Ramanjeet Kaur

Damilola Agbolabori

HR Analytics Understanding Employee Attrition

Table of Content

[Introduction 2](#_Toc164189774)

[Research Objectives and Scope 3](#_Toc164189775)

[**Objective 1** 3](#_Toc164189776)

[**Objective 2** 3](#_Toc164189777)

[**Objective 3** 3](#_Toc164189778)

[**Scope of the Study** 3](#_Toc164189779)

[Data 4](#_Toc164189780)

[**Statistical Summary** 5](#_Toc164189781)

[Data Preprocessing and Exploratory Data Analysis 6](#_Toc164189782)

[**Key Performance Indicators (KPIs) and Evaluation Metrics** 9](#_Toc164189783)

[Findings and Results 11](#_Toc164189784)

[Conclusion 21](#_Toc164189785)

# Introduction

In today's dynamic corporate landscape, organizations are increasingly turning to HR Analytics as a strategic tool to decipher the wealth of data at their disposal and derive actionable insights. This analytical approach empowers HR departments and people managers to identify, understand, and respond to critical people-related trends within their organizations. Among the myriad challenges faced by HR personnel and people managers, employee attrition emerges as a pressing issue that directly impacts organizational stability, productivity, and success.

Employee attrition, whether voluntary or involuntary, presents multifaceted challenges influenced by a complex interplay of factors ranging from macroeconomic trends to individual motivations and job satisfaction levels. The departure of skilled and experienced employees can disrupt operations, erode institutional knowledge, lower morale, and escalate recruitment costs. Thus, it is imperative for organizations to comprehend the root causes of attrition and proactively address them through strategic interventions aimed at enhancing employee engagement, satisfaction, and retention.

HR Analytics holds the key to unlocking insights into the factors driving attrition rates within organizations. By leveraging data analysis techniques, HR specialists and people managers can identify key trends and predictors associated with employee turnover. These may include demographic factors, performance indicators, job satisfaction levels, and organizational culture. Armed with this knowledge, organizations can develop targeted strategies to mitigate attrition, thereby safeguarding their talent pool and sustaining long-term business success.

This study aims to harness the power of HR Analytics to delve into the intricacies of employee attrition. By closely examining people-related trends and factors linked to turnover, we seek to provide organizations with invaluable insights to inform decision-making and drive proactive interventions. Through a comprehensive analysis of data, we aim to shed light on the underlying dynamics of attrition, enabling organizations to implement tailored strategies aimed at fostering employee engagement, satisfaction, and retention.

# Research Objectives and Scope

### **Objective 1**

Explore Demographic Patterns and Attrition Trends

* Analyze the distribution of attrition among employees based on age, gender, marital status, and other demographic variables.
* Identify any significant correlations between demographic factors and employee churn rates.

### **Objective 2**

Assess Job Satisfaction and Performance Indicators

* Evaluate job satisfaction levels across different job roles, departments, and organizational levels.
* Investigate the relationship between performance ratings, job involvement, and employee retention.

### **Objective 3**

Examine Work-Life Balance and Employee Engagement

* Assess the impact of factors such as commute distance, overtime work, and work-life balance ratings on attrition rates.
* Explore patterns in employee engagement and satisfaction to identify areas for improvement.

### **Scope of the Study**

* The study will focus on analyzing a dataset containing information on 1,470 employees across 35 variables.
* Data exploration and analysis will encompass various dimensions, including demographic characteristics, job satisfaction levels, performance indicators, and retention factors.
* The study will utilize HR Analytics techniques to extract meaningful insights and identify actionable trends related to attrition.
* Emphasis will be placed on understanding the interplay between different variables and their impact on employee turnover.

# Data

This IBM data collection was discovered on Kaggle and is openly accessible. There are 35 variables and 1,470 employee records. The description of each variable is provided below:

* Age: Age in years of the employee
* Attrition: People who people leave
* BusinessTravel: How often an employee embarks on a job related travel
* DailyRate: Daily rate at which an employee is paid
* Department: Department where the employee works
* DistanceFromHome: Distance an employee travels from home to work
* Education: Level of education of the employee
* EducationField: What field the employee studied in school
* EmployeeCount: Count of employee
* EmployeeNumber: EMployee number
* EnvironmentSatisfaction: Employee environment satisfaction
* Gender: Gender of the employee
* HourlyRate: Hourly rate of pay of the employee
* JobInvolvement: Employee job involvement ratings
* JobLevel: Employee Job level
* JobRole: Employee Job role
* JobSatisfaction: Employee Job Satisfaction
* MaritalStatus: Employee Marital Status
* MonthlyIncome: Employee monthly income
* MonthlyRate: Employee Monthly rate
* NumCompaniesWorked: Number of companies worked
* Over18: Age over 18 years
* OverTime: Work overtime
* PercentSalaryHike: Salary increment in Percentages
* PerformanceRating: Performance rating
* RelationshipSatisfaction: Relationship satisfaction
* StandardHours: Employee standard hours worked
* StockOptionLevel: Stock options level
* TotalWorkingYears: Total working hours
* TrainingTimesLastYear: Total working years
* WorkLifeBalance: Work life balance rating
* YearsAtCompany: Years at the company
* YearsInCurrentRole: Years in current role
* YearsSinceLastPromotion: Years since last promotion
* YearsWithCurrManager: Years with current manager

### **Statistical Summary**

The Statistical summary of the dataset is described as follows: Count: 1470 entries

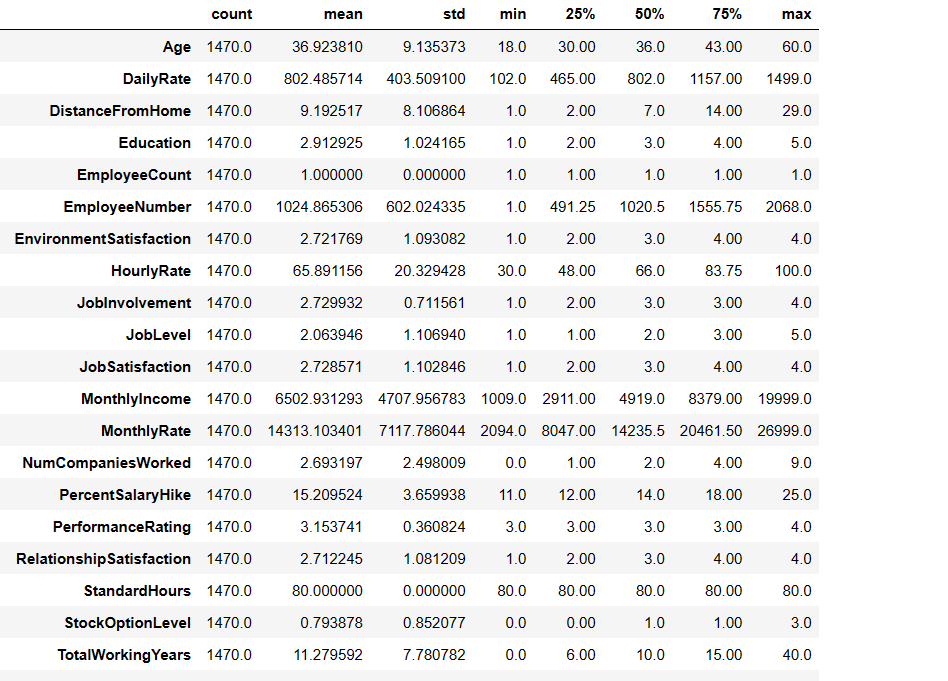
Numeric Variables:

* + Mean age of employees is approximately 36.92 years, with a standard deviation of 9.14 years.
  + DailyRate ranges from 102 to 1499, with a mean of approximately 802.49 and a standard deviation of 403.51.
  + DistanceFromHome ranges from 1 to 29, with a mean of approximately 9.19 and a standard deviation of 8.11.
  + Education levels range from 1 to 5, with a mean of approximately 2.91 and a standard deviation of 1.02.
  + EmployeeCount is always 1.
  + EmployeeNumber ranges from 1 to 2068, with a mean of approximately 1024.87 and a standard deviation of 602.02.
  + EnvironmentSatisfaction ranges from 1 to 4, with a mean of approximately 2.72 and a standard deviation of 1.09.
  + HourlyRate ranges from 30 to 100, with a mean of approximately 65.89 and a standard deviation of 20.33.
  + JobInvolvement ranges from 1 to 4, with a mean of approximately 2.73 and a standard deviation of 0.71.

Categorical Variables:

* + Gender: Two categories (Male and Female)
  + BusinessTravel: Three categories (Non-Travel, Travel\_Rarely, Travel\_Frequently)
  + Department: Multiple departments (e.g., Sales, Research & Development, Human Resources)
  + EducationField: Multiple fields of education (e.g., Life Sciences, Medical, Marketing)
  + JobRole: Multiple job roles (e.g., Sales Executive, Research Scientist, Laboratory Technician)
  + MaritalStatus: Three categories (Married, Single, Divorced)
  + Over18: One category (Yes)
  + OverTime: Two categories (Yes and No)

This summary provides a concise overview of the dataset, including its size, the range and distribution of numeric variables, and the categories present in categorical variables.



# Data Preprocessing and Exploratory Data Analysis

Based on the preliminary data profiling, it is evident that the dataset is generally in a clean state, showing minimal issues with missing values or inconsistencies. However, to further prepare the data for analysis, we performed a transformation on the 'Attrition' column, converting its values to binary ('YES' or 'NO'). This transformation simplifies the interpretation of attrition status, facilitating clearer analysis and insights.

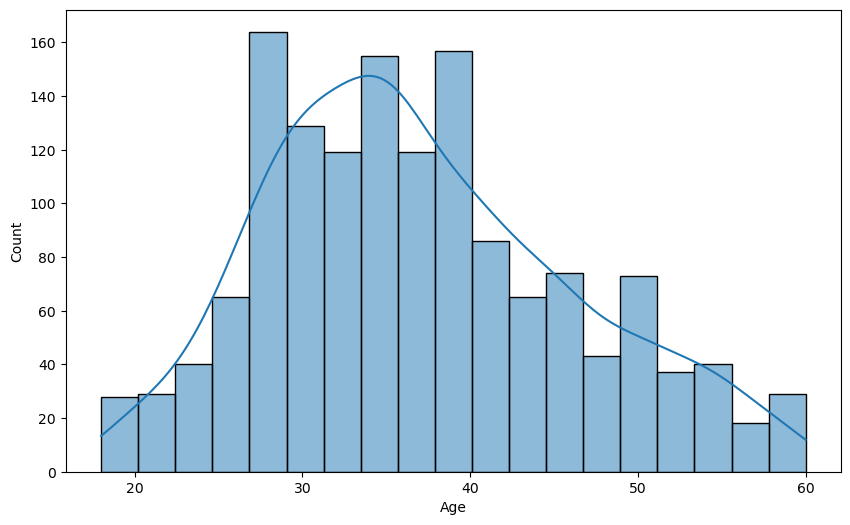
* Age\_Group was derived from the Age column.
* The Attrition column value was transformed into 1 & 0 from Yes & No for some of the analysis.
* MonthlyIncomeCat was derived from the MonthlyIncome column.

**Univariate Analysis: Distribution of Data**

During the exploratory data analysis (EDA) phase, we conducted univariate analysis to examine the distribution of data across several key columns. The following variables were included in our analysis:

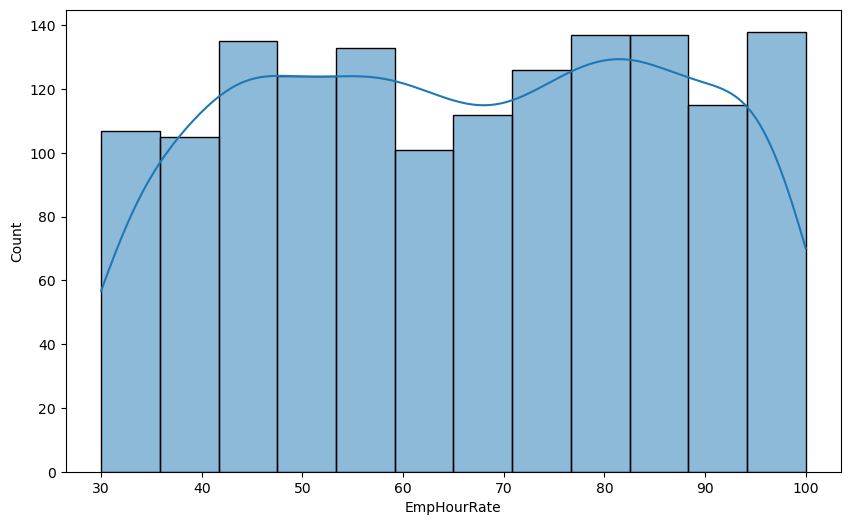
**Age**:

* We investigated the distribution of ages among employees to understand the demographic composition of the workforce.
* The analysis revealed Data my data is normally distributed and we find that range of age between 18 to 60 most of the employees have age between 25 to 40.



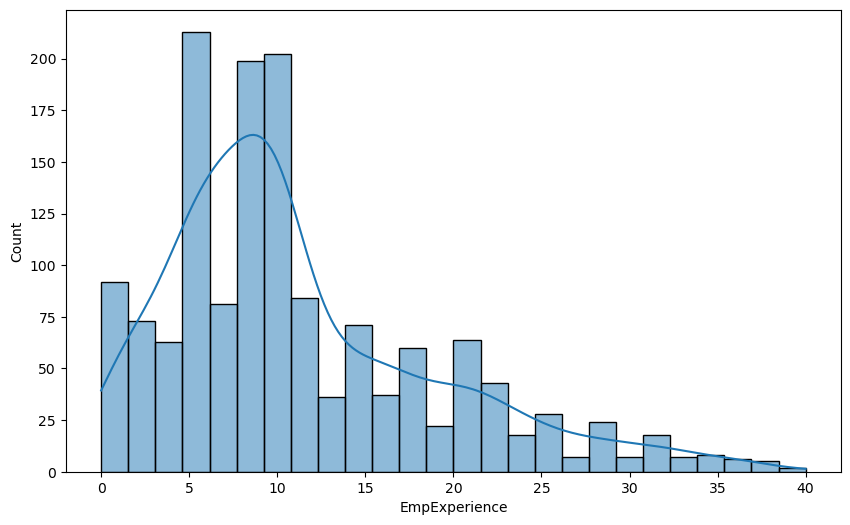
**Hourly Rate**:

* Our analysis of hourly rates provided insights into the range of hourly rate is between 30 to 100 but most of employees hourly rate is 45.



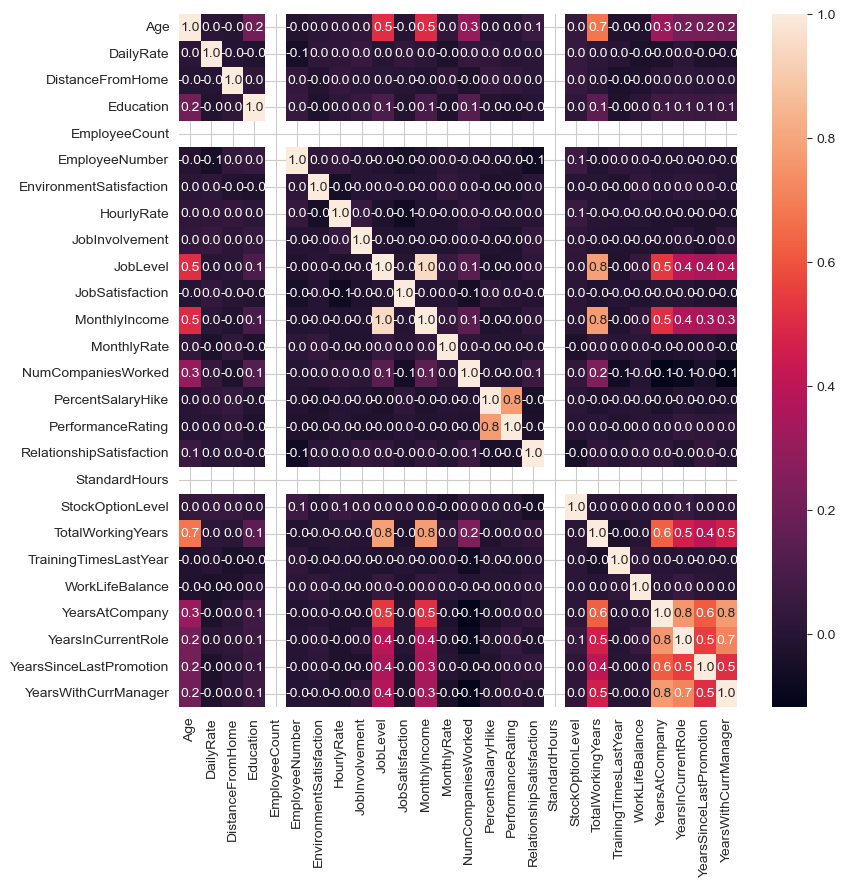
**Total Working Experience**:

* By exploring the total working experience of employees, we aimed to understand the level of expertise present in the workforce and work experience range between 0 to 40, most of the employees experience ranges between 5 to 10.

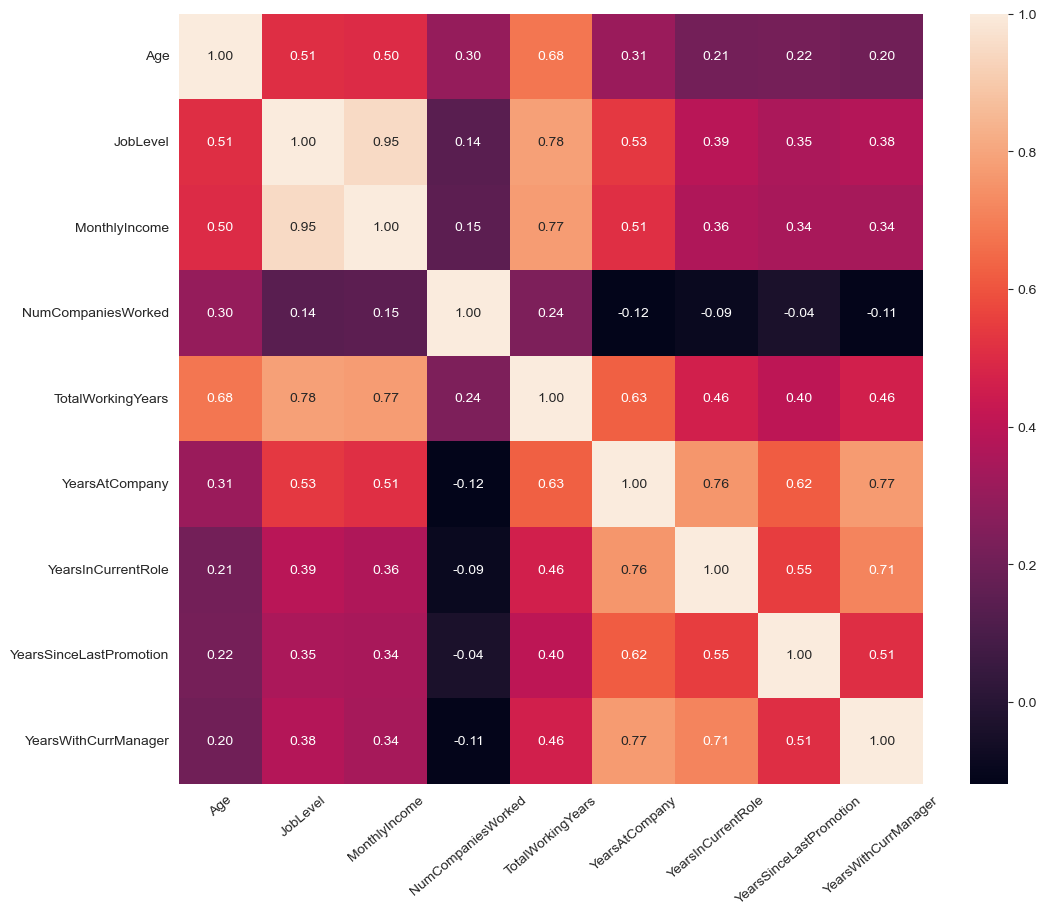


**Correlation Analysis with Heatmap:**

After conducting univariate analysis on various numeric columns, we proceeded to explore the relationships between these variables by calculating their correlation coefficients. This correlation analysis provides insights into the degree of association between different numerical features within the dataset. Using the Pandas library, we computed the correlation matrix, which measures the pairwise correlations between all numeric columns. We then visualized this correlation matrix using a heatmap to better understand the strength and direction of these relationships. The heatmap allows quickly identifying patterns and dependencies among variables, with brighter colors indicating stronger correlations.



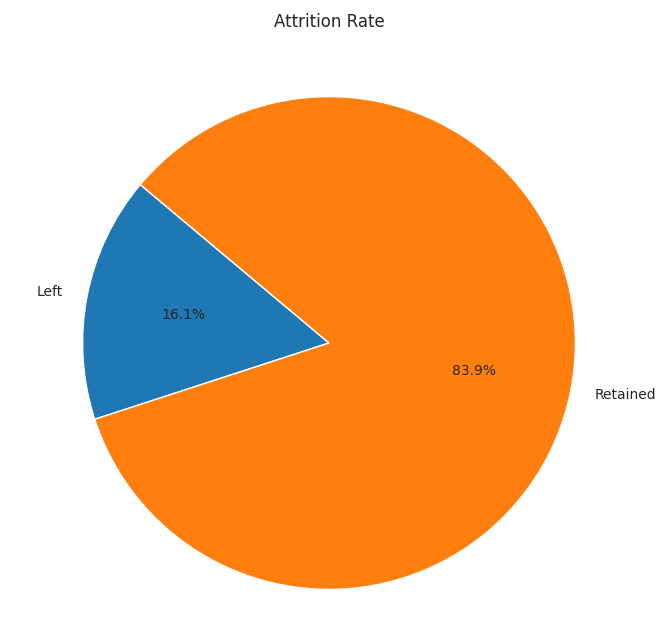
And after the careful examination of heatmap we find the co relation among various columns, like Age, JobLevel, MonthlyIncome, NumCompaniesWorked, TotalWorkingYear, YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, YearsWithCurrManager.



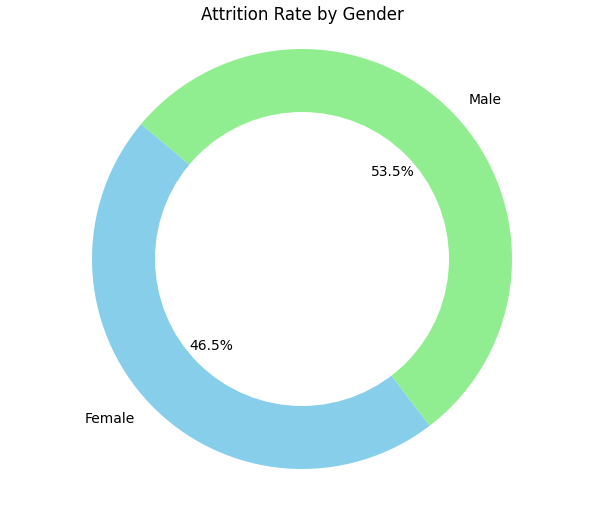
## **Key Performance Indicators (KPIs) and Evaluation Metrics**

**Attrition Rate**: The Attrition rate we calculated as The attrition rate is a measure used by organizations to understand the rate at which employees are leaving the company over a specified period of time. It's an important metric for assessing workforce stability and retention efforts.

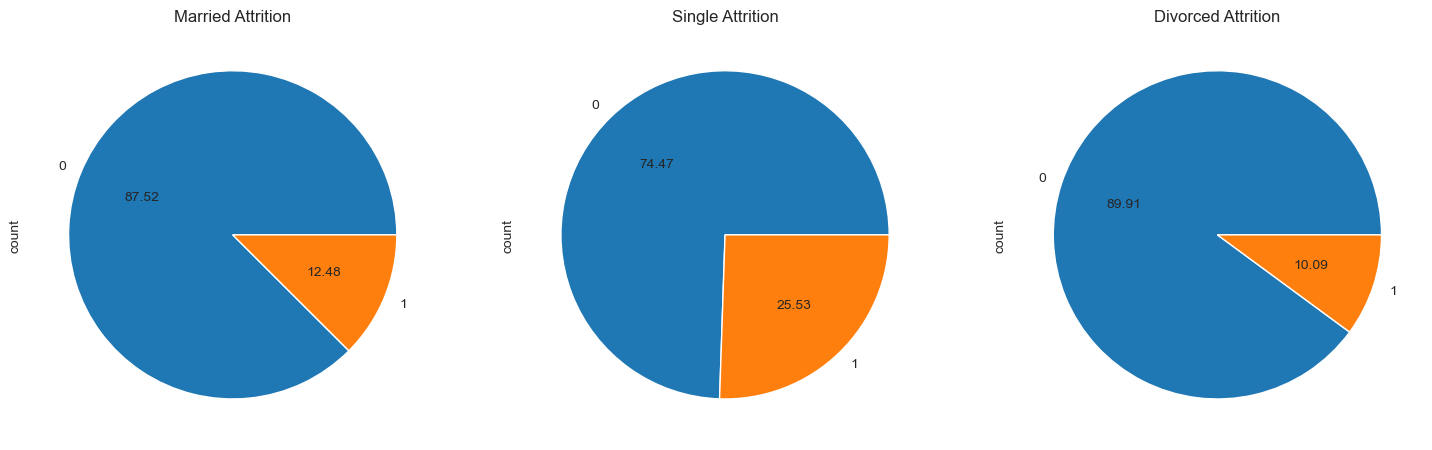
The formula to calculate attrition rate is: **Attrition Rate** = **(Number of employees who left during the period / Average number of employees during the period) x 100**

****

**Attrition by Gender:** We calculate the attrition rate for each gender from 'Attrition' column indicates whether an employee has left the company or not. We grouped the data by 'Gender' and counted the occurrences of attrition for each category. **Attrition Rate By Gender:( Number of employees who left of a particular gender / Total number of employees of that gender) \* 100**

****

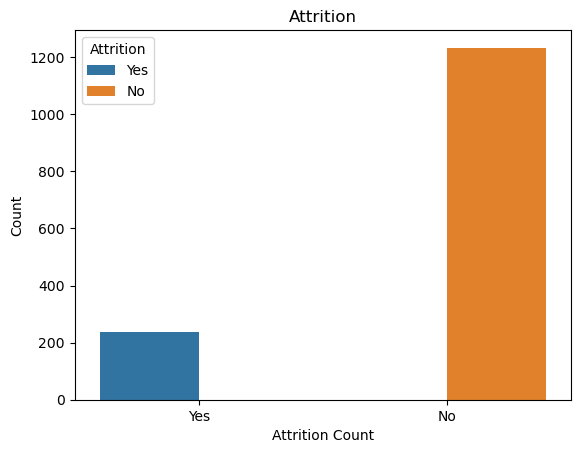
**Attrition By marital Status:** We calculate the attrition rate for each Marital Status from 'Attrition' column indicates whether an employee has left the company or not. We grouped the data by 'Marital Status’' and counted the occurrences of attrition for each category. **Attrition Rate by Marital Status:( Number of employees who left of a particular Marital / Total number of employees of that marital status) \* 100**

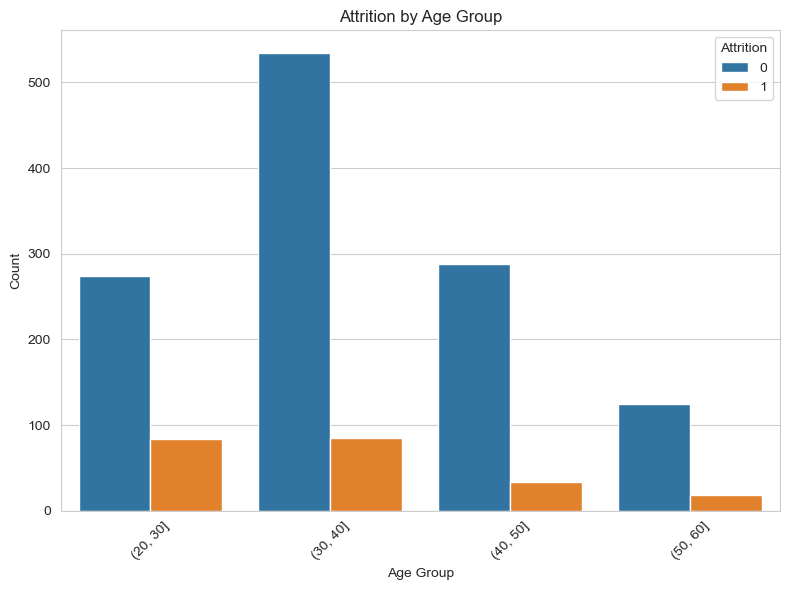
****

# Findings and Results

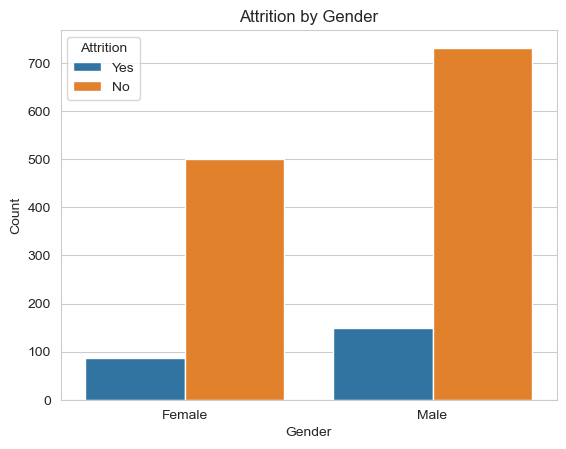
**Attrition Count**: The number of attrition cases in the organization is as follows:

* count of attrition cases for each category (No and Yes) in the organization. If "No" represents employees who did not experience attrition and "Yes" represents employees who did experience attrition.
* No: 1233, Yes: 237"

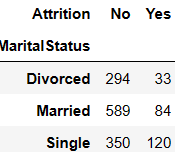
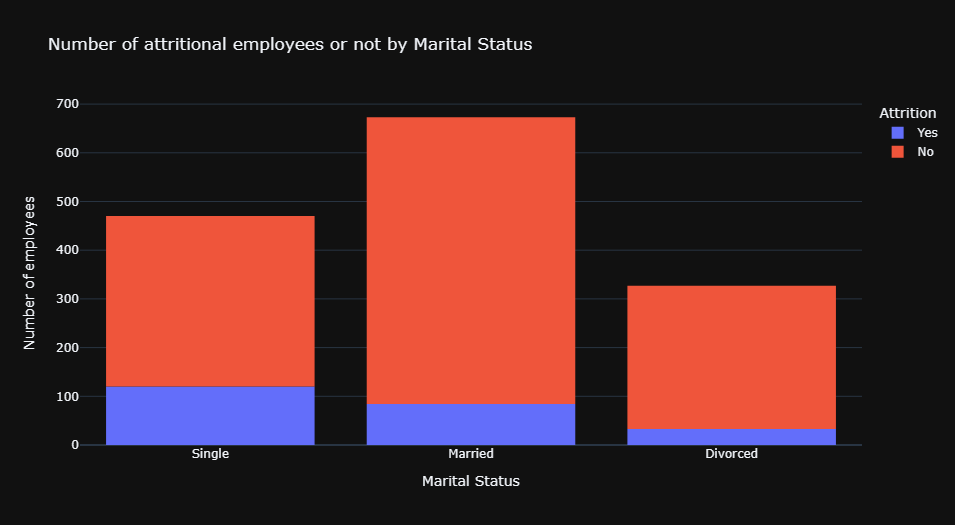


**Attrition Count by Age**: The attrition count by age is analyzed by categorizing ages into different age groups. These new columns were created to detect which age groups like (20,30,40,50,60) have a higher number of attrition cases. The aim is to identify any trends or patterns in attrition rates across different age demographics. 

**Attrition Count by Gender**: The attrition count based on gender indicates that there are 882 males and 588 females in the organization. Therefore, it is evident that males have a higher count in the organization which is 150 as compared to female has87.



**Attrition Count by Marital Status**: Attrition count on the Marital status has single has count of 470, Married has count of 673, and divorced has count of 327 in the organization. The attrition count is higher among single employees which is 120.



**Attrition count by Business Travel**: Attrition count based on business travel has three columns like Travel Rarely Travel\_Rarely has count 1043, Travel\_Frequently has count 277 Non\_Travel 150 . Travel rarely has higher attrition rate.

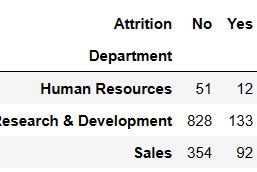
A graph of a graph with different colored bars

Description automatically generated with medium confidence A screenshot of a travel account

Description automatically generated

**Attrition count on Departments**: The attrition count on departments it has three departments like human Resources, Research& Development, Sales. The count for research and development 961.The higher Attrition rate of Research & Development is 133.

A pie chart with numbers and a red triangle

Description automatically generated 

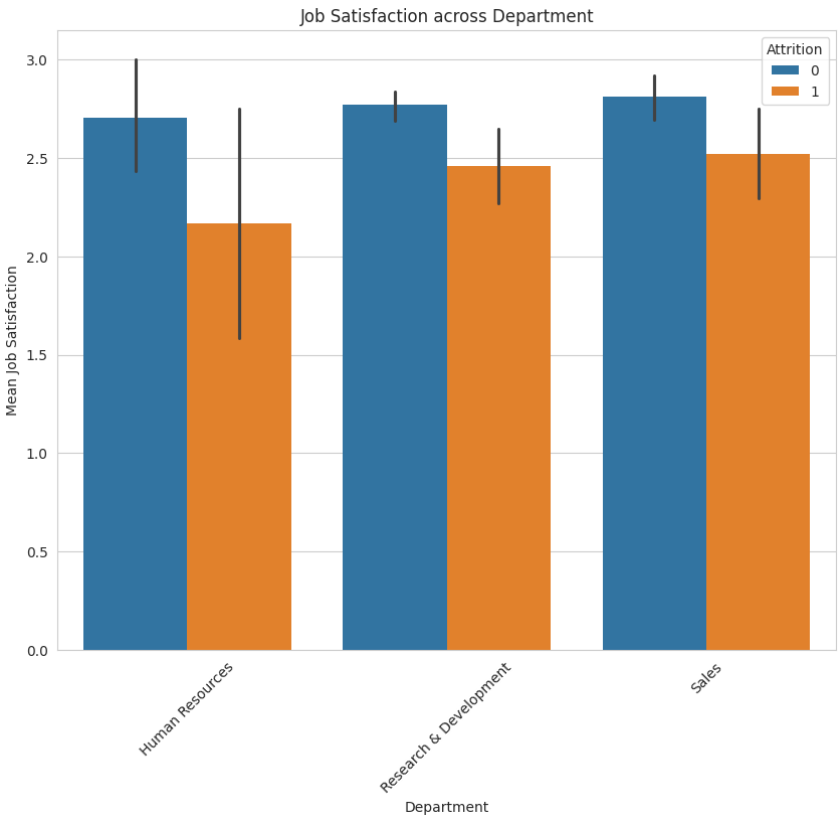
* **Attrition Count on JobRoles**: The attrition count on job roles we have nine Job Roles among nine job roles and higher count of job role is Sales Executive 326 and least count one is Human Resources 52.

A graph of a graph with different colored bars

Description automatically generated with medium confidence A screenshot of a table

Description automatically generated

**Departments vs Job Satisfaction**: The Research & Development and Sales departments have about the same average job satisfaction level and for employees that decided to leave and stay.



**Attrition count on Job satisfaction**: The job satisfaction analysis reveals attrition patterns: 2-star satisfaction has the lowest attrition, while 1 and 3-star satisfaction have the highest rates. 4-star satisfaction correlates with reduced departures, highlighting the relationship between job satisfaction and attrition.

A graph of a number of people

Description automatically generated

**Job Satisfaction and Job Role:** The Research Director role has lower attrition rates and higher job satisfaction levels compared to other job roles depicted in the visualization. This suggests that employees in the Research Director role may be more satisfied with their jobs and are less likely to leave the company, which could indicate a positive aspect of that particular role within the organization.

A graph of a number of people

Description automatically generated

**Department, Job Satisfaction and Attrition:** The Research & Development and Sales departments have about the same average job satisfaction level and for employees that decided to leave and stay.

A graph of different colored bars

Description automatically generated with medium confidence

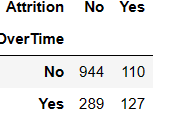
**Performance ratings, job involvement, and employee retention**: The performance rating and job involvement and employee retention rate which is 84% despite differences in attrition status, employees with performance ratings 3 and 4 have similar distributions of job involvement. This insight could indicate that job involvement may not be a significant factor influencing attrition among employees with these performance ratings. However, further analysis and consideration of other factors are needed to make conclusive interpretations about the relationship between performance ratings, job involvement, and employee retention.

A diagram of a performance rating

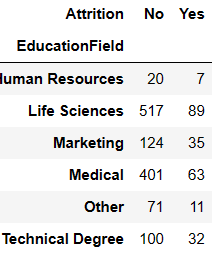
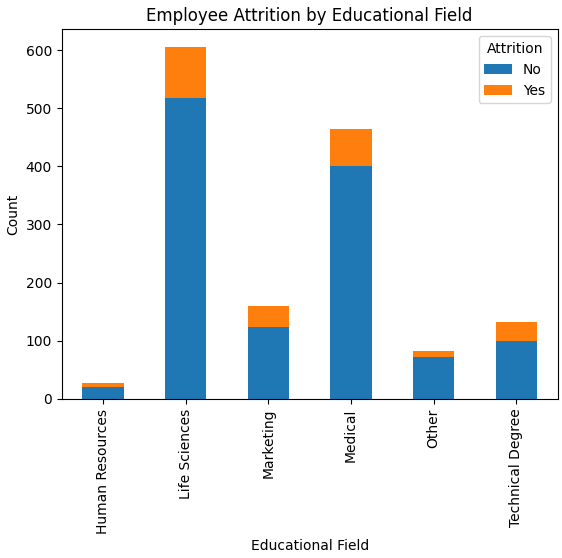
Description automatically generated

**Attrition Count on Work time and Attrition rate:** The attrition rate based on overtime has values of 'Yes' and 'No', with 416 'Yes' and 1054 'No' counts. The attrition rate for 'Yes' is 28.30%.

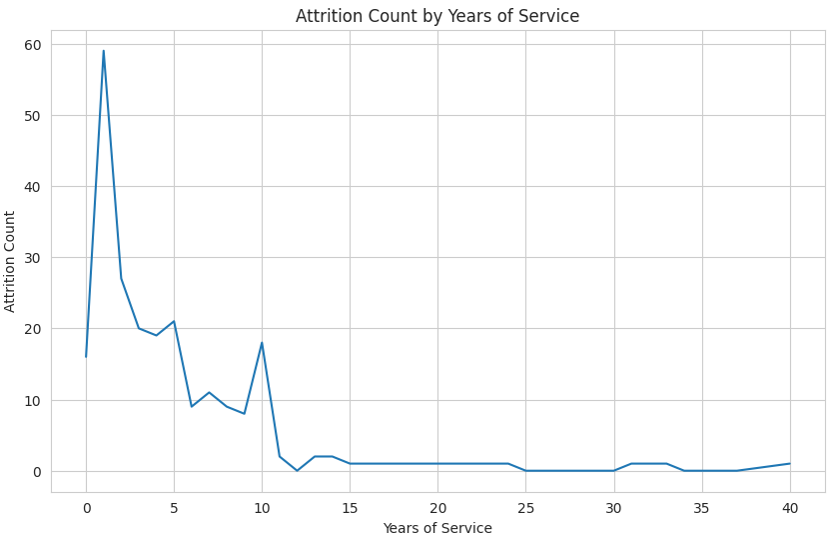
A graph of an attrition rate

Description automatically generated 

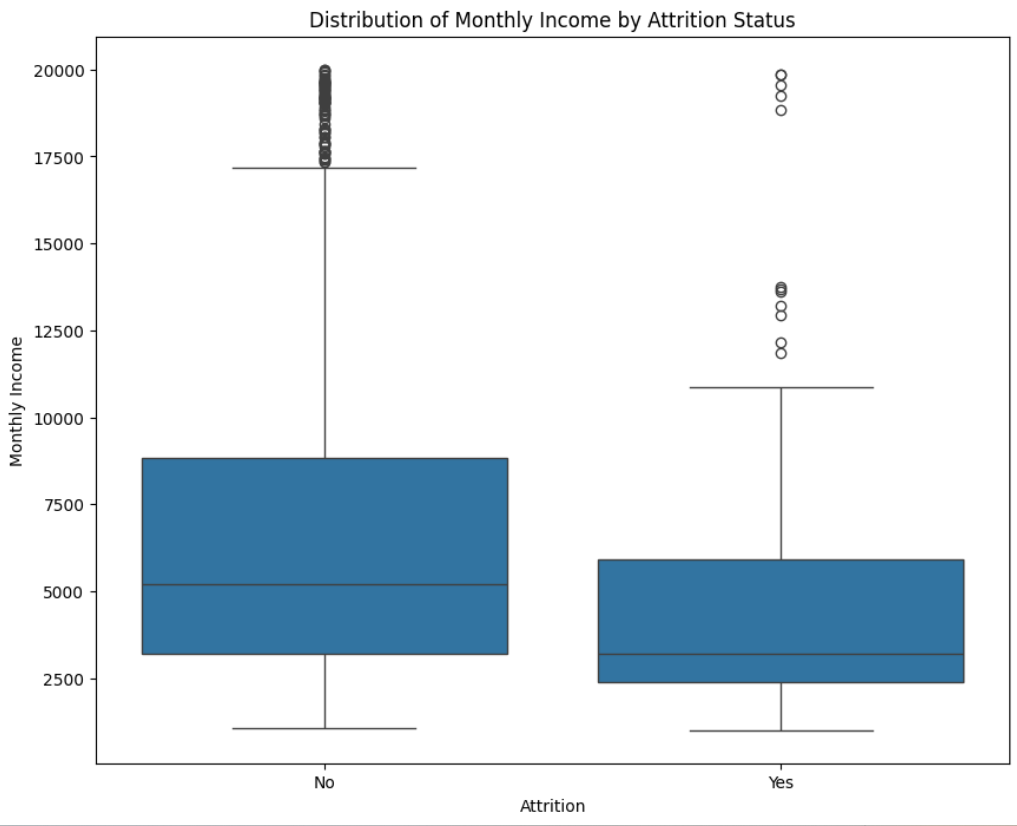
**Attrition Rate by Educational Field**: There is more attrition among employees in the Life Science and Medical educational field than others which has count of 89.

****

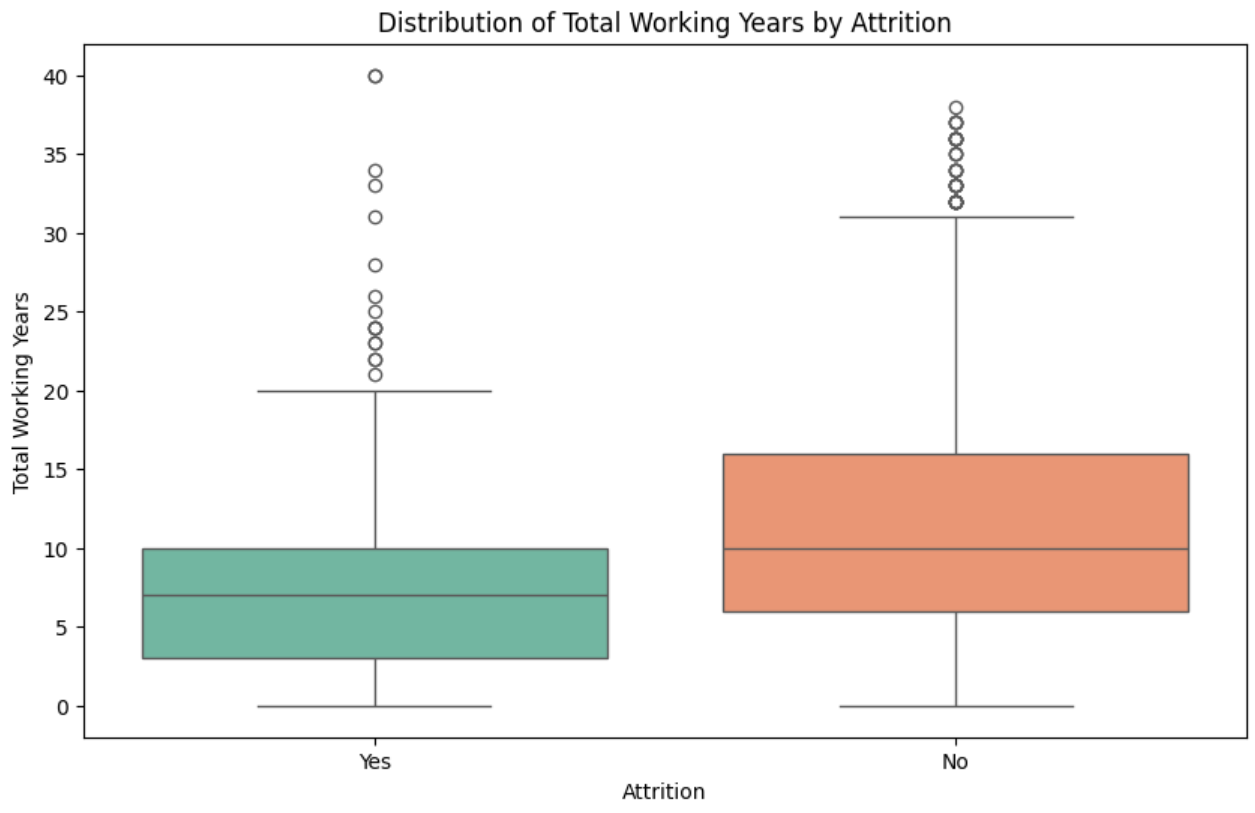
**Year of Service and Attrition:** The more the number of years spent in service the lower the chances of attrition, this is seen from the steady decrease in attrition count as years in service increase. It is also noticed that most attrition occur between one to three years in service which is probably enough time for an employee to decide whether to stay or not.



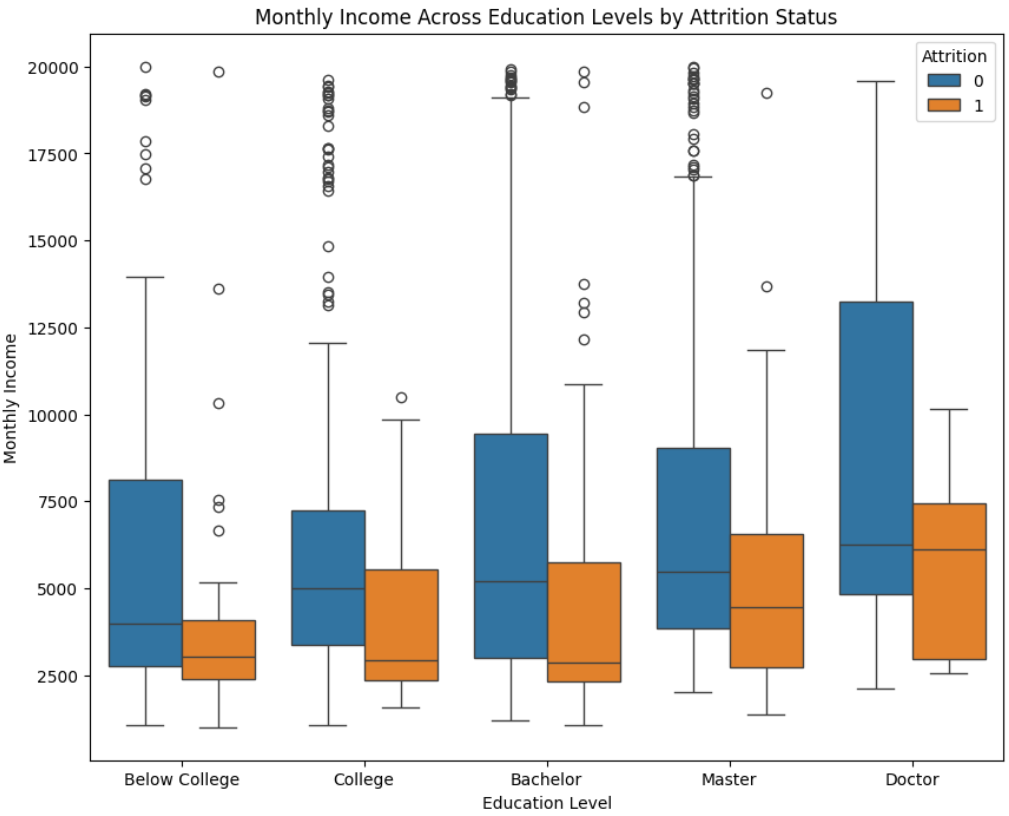
**Monthly Income and Attrition:** Employees that leave have a lower median income than those who choose to stay.



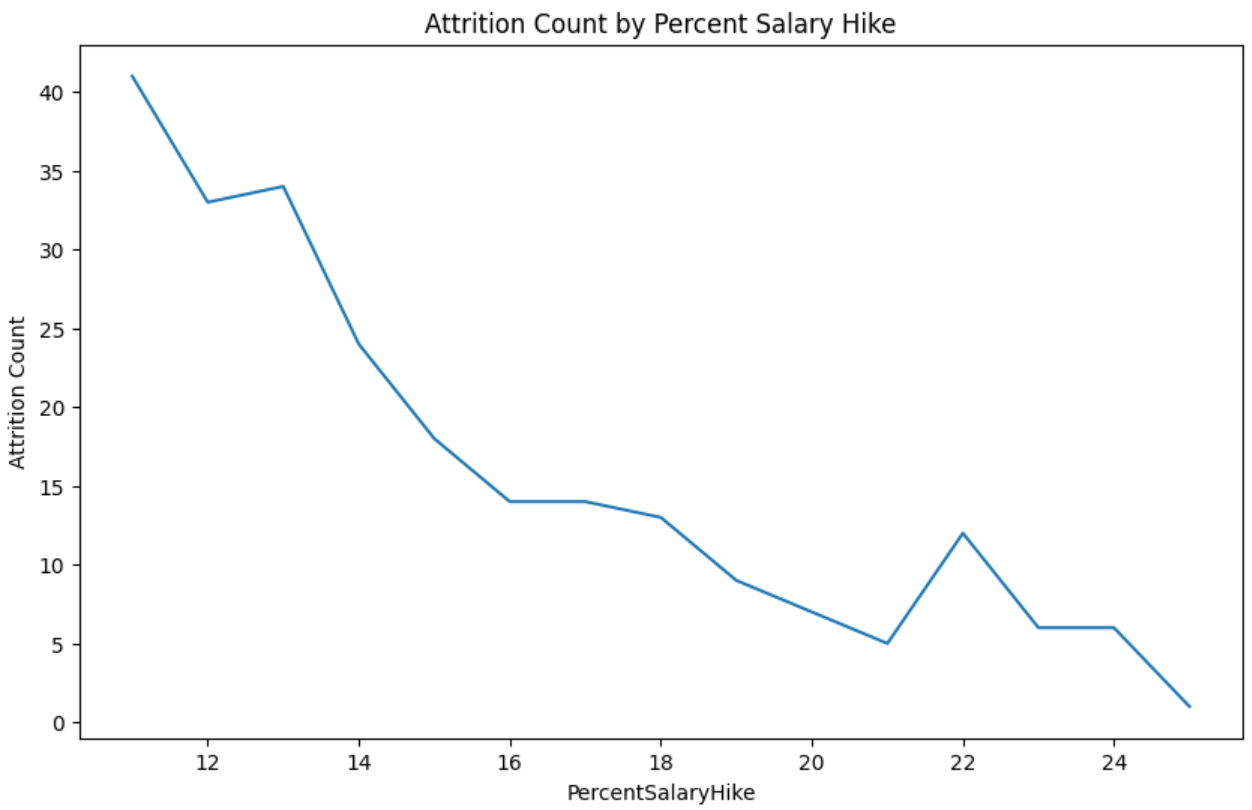
**Total Working Years and Attrition:** The median of total working years for employees that leave is lower than that of employees that stay. The range of total working years of these employees is also lower.



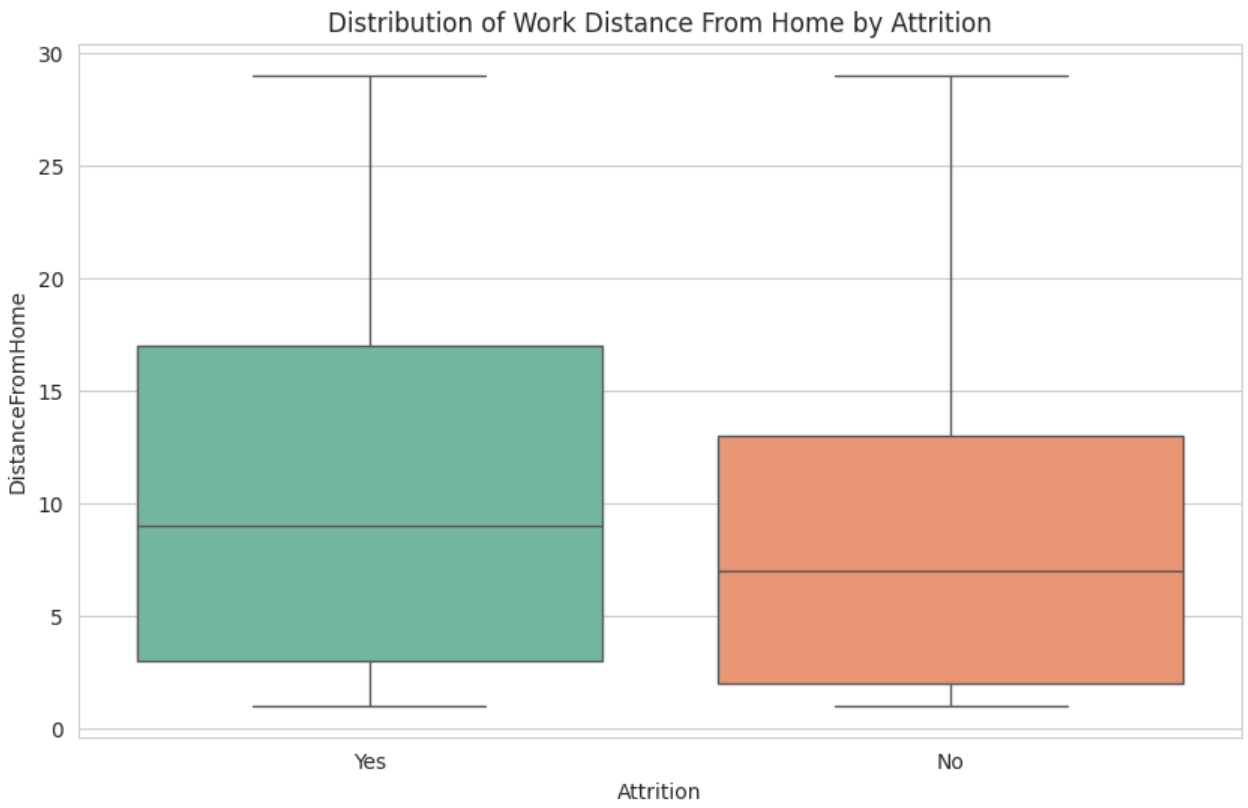
**Educational Level, Monthly Income and Attrition:** The median income for employees with PhD is highest but not significantly higher than other educational levels. The median income for PhD employees that leave is about the same as employees that stay, which is quite different from the pattern of others where it is usually lesser. Monthly Income might not be enough reason to make PhD employees stay.



**Percentage Salary Hike and Attrition:** The number of employees that leave reduces as the percentage salary hike increases. Employees can be motivated to stay longer in their current organization by increasing salary.



**Work Distance from Home and Attrition:** The median distance between work and home for employees that leave is higher than that of employees that stay. Employees might be leaving for jobs closer to their homes.



# Conclusion

The analysis of the HR dataset revealed several key insights regarding employee attrition and related factors. The dataset contained information on various employee characteristics, such as age, job role, marital status, job satisfaction, and department. Here is a summary of the findings:

* **Attrition Rate**: The stacked bar graph indicated that 16.1% of employees chose to leave the company, while 83.9% decided to stay, highlighting a significant contrast in employee preferences.
* **Gender Distribution**: The donut chart showed that males constituted 60% of the workforce, while females made up the remaining 40%, illustrating a notable male majority within the organization.
* **Tenure and Monthly Income**: The line graph revealed a positive correlation between employee tenure and monthly income, indicating that longer-serving employees tend to earn higher incomes.
* **Age and Attrition**: The histogram analysis highlighted that the age range of 25 to 35 exhibited a higher attrition rate, while attrition declined notably beyond the age of 40, suggesting a period of stability and reduced turnover after reaching 40 years of age.
* **Marital Status Distribution**: The pie chart demonstrated that 45.8% of employees were married, 32% were single, and 22.2% were divorced, providing insights into the organization's marital demographics.
* **Marital Status and Attrition**: The histogram analysis revealed distinct attrition patterns based on marital status. Singles had the highest attrition rate, followed by married individuals, while divorced individuals displayed the lowest attrition rate.
* **Job Satisfaction and Attrition**: The bar graph indicated that employees with a 2-star job satisfaction rating had the lowest attrition, while those with 1 and 3-star ratings experienced the highest attrition rates. This finding emphasized the importance of job satisfaction in reducing voluntary departures.
* **Department-wise Attrition**: The bar chart identified that the Research & Development Department had the highest attrition rate, followed by the Sales and HR Departments. This finding suggested a need for further investigation into the reasons behind the high attrition rates in these departments.
* **Over-Time Attrition**: The attrition rate based on overtime has values of 'Yes' and 'No', with 416 'Yes' and 1054 'No' counts. The attrition rate for 'Yes' is 28.30%.

These insights provide valuable information for organizations aiming to address attrition challenges and implement targeted strategies to improve employee retention and satisfaction. By considering factors such as job satisfaction, age demographics, and department-specific attrition rates, organizations can take proactive measures to enhance employee engagement, productivity, and overall organizational success.