Volume 10, Issue 11, November – 2025 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

Exploratory Data Analysis of Employee Attrition Trends

P. S. Sudharshini[1]; R. B. I. V. Rathnamalala[2]; A. I. Udeshika[3]; D. C. Dahanayaka[4]

[1]Nawaloka College of Higher Studies
[2]Asiri Central Hospital
[3]Zone 247
[4]Habaraduwa Pradeshiya Sabha

Publication Date: 2025/12/03

Abstract: This exploratory data analysis (EDA) examines the factors influencing employee attrition using a Kaggle dataset comprising 1000 employee records with 26 demographic, job-related, compensation, and satisfaction variables. The target variable, Attrition (Yes/No), reveals a relatively low overall turnover rate of 18.9%, with a marked class imbalance (81.1% No vs. 18.9% Yes).

Descriptive statistics and visualizations highlight a mid-career workforce (median age 41 years, average tenure approximately 15 years), male dominated gender composition, and finance as the largest department. Key findings indicate that lower monthly income, overtime work, and certain high pressure job roles (particularly Executive and Manager positions) are strongly associated with higher attrition rates. Employees who left the organization typically earned less, worked overtime more frequently, and had slightly shorter tenure, suggesting compensation dissatisfaction, burnout, and early career vulnerability as prominent drivers.

Surprisingly, job satisfaction alone showed weak predictive power, with attrition occurring across all satisfaction levels, including high ratings, underscoring the multifaceted nature of turnover. While age and tenure provide contextual insights, compensation, overtime, and job role emerged as the strongest correlates of attrition.

These results offer actionable insight for organizations to prioritize competitive salary structures, overtime reduction policies, and targeted retention programs for leadership roles, thereby reducing turnover costs and fostering a more stable, engaged workforce.

How to Cite: P. S. Sudharshini; R. B. I. V. Rathnamalala; A. I. Udeshika; D. C. Dahanayaka (2025). Exploratory Data Analysis of Employee Attrition Trends. *International Journal of Innovative Science and Research Technology*, 10(11), 2371-2378. https://doi.org/10.38124/ijisrt/25nov1298

I. INTODUCTION

In today's dynamic business environment, retaining skilled and experienced employees is a critical challenge for organizations. Employee attrition, defined as the voluntary or involuntary departure of employees from a company, can significantly impact an organization's productively, financial stability, and overall morale. The costs associated with high turnover are substantial, encompassing expenses related to recruitment, onboarding, training new hires, and the loss of institutional knowledge and expertise. Furthermore, frequent employee departures can disrupt team unity and negatively affect the morale of remaining staff.

➢ *Problem Statement:*

This project aims to conduct an in-depth analysis of the "Employee Attrition Prediction Dataset" to identify and understand the various factors that influence employee attrition. The core problem is to ascertain which demographic, job-related, compensation, and satisfaction variables are most strongly associated with an employee's decision to leave the organization.

➢ *Objectives:*

The specific objectives of this study are:

- To identify trends and patterns in employee attrition.
- To investigate the relationship between attrition and factors such as salary, job satisfaction, work-life balance, and years at the company.
- To explore the distribution of key demographic and job-related variables.

➢ *Significance:*

The insights derived from this analysis hold significant value for the organization. By pinpointing the variables that

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

most strongly correlate with attrition, management and Human Resources (HR) departments can develop targeted retention strategies. For instance, if specific job roles or compensation levels are identified as high-risk areas, tailored programs such as salary adjustments, career development opportunities, or stress management initiatives could be implemented. Understanding the impact of factors like overtime and work-life balance can lead to policy change that promote a healthier and more sustainable work environment. Ultimately, this report provides a data-driven foundation for improving employee engagement, reducing turnover costs, and fostering a stable and productive workforce, contributing to the long-term success and competitiveness of the organization.

II. DATA

The dataset used for this comprehensive analysis is the "Employee Attrition Prediction Dataset", which was sourced from Kaggle. This dataset provides strong foundation for understanding employee turnover dynamics within a simulated organizational context.

➢ *Dataset Structure:*
The dataset contains records for exactly 1000 employees, with each row representing a unique employee observation. For each employee, there are 26 distinct attributes, offering a multi-dimensional view of various aspects of their employment and personal characteristics.

➢ *Target Variable:*
The primary variable of interest and the focus of this predictive analysis is Attrition. This is a binary categorical variable that indicates whether an employee has left the organization (Yes) or remained (No). Understanding the drivers behind this variable is the central goal.

➢ *Variables Categories:*
The 26 attributes can be logically grouped into several key categories to facilitate analysis:

- *Demographic Variables:*
These describes the personal characteristics of the employees.

✓ *Age:*
The age of the employee in years. (Numerical, Continuous)

✓ *Gender:*
Categorical, indicating whether the employee is Male or Female.

✓ *Marital Status:*
Categorical, representing the employee's marital status.

- *Job-Related Variables:*
These variables relate to the employee's role, length of service, and work conditions within the company.

✓ *Department:*
Categorical, the specific department an employee works in (e.g., Finance, IT, HR).

✓ *Job Role:*
Categorical, the specific position held by the employee (e.g., Analyst, Manager, Executive)

✓ *Job Level:*
An ordinal numerical variable, likely representing ranked seniority (e.g., 1-5)

✓ *Years at Company:*
Numerical (Continuous), the total number of years the employee has been with the organization.

✓ *Years in Current Role:*
Numerical (Continuous), the duration an employee has held their current position.

✓ *Overtime:*
Categorical, indicating whether an employee regularly works overtime ("Yes" or "No").

✓ *Project Count:*
Numerical (Discrete), the number of projects an employee is currently involved in.

✓ *Absenteeism:*
Numerical (Discrete), a measure of how often an employee is absent.

- *Compensation Variable:*
These variables provide insight into the financial aspects of employment.

✓ *Monthly Income:*
Numerical (Continuous), the employee's gross monthly earnings.

✓ *Hourly Rate:*
Numerical (Continuous), the employee's hourly compensation.

✓ *Stock Option Level:*
Ordinal numerical, indicating the level of stock options.

- Satisfaction & Performance Variable:

✓ *Job Satisfaction:*
Ordinal, rated on a scale (e.g., 1-5), indicating satisfaction with their job.

✓ *Work Life Balance:*
Ordinal, rated on a scale (e.g., 1-4), reflecting perceived balance between work and personal life.

✓ *Performance Rating:*
Numerical (Discrete/Ordinal), an evaluation of the employee's performance.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

- *Other Variables:*

✓ *Distance From Home:*
Numerical (Continuous), the travel distance.

✓ *Variable Types:*
The dataset's variables are a mix of:

✓ *Categorical:*
Gender, Department, Marital Status, Job Role, Attrition, Overtime. These represent distinct, non-numeric categories.

✓ *Numerical (Continuous/ Discrete):*
Age, Monthly Income, Years at Company, Absenteeism, Hourly Rate, Distance from Home, Project Count. These are quantitative measures.

✓ *Ordinal:*
Job Satisfaction, Work Life Balance, Job Involvement, Job Level, Stock Option Level, Performance Rating. These are categorical variables with a basic order or ranking.

➢ *Data Source and Quality:*
The dataset was obtained from Kaggle, a reputable platform for data sharing. For the purpose of the exploratory analysis, the dataset was assumed to be clean and ready for direct use. No specific data cleaning or transformation steps were performed prior to this analysis, though such steps would be crucial for a more advances predictive modelling phase.

III. ANALYSIS

The Analysis phase involved a systematic examination of the dataset to uncover patterns and relationships that might explain employee attrition. This section presents the findings from descriptive statistics and various graphical representations.

➢ *Basic Descriptive Statistics of Key Variables*
To initiate the exploratory analysis, a statistical summary of five critical variables was conducted. This provides a foundational understanding of the central tendencies and spread within the employee population.

- *Age:*
The Age of employees within the dataset covers a range from 20 years to 59 years. The median age is 41 years, with the mean being very close to this value. This indicates that the workforce is primarily composed of individuals in their mi-career stages, suggesting a degree of experience and stability within the company. The distribution appears fairly balanced around the central tendency.

- *Monthly Income:*
Monthly Income exhibits a wide range, from a minimum of Rs.3,001 to a maximum of Rs 19,999. The median monthly income is Rs. 11,256. Notably, the mean income (Approximately Rs. 11.992), is slightly higher than the median. This suggests a slight right-skewness in the income distribution, implying that a smaller number of employees earn significantly higher incomes, pulling the mean upwards. This could be indicative of a rank pay structure.

- *Years at Company:*
On average, employees have demonstrated considerable loyalty, with the Years at Company nearing 15 years. The maximum length of service observed in the dataset is an impressive 29 years. This indicates a significant proportion of long-term employees, which can be beneficial for institutional knowledge and stability.

- *Job Satisfaction:*
Job Satisfaction is measured on an ordinal scale from 1 to 5, where higher values indicate greater satisfaction. The mean rating is 3, while the average is 3.15. These figures suggest that, on the whole, employees exhibit moderate levels of job satisfaction. While not overwhelmingly high, it also does not indicate widespread dissatisfaction.

- *Work-Life Balance:*
Similarly, Work Life Balance is measured on an ordinal scale, typically from 1 to 4. The median rating is 3. The mean of 2.5 suggests that while many employees experience an acceptable balance, there might be a portion of the workforce perceiving an imbalance, as the average leans slightly towards the lower end of the scale. This is a crucial area for HR to monitor, as imbalance can contribute to stress and eventual attrition.

➢ *Analysis of the Target Variable: Attrition*
The core of this analysis turns around the Attrition variable, which distinguishes between employees who have left the organization ("Yes") and those who have remained ("No"). Understanding its distribution is primary.

Table 1 Attrition Distribution in the Dataset

Attrition Status	Count	Percentage
No	811	81.1%
Yes	189	18.9%

As per the statistics appeared in the Table 1, substantial majority of the employees, precisely 81.1%, are still with the organization, indicating a generally stable workforce. Only 18.9% of the employees have experienced attrition. This suggests that the overall attrition rate within this dataset is comparatively low. It is crucial to note the significant class imbalance between the 'No' and 'Yes' categories. The 'No' class is considerably larger than the 'Yes' class. This imbalance is an important consideration for any future predictive modelling, as it can bias models toward the majority class and potentially lead to misleading performance metrics if not addressed appropriately. For exploratory analysis, this imbalance means that direct comparisons of counts might sometimes unclear underlying proportions.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

➢ *Visual Analysis and Key Findings*

To further explore the relationships between various attributes and attrition, several visualizations were generated.

- *Employee Age Distribution*

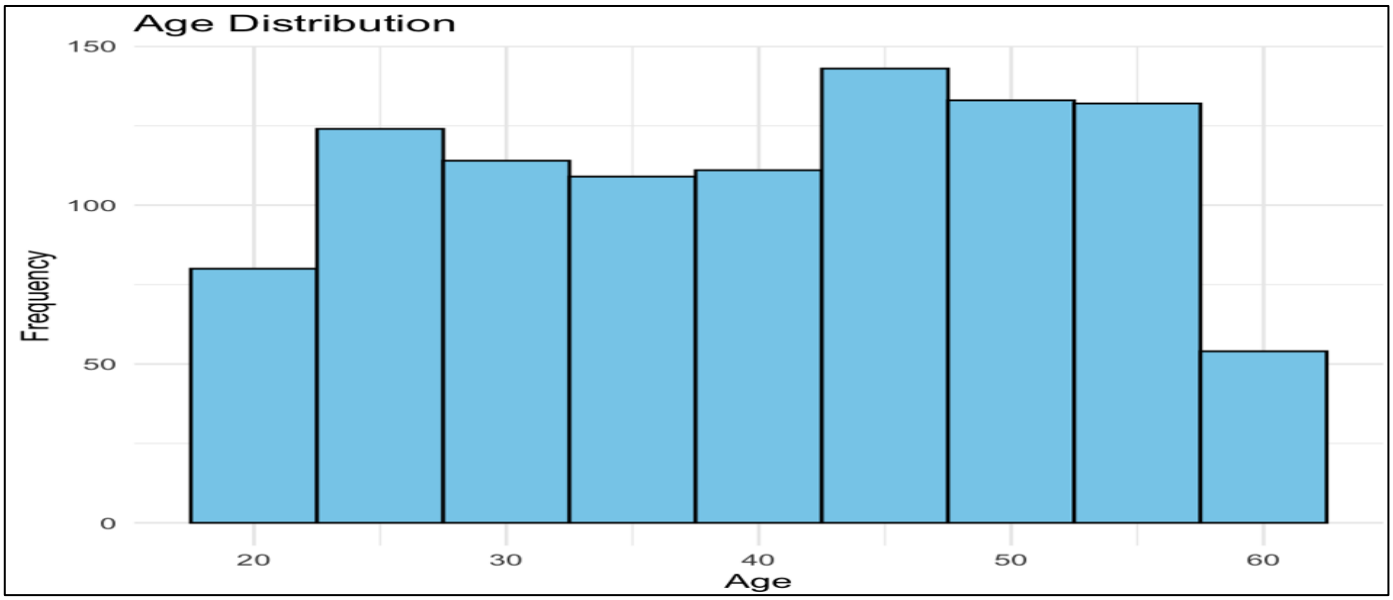

Fig 1 Histogram of Employee

✓ *Age Distribution*

Figure1 visually confirms that the age distribution of employees is approximately symmetric, without extreme skewness. There is a noticeable concentration of employees within the 35-50 age range. This reinforces the earlier observation that the company's workforce is largely dominated by mid-career individuals. Very few employees are observed in the youngest age group (under 25) or the oldest age group (above 55). This suggests that the company employs fewer early-career individuals who might just be starting out, and also fewer near-retirement individuals. This age profile could impact talent obtaining and succession planning.

- *Gender Composition*

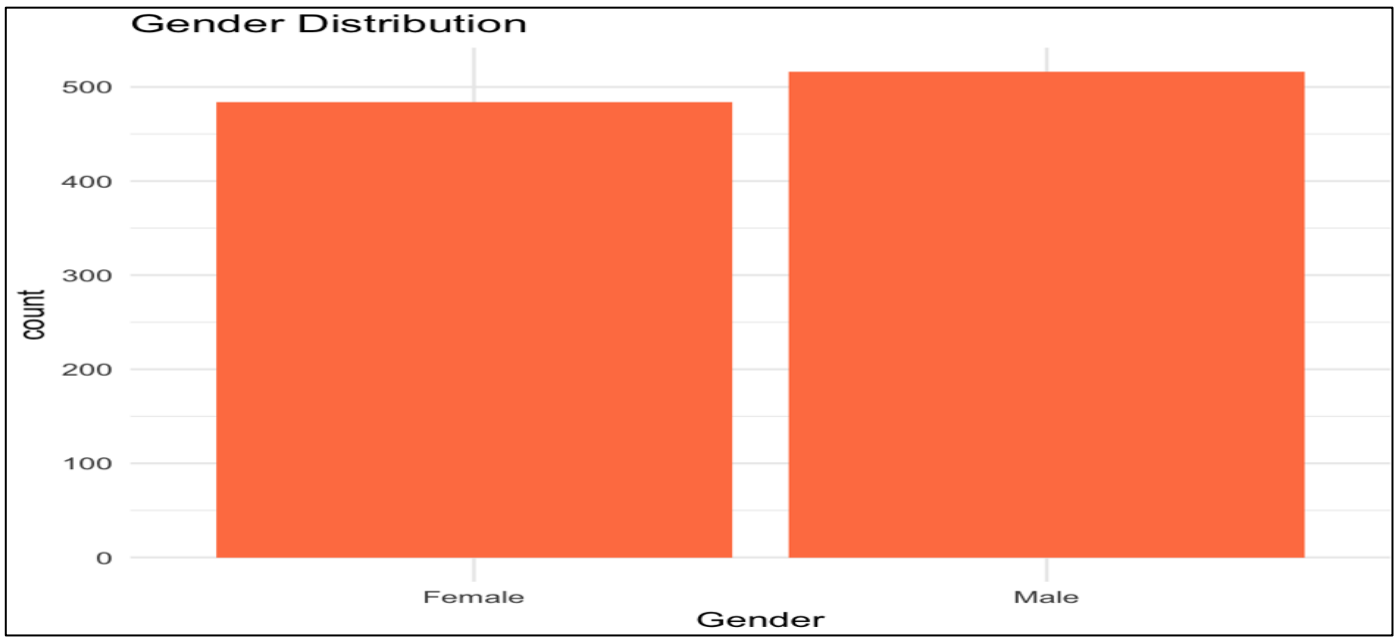

Fig 2 Bar Chart of Employees

✓ *Gender Composition*

Figure 2 clearly illustrates a higher proportion of male employees compared to female employees in the dataset. This indicates a gender imbalance within the workforce, with males forming the majority. Understanding this gender distribution is essential for any deeper demographic analysis, especially when examining attrition trends. For instance, future analysis might explore if attrition rates differ

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

significantly between genders or across departments with varying gender compositions.

- *Departmental Distribution*

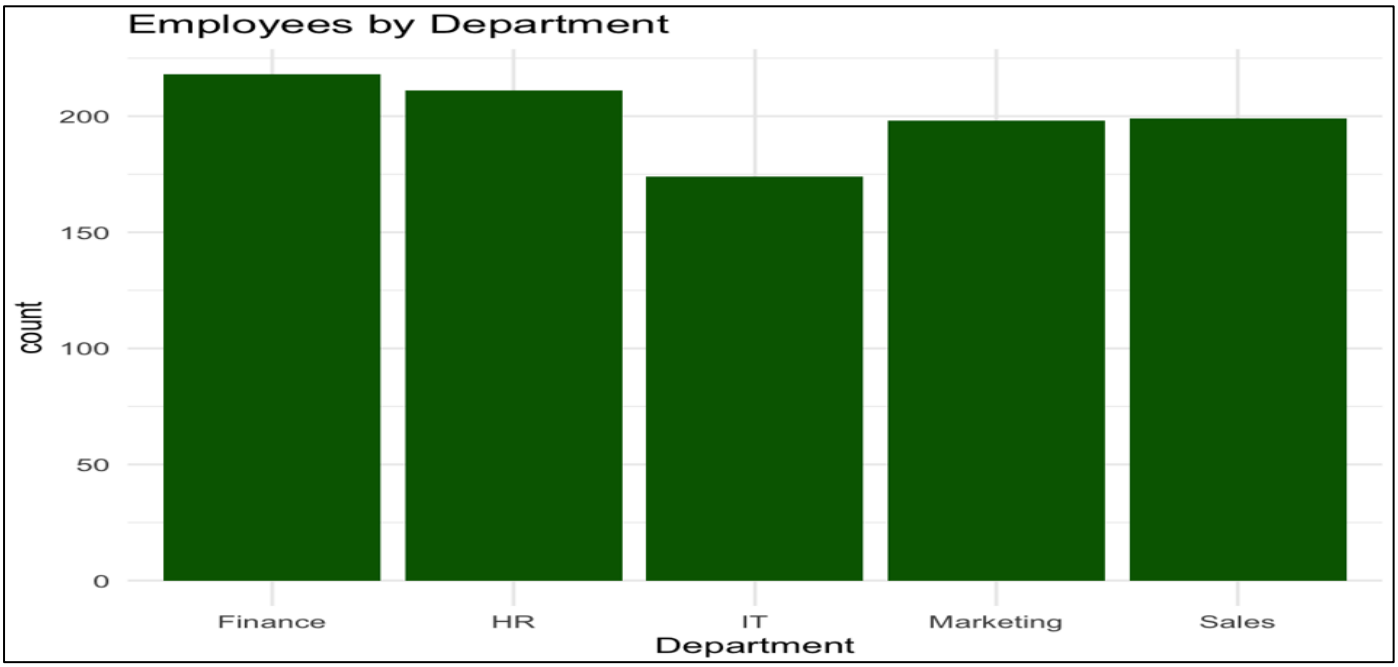

Fig 3 Bar Graph of Employees by Department

Bar graph in Figure 3 interprets the result that the financial department has the higher number of employees, suggesting it is coring operational area and possibly the largest functional unit within the organization. Conversely, the IT department has the smallest number of employees, indicating it might operate as a leaner, more specialized support function rather than a large operational department. This distribution highlights the organizational structure and resource allocation across different functions.

- *Monthly Income vs. Attrition*

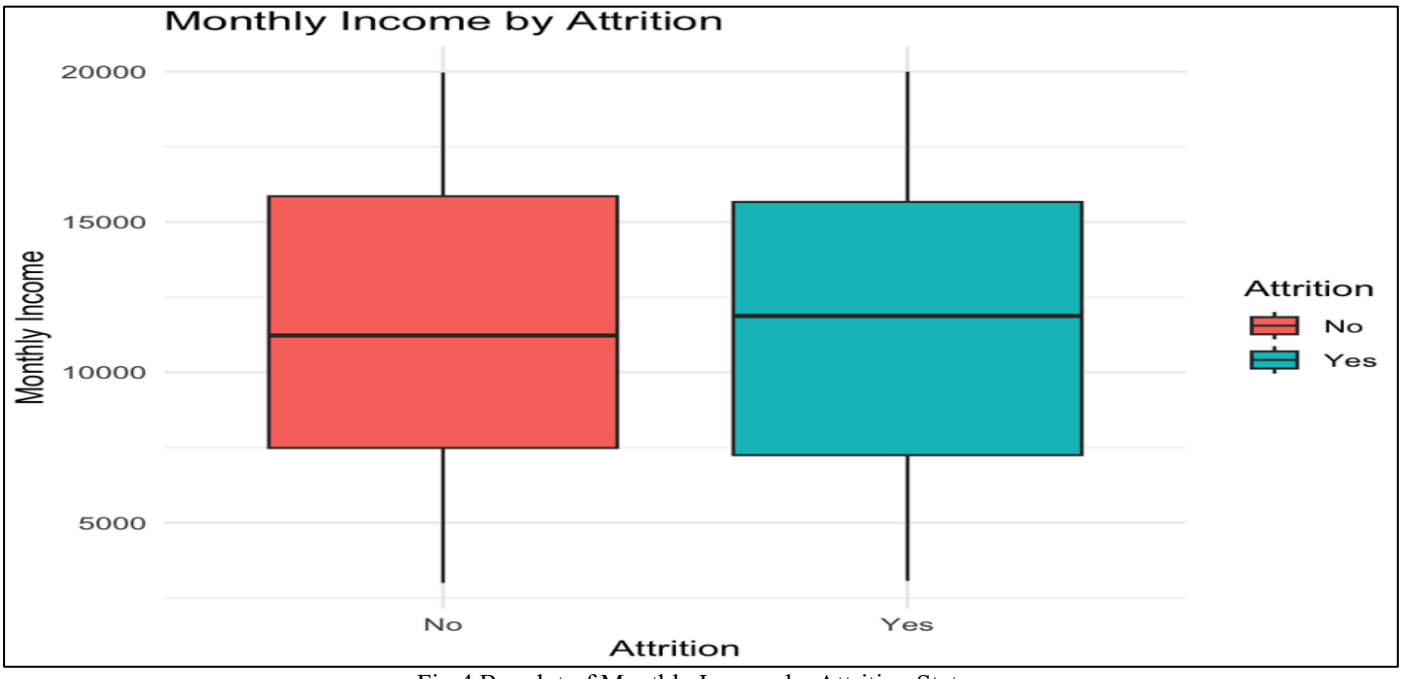

Fig 4 Boxplot of Monthly Income by Attrition Status

Figure 4 clearly reveals a distinct pattern emerges from the boxplot, employees who left the organization generally exhibit lower monthly incomes compared to those who stayed. The median monthly income for employees who did not attrite is visibly higher than that for employees who did. This suggests a potential association between lower compensation and higher attrition rates. This finding strongly indicated that dissatisfaction with compensation may be a

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

significant contributing factor to employee turnover. Further analysis could explore income threshold below which risk substantially increases.

- *Attrition Across Job Roles*

Fig 5 Bar Chart of Attrition Across Job Roles

The data in figure 5 shows the Executive and Manager job roles show a relatively higher absolute count of attrition cases compared to other roles, indicating a potential higher turnover in leadership and supervisory positions. Despite the Analyst role having the highest overall number of employees, it appears to have a lower relative attrition rate compared to its size. This suggests that the Analyst role might be more stable or that factors affecting attrition are less noticeable here.

The Assistant role also displays moderate attrition, with slightly fewer employees leaving compared to Executives and Managers. This trend might suggest that executive and managerial positions experience more pressure, higher demands, or increased workload, which could contribute to higher turnover in these critical roles. Further investigation into the specific challenges of these roles is warranted.

- *Overtime Work and Attrition*

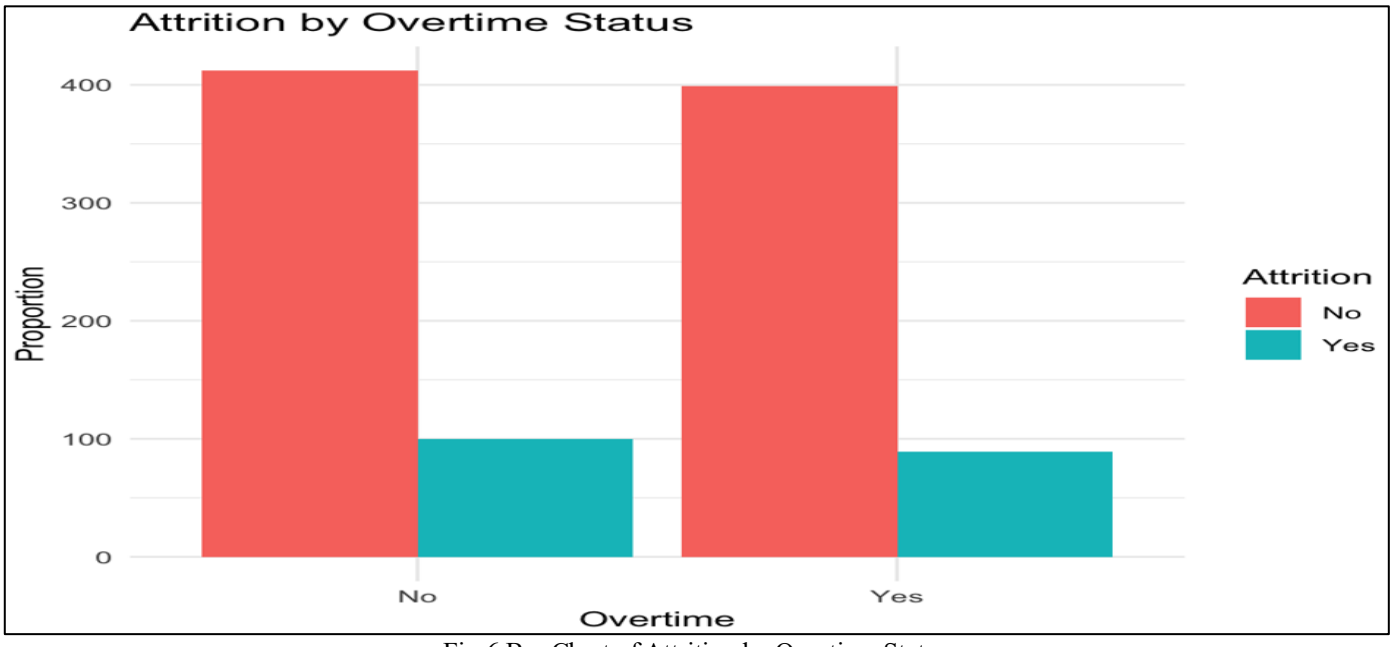

Fig 6 Bar Chart of Attrition by Overtime Status

From Figure 6, it can be observed that employees who regularly work overtime show a markedly higher proportion of attrition compared to those who do not work overtime. Even if the total number of employees in both overtime and non-overtime groups were similar, the attrition rate is noticeably higher among those who engage in overtime.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

This finding strongly suggests a possible negative impact of sustained overtime on employee retention. This could be due to factors such as increased burnout, compromised work-life balance, or general job dissatisfaction stemming from excessive hours. Addressing overtime policies could be a crucial strategy for retention.

- *Employee Tenure and Attrition*

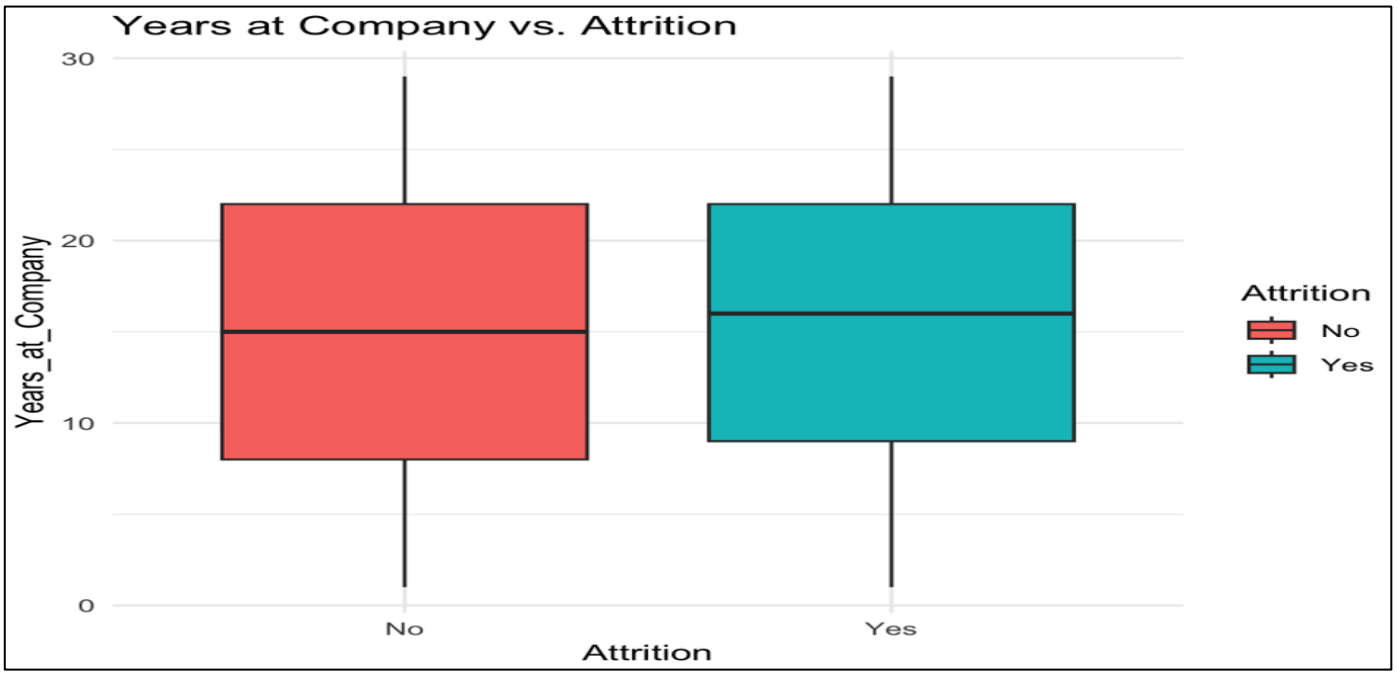

Fig 7 Boxplot of Years at Company by Attrition Status

Analysis of Figure 7 confirms employees who left the organization tend to have slightly fewer years of service compared to those who stayed. The median year at the company is somewhat lower for employees who attired, implying that newer employees or those with relatively less tenure might be more liable to leaving. However, the overall spread of tenure is fairly wide in both groups, and there is significant overlap between the two distributions. This indicates that while tenure might play a role, it is likely not a independent strong predictor of attrition. This suggests that tenure might become more influential when combined with other factors, such as job satisfaction.

- *Job Satisfaction and Attrition*

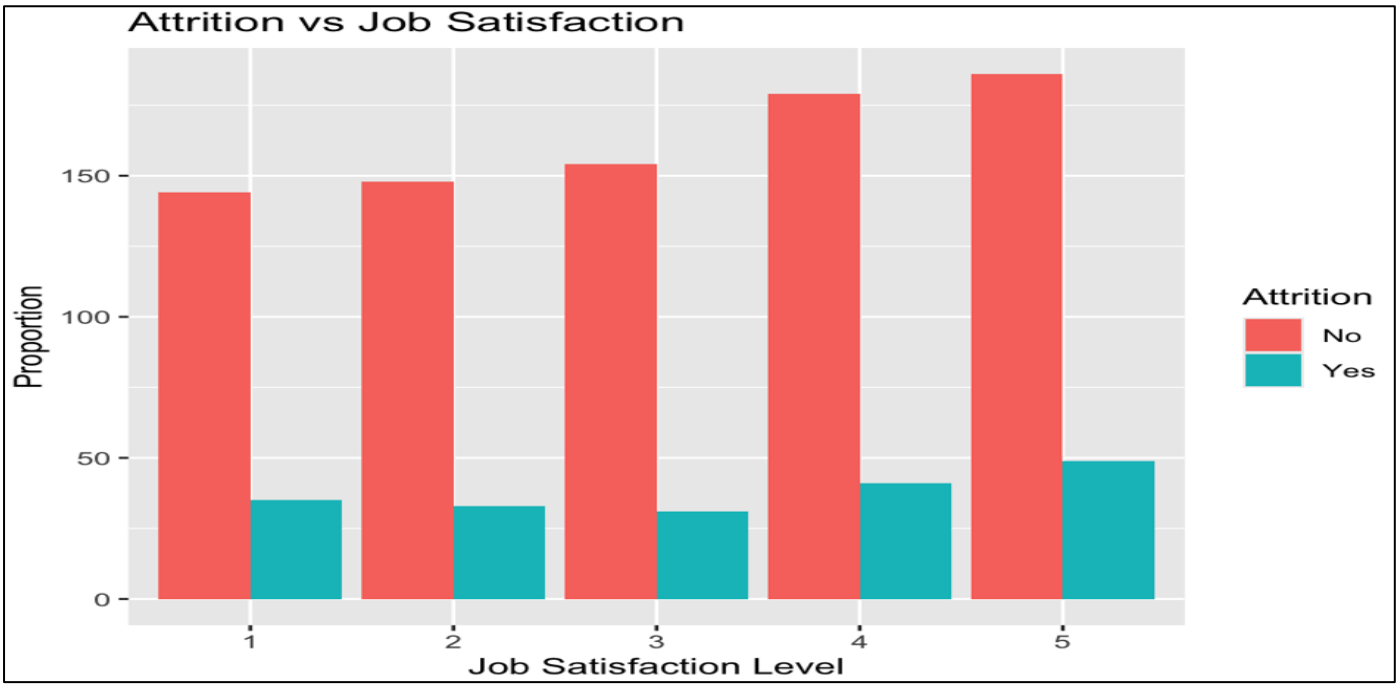

Fig 8 Bar Chart of Attrition by Job Satisfaction Level

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25nov1298

Surprisingly, attrition is not exclusively limited to low job satisfaction levels. Evan employees who reported higher satisfaction levels (4 and 5) show notable instances of attrition. The absolute count of attrition cases remains relatively stable across satisfaction level 1 to 3 but appears to increase slightly at levels 4 and 5. This counter-illogical finding warrants further investigation. This suggests that job satisfaction alone may not be a strong independent predictor of attrition in this organization. Other underlying factors, even for highly satisfied employees, could be driving their departure. This finding underscores the complex nature of attrition.

IV. CONCERNS AND LIMITATIONS

As the dataset is from Kaggle, its specific real-world context (e.g., industry, company size, exact time frame of data collection) is not provided. This limits the generalizability of findings to other organizations without further validation.

While comprehensive, the dataset may not include all possible variables that could influence attrition (e.g., direct manager relationships, company culture, personal life events).

This analysis primarily identifies correlations between variables and attrition. It does not establish definitive causal relationships. For example, while lower income correlated with attrition, it does not definitively mean lower income causes attrition in every case, other puzzling factors could be at play.

V. CONCLUSIONS

The organization primarily comprises a mid-career workforce (median age 41), with substantial proportion of long-tenured employees (average of nearly 15 years at company). However, there is a notable gender imbalance, with more male employees. The Finance department is the largest, while IT is the smallest. The overall attrition rate is relatively low at 18.9%, indicating that the majority of employees remain with the organization. However, the significant class imbalance in the attrition variable (81.1% 'No' vs. 18.9% 'Yes') highlights a data challenge that needs to be carefully managed in any subsequent predictive modelling efforts to avoid biased results.

Lower monthly income appears to be a strong indicator of higher attrition. Employees who left the organization generally had lower earnings, suggesting that dissatisfaction with compensation is a significant contributing factor turnover. Executive and Managerial roles show a disproportionately higher number of attrition cases. This suggest that these positions may be associated with increased pressure, workload, or expectations that contribute to higher turnover. Working overtime is strongly correlated with higher attrition rates. This indicates that excessive working hours likely lead to burnout, poor work-life balance, and ultimately, a higher tendency for employees to leave. While less impactful as a independent factor, employees with slightly shorter tenures tend to attrite more. This suggests that the initial years at the company might be a critical period for retention.

Unexpectedly, job satisfaction alone does not appear to be a conclusively independent predictor of attrition. Employees across all satisfaction levels, including those reporting high satisfaction, have shown instances of attrition. This implies that factors beyond simple job satisfaction, such as compensation, workload, external opportunities, or personal circumstances, play a significant role in an employee's decision to leave, even if they are generally content with their role.

In answering the research question, this analysis identifies compensation, job role, and overtime as primary divers of employee attrition within this dataset. While employee age and overall tenure provide relating to context information, their direct correlation with attrition is less noticeable than income or overtime. The unexpected finding regarding job satisfaction underscores the complex and multi-faceted nature of employee retention, suggesting that organizations must consider a comprehensive view of employee well-being and motivations beyond just job satisfaction. These findings provide actionable insights for the organization to develop targeted strategies for improving employee retention and promoting more stable workforce.

REFERENCES

[1]. Andrew T. Jebb., Scott Parrigon., & Sang Eun Woo. (2016). *Exploratory Data Analysis as a Foundation of Inductive Research*. Human Resource Management Review. 27(2). https://doi.org/10.1016/j.hrmr.2016.08.003

[2]. Ayesha Banu Mohd., Sharmila Reddy., & Rama M A. (2022). *Exploratory Data Analysis (GEDA): A Case Study on Employee Attrition*. 10.46243/jst. 2022.v7. i09.pp01-11

[3]. Fatbardha Maloku., Besnik Maloku. (2024). *Analysing IBM HR Data: Employee Attrition and Performance Insights*. Journal of Engineering and Applied Sciences Technology. 6(8): 1-10:1-10. https://doi.org/10.47363/JEAST/2024(6)268

[4]. Hardik I., Dharmendra Patel. (2024). *Exploratory Data Analysis and Feature Selection for Predictive Modelling of Student Academic Performance Using a Proposed Dataset*. International Journal of Engineering Trends and Technology. 72(11):131-143. https://doi.org/10.14445/22315381/IJETT-V72I11P116

[5]. Mohammed Salmanuddin., Rushikesh Kulkarani., Atharva Mohite., & Mahendra Patil. (2023). *Exploratory Data Analysis*. ICSTEM. https://doi.org/10.35629/5252-050413881392

[6]. Simon Gim., & Eun Tack Im. (2023). *A Study on Predicting Employee Attrition Using Machine Learning*. Studies in Computational Intelligence. https://doi.org/10.1007/978-3-031-19608-9_5

[7]. Kaggle Dataset: https://www.kaggle.com/datasets/ziya07/employee-attrition-prediction-dataset