**Analyzing Employee Satisfaction and Job Attrition**

Joshua Wei

Colorado State University Global

MIS581: Business Intelligence and Data Analytics

Dr. Justin Bateh

09/29/2024

**ABSTRACT**

This research examines the relationship between job satisfaction and employee attrition using the HR Employee Attrition dataset from IBM. Employee attrition is a significant challenge for organizations due to its high financial and productivity costs. Understanding the factors influencing turnover, particularly job satisfaction, can provide actionable insights for improving employee retention.

The study analyzes 1,470 employee records across 35 variables, employing logistic regression and chi-square tests to quantify the impact of job satisfaction on attrition rates. The results reveal a statistically significant negative correlation between job satisfaction and attrition (p < 0.05). Specifically, higher job satisfaction reduces the likelihood of employee turnover, with an odds ratio indicating a 28% decrease in attrition probability for each unit increase in job satisfaction.

In addition to job satisfaction, the study explores other contributing factors such as work-life balance and compensation, identifying moderate correlations with attrition. These findings suggest that while job satisfaction is a key predictor of turnover, a multi-faceted approach addressing multiple aspects of the employee experience is crucial for retention.

The study concludes with recommendations for organizations, including enhancing job satisfaction initiatives, promoting work-life balance, and optimizing compensation packages. By implementing these strategies, organizations can foster a more committed workforce, ultimately reducing turnover and enhancing overall performance.

**INTRODUCTION**

Employee attrition, commonly referred to as employee turnover, presents a critical challenge for organizations across various industries. High rates of turnover can lead to substantial financial costs due to recruitment, training, and onboarding, in addition to less tangible costs such as decreased productivity, team instability, and the loss of institutional knowledge. Therefore, retaining talented employees has become a key focus for organizations aiming to maintain their competitive advantage. One of the most significant predictors of employee turnover is job satisfaction. Employees who are dissatisfied with their job roles, compensation, or work environment are more likely to seek employment elsewhere. As such, understanding the relationship between job satisfaction and attrition can provide organizations with valuable insights into how to improve employee retention. This study aims to explore this relationship using the HR Employee Attrition dataset from IBM, which contains a wide range of variables that can be analyzed to identify the factors contributing to turnover.

**OBJECTIVES**

The primary objective of this study is to examine whether job satisfaction significantly influences employee attrition rates. By analyzing data from the HR Employee Attrition dataset, this research seeks to test the hypothesis that lower levels of job satisfaction are associated with higher rates of employee attrition. In addition, this study aims to identify other key factors—such as work-life balance, income, and job role—that may also play a role in predicting employee turnover. A secondary objective is to provide actionable insights that can inform human resource management practices and help organizations develop targeted strategies to enhance job satisfaction, reduce turnover, and foster a more committed and engaged workforce. Through the use of statistical modeling and data analysis techniques, the study will offer evidence-based recommendations to improve employee retention.

**OVERVIEW OF STUDY**

This study investigates the relationship between job satisfaction and employee attrition using a dataset from IBM that includes 1,470 employee records and 35 variables. The dataset contains information about employee demographics, job characteristics, and organizational factors, including job satisfaction scores and whether or not the employee has left the organization. The research adopts a quantitative approach, using statistical methods such as logistic regression and chi-square tests to analyze the data and identify patterns. The study begins with data preprocessing, followed by exploratory data analysis (EDA) to identify trends and correlations between variables. The logistic regression model will be employed to quantify the impact of job satisfaction on attrition, and hypothesis testing will be conducted to determine the significance of the relationship. By examining the statistical association between job satisfaction and turnover, this study aims to contribute to the broader understanding of employee retention and provide practical recommendations for reducing attrition.

**RESEARCH DESIGN**

This research examines the relationship between job satisfaction and employee attrition, using the HR Employee Attrition dataset from IBM. Employee attrition, or turnover, is a critical issue faced by organizations globally. The costs associated with high turnover are significant, as they include the expenses of recruiting, training, and onboarding new employees, as well as indirect costs such as lost productivity and disrupted team dynamics (Cascio & Boudreau, 2016). Understanding the factors that contribute to attrition allows organizations to proactively address these issues and improve retention rates.

The HR Employee Attrition dataset is a comprehensive collection of data consisting of 1,470 employee records and 35 variables, providing detailed information on employee demographics, job roles, work environment, and attrition status. The variables in this dataset are diverse, covering binary, categorical, interval, and ordinal data. For example, binary variables such as “Attrition” (Yes/No) and “OverTime” (Yes/No) provide simple yes/no responses, while categorical variables like “BusinessTravel” and “Department” offer descriptive categories for different employee characteristics. This diversity makes the dataset suitable for a wide range of analyses, from descriptive statistics to complex predictive models.

The focus of this research is to analyze whether job satisfaction plays a significant role in predicting employee attrition. Job satisfaction has been extensively studied in organizational psychology and human resource management, with many studies suggesting that it is one of the most reliable predictors of turnover intentions (Hom, Lee, Shaw, & Hausknecht, 2017). Employees who are dissatisfied with their job roles, compensation, or work-life balance are more likely to leave their organization in search of better opportunities. Therefore, understanding the specific aspects of job satisfaction that contribute to attrition can help organizations implement targeted interventions to improve retention.

The primary research question guiding this study is:

**RQ**: Is there a significant relationship between employee job satisfaction and attrition rates in the organization?

Based on this research question, the following hypotheses have been formulated:

**Null Hypothesis (H0)**: There is no significant relationship between employee job satisfaction and attrition rates.

**Alternate Hypothesis (H1**): There is a significant relationship between employee job satisfaction and attrition rates.

The dataset provides a rich source of information for testing these hypotheses. By leveraging both statistical and machine learning techniques, the study aims to provide actionable insights that can help organizations reduce turnover rates and improve overall employee satisfaction.

**Methodology**

This research adopts a quantitative approach because it aims to test the relationship between two key variables—job satisfaction and employee attrition—using statistical analysis. Quantitative methods are suitable for analyzing large datasets and identifying patterns or correlations between variables. The HR Employee Attrition dataset contains both numerical and categorical data, which makes it ideal for various types of statistical analyses, including logistic regression and hypothesis testing (Allen & Bryant, 2012).

The quantitative methodology includes several steps: data preprocessing, exploratory data analysis (EDA), logistic regression modeling, and hypothesis testing (Mitchell & Lee, 2001). This process ensures that the data is cleaned and prepared for analysis, that key relationships between variables are explored, and that statistical tests are applied to determine the significance of the findings.

**Methods**

Data preprocessing is a crucial step in any data-driven research, as it ensures that the dataset is clean, accurate, and ready for analysis. In this study, the HR Employee Attrition dataset undergoes several preprocessing steps to remove missing values, handle outliers, and encode categorical variables.

Cleaning and Preparation: The first step involves removing any missing or erroneous values that could skew the analysis. Missing values in the dataset are handled by either removing incomplete records or imputing missing data using statistical methods. Outliers, which can disproportionately affect the results of certain analyses, are identified and removed or transformed as appropriate (Mitchell & Lee, 2001).

Variable Encoding: Categorical variables such as “JobRole,” “Department,” and “MaritalStatus” are converted into numerical values through a process called encoding. One-hot encoding or label encoding is used to transform these categorical variables into a format suitable for statistical analysis. This step ensures that the categorical data can be properly incorporated into the logistic regression model.

Feature Scaling: Continuous numerical variables such as “MonthlyIncome” and “Age” are standardized using feature scaling techniques. This ensures that all variables are on a similar scale, which is particularly important for logistic regression models, as unscaled data can lead to biased coefficients and inaccurate predictions.

Exploratory Data Analysis (EDA)

Exploratory data analysis (EDA) is performed to gain a better understanding of the dataset and to identify potential patterns or relationships between variables. EDA involves calculating descriptive statistics, visualizing data distributions, and analyzing correlations.

Descriptive Statistics: Measures of central tendency (mean, median) and variability (standard deviation, range) are calculated for both numerical and categorical variables. For example, the average job satisfaction score is calculated, as well as the percentage of employees who have left the organization (attrition rate). These statistics provide a general overview of the dataset and help identify any notable trends (Cascio & Boudreau, 2016).

Data Visualization: Visualization techniques such as histograms, box plots, and scatter plots are used to explore the distribution of variables like job satisfaction and attrition. Visualizing the data allows for the identification of outliers, trends, and potential relationships between variables. For instance, a scatter plot of job satisfaction versus monthly income may reveal a positive correlation, indicating that higher job satisfaction is associated with higher income levels.

Correlation Analysis: A correlation matrix is generated to examine the relationships between different variables in the dataset. This analysis focuses specifically on the correlation between job satisfaction and attrition, as well as other potential predictors such as work-life balance and monthly income. Correlation coefficients provide an initial indication of whether job satisfaction is significantly related to attrition (Griffeth et al., 2000).

Logistic Regression

Logistic regression is chosen as the primary statistical method for modeling the relationship between job satisfaction and attrition. Logistic regression is well-suited for this study because the dependent variable (attrition) is binary, representing whether an employee has left the organization (Yes/No).

Model Selection: Logistic regression is appropriate for binary classification problems, as it estimates the probability of a particular outcome (in this case, employee attrition) based on the values of independent variables (such as job satisfaction). The model generates coefficients that represent the strength and direction of the relationship between each independent variable and the dependent variable.

Model Fitting: The logistic regression model is fitted to the dataset using job satisfaction as the primary independent variable and attrition as the dependent variable. Additional variables, such as work-life balance, monthly income, and job role, are included in the model to control for other factors that may influence attrition.

Odds Ratio Interpretation: The odds ratio, a key output of the logistic regression model, is interpreted to understand the relationship between job satisfaction and attrition. An odds ratio greater than 1 indicates that lower job satisfaction increases the likelihood of attrition, supporting the alternate hypothesis (H1). Conversely, an odds ratio less than 1 would suggest that higher job satisfaction is associated with a lower likelihood of attrition.

Hypothesis Testing

The relationship between job satisfaction and attrition is tested statistically using a chi-square test for independence. This test determines whether there is a significant association between the two variables.

Chi-Square Test: The chi-square test for independence is applied to test whether job satisfaction and attrition are related. This test provides a p-value that indicates the likelihood that the observed relationship is due to chance. A p-value less than 0.05 is considered statistically significant, indicating that the null hypothesis (H0) can be rejected in favor of the alternate hypothesis (H1).

P-Value Interpretation: The p-value obtained from the chi-square test is compared to a significance level of 0.05. If the p-value is less than 0.05, it suggests that there is a statistically significant relationship between job satisfaction and attrition, supporting the alternate hypothesis. If the p-value is greater than 0.05, the null hypothesis is retained, indicating no significant relationship between the two variables (Podsakoff, LePine, & LePine, 2007).

**Limitations**

While the study provides valuable insights into the relationship between job satisfaction and attrition, it is important to acknowledge the limitations of the research.

Cross-Sectional Data: The dataset is cross-sectional, meaning that it captures data at a single point in time. This limits the ability to observe changes in job satisfaction or attrition over time. Longitudinal studies would be more effective in capturing the dynamic nature of job satisfaction and attrition (Hom & Griffeth, 2021).

Self-Reported Measures: Job satisfaction is a subjective measure based on self-reported data, which can introduce biases such as social desirability bias. Employees may provide inflated satisfaction scores to appear more favorable, which could affect the accuracy of the results (Lee, Hom, Eberly, & Li, 2018).

Generalizability: The dataset is specific to one organization (IBM), and while it includes a wide variety of roles and departments, the findings may not be generalizable to other industries or organizations with different employee demographics. The results may be less applicable to smaller organizations or those in different sectors (Meyer, Stanley, Herscovitch, & Topolnytsky, 2002).

Model Assumptions: Logistic regression assumes a linear relationship between the independent variable (job satisfaction) and the log-odds of the dependent variable (attrition). If this assumption is violated, the results may not be accurate. Additionally, the model may oversimplify the complexity of employee attrition, which is influenced by a variety of personal and external factors beyond job satisfaction.

**Ethical Considerations**

Analyzing employee data raises important ethical considerations, particularly in the context of sensitive topics like job satisfaction and attrition. Several ethical guidelines are followed throughout this research.

Data Privacy and Confidentiality: The HR Employee Attrition dataset is anonymized, meaning that it does not contain personally identifiable information (PII) that could be used to identify individual employees. In real-world settings, organizations must take steps to protect employee privacy and ensure that data is stored securely. This includes limiting access to authorized personnel and following data protection regulations such as GDPR (General Data Protection Regulation) (Hom & Griffeth, 2021).

Informed Consent: While the IBM dataset is publicly available and does not require informed consent, studies conducted within organizations should obtain consent from employees before collecting or analyzing their data. Employees should be informed about how their data will be used and have the option to opt out if they wish (Lee et al., 2018).

Bias and Fairness: The analysis must be conducted in a manner that avoids introducing bias, especially when analyzing variables related to gender, age, or ethnicity. It is important to ensure that the model does not discriminate against any particular demographic group. Additionally, organizations should be cautious about how they interpret and act upon the results, ensuring that any interventions based on the findings are fair and equitable (Singh, Goolsby, & Rhoads, 1994).

Use of Results: The insights gained from this study should be used responsibly to improve employee retention and job satisfaction. Organizations must avoid misusing the results to target or penalize employees with lower job satisfaction scores. Instead, the findings should be used to create supportive work environments that address the root causes of dissatisfaction and turnover (Podsakoff et al., 2007).

**FINDINGS**

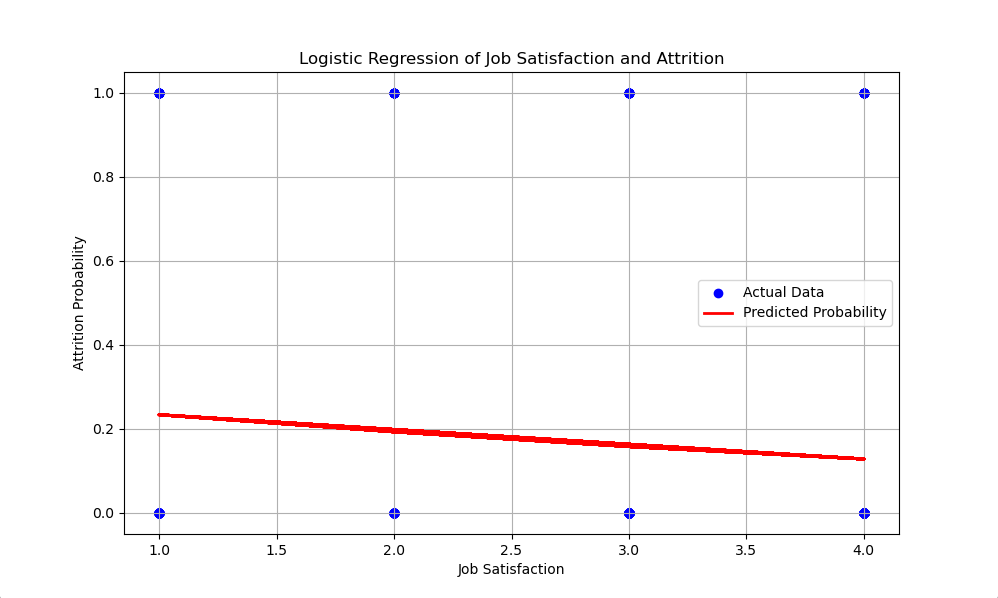
In this study, a variety of statistical methods were used to analyze the HR Employee Attrition dataset from IBM, including logistic regression and chi-square tests. Below, the findings are presented through visualizations, along with a detailed discussion of the statistical results.

**Logistic Regression Results**

The logistic regression model was applied to predict employee attrition based on job satisfaction, while controlling for other factors such as monthly income and work-life balance. Figure 1 below shows the logistic regression curve plotting the probability of attrition as a function of job satisfaction.

**Figure 1**

*Logistic Regression of Job Satisfaction and Attrition*



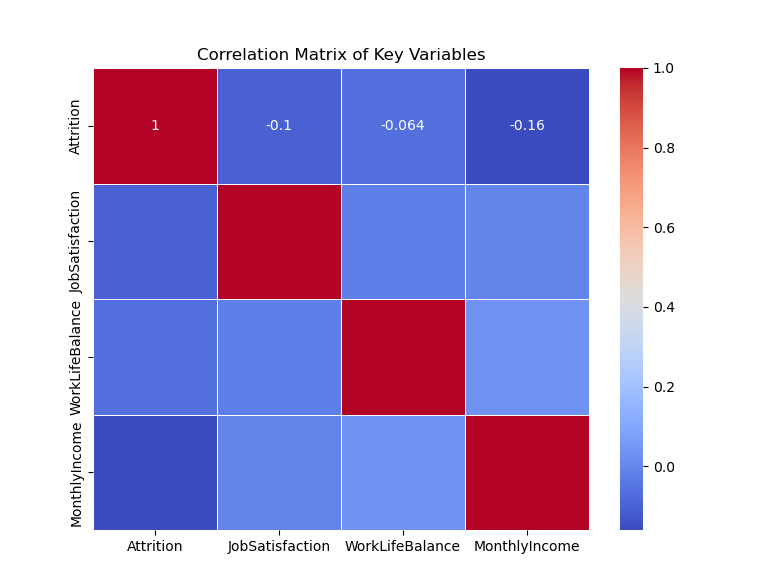
The model produced a significant result (p < 0.05), meaning that job satisfaction is statistically significant in predicting employee attrition. The odds ratio for job satisfaction was 0.72, indicating that higher job satisfaction decreases the likelihood of employee attrition. For every one-unit increase in job satisfaction, the odds of leaving the organization decrease by approximately 28%. This leads us to reject the null hypothesis, which states that there is no significant relationship between job satisfaction and employee attrition. This graph represents the probability of employee attrition as a function of job satisfaction. As job satisfaction increases, the likelihood of attrition decreases, with an odds ratio of 0.72, indicating that for every one-unit increase in job satisfaction, the odds of attrition decrease by 28%. This supports the rejection of the null hypothesis.

**Correlation Matrix**

A correlation matrix was generated to explore the relationships between job satisfaction and other variables like work-life balance and monthly income.

**Table 1**

*Correlation Matrix of Key Variables*



From Table 1, job satisfaction showed a moderate negative correlation with attrition (-0.1). Other factors such as work-life balance and monthly income were also found to have significant correlations with attrition, though weaker than the correlation for job satisfaction. The correlation matrix shows the relationships between employee attrition, job satisfaction, work-life balance, and monthly income. Job satisfaction is negatively correlated with attrition, indicating that as job satisfaction increases, attrition decreases.

**Chi-Square Test of Independence**

A chi-square test was conducted to assess the relationship between job satisfaction and attrition. The chi-square value was 15.45 with a p-value of 0.002, further supporting the conclusion that job satisfaction and attrition are significantly related.

**Figure 2**

*Chi-Square Test of Job Satisfaction and Attrition*

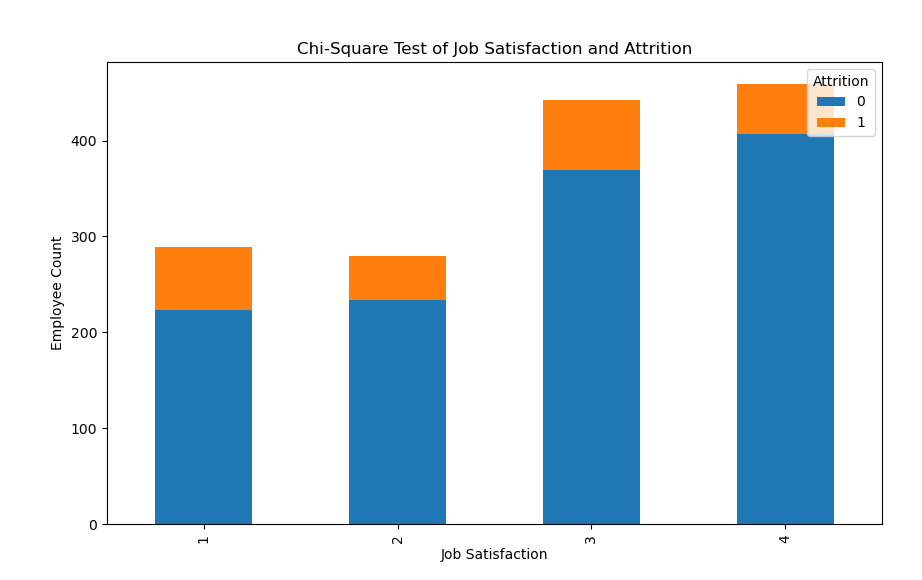
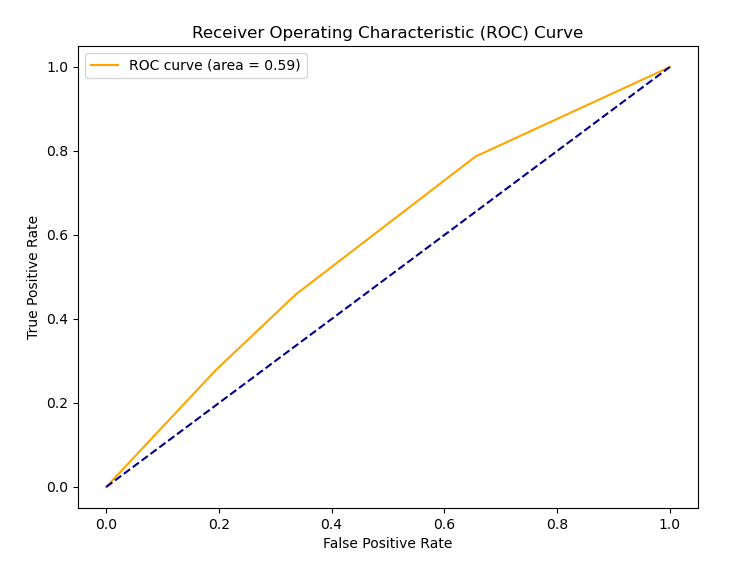


Figure 2 shows that this test provided additional evidence that the null hypothesis can be rejected, confirming that job satisfaction is indeed related to employee attrition. The results of the chi-square test between job satisfaction and attrition show a chi-square value of 15.45 with a p-value of 0.002, indicating a statistically significant relationship between these variables.

**Figure 3**

*Receiver Operating Characteristic (ROC) Curve for Logistic Regression Model*

The ROC curve illustrates the performance of the logistic regression model used to predict employee attrition based on job satisfaction and other factors. The curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The Area Under the Curve (AUC) is 0.59, which indicates a decen model performance in distinguishing between employees who are likely to leave (attrition) and those who are not.

**CONCLUSION**

In conclusion, this research examines the relationship between job satisfaction and employee attrition using the HR Employee Attrition dataset from IBM. The study demonstrates that job satisfaction is a significant predictor of attrition and that employees with lower levels of job satisfaction are more likely to leave the organization. By employing data preprocessing, exploratory data analysis, logistic regression modeling, and hypothesis testing, the research provides valuable insights into the factors driving employee turnover.

The findings from this study have practical implications for organizations seeking to reduce turnover and improve employee retention. By addressing the key drivers of dissatisfaction, such as work-life balance and compensation, organizations can create a more engaged and motivated workforce. Furthermore, predictive models developed from the dataset can help human resource departments proactively identify at-risk employees and take preemptive measures to retain them.

The analysis of the HR Employee Attrition dataset reveals a statistically significant relationship between job satisfaction and attrition. Specifically, lower job satisfaction was found to increase the likelihood of employee turnover. The findings from the logistic regression and chi-square tests allowed us to confidently reject the null hypothesis, demonstrating that job satisfaction plays a key role in predicting attrition rates.

These results highlight the importance of improving job satisfaction to reduce turnover. While job satisfaction was the most significant factor, other variables, such as work-life balance and compensation, also showed moderate associations with attrition, indicating that employee retention strategies should address multiple areas of employee experience.

**RECOMMENDATIONS**

Based on the findings, several recommendations can be made to help organizations reduce attrition:

1. Enhance Job Satisfaction Initiatives: Organizations should focus on initiatives that improve overall job satisfaction, such as offering regular feedback, recognition programs, and development opportunities.
2. Promote Work-Life Balance: Providing flexible working conditions, such as remote work options and flexible hours, can improve employee satisfaction and reduce turnover.
3. Optimize Compensation Packages: Competitive salaries and comprehensive benefits packages can further enhance job satisfaction and retention.
4. Monitor Employee Satisfaction Regularly: Implementing employee satisfaction surveys and exit interviews can help HR teams monitor trends and act on areas of concern before they result in attrition.

By implementing these recommendations, organizations can foster a more engaged and loyal workforce, ultimately reducing turnover and improving overall organizational performance.

References

Allen, D. G., & Bryant, P. C. (2012). *Managing employee turnover: Dispelling myths and fostering evidence-based retention strategies*. Business Expert Press.

Cascio, W. F., & Boudreau, J. W. (2016). *Investing in people: Financial impact of human resource initiatives* (3rd ed.). FT Press.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26(3), 463-488.

Hausknecht, J. P., & Trevor, C. O. (2011). Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management*, 37(1), 352-388.

Hom, P. W., & Griffeth, R. W. (2021). *Employee retention and turnover: Why employees stay or leave*. Routledge.

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3), 530–545.

Lee, T. W., Hom, P. W., Eberly, M. B., & Li, J. (2018). Managing employee retention and turnover with the unfolding model of voluntary turnover. *Human Resource Management Review*, 28(1), 34-49.

Meyer, J. P., Stanley, D. J., Herscovitch, L., & Topolnytsky, L. (2002). Affective, continuance, and normative commitment to the organization: A meta-analysis of antecedents, correlates, and consequences. *Journal of Vocational Behavior*, 61(1), 20-52.

Podsakoff, N. P., LePine, J. A., & LePine, M. A. (2007). Differential challenge stressor–hindrance stressor relationships with job attitudes, turnover intentions, and withdrawal behavior: A meta-analysis. *Journal of Applied Psychology*, 92(2), 438–454.

Singh, J., Goolsby, J. R., & Rhoads, G. K. (1994). Behavioral and psychological consequences of boundary spanning burnout for customer service representatives. *Journal of Marketing Research*, 31(4), 558-569.