IBM HR Employee Attrition Analysis

Statistical Computing Group Project

Group Members 1, 2, & 3

2025-12-30

Contents

1 Executive Summary

This report analyzes the IBM HR Employee Attrition dataset to identify factors contributing to employee turnover. The analysis is divided into three thematic areas:

- **Work Environment:** Employee sentiment and work-life balance.
- **Financial Structure:** Compensation and job hierarchy.
- **Demographics:** Age profiles and tenure modeling.

2 1. Introduction and Data Preparation

We analyze 8 key variables split into three categories.

3 2. Univariate Analysis

This section explores the distribution of variables within each thematic area.

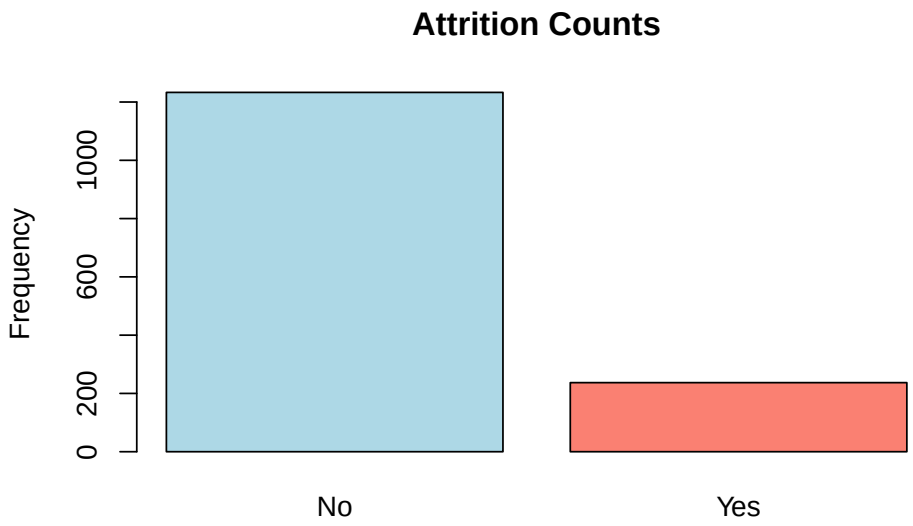
3.1 2.1 Work Environment (Member 1)

Focus: Attrition (Target), Job Satisfaction, Work-Life Balance.

3.1.1 Attrition Status

Table 1: Target Variable: Attrition Distribution

	Count	Proportion
No	1233	0.8388
Yes	237	0.1612



3.1.2 Employee Sentiment (Satisfaction & Balance)

3.2 2.2 Financial & Structural (Member 2)

Focus: Job Level, Overtime, Monthly Income.

3.2.1 Structural Variables

Table 2: Distribution of Job Levels

Var1	Freq
1	543
2	534
3	218
4	106
5	69
Sum	1470

Table 3: Overtime Status

	Count	Percent
No	1054	71.7
Yes	416	28.3

3.2.2 Monthly Income Distribution

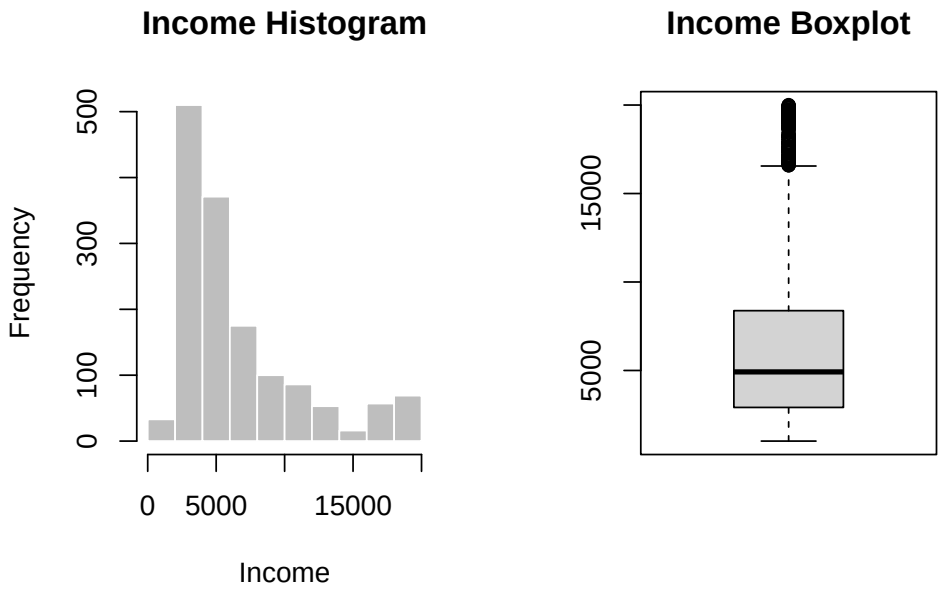
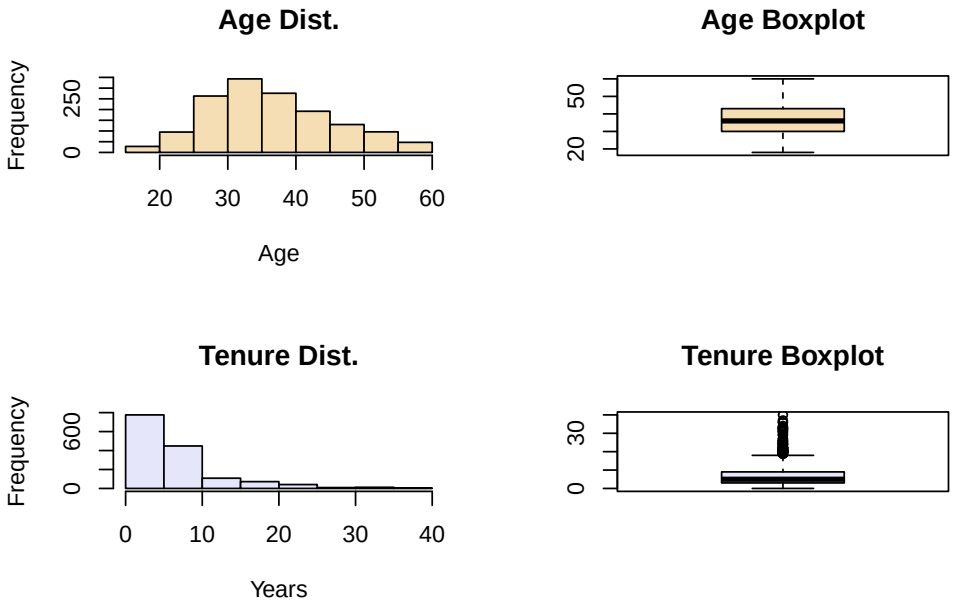


Table 4: Descriptive Statistics: Monthly Income

Mean	Median	SD	IQR
6502.931	4919	4707.957	5468

3.3 2.3 Demographics (Member 3)

Focus: Age, Years at Company.



4 3. Bivariate Analysis

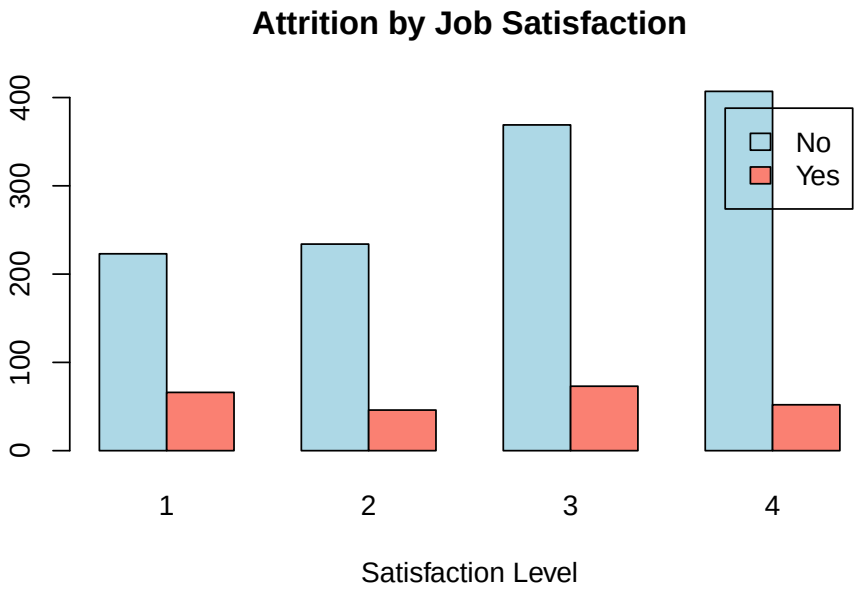
We now investigate how these factors relate to employee attrition.

4.1 3.1 Impact of Sentiment on Attrition (Member 1)

Does low satisfaction or poor work-life balance drive attrition?

Table 5: Attrition by Job Satisfaction Level

	1	2	3	4
No	223	234	369	407
Yes	66	46	73	52



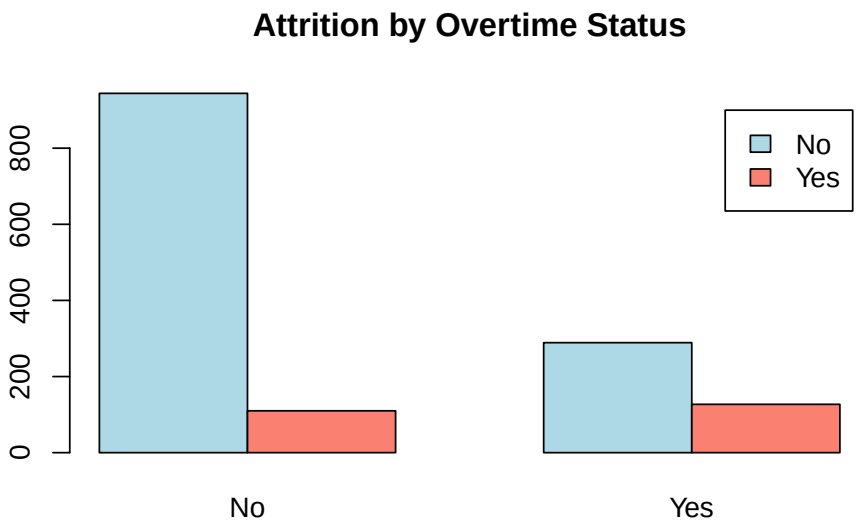
4.2 3.2 Financial Factors & Attrition (Member 2)

Are over-worked or under-paid employees leaving?

4.2.1 Overtime and Attrition

Table 6: Proportion of Attrition within Overtime Groups

	No	Yes
No	0.9	0.69
Yes	0.1	0.31



4.2.2 Income and Attrition

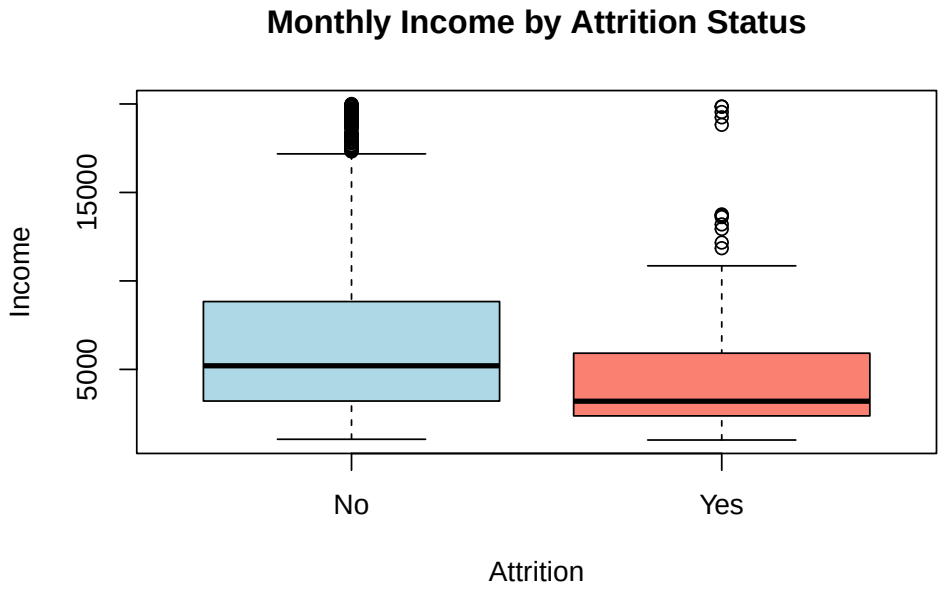
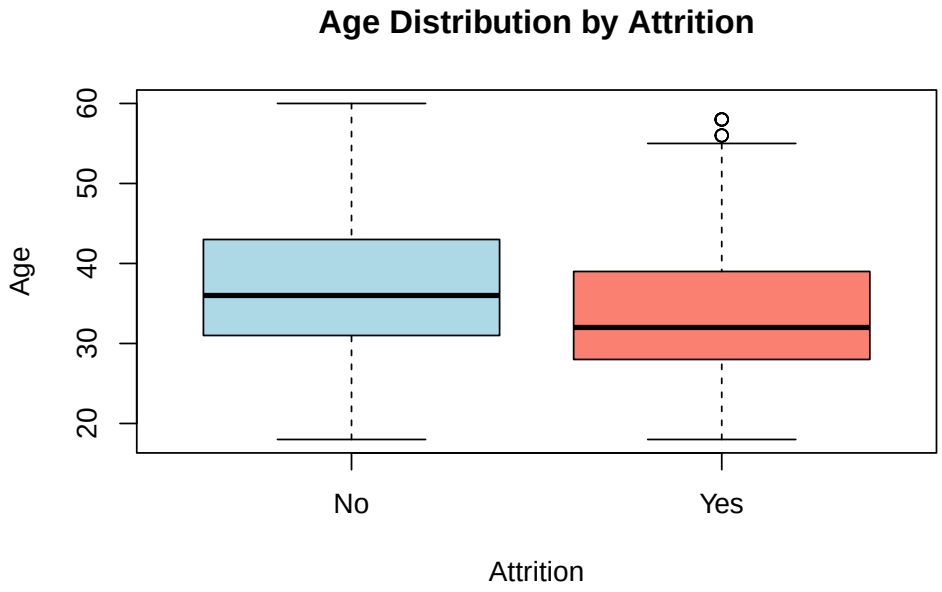


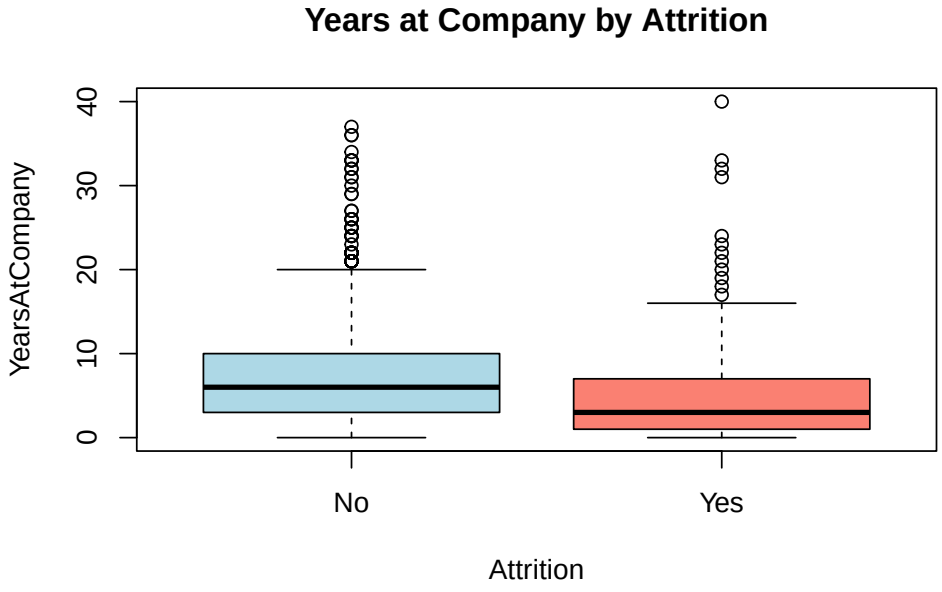
Table 7: Median Income by Attrition Group

	Median_Income
No	5204
Yes	3202

4.3 3.3 Demographics & Attrition (Member 3)

Are younger or newer employees more likely to leave?





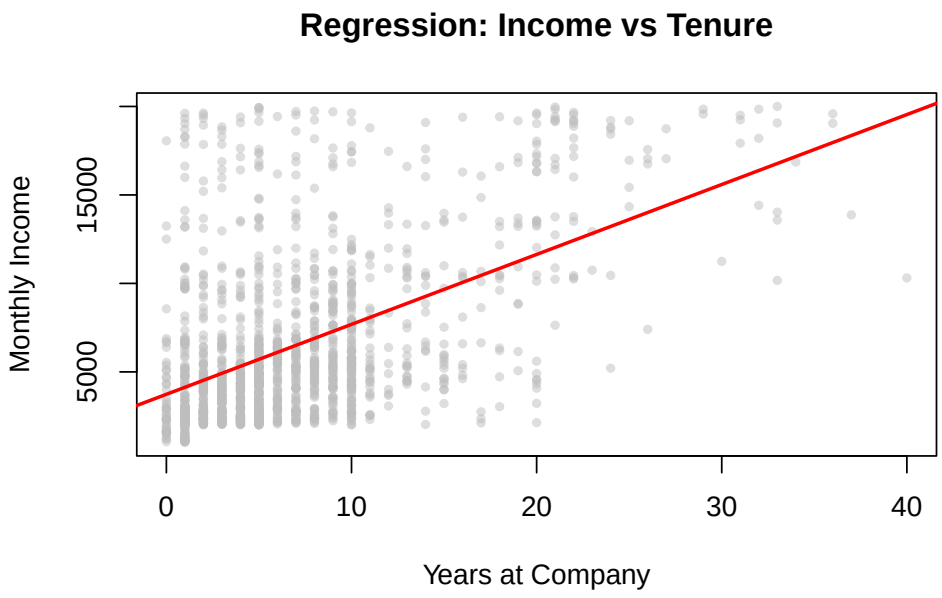
5 4. Linear Regression Analysis (Member 3)

Finally, we model the relationship between Tenure (Predictor) and Income (Response) to understand how compensation grows with time at the company.

Model: $MonthlyIncome = \beta_0 + \beta_1(YearsAtCompany) + \epsilon$

```
##
## Call:
## lm(formula = MonthlyIncome ~ YearsAtCompany, data = hr_sub)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9504  -2499  -1188   1393  15484
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3733.3      160.1   23.32   <2e-16 ***
## YearsAtCompany    395.2       17.2   22.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4039 on 1468 degrees of freedom
## Multiple R-squared:  0.2645, Adjusted R-squared:  0.264
## F-statistic: 527.9 on 1 and 1468 DF,  p-value: < 2.2e-16
```

5.1 Visualization



5.2 Predictions

We predict the expected income for employees with 2, 5, 10, and 20 years of tenure.

Table 8: Income Predictions based on Tenure

Tenure_Years	Predicted_Income
2	4523.68
5	5709.30
10	7685.32
20	11637.36

knitr::kable(round(tab_ot_prop, 2), caption = “Proportion of Attrition within Overtime Groups”)

barplot(tab_ot_attr, beside=TRUE, main=“Attrition by Overtime Status”, legend=rownames(tab_ot_attr), col=c(“lightblue”, “salmon”)) Income and AttritionCode snippet# Boxplot of Income vs Attrition boxplot(MonthlyIncome ~ Attrition, data=hr_sub, main=“Monthly Income by Attrition Status”, col=c(“lightblue”, “salmon”), ylab=“Income”)

6 Stats

inc_by_attr <- tapply(hr_sub$MonthlyIncome, hr_sub$Attrition, median) knitr::kable(data.frame(Median_Income=inc caption=“Median Income by Attrition Group”) 3.3 Demographics & Attrition (Member 3)Are younger or newer employees more likely to leave?Code snippet# Age vs Attrition boxplot(Age ~ Attrition, data=hr_sub, main=“Age Distribution by Attrition”, col=c(“lightblue”, “salmon”))

7 Tenure vs Attrition

boxplot(YearsAtCompany ~ Attrition, data=hr_sub, main=“Years at Company by Attrition”, col=c(“lightblue”, “salmon”)) 4. Linear Regression Analysis (Member 3)Finally, we model the relationship between Tenure (Predictor) and Income (Response) to understand how compensation grows with time at the company.Model: $MonthlyIncome = \beta_0 + \beta_1(YearsAtCompany) + \epsilon$Code snippet# Fit Linear Model lm_model <- lm(MonthlyIncome ~ YearsAtCompany, data = hr_sub) summary(lm_model) Visualization:Code snippetplot(hr_sub$YearsAtCompany, hr_sub$MonthlyIncome, main = “Regression: Income vs Tenure”, xlab = “Years at Company”, ylab = “Monthly Income”, pch = 19, col = alpha(“gray”, 0.5), cex = 0.6) abline(lm_model, col = “red”, lwd = 2) Predictions:We predict the expected income for employees with 2, 5, 10, and 20 years of tenure.Code snippetnew_data <- data.frame(YearsAtCompany = c(2, 5, 10, 20)) preds <- predict(lm_model, newdata = new_data)

pred_results <- data.frame(Tenure_Years = new_data$YearsAtCompany, Predicted_Income = round(preds, 2)) knitr::kable(pred_results, caption = “Income Predictions based on Tenure”)