

EMPLOYEE ATTRITION PREDICTION PROJECT

Department of Human Resources

Presented by Reinhardt Kefa



TABLE

of contents

01. Introduction

05. Methodology

02. Objectives

06. Objectives

03. Hypothesis

07. Analysis

04. Goals

08. Insights and Recommendations

Introduction

Larana Company currently faces three major challenges:

- No automated system to predict which employees are likely to leave
- HR decisions are mostly reactive rather than proactive
- High turnover costs due to late interventions

As a result:

- Critical roles become vacant unexpectedly
- Productivity declines
- Recruitment and training expenses increase

The company needs a reliable way to answer one key question:

“Can we predict which employees are likely to leave before they actually do”



Objectives

1

Primary objective is to build a machine learning model that can accurately predict employee attrition using historical HR data

2

Perform deep exploratory data analysis on employee records and identify factors most strongly associated with attrition

3

Recommend the best performing model for deployment

4

Provide actionable insights to HR management



Methodology

To ensure a structured and professional approach, this project followed a complete end-to-end data science workflow, consisting of the following stages:

- Data collection
- Data Understanding
- Exploratory Data Analysis(EDA)
- Feature Engineering
- Model Development
- Model Evaluation
- Insights and Recommendations



Data Collection

The dataset used in this project was sourced from the company's HR database.

It contains historical employee records with attributes such as:

- Job role
- Salary level
- Work environment satisfaction
- Performance ratings
- Years at company
- Overtime status
- Training and Promotion history

The target variable: Employee attrition



Exploratory Data Analysis

EDA was performed to uncover patterns and relationships within the data, including:

- Distributions of attrition across departments
- impact of salary, job satisfaction and overtime
- correlation between experience and employee turnover
- identification of important features affecting attrition

Key findings from EDA helped guide feature selection and modeling decisions



Feature Engineering

EDA was performed to uncover patterns and relationships within the data, including:

- Distributions of attrition across departments
- impact of salary, job satisfaction and overtime
- correlation between experience and employee turnover
- identification of important features affecting attrition

Key findings from EDA helped guide feature selection and modeling decisions



Model Development

a

Multiple classificationalgorithms were trained and compared including:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- K-Nearest Neighbors
- Gradient Boosting
- XGBoost
- Adaboost
- Stochastic Gradient descent

This multi-model approach ensured that the best-performing algorithm was selected



Model Evaluation

a

Models were evaluated using several performance metrics:

- Accuracy
- Precision
- Recall
- F1 Score
- ROC-AUC Score

A train-test split strategy was applied to ensure fair evaluation and models were ranked based on their ability to correctly predict employee attrition



Deployment Concept

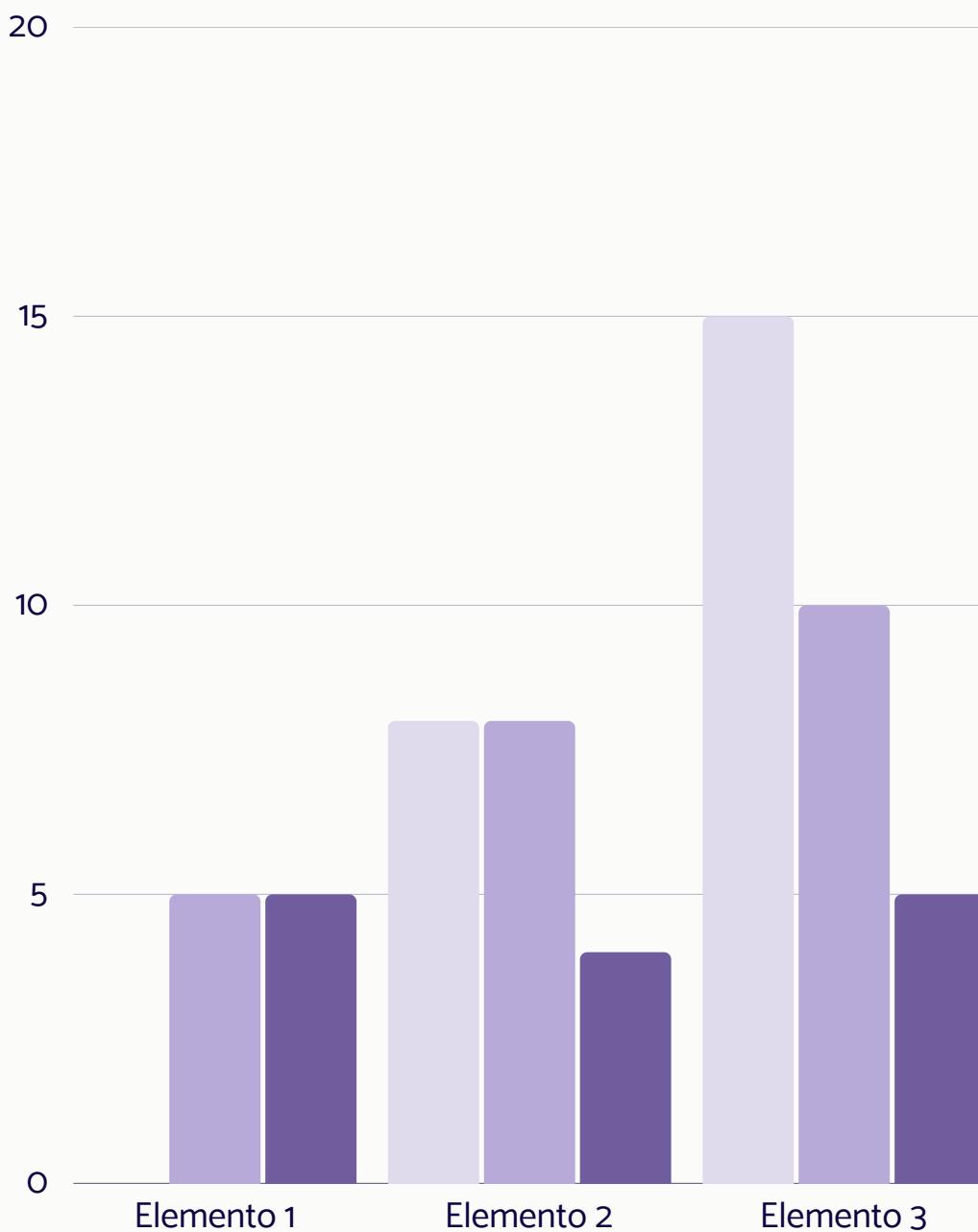
a

Although this project is a proof of concept, the final model can be integrated into:

- HR decision-support systems
- Employee monitoring dashboards
- Retention strategy planning tools

This enables the company to shift from reactive HR management to proactive talent retention

Data Analysis



Analysis focused on understanding employee behavior and identifying factors that influence attrition

Using statistical exploration and visual analysis, we examined:

- Attrition distribution
- Relationships between key HR variables
- Trends across departments
- Impact of salary, overtime and job satisfaction
- Employee experience and career progression



Basic concepts

Concept 1

 Lorem ipsum dolor sit amet,
 consectetur adipiscing elit. Ut a enim
 nec nisl ullamcorper eleifend.
 Praesent risus leo, fringilla et nulla at,
 egestas euismod orci. Suspendisse
 porttitor diam eu condimentum
 aliquam. Fusce interdum cursus nisl
 ut rutrum. Donec et sapien sit amet
 nisl pretium efficitur.

Concept 2

 Lorem ipsum dolor sit amet,
 consectetur adipiscing elit. Ut a enim
 nec nisl ullamcorper eleifend.
 Praesent risus leo, fringilla et nulla at,
 egestas euismod orci. Suspendisse
 porttitor diam eu condimentum
 aliquam. Fusce interdum cursus nisl
 ut rutrum. Donec et sapien sit amet
 nisl pretium efficitur.

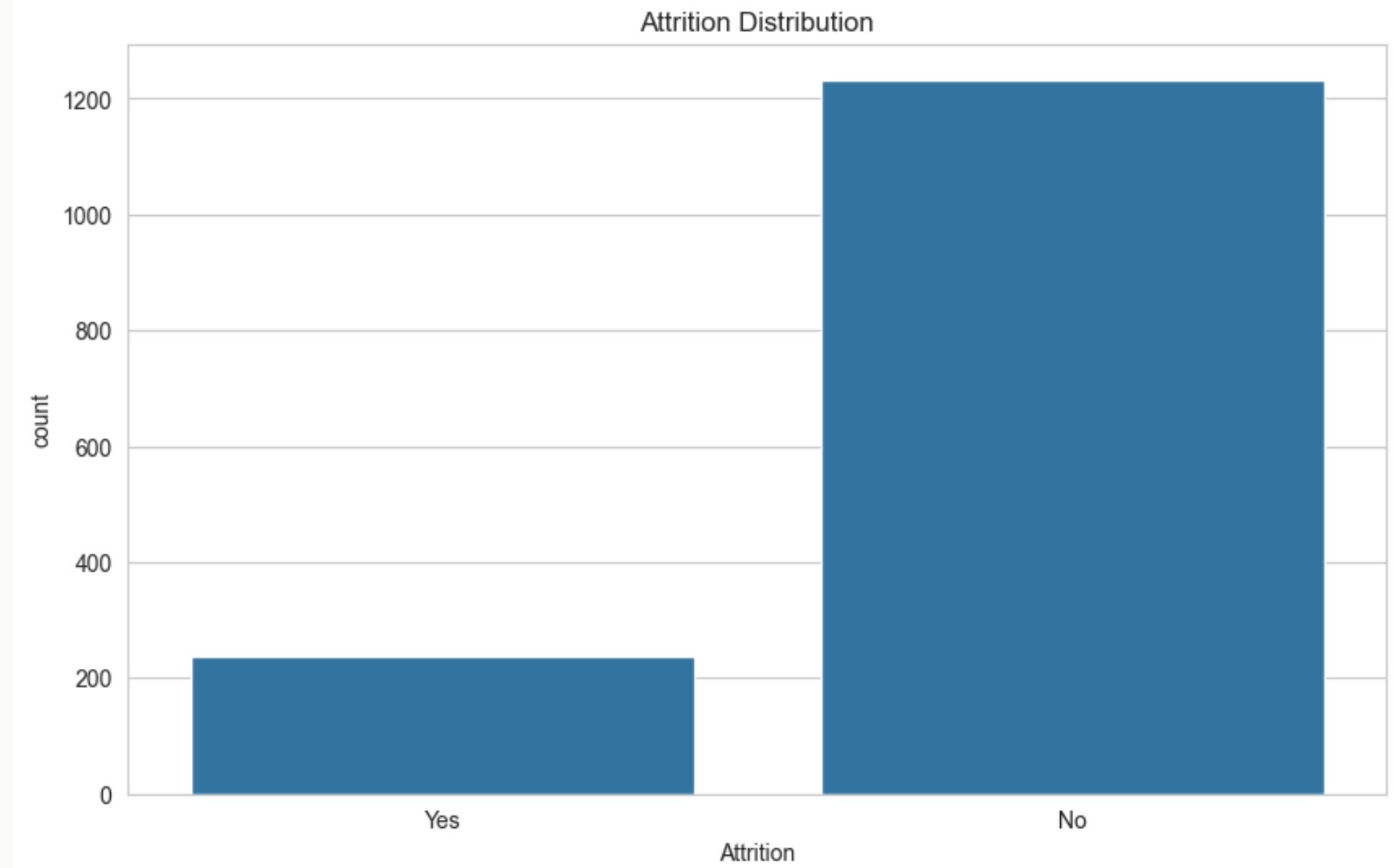
Concept 3

 Lorem ipsum dolor sit amet,
 consectetur adipiscing elit. Ut a enim
 nec nisl ullamcorper eleifend.
 Praesent risus leo, fringilla et nulla at,
 egestas euismod orci. Suspendisse
 porttitor diam eu condimentum
 aliquam. Fusce interdum cursus nisl
 ut rutrum. Donec et sapien sit amet
 nisl pretium efficitur.

Attrition Distribution

Employee Attrition Overview

- The dataset Showed class imbalance between employees who left anf those who stayed
- Majority of employees remain with the comapny while a smaller portion leave
- This imbalance was considered during modeling to avoid biased predictions



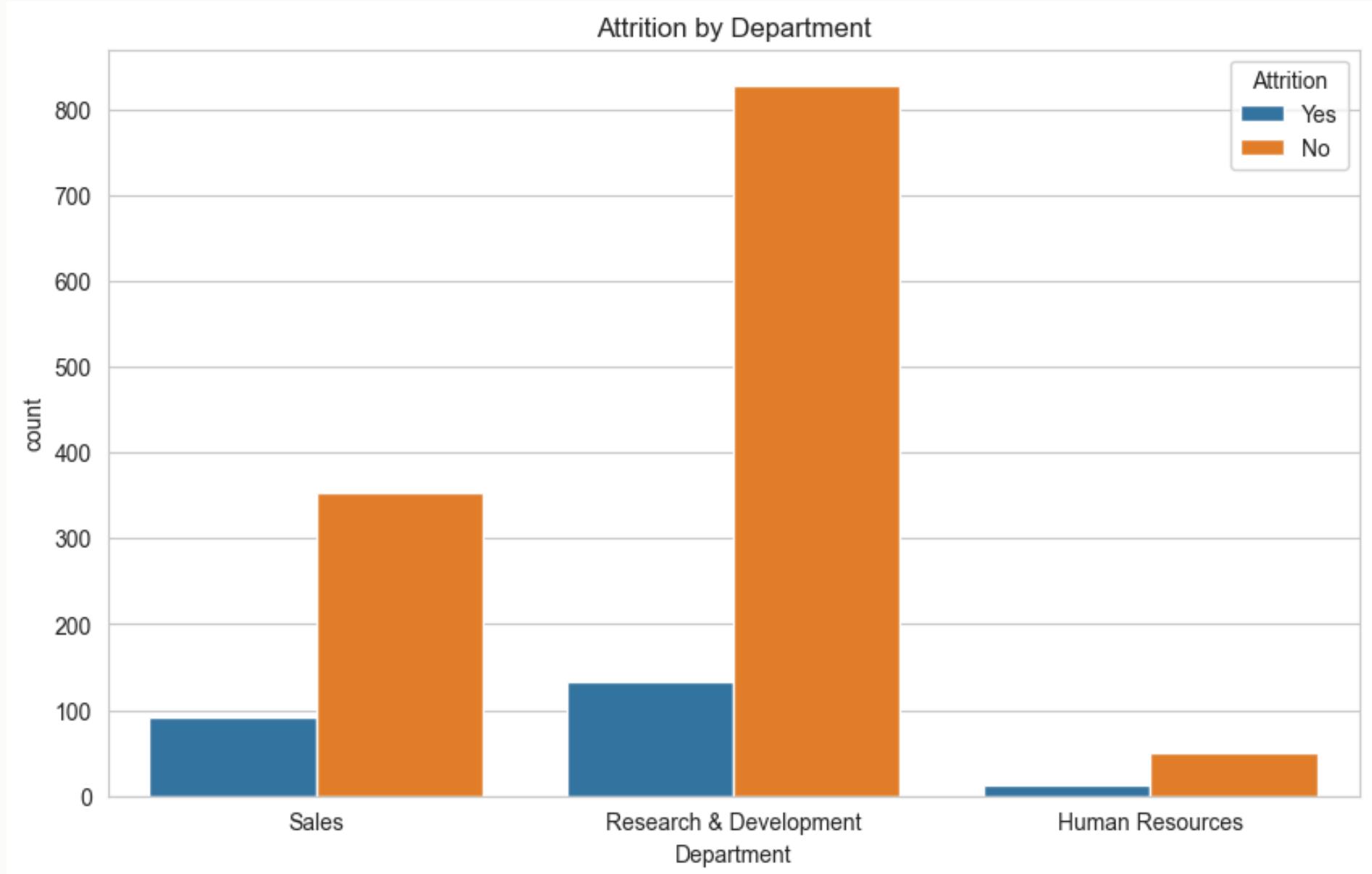
Departmental Attrition

Attrition by department

- Attrition rates vary across departments
- Sales and R&D department show higher employee turnover than others
- Roles with higher workload and pressure exhibit increased attrition

Key Insight

Certain units require stronger retention strategies than others



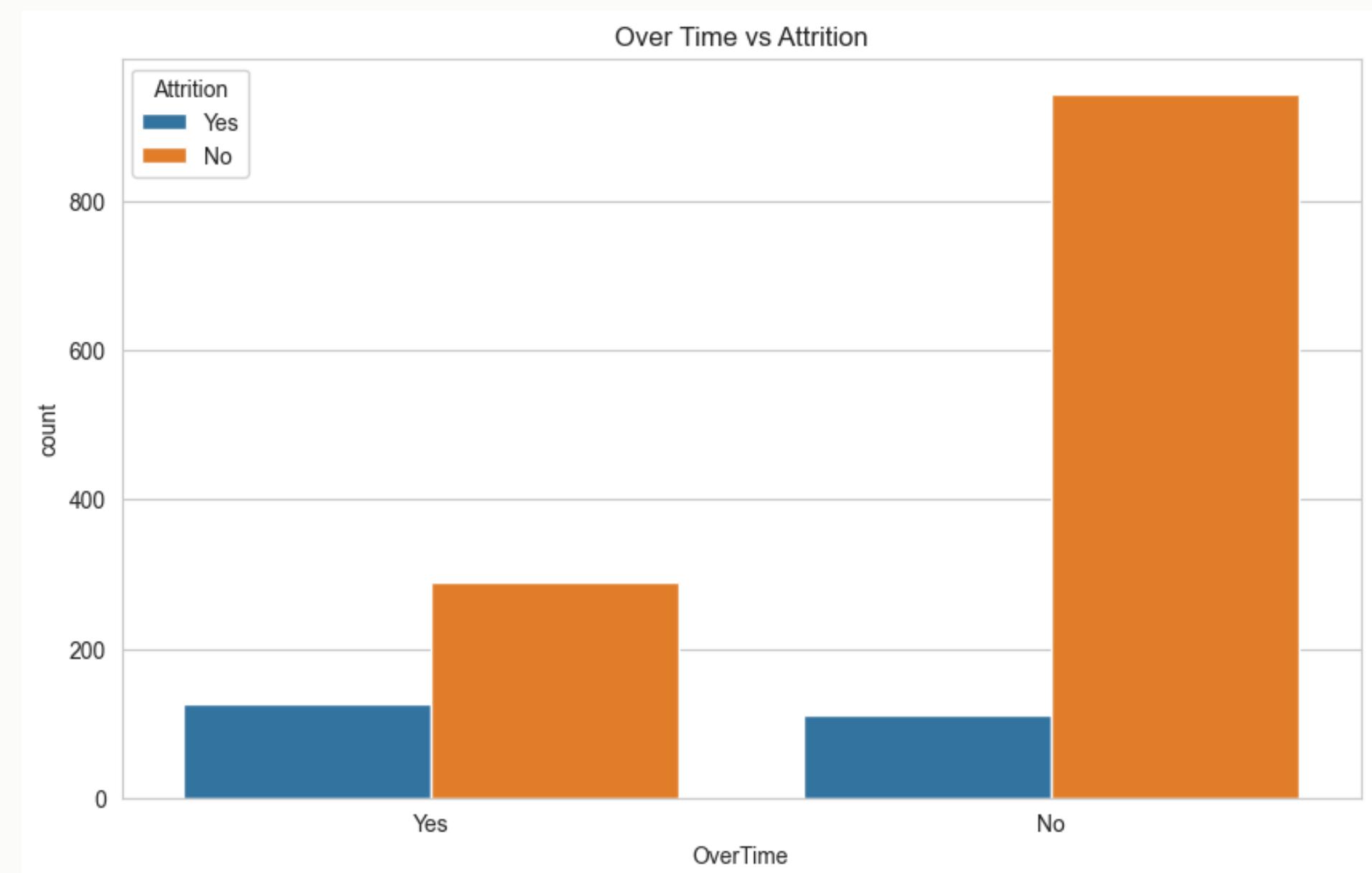
Impact of overtime

Overtime and Attrition

- Employees who frequently work overtime are far more likely to leave
- High Overtime correlates with burnout and dissatisfaction
- Overtime emerged as one of the strongest predictors of attrition

Business Takeaway

Work-life balance plays a critical role in employee retention



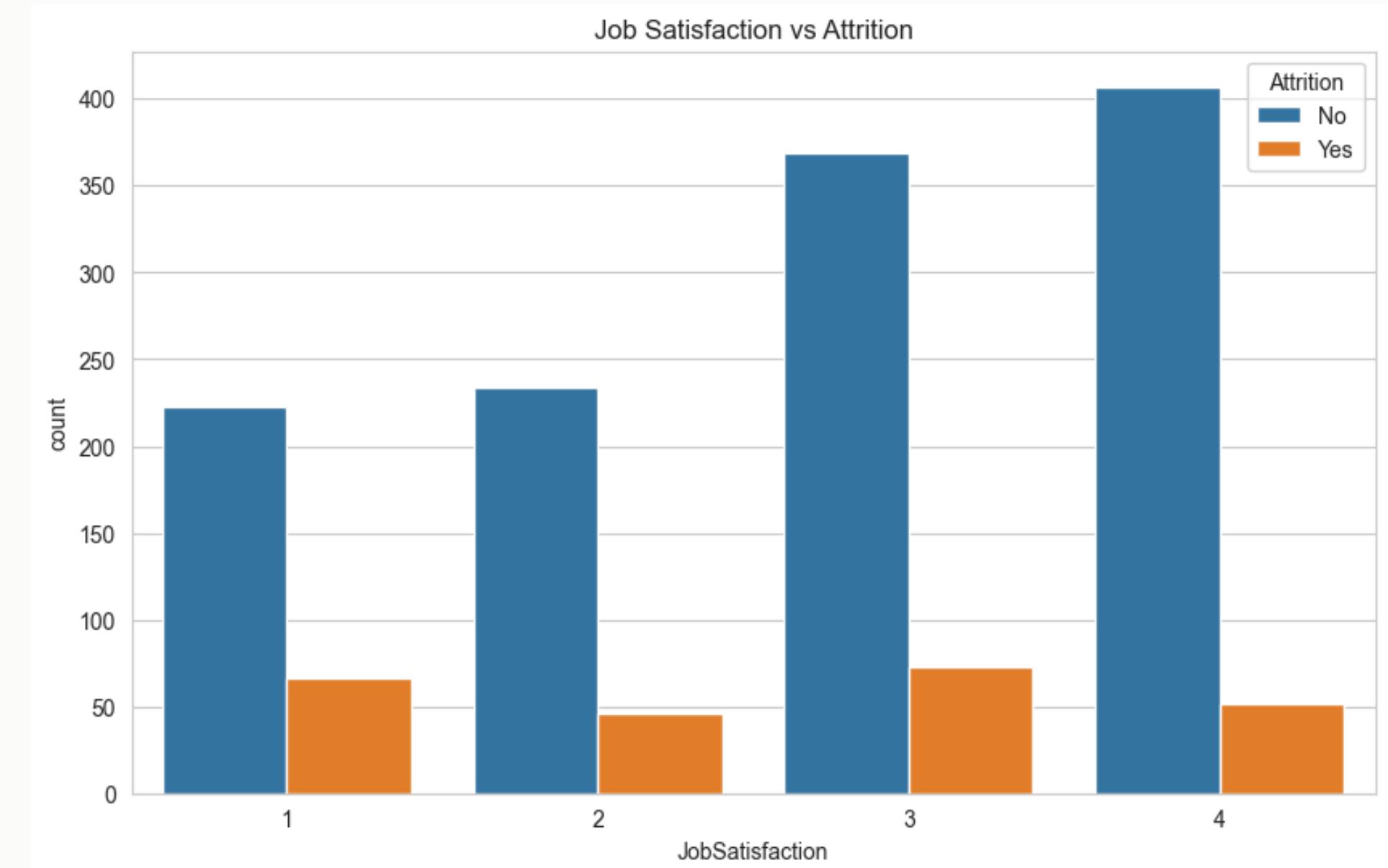
Job satisfaction and Environment

Job satisfaction factors

- Analysis revealed that employees with job satisfaction of 3 had the highest turnover and as expected the ones with the lowest job ratings.

Insight

Employee engagement and satisfaction are critical drivers of loyalty



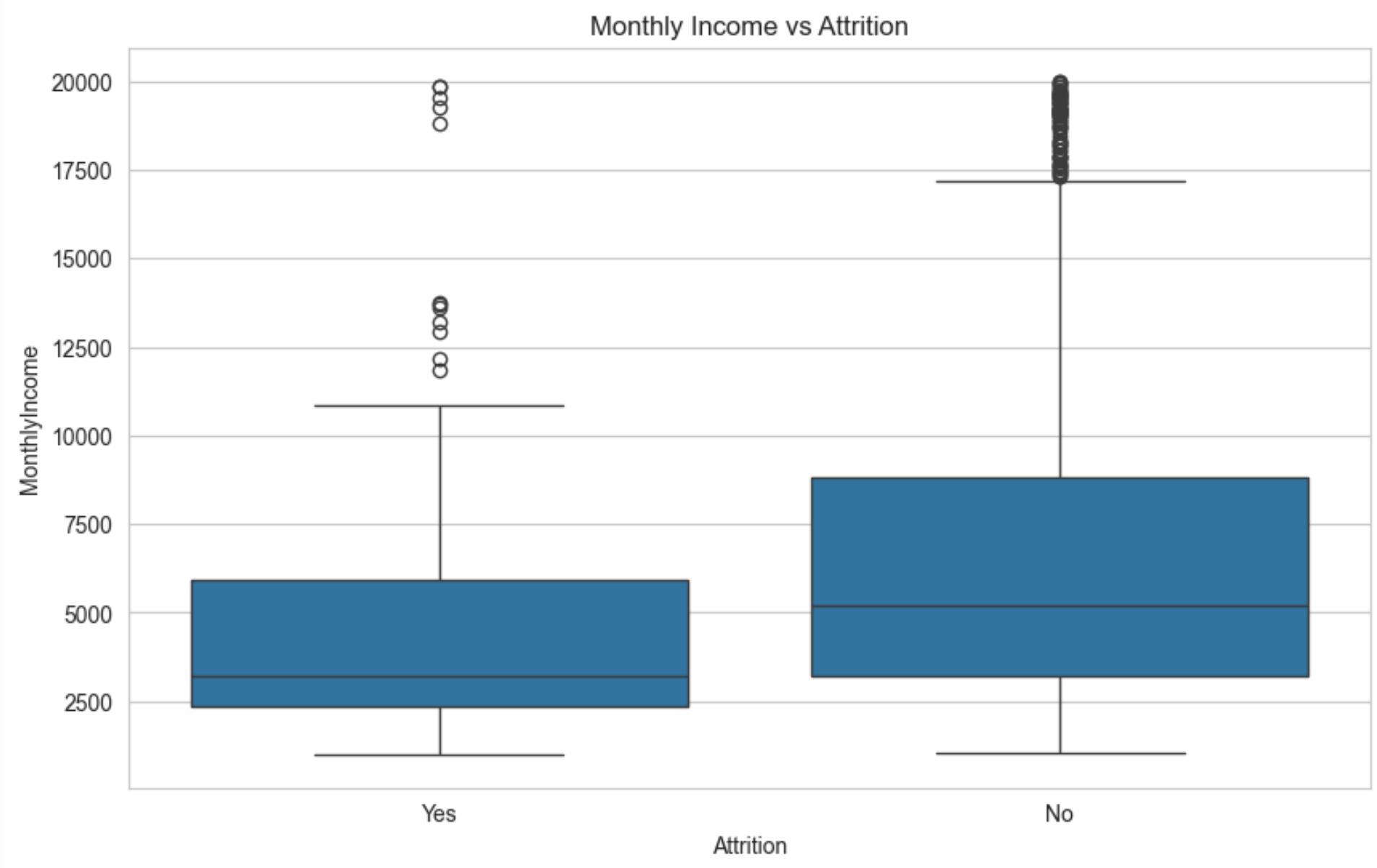
Salary and Attrition

Monthly income vs Attrition

- Lower income employees show higher turnover rates
- Competitive compensations reduces likelihood of resignation
- Pay structure is a key factor in retaining top talent

Recommendation Insight

Strategic salary reviews could help reduce turnover



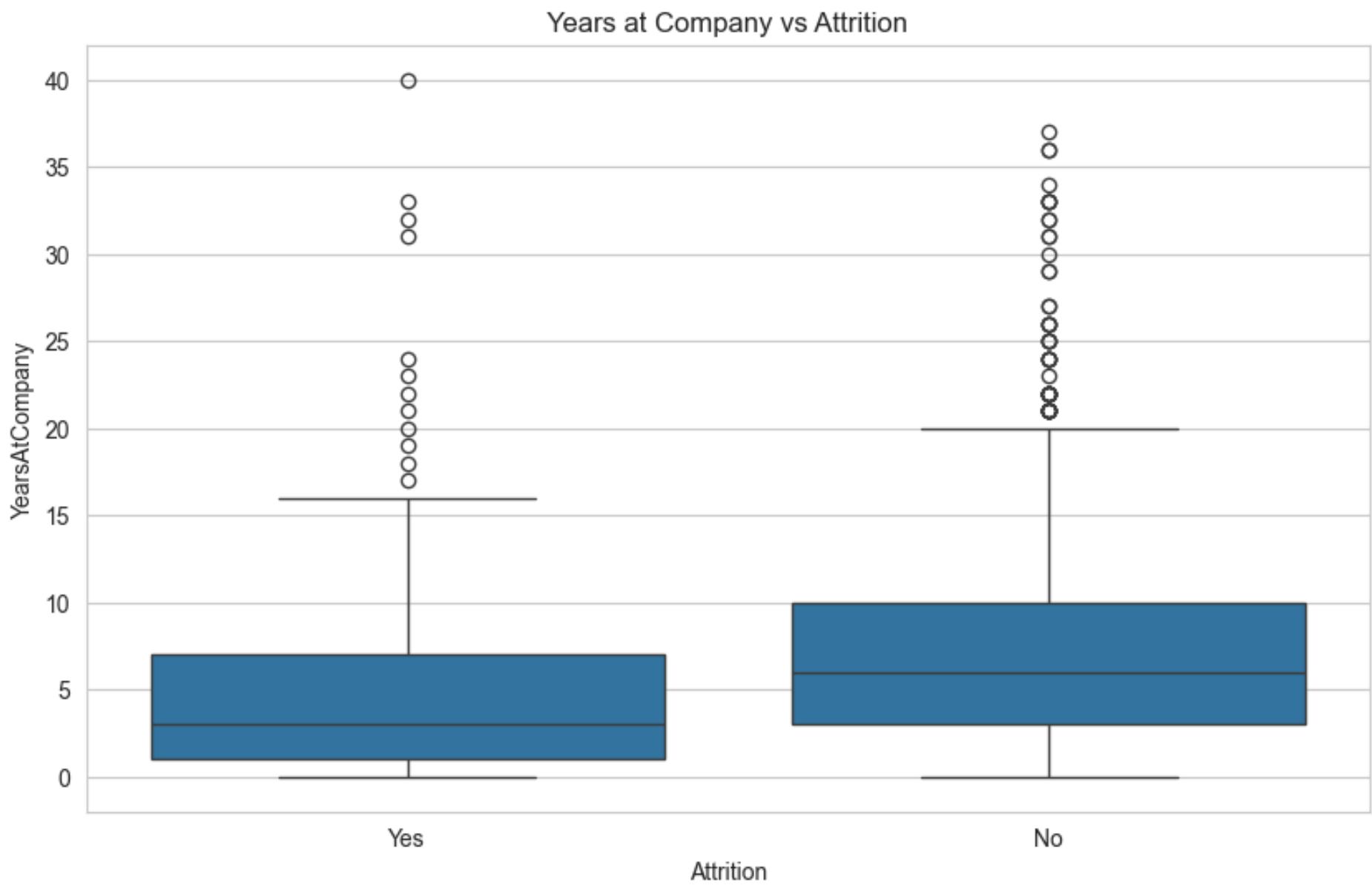
Experience and Career Progression

Experience based trends

- Employees with fewer years at the company are more likely to leave
- Attrition decreases as tenure increases
- Lack of promotions correlates with higher resignation rates

Insight

Clear career paths improve long term retention

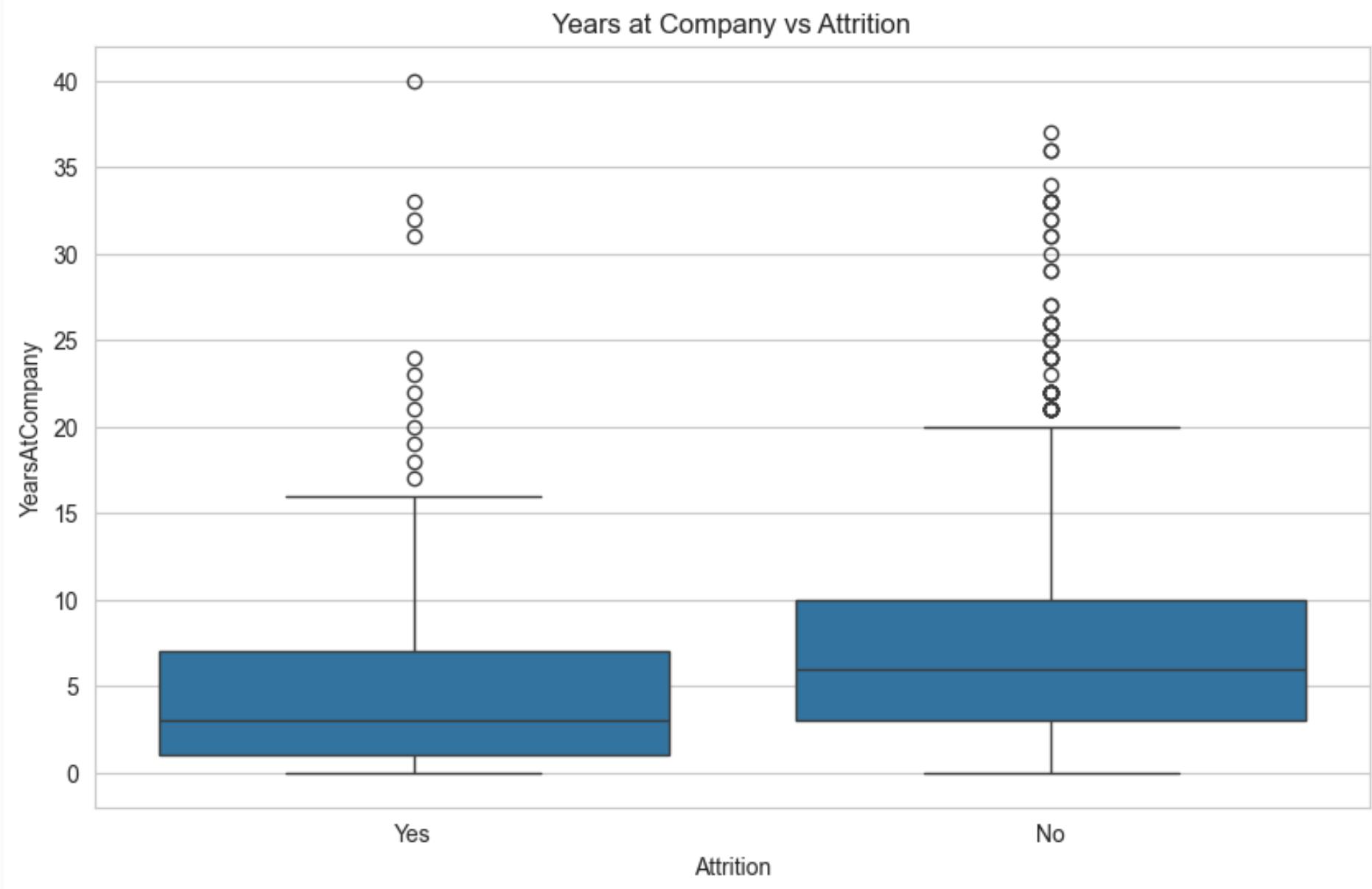


Correlation Analysis

Feature Relationships

Correlation analysis showed that the most influential factors include:

- Overtime
- Job satisfaction
- Years at company
- Monthly income
- Work environment



Models & Evaluation

Model Approach

To predict model employee attrition, multiple classification models were trained and compared

Algorithms Tested:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- K-Nearest Neighbors
- Gradient Boosting
- XGBoost
- Adaboost
- Stochastic Gradient descent

Metrics used: Accuracy, Precision, Recall, F1-score, ROC-AUC

Goal: "Select the model that best balances prediction accuracy and business usefulness"

Performance Summary

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	87.07%	68.00%	36.17%	47.22%	80.84%
SGD Classifier	62.58%	27.34%	80.85%	40.86%	81.09%
XGBoost	84.69%	55.00%	23.40%	32.83%	76.92%
Decision Tree	80.27%	37.77%	36.17%	36.95%	62.41%
Random Forest	84.01%	50.00%	10.63%	17.54%	79.67%

Key Finding:

- Highest accuracy not equal to best business model
- Recall and F1-score were prioritised
- Interpretability also considered

Final Model Selection

Selected Model: Logistic Regression

Why?

Best Overall balance of metrics

- Highest F1 Score(47.22%)
- High accuracy (87%)
- Good ROC-AUC (80.08%)
- Most interpretable for HR decisions

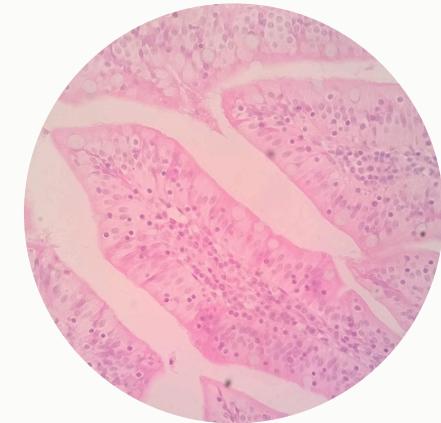
Most important Predictors identified:

- Overtime
- Monthly income
- Job Satisfaction
- Years at Company
- Work Environment

Conclusion

The Logistic Regression model provides a reliable and practical tool for identifying employees at risk of leaving and supports a proactive retention strategies

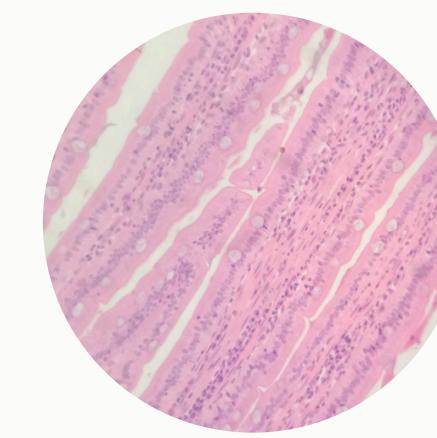
Key Insights from Analysis



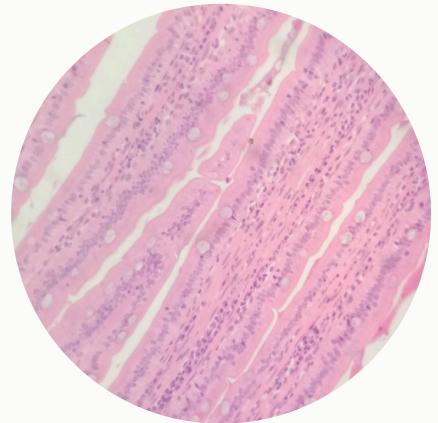
Overtime-Strongest attrition driver



Low job satisfaction increases turnover



Income level Matters



Early-career employee are most vulnerable

Recommendations

1. Control Overtime

- Redistribute workload
- Introduction of overtime monitoring dashboards
- Encourage work-life balance

2. Improve Job satisfaction

- Regular feedback surveys
- Recognition programs
- Clear career development paths

3. Compensation Review

- Benchmark salaries
- Performance-based incentives
- Transparent promotion policies

4. Targeted Retention Programs

- Focus on new employee (First 2 Years)
- Mentorship and training programs
- Personalized retention startegies for high risk staff

