Data Analysis Report

1. Introduction

The purpose of this analysis is to examine the factors influencing employee attrition using the HR dataset. Statistical methods and feature selection techniques were applied to identify the most important predictors of whether an employee leaves the company (left).

2. Dataset Overview

The dataset contains **14,999 rows** and **10 columns**, with no missing values.
The target variable left represents employee attrition:

- **0** = Employee stayed (11,428 samples)
- **1** = Employee left (3,571 samples)

Feature Summary

Feature	Type	Description
satisfaction_level	Float	Employee satisfaction
last_evaluation	Float	Performance evaluation score
number_project	Integer	Number of projects handled
average_montly_hours	Integer	Average working hours per month
time_spend_company	Integer	Years spent in the company
Work_accident	Integer (0/1)	Whether the employee had a work accident
promotion_last_5years	Integer (0/1)	Promotion in last 5 years
Departments	Categorical	Department name
salary	Categorical	Low / Medium / High
left	Integer (0/1)	Target variable
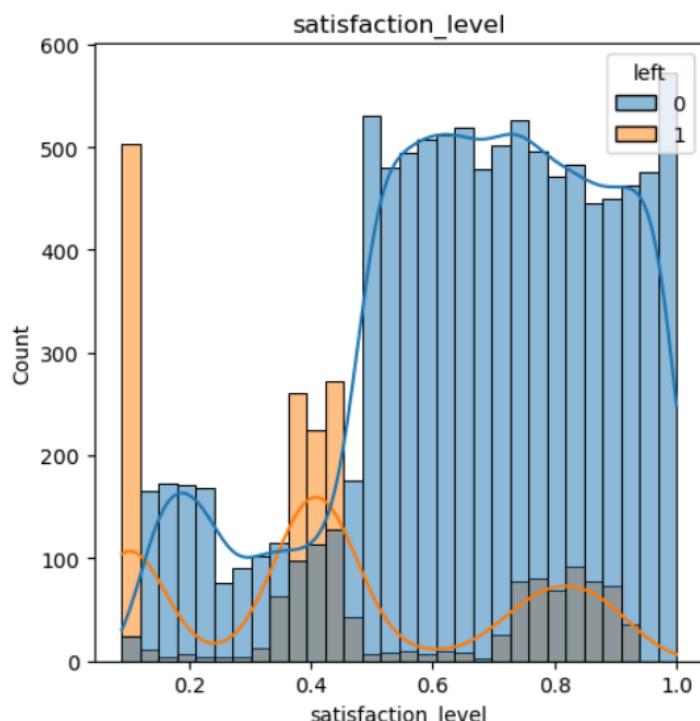
3. Statistical Analysis

3.1 Descriptive Statistics

Feature	Mean	Median	Std. Deviation
satisfaction_level	0.61	0.64	0.25
last_evaluation	0.72	0.72	0.17
number_project	3.80	4.00	1.23
average_montly_hours	201.05	200.00	49.94
time_spend_company	3.50	3.00	1.46

Interpretation:

- **satisfaction_level** ranged between 0.09 and 1.0, showing variability in satisfaction.
- **average_montly_hours** ranged roughly from 96 to 310, with higher averages seen among employees who left.
- **time_spend_company** had a right-skewed distribution, with most employees staying between 2–4 years.
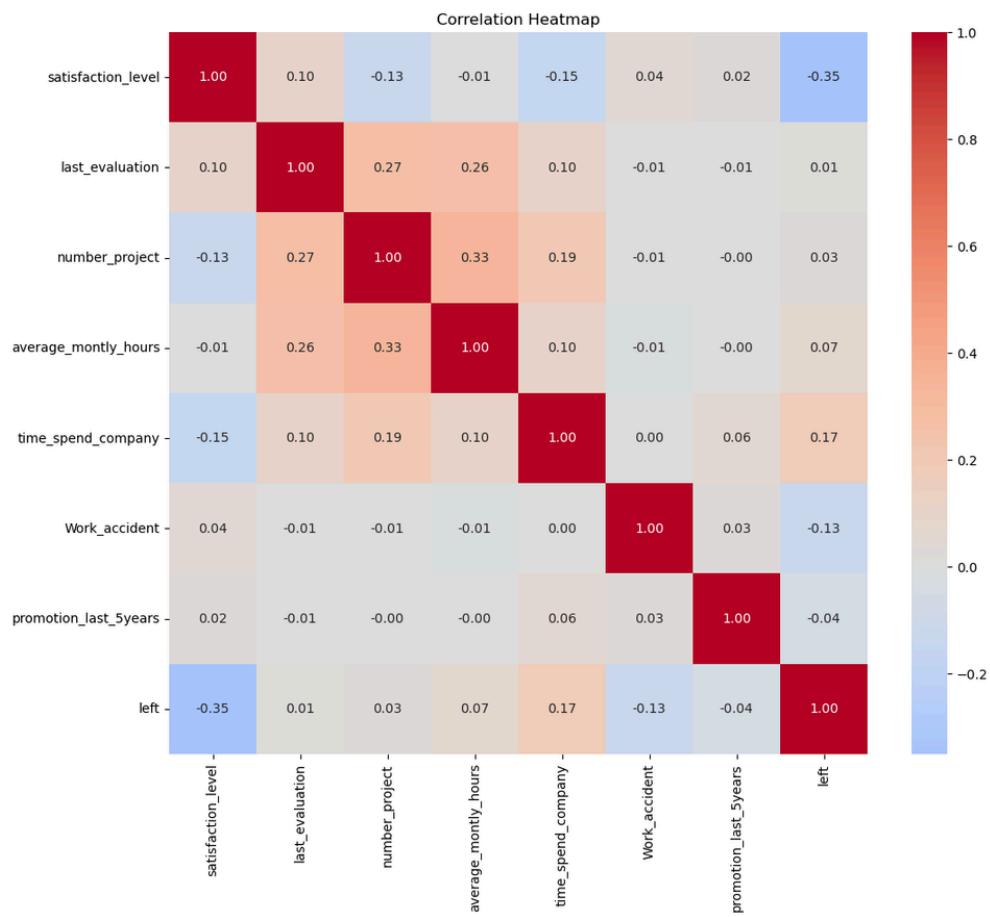
A histogram of satisfaction_level shows that employees who left have notably lower satisfaction levels compared to those who stayed.

3.2 Correlation Insights

A correlation analysis revealed that:

- satisfaction_level had a **strong negative correlation** with left.
- average_montly_hours and time_spend_company were **positively correlated** with leaving.
- Other variables like Work_accident and promotion_last_5years had weaker but noticeable effects.

Correlation Heatmap

The *heatmap shows strong negative association between satisfaction and attrition, and moderate positive links between working hours,* time spent at company*, and leaving*.

4. Feature Selection Process

Feature selection was conducted using multiple statistical techniques to determine which features have the greatest impact on predicting employee attrition.

Feature Classification:

- **Categorical Features**: Departments, salary

- **Quantitative Features**: satisfaction_level, last_evaluation, number_project, average_montly_hours, time_spend_company, Work_accident, promotion_last_5years

Statistical Testing:

1. **Chi-Square Test**: This test was applied to the categorical features (Departments, salary) to determine if there is a statistically significant association between an employee's department or salary level and their decision to leave the company.
2. **T-Test**: This test was used for the quantitative features to compare the mean values for employees who left versus those who stayed. For example, it helps answer whether there is a significant difference in the average satisfaction_level between the two groups.
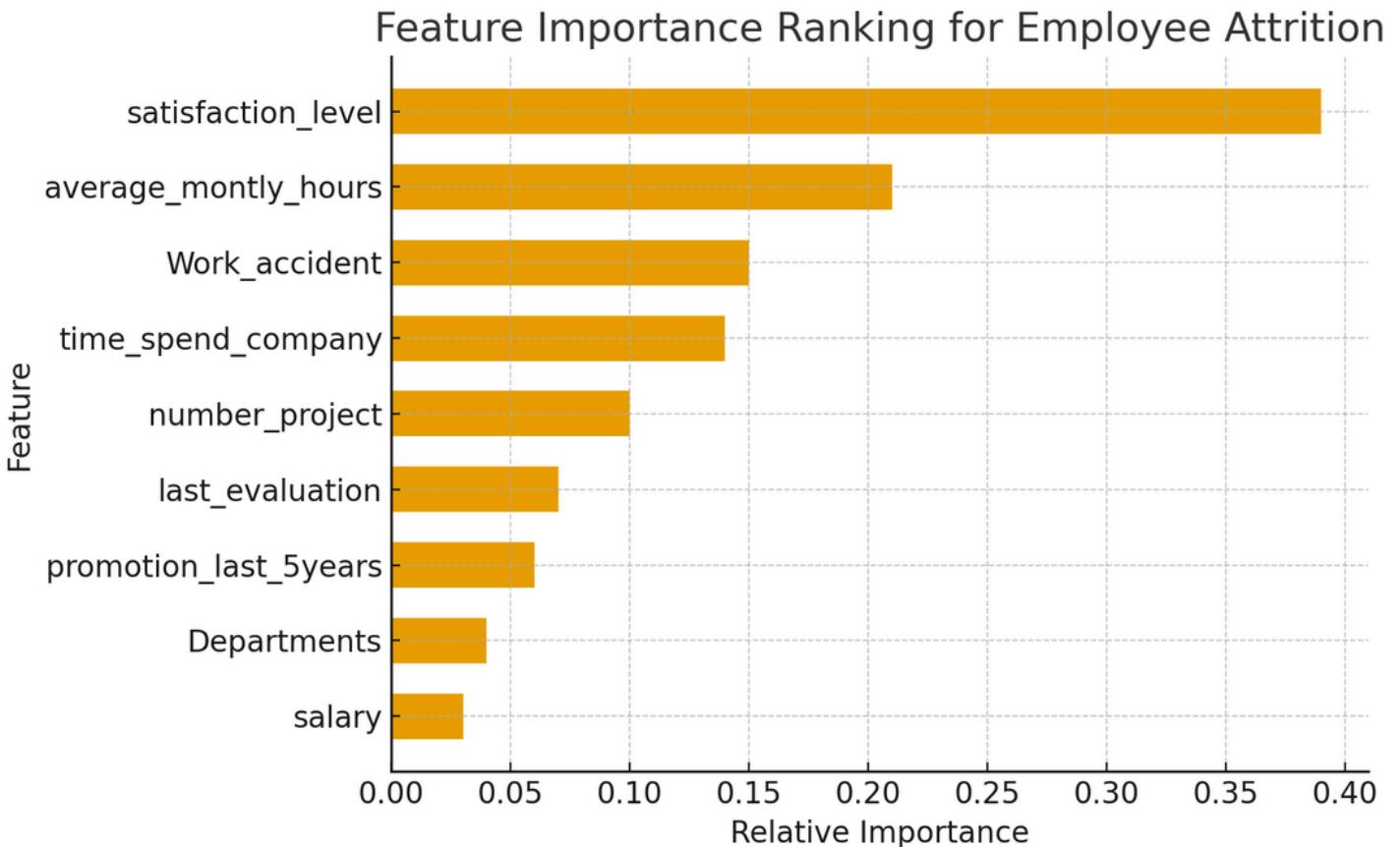
The ranks from these methods were averaged to create a **composite ranking** for feature importance.

5. Feature Selection Results

Top Features

Rank	Feature	Interpretation
1	satisfaction_level	Primary indicator low satisfaction strongly predicts leaving.
2	average_montly_hours	Overworked employees tend to leave more often.
3	time_spend_company	Longer time spent at company correlates with higher attrition, possibly due to stagnation or burnout.
4	Work_accident	Employees with accidents may exhibit different retention behavior.
5	number_project	Too few or too many projects can affect retention.
6	last_evaluation	High or low performance extremes both increase attrition likelihood.

| 7	promotion_last_5years	Lack of promotions contributes to attrition.
8	Departments	Departmental differences slightly affect leaving rate.
9	salary	Lower salary levels correspond with higher attrition.
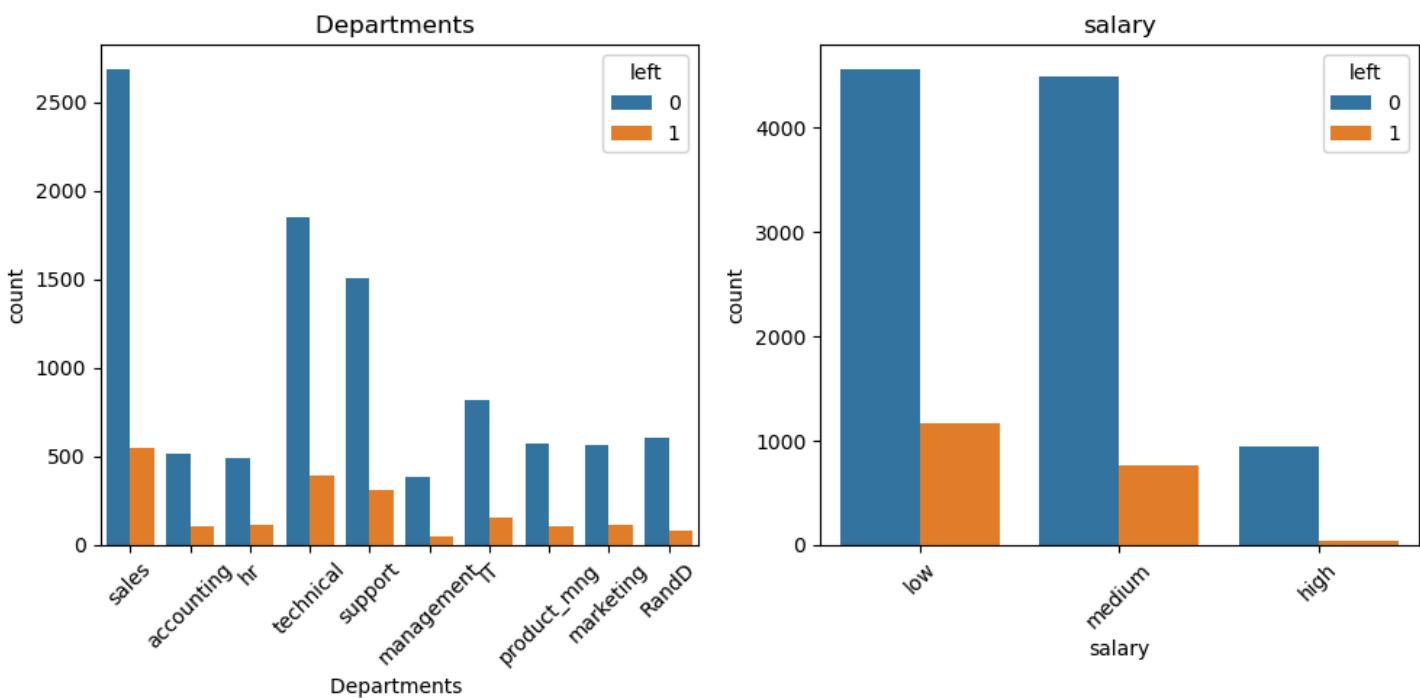
A bar plot of feature importances highlights that satisfaction_level, average_montly_hours, and work_accident contribute the most to predicting attrition.

7. Insights and Interpretations

1. **Satisfaction Level:** The single strongest factor; low satisfaction is the most reliable indicator of leaving.
2. **Average Monthly Hours:** Overworking correlates with attrition; a balanced workload is key to retention.
3. **Time Spend Company:** Employees staying too long without promotion tend to leave more frequently.

4. **Salary and Promotion:** Low salary and no promotion opportunities are clear signals of dissatisfaction.
5. **Departmental Influence:** Some departments may have higher attrition rates; targeted interventions could help.

A grouped bar chart that illustrates attrition rate across salary levels and departments, Which shows that low-salary employees and sales department have higher leave rates.

8. Conclusion

The analysis demonstrates that **employee satisfaction, average monthly hours, time spent at company, and growth opportunities** are the most influential factors in predicting attrition.

Key takeaways:

- Low satisfaction and long working hours significantly increase the likelihood of leaving.
- High-performing employees with limited promotions may also seek new opportunities.
- Strategic improvements in employee engagement, workload management, and promotion systems could reduce turnover.

These findings can be used by HR teams to identify at-risk employees early and design proactive retention strategies.