Charles University
Faculty of Social Sciences, Institute of Economic Studies
Data Analysis in R - JEB157
WS 2025/2026

What Drives Job Satisfaction? An Empirical Analysis of Employee Data.

Alicja Rączka, 19627526

Abstract

How employees feel about their jobs directly impacts how they perform and whether they choose to stay with a company. Because of this, understanding what actually drives job satisfaction is a major part of effective HR management. In this study, I investigated these factors using employee-level data from a large corporate dataset.

My analysis focuses on the *IBM HR Analytics* dataset, covering 1470 employees. It includes everything from basic demographics to specific job roles and performance outcomes. To make sense of the data, I used a mix of descriptive statistics, charts, and multiple linear regression to see how different personal and professional factors link up with job satisfaction.

The results show that satisfaction levels aren't the same for everyone; they vary significantly across different groups and are closely tied to employee retention. Simply put, when satisfaction drops, the likelihood of someone leaving the company goes up. These findings suggest that if organizations want to keep their talent, they need to focus more on workplace conditions and overall employee well-being.

1. Introduction

It is well known that how people feel about their work directly shapes their behaviour and, ultimately, the success of the whole organization. When people are happy at work, they are usually more productive and loyal. On the flip side, low job satisfaction often leads to disengagement and higher turnover, which can be incredibly costly for any business.

Recently, more companies have started using data-driven methods to get a clearer picture of workplace dynamics. Thanks to HR analytics, we can now look at large datasets that cover everything from demographics to specific job roles. This gives us a great opportunity to pinpoint what actually makes people satisfied or frustrated at work.

My research focuses on identifying these key drivers of satisfaction. Specifically, I wanted to answer a few main questions: Does satisfaction vary between different departments? How do personal factors like age, income, and seniority play a role? Also, I looked at whether working conditions, such as frequently working overtime, make a noticeable difference. Finally, I examined the link between satisfaction and attrition to see if unhappy employees are truly more likely to quit.

The goal of this report is to move beyond guesswork and use empirical data to evaluate what keeps employees satisfied. By using a mix of descriptive statistics and regression analysis, I aim to provide clear, evidence-based insights that can help managers improve retention and better understand their teams.

2. Data and Methods

2.1 Data Source and Data Overview

The data for this study comes from the *IBM HR Analytics Employee Attrition & Performance* dataset, which I sourced from Kaggle. This dataset is a standard resource in HR research, as it provides a detailed look at employee behaviour within a corporate setting.

In total, the dataset covers 1470 employees. For my analysis, I specifically focused on variables that could directly or indirectly influence how people feel about their work. This includes basic demographics like age and gender, but also more specific job-related factors such as department, job role, monthly income, and how many years they've been with the company. I also included "overtime status" as a potential factor, along with "attrition" to see how satisfaction links to people leaving.

To keep the analysis consistent, it's important to note how these things are measured. Job satisfaction is recorded on a scale from 1 to 4 (where 4 is the highest), and attrition is a simple "yes/no" (binary) variable. Having a mix of both numerical and categorical data like this allowed me to use a variety of methods, from basic descriptive statistics to more complex regression models.

2.2 Data Cleaning

Before diving into the actual analysis, I thoroughly inspected the dataset to make sure the data was clean and consistent. I ran column-wise checks on all the selected variables to look for any missing values. Luckily, there were no missing observations in the parts of the data I used, so I didn't have to worry about data imputation or excluding any records.

To prepare the data for statistical testing, I handled the variables based on their type. I converted categorical fields- like gender, department, job role, and overtime status- into factor variables to ensure they would be processed correctly in the regression models. As for the numerical variables, I kept them

in their original form; they all fell within reasonable ranges, and I didn't find any extreme outliers that would have skewed the results.

Overall, the dataset was very well-structured from the start, which meant it was ready for analysis without needing any heavy preprocessing or major adjustments.

2.3 Data Transformation

To make the results easier to interpret, I decided to focus only on a specific subset of variables that directly related to my research questions. This helped reduce unnecessary complexity while still giving me plenty of information to work with.

I chose not to apply any major transformations like normalization or scaling. Since job satisfaction is an ordinal variable and the other predictors are already easy to understand in their original units, keeping them as they were made the most sense. The only real change I made was explicitly converting categorical variables into factor types, which was necessary for running group comparisons and making sure they worked correctly in the regression models.

By keeping the data close to its original form, I wanted to ensure the findings remained straightforward and easy to apply in a real-world corporate context.

2.4 Statistical Methods

For the analysis, I used a combination of descriptive statistics, visual tools, and formal statistical modelling. I conducted all the work within the R statistical environment.

I started by looking at the basic distributions of key numerical variables like age, income, and seniority. To get a better feel for the data, I used histograms and boxplots, which helped me spot any immediate patterns or differences between various groups of employees.

To dig deeper, I moved on to hypothesis testing. Specifically, I ran a two-sample t-test to see if there was a real gap in satisfaction between those who stayed at the company and those who left. I also used a one-way ANOVA to check if satisfaction levels varied significantly across different departments.

After that, I explored how these variables relate to one another using correlation matrices and scatterplots. This provided a clear visual and numerical summary of the linear relationships in the data. Finally, I built a multiple linear regression model to identify which factors- like monthly income, age, or overtime- actually drive job satisfaction when holding everything else constant. This allowed me to see which variables were statistically significant and which had the biggest impact.

3. Results

3.1 Descriptive Statistics

Descriptive statistics were computed for key numerical variables, including age, monthly income, years at the company, and job satisfaction. Job satisfaction is measured on a four-point ordinal scale, with higher values indicating greater satisfaction.

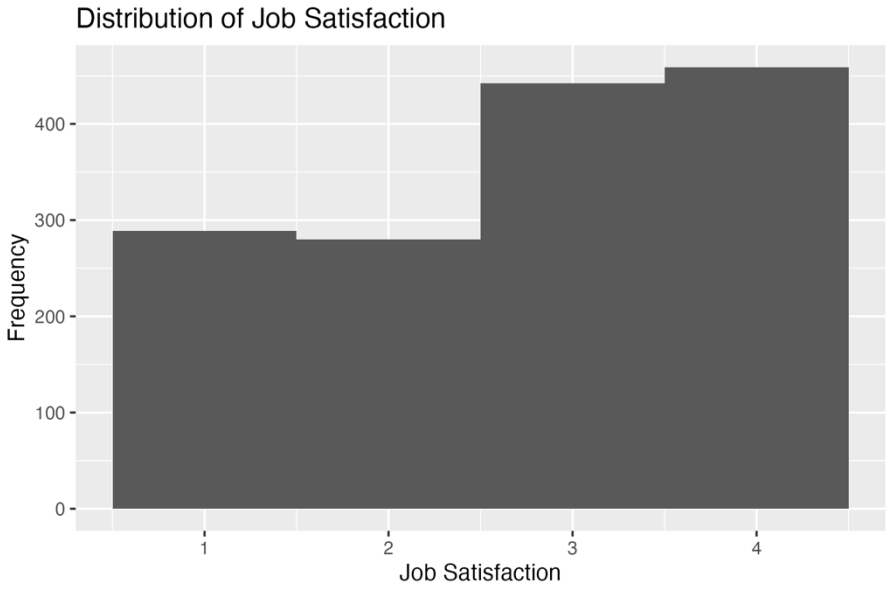

Figure 1: *Distribution of job satisfaction across employees. The histogram shows that most employees report moderate levels of job satisfaction, with no extreme skewness in the distribution.*

The distribution of job satisfaction shows moderate variation across employees, with most observations concentrated around the middle values of the scale. Graphical analysis using histograms suggests no extreme skewness in the distribution. Boxplots were used to examine differences in job satisfaction across departments, indicating relatively similar distributions between organizational units.

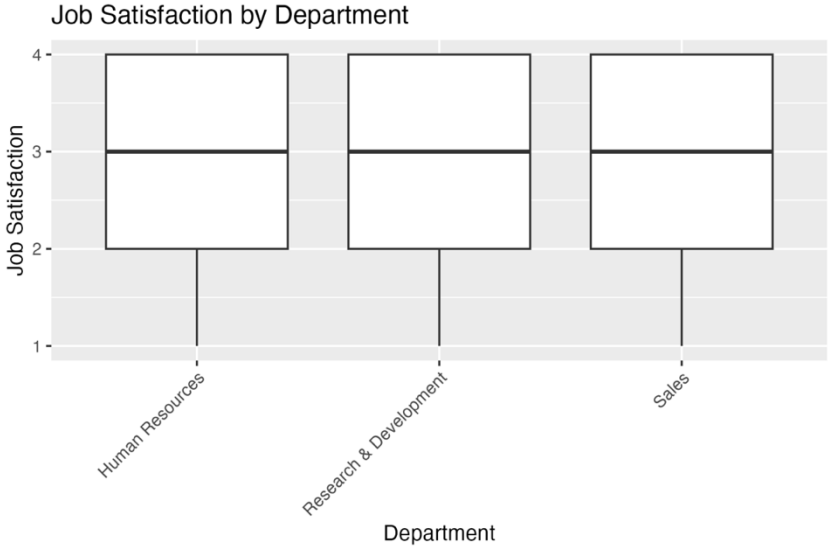

Figure 2: *Job satisfaction across departments. The boxplots indicate similar distributions of job satisfaction across organizational units, suggesting limited department-level differences.*

Overall, the descriptive analysis provides an initial overview of employee characteristics and suggests that job satisfaction varies across individuals but remains broadly comparable across departments.

3.2 Job Satisfaction and Employee Attrition

To examine whether job satisfaction differs between employees who left the company and those who remained, a two-sample Welch t-test was conducted. The test compares mean job satisfaction between employees with attrition status “Yes” and “No”.

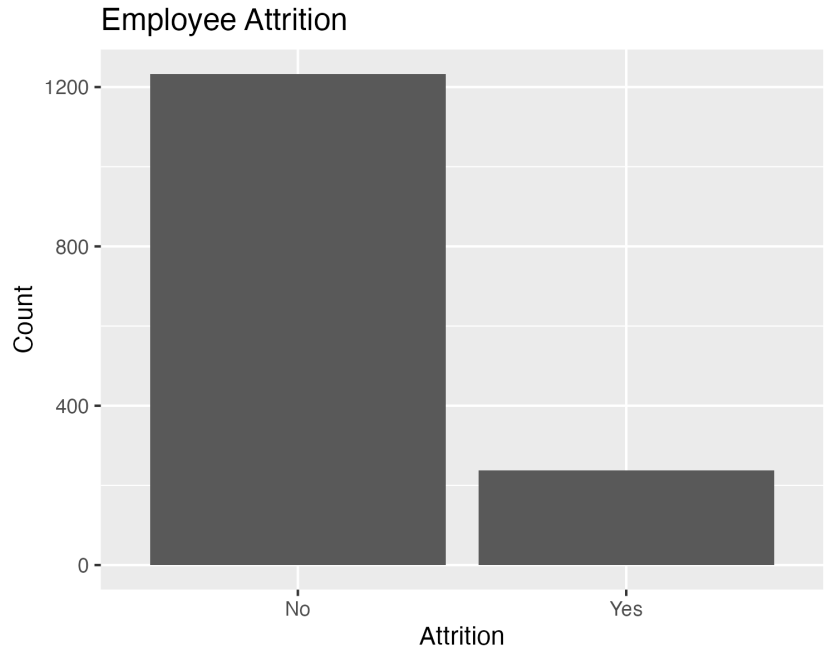

Figure 3: *Distribution of employee attrition. The bar chart shows the number of employees who stayed with the company compared to those who left.*

The results indicate a statistically significant difference in job satisfaction between the two groups ($t = 3.93$, $p < 0.001$). Employees who remained with the company report a higher average level of job satisfaction (mean = 2.78) compared to employees who left the organization (mean = 2.47). The 95% confidence interval for the difference in means does not include zero, further supporting the presence of a meaningful difference.

These findings suggest that lower job satisfaction is associated with a higher likelihood of employee attrition.

3.3 Job Satisfaction Across Departments

A one-way analysis of variance (ANOVA) was conducted to test whether mean job satisfaction differs across departments. The null hypothesis assumes equal mean job satisfaction across all departments.

The ANOVA results do not indicate statistically significant differences in job satisfaction between departments ($F = 0.50$, $p = 0.605$). This suggests that, within the organization, job satisfaction levels are broadly similar across departments and that departmental affiliation alone does not explain variation in job satisfaction.

3.4 Regression Analysis

To further investigate the determinants of job satisfaction, a multiple linear regression model was estimated. Job satisfaction was regressed on monthly income, years at the company, age, and overtime status.

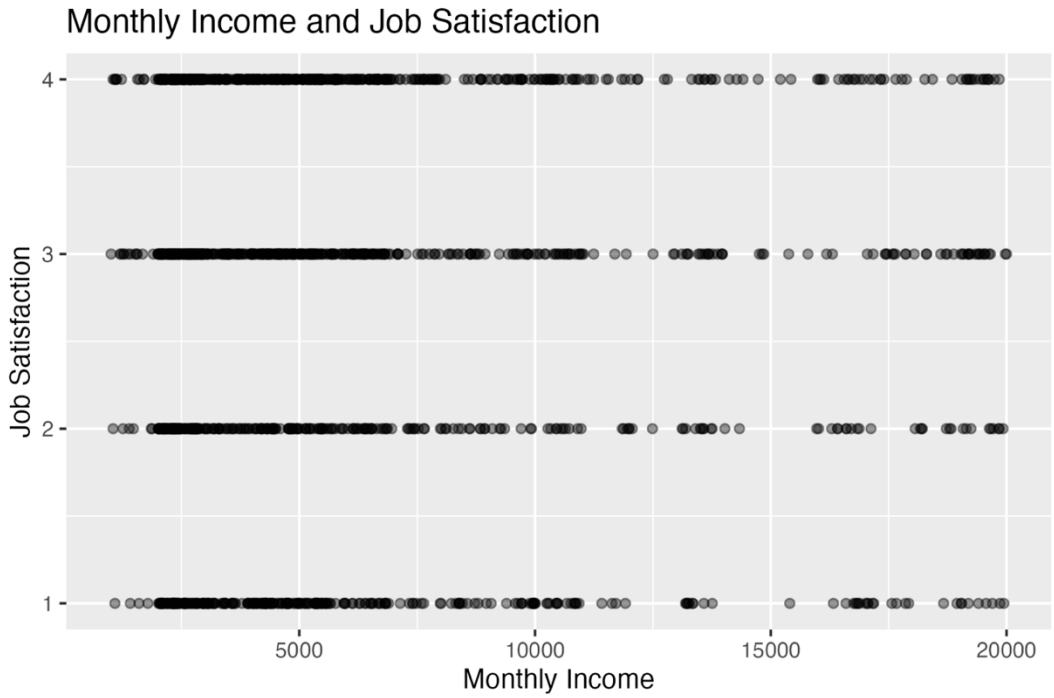

Figure 4: *Relationship between monthly income and job satisfaction. The scatter plot suggests no strong linear association between income and job satisfaction.*

The regression results indicate that none of the included explanatory variables have a statistically significant effect on job satisfaction at conventional significance levels. Monthly income, tenure, age, and overtime status all exhibit small coefficient estimates with high p-values, suggesting weak linear relationships with job satisfaction when controlling for other variables.

The overall explanatory power of the model is very low, with an R-squared value close to zero. This indicates that the included variables explain only a negligible proportion of the variation in job satisfaction. The F-test for overall model significance is also not statistically significant.

These results suggest that job satisfaction is not strongly driven by the selected demographic and job-related variables and may instead depend on unobserved factors such as workplace culture, management practices, or individual preferences.

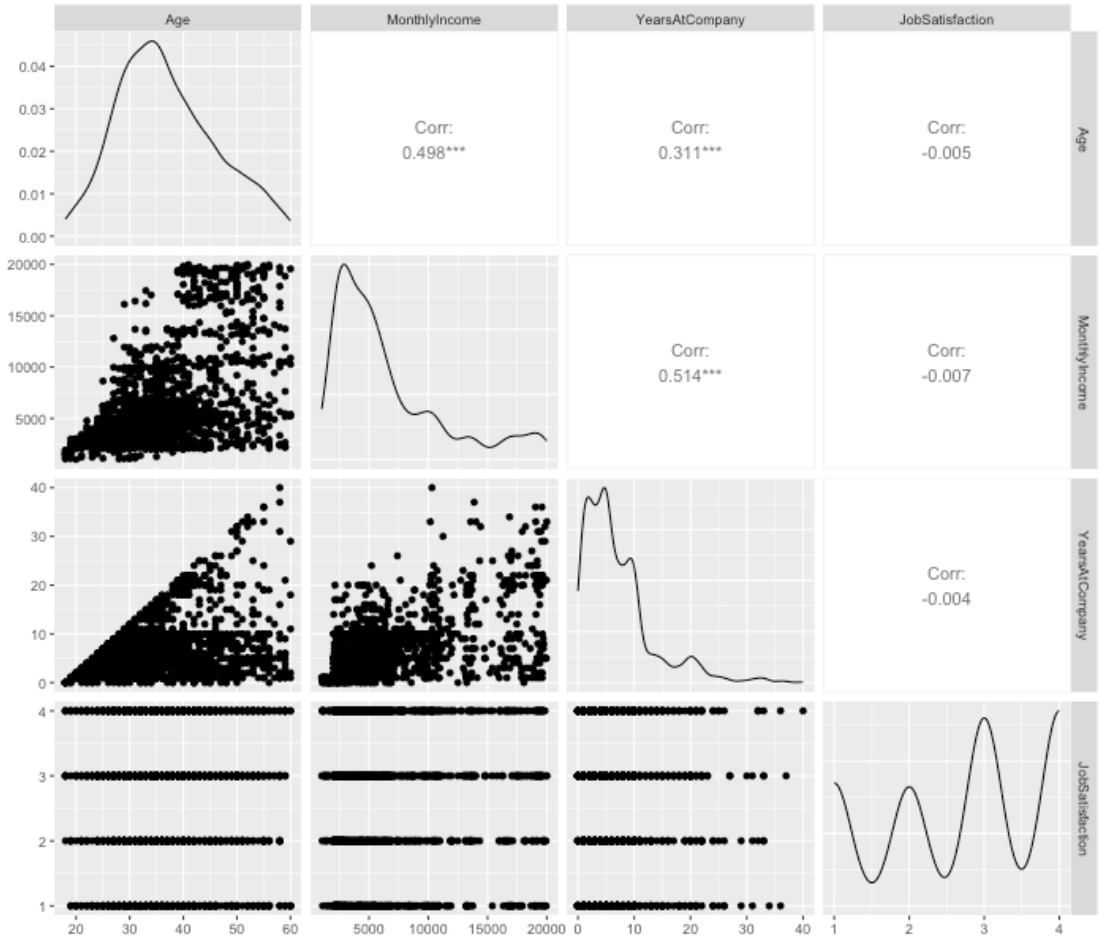

Figure 5: *Scatterplot matrix of numerical variables. The figure provides a visual overview of pairwise relationships and confirms weak linear associations among the selected variables.*

4. Discussion and Conclusion

In this study, I set out to identify what actually drives job satisfaction by looking at everything from demographics to specific working conditions. My main goal was to see how these factors connect and, more importantly, how they influence whether an employee decides to stay or leave the company.

The most striking finding is the clear link between satisfaction and attrition. The data shows that employees with lower satisfaction scores were much more likely to leave, which isn’t exactly a surprise, but it strongly confirms that satisfaction is a major red flag for turnover. It proves that once someone starts feeling unhappy with their role, they’re already halfway out the door.

Interestingly, I didn't find any significant differences in satisfaction levels between departments. This suggests that the "department" itself isn't the problem- instead, it might be the overall company culture or broad organizational policies that set the tone for everyone.

What I found most surprising, however, was the regression model. Despite including variables like income, age, and overtime, the model had surprisingly low explanatory power. It turns out that these "hard" factors don't tell the whole story. This suggests that job satisfaction is likely driven by things this dataset didn't capture- like how much someone likes their boss, the quality of their workplace

relationships, or simply feeling recognized for their work. It’s a reminder that satisfaction is far more complex and personal than just a paycheck or a job title.

Of course, this study has its limits. Since it's based on a single dataset, we can't automatically assume the same patterns apply to every industry. Also, treating an ordinal scale like a continuous variable in a regression is a bit of a shortcut, and we are missing data on psychological factors.

Nevertheless, the takeaways for HR managers are clear. If you want to keep your best people, you can't just look at compensation. You must monitor how they actually feel. Moving forward, it would be great to see research that includes more "human" variables or tracks these changes over time to get an even clearer picture of why people stay or go.