Predicting Employee Attrition in HR Analytics

Eliana Alon | Lauren Furman | Joseph Aytch | Michelle Valencia

DSC 450 Applied Data Science

Bellevue University- Fall 2024

Table of Contents

[Introduction 2](#_Toc182131330)

[Business Problem/Hypothesis 3](#_Toc182131331)

[Methods/Analysis 3](#_Toc182131332)

[Results 11](#_Toc182131333)

[Recommendations/Ethical Considerations 12](#_Toc182131334)

[Conclusion 13](#_Toc182131335)

[References 13](#_Toc182131336)

[Appendix 13](#_Toc182131337)

# Introduction

This project uses the [HR Analytics Dataset](https://www.kaggle.com/datasets/rishikeshkonapure/hr-analytics-prediction) from Kaggle, which provides comprehensive information about

employees within a company, covering key aspects such as age, department, job role, and marital status. The

dataset includes the target variable "Attrition" indicating whether an employee has left the company. This

particular dataset is invaluable for analyzing factors influencing employee turnover and developing effective

retention strategies.

# Business Problem/Hypothesis

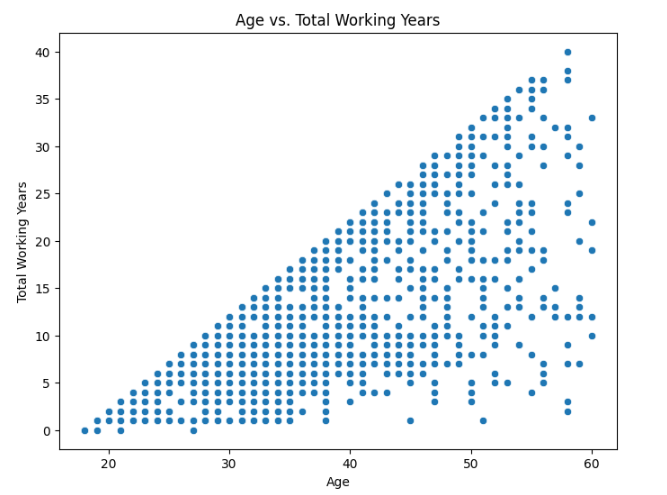
The business problem is to understand the factors that lead to employee attrition. The **hypothesis** is that certain

demographics and job characteristics, such as age, job level, and tenure with the current manager, are significant

predictors of whether an employee will leave the company.

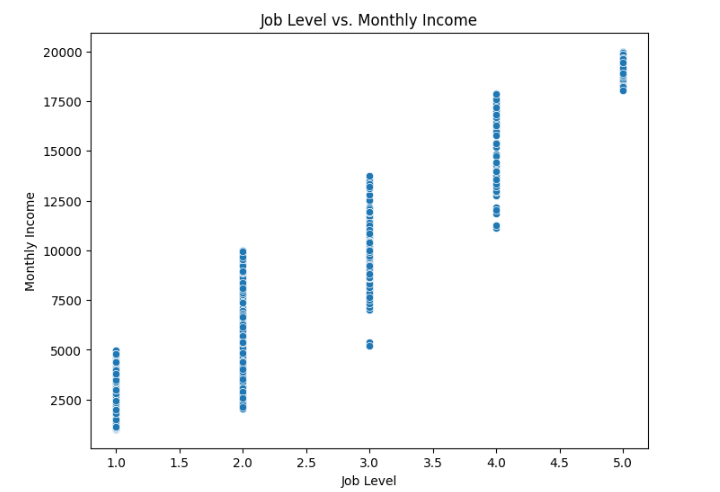
# Methods/Analysis

Figure 1:



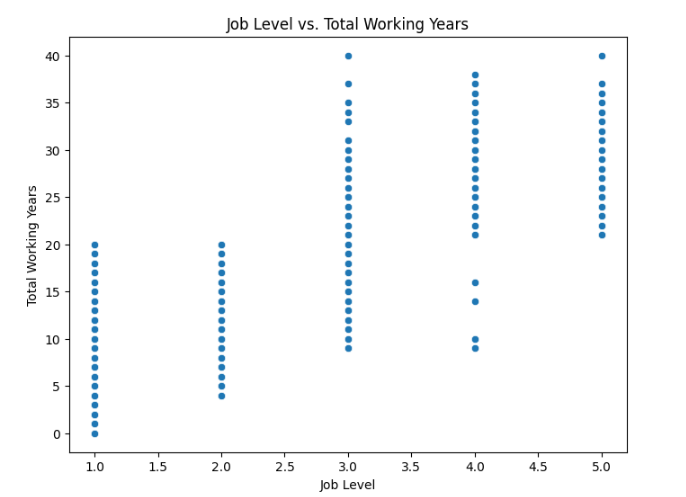
The scatterplot generally shows an upward trend, indicating a positive correlation between age and total working years. This suggests that as employees get older, they tend to have accumulated more years of work experience. This is a logical and expected relationship in the workforce. While there's a clear positive trend, there's also variability in the data. This means that for a given age, there's a range of total working years. This variability can be attributed to factors such as career breaks, career changes, and individual career paths. Many factors play into how many years a person ends up working over their lifetime, many of which are not up to the company.

Figure 2:



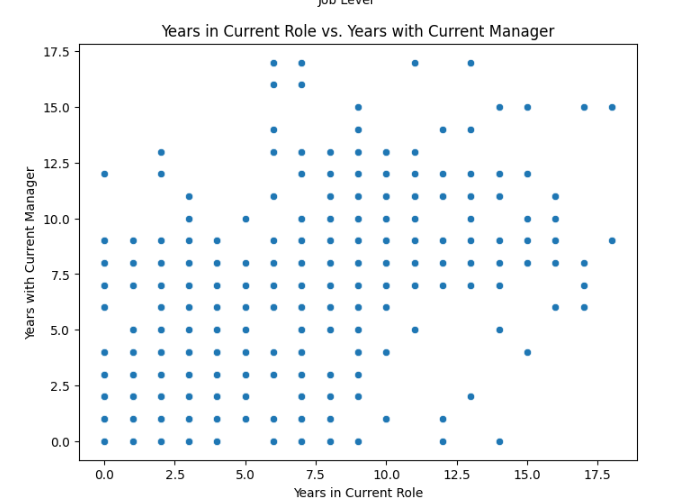
The scatterplot reveals a clear positive correlation between job level and monthly income. As job level increases, monthly income tends to increase as well. This is expected, as higher job levels typically come with greater responsibilities and require more experience, leading to higher compensation. While there's clustering, you'll also observe some variation in income within each job level. This variation can be attributed to factors such as experience, performance, and negotiation skills. Some employees within the same level might have higher incomes due to their longer tenure, better performance ratings, or successful salary negotiations.

Figure 3:



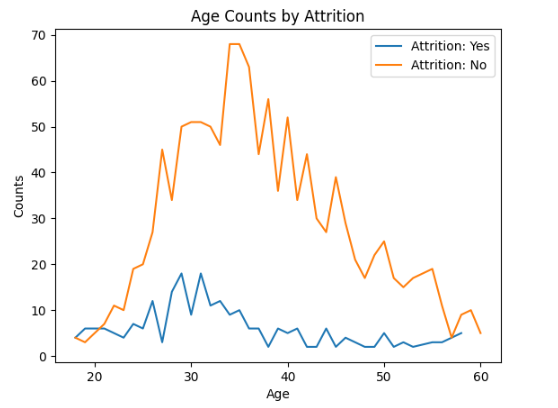
The scatterplot generally demonstrates a positive correlation between job level and total working years. This suggests that as employees gain more work experience, they tend to progress to higher job levels within the company. This aligns with the expectation that career advancement often involves accumulating years of experience and developing skills.

Figure 4:



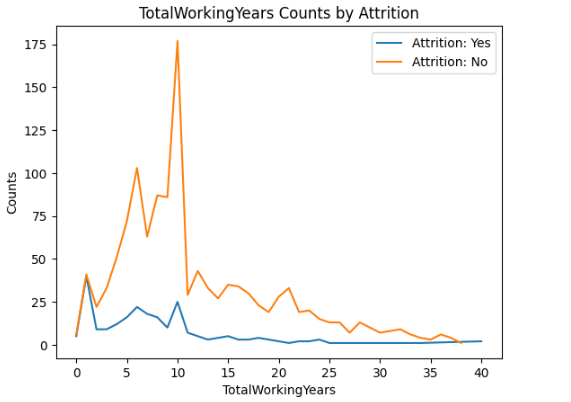
This scatterplot reveals a slight positive correlation between years in current role and years with current manager. This indicates that employees who have been in their current role for a longer period tend to have also been with their current manager for a longer time. This relationship is logical, as employees often stay in roles where they have a good working relationship with their manager. This may also imply that teams that stick with similar management and team members may positively contribute to employee retention.

Figure 5:



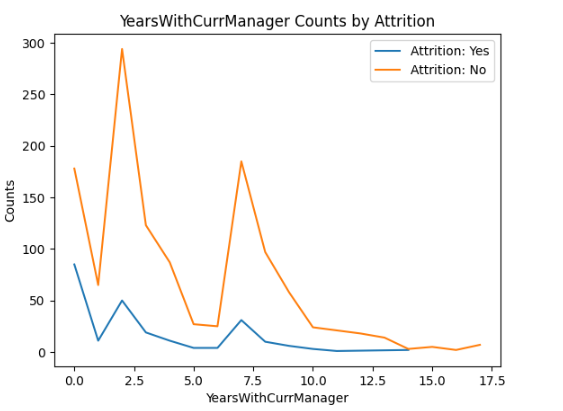
The graph shows a peak in attrition around the age of 30-35 for employees who left the company (Attrition = "Yes"). This suggests that employees in this age range are more likely to leave the company compared to other age groups. As employee age increases beyond 35, the attrition rate appears to decrease for both groups (Attrition = "Yes" and "No"). This could indicate that employees become more stable in their roles and less likely to leave as they gain more experience and seniority.

Figure 6:



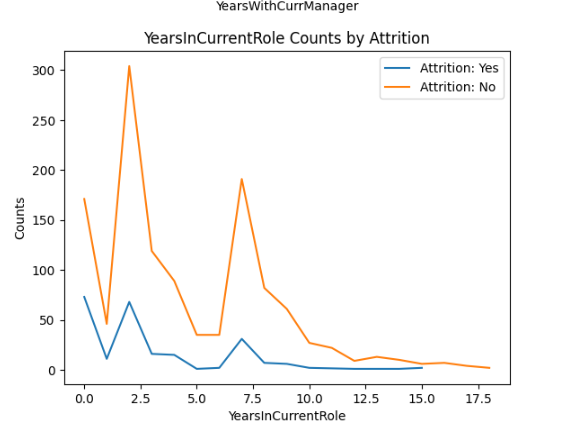
The graph shows a peak in attrition for employees with 0-2 years of total working experience. This suggests that individuals who are relatively new to the workforce are more likely to leave the company compared to those with more experience. The graph also shows a peak in retention (Attrition = "No") around 8-10 years of total working experience. Employees with more experience tend to be more stable in their roles and less likely to leave. While this graph may imply internal factors, external factors, such as economic conditions or industry trends, can also influence attrition rates at different experience levels.

Figure 7:



This graph shows that there is a significantly high attrition rate for employees in their first year, which, like the graph in figure 6, may be attributed to external factors. However, there is a slight increase in attrition around the 2.5-year mark with the current manager. This could indicate a potential point where employees may reassess their career trajectory or seek new opportunities, even if they have a positive relationship with their manager. There is also another significant spike at the 7-year mark, which may imply a similar scenario. Overall, as the number of years with the current manager increases, the attrition rate generally decreases for both groups (Attrition = "Yes" and "No").

Figure 8:



Analyzing the data, we notice key decision points at around 2.5 and 7.5 years, where employees often choose whether to stay or move on. This pattern is clear when looking at tenure in their current role and time with their manager. Even long-term employees tend to reassess their career every five years, unless they are exceptionally satisfied, as shown by another drop at the 10-year mark. The average loyal employee's tenure is about 10 years, while the peak age in the dataset is between 30 and 35, indicating that employees in this age range value stability and are more likely to stay in their current positions.

# Results

**Distributions of Features**: Age and Education are unimodal and symmetric, Daily Rate and Hourly Rate are

symmetric, Years in Current Role is bimodal, and features like Distance from Home, Monthly Income, and Years

Since Last Promotion show a right-skewed distribution.

**Categorical Insights**: Most employees are in their current roles, rarely travel, work in R&D, are male, and married.

**Correlation Analysis**: High correlations observed between Total Working Years and Job Level, Monthly Income,

Years at Company, etc., indicating potential multicollinearity.

Based on our analysis, it seems that our hypothesis regarding employee turnover prediction holds true to a certain

extent. We identified key factors associated with employee attrition through visualizations and exploratory data

analysis techniques, including age, job level, total working years, and tenure in the current role. The observed

correlations and patterns suggest that specific demographics and employment characteristics are linked to the

likelihood of employees leaving the company. Nevertheless, further validation via predictive modeling and

evaluation of model performance metrics is required to fully confirm the hypothesis and gauge the predictive

power of these identified factors.

# Recommendations/Ethical Considerations

Based on the findings, recommended actions include focusing on identified predictors to improve employee

retention strategies. Ethical considerations include ensuring data privacy and avoiding any potential biases or

discriminatory practices in the predictive models. Privacy and Confidentiality: All data collected in the survey was anonymized and stored securely to ensure confidentiality. Participants were made aware that their responses would be used only for academic analysis.

**Informed Consent:** Prior to participation, all respondents gave informed consent and were fully aware of the scope and purpose of the research.

# Conclusion

The analysis confirms the hypothesis that certain job characteristics and demographics are linked to employee

attrition. Further validation through predictive modeling is required to fully confirm these findings and ensure the

model's effectiveness and ethical application.

# References

**Dataset Source**: Include the citation for the HR Analytics Dataset from Kaggle.

* Kaggle. (n.d.). HR Analytics Dataset. Retrieved from [https://www.kaggle.com/code/paramarthasengupta/hr-analytics-prediction-why-do-people-resign/notebook](https://www.kaggle.com/code/paramarthasengupta/hr-analytics-prediction-why-do-people-resign/notebook?form=MG0AV3)

**Software Documentation**:

* Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
* Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

# Appendix



