# IBM HR Employee Attrition Analysis
## Statistical Computing Project

Amaan Shaikh | Danish Ali Shaikh | Rajat Pandey

2025-12-29

## Contents

## 1 Introduction

The IBM HR Employee Attrition dataset serves as a sample of workforce data, organized as a data matrix of observations and qualitative and quantitative variables. This report performs a comprehensive descriptive statistical analysis to summarize the distribution of individual variables using frequencies, measures of location, and measures of dispersion. Bivariate techniques, including contingency tables and correlation coefficients, are applied to investigate the joint behavior and co-dependency between employee attributes and attrition. Furthermore, simple linear regression is utilized to model the linear relationship between tenure and monthly income by estimating the regression coefficients.

### 1.1 Dataset & Data Preparation

We analyze ten key variables classified into the following datatypes :

- **Nominal**: `Attrition`, `OverTime`
- **Ordinal**: `JobLevel` (1-5), `JobSatisfaction` (1-4), `WorkLifeBalance` (1-4)
- **Numeric (Discrete)**: `Age`, `YearsAtCompany`, `TotalWorkingYears`, `NumCompaniesWorked`
- **Numeric (Continuous)**: `MonthlyIncome`

```r
# Load data and select variables
hr <- read.csv("WA_Fn-UseC_-HR-Employee-Attrition.csv", stringsAsFactors = FALSE)
hr_sub <- hr[, c("Attrition", "OverTime", "JobLevel", "JobSatisfaction",
                 "WorkLifeBalance","Age", "YearsAtCompany", "TotalWorkingYears",
```

```
                  "NumCompaniesWorked", "MonthlyIncome")]

# Convert data types
hr_sub$Attrition <- factor(hr_sub$Attrition)
hr_sub$OverTime  <- factor(hr_sub$OverTime)
hr_sub$JobLevel <- factor(hr_sub$JobLevel, levels = 1:5, ordered = TRUE)
hr_sub$JobSatisfaction <- factor(hr_sub$JobSatisfaction, levels = 1:4, ordered = TRUE)
hr_sub$WorkLifeBalance <- factor(hr_sub$WorkLifeBalance, levels = 1:4, ordered = TRUE)
```

## 1.2 Data Summary

Table 1: Dataset Dimensions

| Rows | Columns |
|------|---------|
| 1470 | 10 |

```
## Attrition  OverTime   JobLevel JobSatisfaction WorkLifeBalance       Age
## No :1233   No :1054   1:543    1:289           1: 80           Min.   :18.00
## Yes: 237   Yes: 416   2:534    2:280           2:344           1st Qu.:30.00
##                       3:218    3:442           3:893           Median :36.00
##                       4:106    4:459           4:153           Mean   :36.92
##                       5: 69                                    3rd Qu.:43.00
##                                                                Max.   :60.00
## YearsAtCompany   TotalWorkingYears NumCompaniesWorked MonthlyIncome
## Min.   : 0.000   Min.   : 0.00     Min.   :0.000      Min.   : 1009
## 1st Qu.: 3.000   1st Qu.: 6.00     1st Qu.:1.000      1st Qu.: 2911
## Median : 5.000   Median :10.00     Median :2.000      Median : 4919
## Mean   : 7.008   Mean   :11.28     Mean   :2.693      Mean   : 6503
## 3rd Qu.: 9.000   3rd Qu.:15.00     3rd Qu.:4.000      3rd Qu.: 8379
## Max.   :40.000   Max.   :40.00     Max.   :9.000      Max.   :19999
```

Table 2: First 6 Observations (Transposed View)

|                    | 1    | 2    | 3    | 4    | 5    | 6    |
|--------------------|------|------|------|------|------|------|
| Attrition          | Yes  | No   | Yes  | No   | No   | No   |
| OverTime           | Yes  | No   | Yes  | Yes  | No   | No   |
| JobLevel           | 2    | 2    | 1    | 1    | 1    | 1    |
| JobSatisfaction    | 4    | 2    | 3    | 3    | 2    | 4    |
| WorkLifeBalance    | 1    | 3    | 3    | 3    | 3    | 2    |
| Age                | 41   | 49   | 37   | 33   | 27   | 32   |
| YearsAtCompany     | 6    | 10   | 0    | 8    | 2    | 7    |
| TotalWorkingYears  | 8    | 10   | 7    | 8    | 6    | 8    |
| NumCompaniesWorked | 8    | 1    | 6    | 1    | 9    | 0    |
| MonthlyIncome      | 5993 | 5130 | 2090 | 2909 | 3468 | 3068 |

# 2 Univariate Analysis

Univariate analysis examines each variable individually to understand its distribution, measures of location & measures of dispersion.

## 2.1 Categorical and Ordinal Variables

### 2.1.1 Attrition Status

Here we tabulate the absolute and relative frequencies of employees who left the company versus those who stayed.

Table 3: Attrition Status: Frequency and Proportion

| Attrition | Frequency | Proportion |
|-----------|-----------|------------|
| No | 1233 | 0.839 |
| Yes | 237 | 0.161 |

### 2.1.2 Overtime Status

This shows the distribution of employees working overtime versus standard hours.

Table 4: Overtime Status: Frequency and Proportion

| OverTime | Frequency | Proportion |
|----------|-----------|------------|
| No | 1054 | 0.717 |
| Yes | 416 | 0.283 |

### 2.1.3 3.1.3 Job Level (Ordinal)

The frequency distributions reveal employee composition. The attrition rate indicates what proportion of employees left the company. Job level cumulative frequencies show the hierarchical distribution from entry-level to senior positions.

Table 5: Job Level Distribution (Ordinal)

| Job Level | Frequency | Percentage | Cumulative Frequency | Cumulative % |
|-----------|-----------|------------|----------------------|--------------|
| 1 | 543 | 36.94 | 543 | 36.94 |
| 2 | 534 | 36.33 | 1077 | 73.27 |
| 3 | 218 | 14.83 | 1295 | 88.10 |
| 4 | 106 | 7.21 | 1401 | 95.31 |
| 5 | 69 | 4.69 | 1470 | 100.00 |