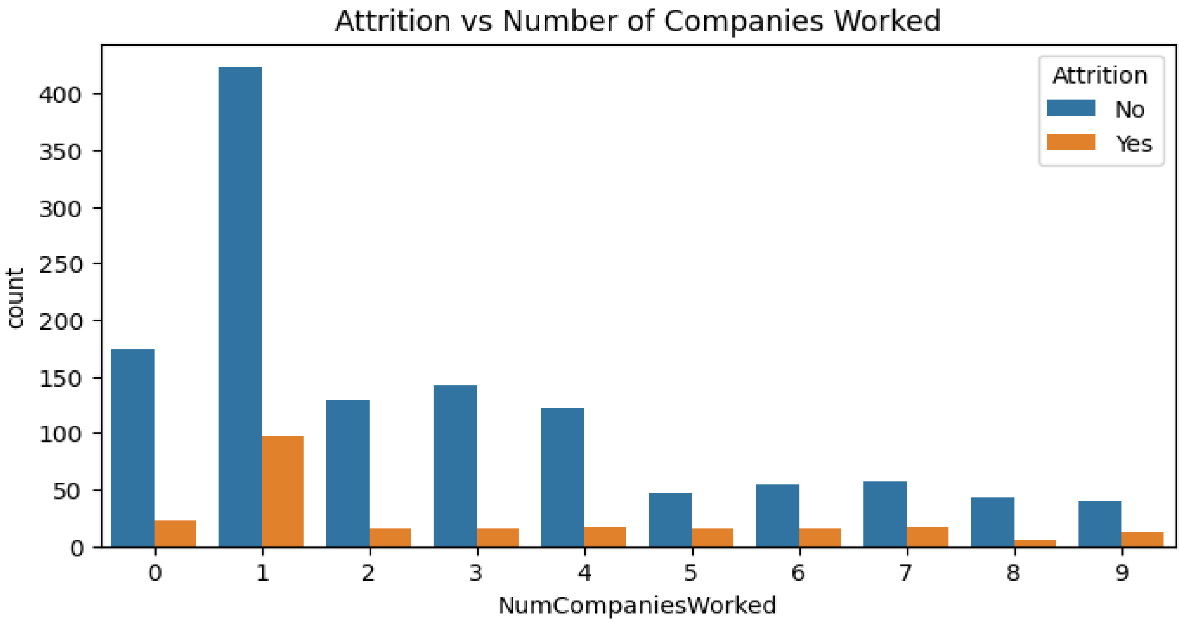
```
Attrition          No   Yes       Attrition percentage
JobInvolvement
1                   55   28               33.73
2                  304   71               18.93
3                  743   125              14.40
4                  131   13               9.02
```



# 18. ATTRITION VS. NUMBER OF COMPANIES WORKED

Employees who have worked at **fewer companies** tend to stay longer, while those who have worked at **many companies** are more likely to leave. This may suggest that employees who have stayed in one place for a longer time are more likely to be loyal.

```
plt.figure(figsize=(10,6))
sns.countplot(data=df, x='NumCompaniesWorked', hue='Attrition')
plt.title('Attrition vs Number of Companies Worked')
plt.show()
```



# 19. ATTRITION VS. EDUCATIONAL FIELD

**Human Resources** and **Technical Degree** fields have the highest attrition percentages (around **25%**). **Life Sciences, Medical, and other fields** have lower attrition percentages (around **13-15%**).

**High attrition in Human Resources and Technical Degree fields** may be driven by factors such as **high stress**, **lack of career growth**, or **skill mismatches**.

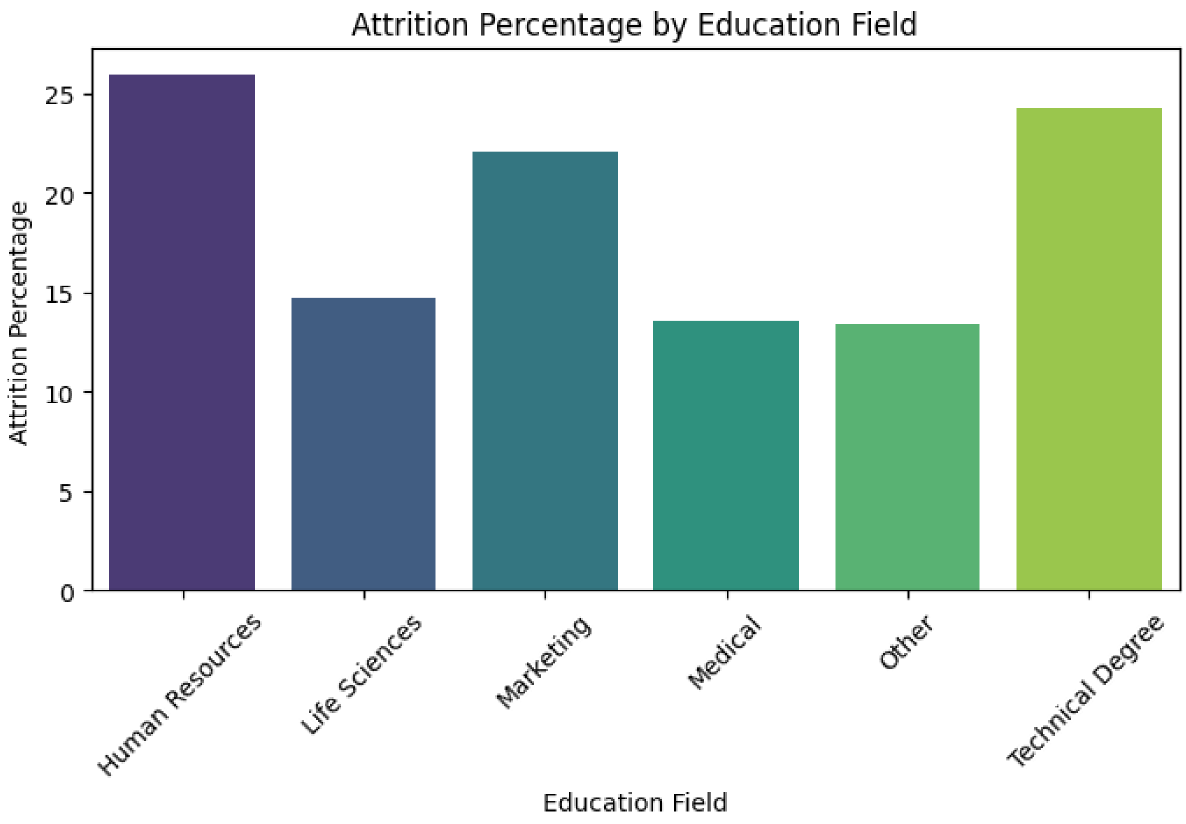
**Technical roles** often face intense competition, burnout, and the constant need for upskilling, which may contribute to the higher turnover. Similarly, roles in **Human Resources** can be emotionally taxing, with employees leaving due to stress or limited growth opportunities.

On the other hand, fields like **Life Sciences** and **Medical** typically offer more structured career paths, job stability, and opportunities for professional development, which may explain their lower attrition rates.

```
attrition_edu = df.groupby(['EducationField',
'Attrition']).size().reset_index(name='Count')

pivot_table = attrition_edu.pivot(index='EducationField', columns='Attrition',
values='Count').fillna(0)

pivot_table['Attrition_Percentage'] = round((pivot_table['Yes'] /
(pivot_table['Yes'] + pivot_table['No'])) * 100, 2)

print(pivot_table)

plt.figure(figsize=(8, 4))

sns.barplot(x='EducationField', y='Attrition_Percentage',
data=pivot_table.reset_index(), palette='viridis')
```

```
plt.title('Attrition Percentage by Education Field')
plt.xlabel('Education Field')
plt.ylabel('Attrition Percentage')
plt.xticks(rotation=45)
plt.show()
```

```
Attrition           No   Yes    Attrition_Percentage
EducationField
Human Resources     20    7           25.93
Life Sciences      517   89           14.69
Marketing          124   35           22.01
Medical            401   63           13.58
Other               71   11           13.41
Technical Degree   100   32           24.24
```

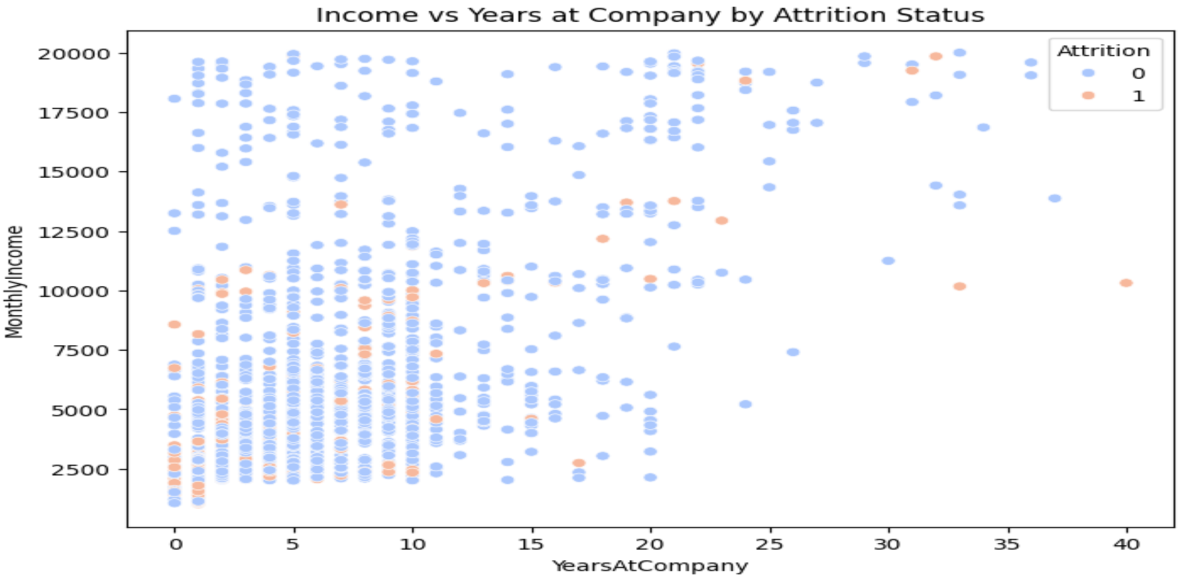


# 20.INCOME VS YEARS AT THE COMPANY

Employees with fewer years at the company and lower income show higher attrition. This isdue to factors such as **lack of career growth**, **unmet expectations**, or the search for better opportunities elsewhere. Early-career employees often feel less committed and more likely to explore new roles to advance their careers.

**Lower-income employees** tend to leave at higher rates, possibly due to **dissatisfaction with compensation** and the desire for higher-paying roles or better financial security. These employees may seek more competitive offers from other organizations to meet their financial goals.

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='YearsAtCompany', y='MonthlyIncome', hue='Attrition', data=df, palette='coolwarm')
plt.title('Income vs Years at Company by Attrition Status')
plt.show()
```



## 21.CORRELATION B/W FEATURES

S**trong Correlations**:

- **TotalWorkingYears** and **JobLevel** have a high positive correlation (0.78), indicating that as total working years increase, job level tends to increase.
- **MonthlyIncome** and **JobLevel** (0.95) also show a strong positive relationship, meaning higher job levels lead to higher incomes.
- **YearsAtCompany** and **YearsInCurrentRole** (0.76) are highly correlated, which makes sense as employees often stay in roles for a significant portion of their company tenure.

**Moderate Correlations**:

- **Age** and **TotalWorkingYears** (0.68): Older employees typically have more working years.
- **MonthlyIncome** and **TotalWorkingYears** (0.77): Longer work experience relates to higher salaries.

**Low/No Correlation**:

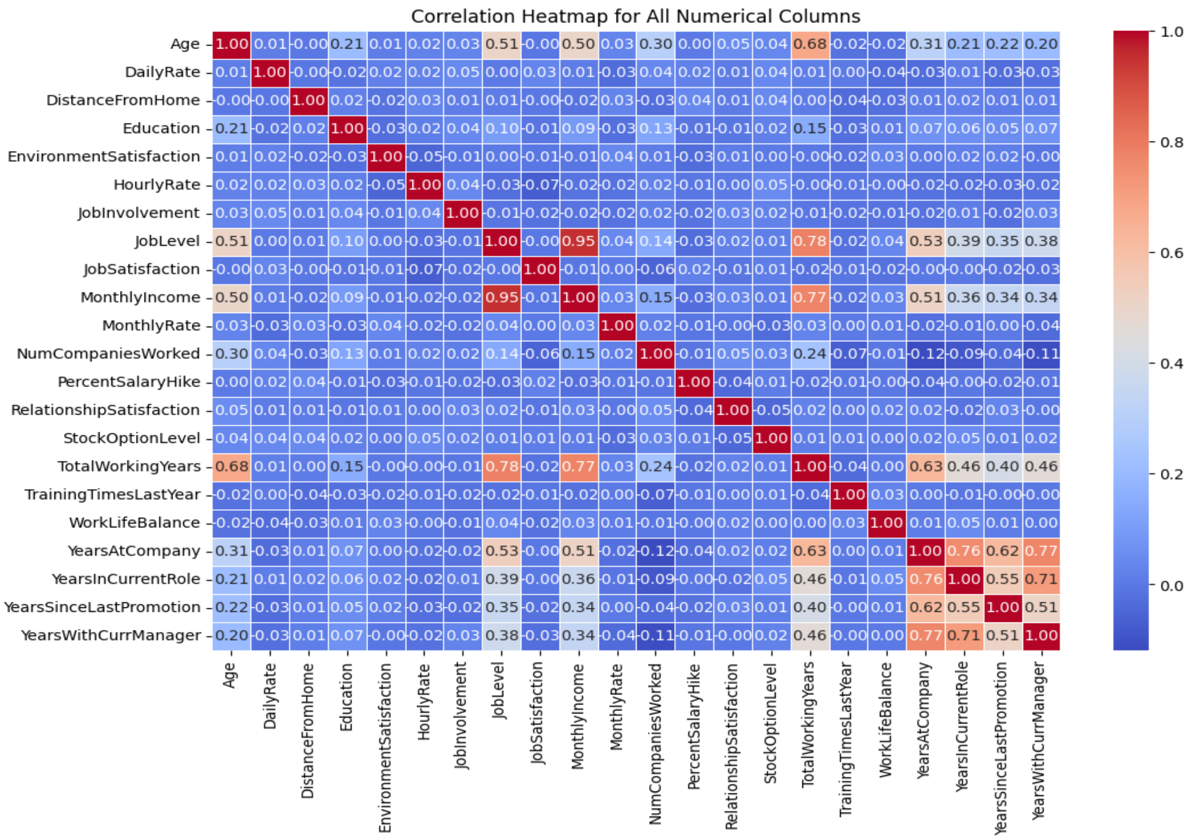
- Columns like **DailyRate**, **HourlyRate**, and **DistanceFromHome** have minimal correlations with other variables (mostly close to 0), suggesting weaker relationships.

**Negative Correlations**:

- **Age** and **JobLevel** (-0.51) show some negative correlation.

- **NumCompaniesWorked** and **TotalWorkingYears** (-0.30): More job changes may slightly reduce overall work experience.

```
df['Attrition'] = df['Attrition'].map({'Yes': 1, 'No': 0})
numerical_df = df.select_dtypes(include=['int','float'])
correlation_matrix = numerical_df.corr()
print("Correlation Matrix for All Numerical Columns:")
print(correlation_matrix)
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",
linewidths=0.5)
plt.title('Correlation Heatmap for All Numerical Columns')
plt.show()
```

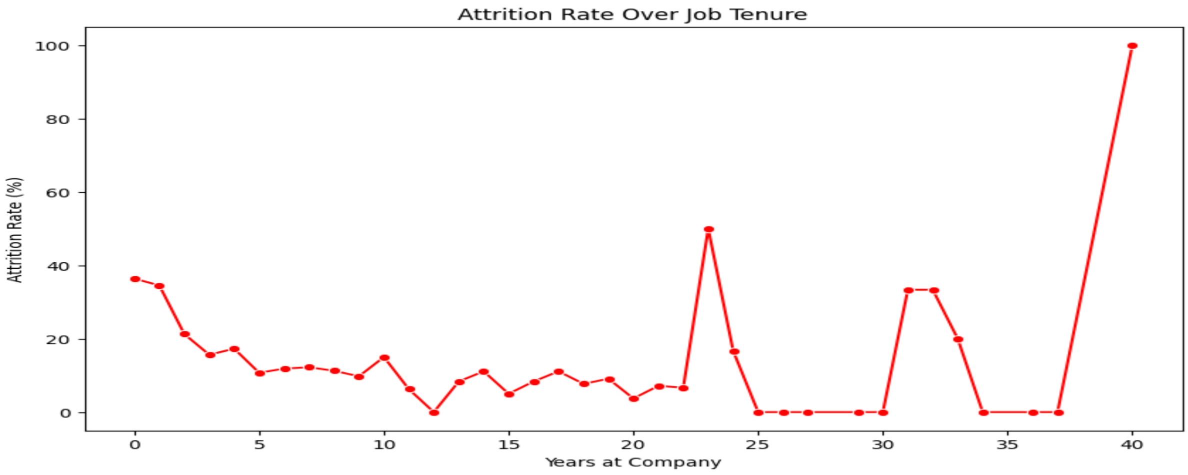


## 22. ATTRITION RATE OVER JOB TENURE

Attrition rates peak for employees with less than 2 years of tenure, then gradually decrease as tenure increases. This is likely due to **early career exploration**, unmet expectations, or lack of strong **organizational commitment**. These employees may leave for better opportunities as they continue to develop professionally.

**Attrition decreases with tenure**, indicating that employees who have been with the company for **longer periods** are more likely to stay, likely due to **increased loyalty**, **career development**, and **familiarity with company culture**.

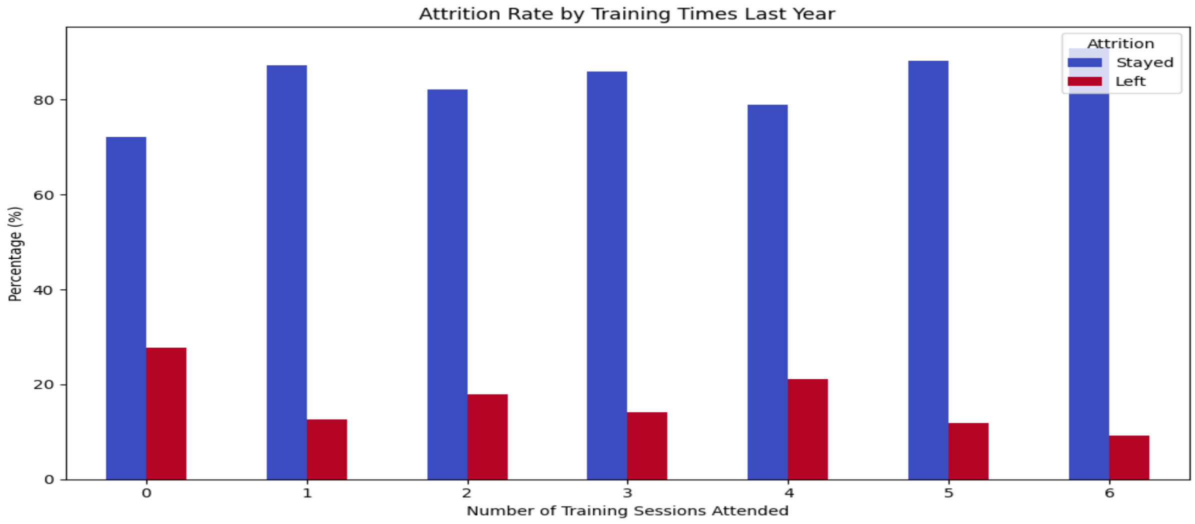
```
attrition_tenure = df.groupby('YearsAtCompany')['Attrition'].value_counts(normalize=True).unstack() * 100
plt.figure(figsize=(10, 6))
sns.lineplot(data=attrition_tenure[1], color='red', marker='o') # Access using 1 instead of 'Yes'
plt.title('Attrition Rate Over Job Tenure')
plt.xlabel('Years at Company')
plt.ylabel('Attrition Rate (%)')
plt.show()
```



# 23. ATTRITION VS. TRAINING TIME

Employees who attended fewer than **2** training sessions show a significantly higher attrition rate (around **18-20%**) compared to those attending **6+** sessions (around **8-10%**). This highlights the need for regular training programs to improve retention.

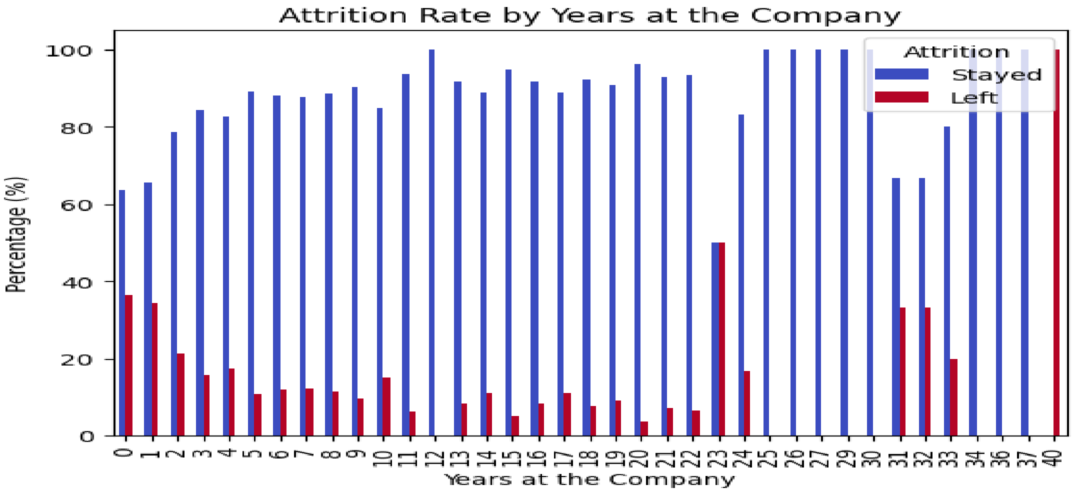
```
training_attrition = df.groupby('TrainingTimesLastYear')['Attrition'].value_counts(normalize=True).unstack() * 100
plt.figure(figsize=(10, 6))
training_attrition.plot(kind='bar', stacked=False, colormap='coolwarm', figsize=(10, 6))
plt.title('Attrition Rate by Training Times Last Year')
plt.xlabel('Number of Training Sessions Attended')
plt.ylabel('Percentage (%)')
plt.legend(title='Attrition', loc='upper right', labels=['Stayed', 'Left'])
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```



# 24.ATTRITION VS. YEARS AT THE COMPANY

Employees with less than **2** years at the company exhibit the highest attrition rates **(~28%).** Attrition drops steadily for employees with **5+ years**, indicating improved loyalty and organizational commitment over time.

```
years_attrition = df.groupby('YearsAtCompany')['Attrition'].value_counts(normalize=True).unstack() * 100
plt.figure(figsize=(15, 12))
years_attrition.plot(kind='bar', stacked=False, colormap='coolwarm')
plt.title('Attrition Rate by Years at the Company')
plt.xlabel('Years at the Company')
plt.ylabel('Percentage (%)')
plt.legend(title='Attrition', loc='upper right', labels=['Stayed', 'Left'])
plt.xticks(rotation=90)
plt.show()
```



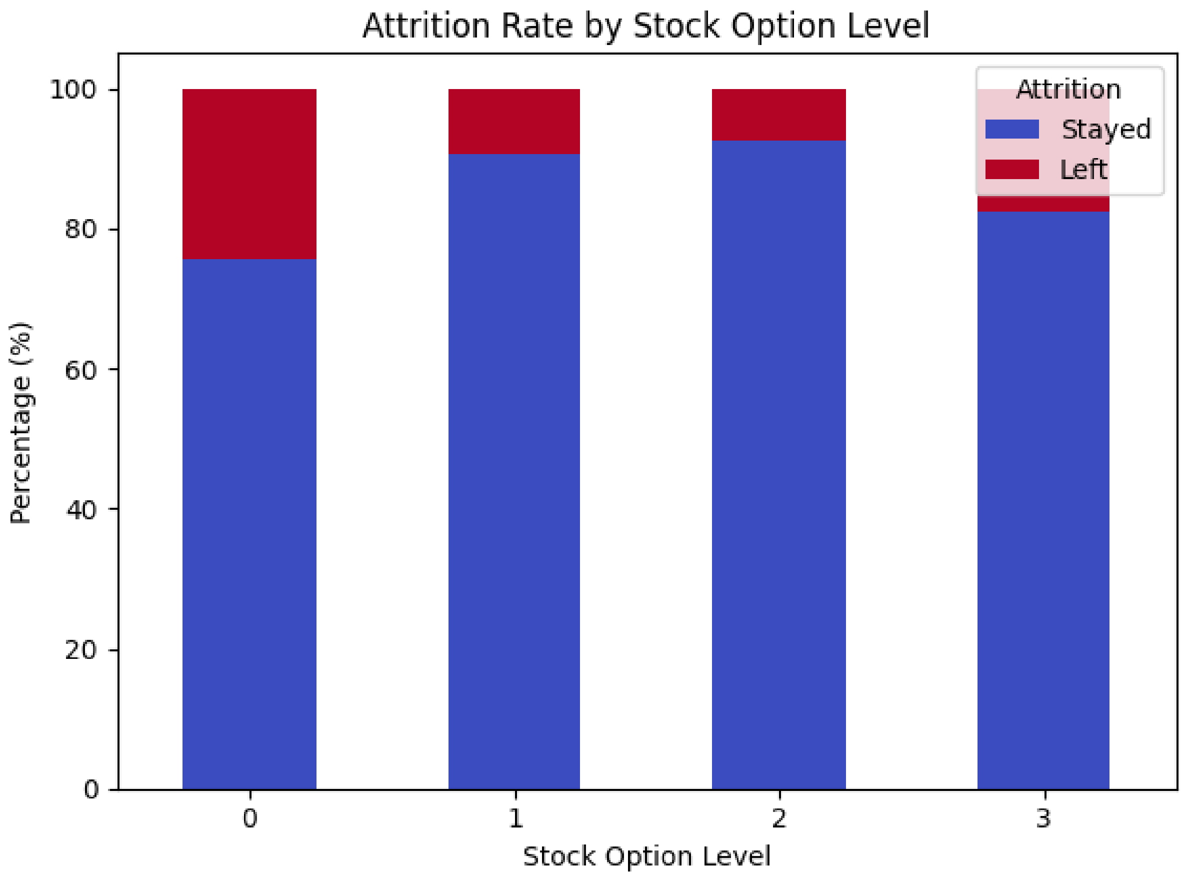
## 25.ATTRITION VS. STOCK OPTION LEVEL

Employees with no stock options (**Level 0**) show the highest attrition rate (~**24%**), indicating dissatisfaction with financial incentives.

Employees with **Level 2** and **Level 3** stock options exhibit significantly lower attrition rates **(~10-12%**), highlighting the effectiveness of stock options in retaining employees.

Surprisingly, even employees with **Level 3** stock options experience some attrition, suggesting that stock options alone are not enough to guarantee retention.

```
stock_attrition = df.groupby('StockOptionLevel')['Attrition'].value_counts(normalize=True).unstack() * 100

plt.figure(figsize=(10, 6))

stock_attrition.plot(kind='bar', stacked=True, colormap='coolwarm')

plt.title('Attrition Rate by Stock Option Level')

plt.xlabel('Stock Option Level')

plt.ylabel('Percentage (%)')

plt.legend(title='Attrition', loc='upper right', labels=['Stayed', 'Left'])

plt.xticks(rotation=0)

plt.tight_layout()

plt.show()
```


Attrition Rate by Stock Option Level
Percentage (%)
0
20
40
60
80
100
0
1
2
3
Stock Option Level
Attrition
Stayed
Left

# KEY FINDINGS AND INSIGHTS

1. **Demographics:**
   - **Age:** Younger employees (under 30) and employees with less than 2 years of tenure have higher attrition rates.
   - **Marital Status:** Single employees are more prone to leaving, while married employees exhibit greater stability.
   - **Education Field:** Employees with education in the "Human Resources" field have slightly higher attrition rates compared to other fields.
2. **Job-Related Factors:**
   - **Job Satisfaction:** Employees dissatisfied with their jobs are significantly more likely to leave.
   - **Department:** Higher attrition rates were observed in the **Sales** and **HR** departments compared to others.
   - **Overtime:** Employees frequently working overtime are at a much higher risk of attrition.
   - **Promotions:** Employees who **haven't received a promotion in 5+ years** exhibit a significantly higher attrition rate compared to those recently promoted.
3. **Compensation and Benefits:**
   - **Income:** Lower-income employees (e.g., earning below the median income) are at a significantly higher risk of leaving.

4. **Work-Life Balance:**

- Employees with poor work-life balance (indicated by high overtime) and long commuting distances face higher attrition risks.
- Training time doesn't have a significant influence on attrition but maintaining development programs still supports retention.

# SURPRISING TRENDS

**Gender Neutrality**:

- Gender was not a significant factor in attrition. Both males and females displayed similar attrition patterns, contradicting common assumptions.

**Distance from Home**:

- Employees living closer to their workplace **(<5 km**) were surprisingly at slightly higher risk of leaving compared to those with moderate commuting distances.

**Job Role Differences**:

- Employees in job roles such as Sales Executives and Laboratory Technicians had significantly higher attrition rates, even when controlling for compensation and satisfaction levels.

**Overtime's Impact:**

- A staggering **60%** attrition rate was observed among employees working overtime, far higher than other factors.

# RECOMMENDATIONS

**Improve Job Satisfaction:**

- Regular surveys to gauge employee satisfaction and take corrective actions based on feedback.
- Provide clear career growth opportunities and structured paths for promotions.

**Work-Life Balance:**

- Limit overtime hours to avoid employee burnout.
- Implement flexible work arrangements or hybrid working options for employees facing commuting difficulties.

**Employee Retention Programs:**

- Increase engagement and retention efforts, especially for younger and single employees, through tailored mentorship and career development programs.
- Enhance rewards such as performance-based bonuses, stock options, or recognition awards for high-performing employees.

**Focus on High-Risk Groups:**

- Prioritize efforts in high-attrition departments like **Sales** and **HR**.
- Introduce additional training and development initiatives for roles such as **Sales Executives** and **Technicians** to boost morale and skills.

**Promotion Framework:**

- Develop clear policies for timely promotions and communicate them transparently to employees.
- Offer alternatives like lateral moves or increased responsibilities to employees who may not immediately qualify for promotion.

**Fair Compensation:**

- Conduct market comparisons to ensure competitive salaries, especially for lower-income groups.
- Provide annual raises to keep compensation aligned with employee expectations.

# CONCLUSION

The analysis highlights that attrition in the organization is influenced by a combination of demographic, job-related, and compensation factors. Key contributors to attrition include poor job satisfaction, frequent overtime, lack of career advancement opportunities, and inadequate compensation. Surprisingly, gender and training time were found to have minimal impact on attrition.

By implementing the recommendations above, the organization can improve retention rates, reduce turnover costs, and foster a more satisfied and engaged workforce. Specifically, addressing issues related to promotions, overtime, and employee recognition can have a profound impact on improving employee loyalty and overall organizational performance.

# REFERENCES

Thathvar, Pavan Subhash. "IBM HR Analytics Employee Attrition Dataset." *Kaggle*, https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset.