

Predicting Employee Attrition Using Statistical Modeling:

An HR Analytics Approach to Workforce Retention



Addressing Employee Turnover: A Critical Business Challenge

Employee turnover significantly impacts organizations, incurring substantial direct and indirect costs. Our goal is to identify key factors and build a predictive model to mitigate these challenges.

Primary Research Question

Identify workplace and demographic factors influencing attrition. Develop a reliable statistical model for high-risk employees.

Business Objectives

- Identify significant turnover factors
- Develop predictive attrition model
- Provide actionable HR insights
- Create proactive decision support tool

Expected Impact

Reduce turnover costs by 15-30% through data-driven retention strategies and early interventions.



Statistical Models for Attrition Prediction

Problem Type

- Binary Classification (Employee Leaves: Yes/No)
- Regression Analysis (Workplace Factor Relationships)
- Comparative Analysis (Group Differences)

Primary Model

- Logistic Regression: Predicts binary outcome (Attrition Yes/No)
- Identifies odds ratios for risk factors

Supporting Models

- Multiple Linear Regression
- ANOVA
- Chi-Square Tests

Analytical Techniques

- Correlation Analysis
- Descriptive Statistics
- Survival Analysis



Deep Dive: Logistic Regression for Attrition



Binary Outcome

Attrition is a clear Yes/No.



Probability Scores

Provides 0-1 risk categorization.



Interpretability

Odds ratios show factor impact.



Multiple Predictors

Handles various independent variables.

$$\text{logit}(P(\text{Attrition} = \text{Yes})) = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{MonthlyIncome}) + \beta_3(\text{JobSatisfaction}) + \beta_4(\text{OverTime}) + \beta_5(\text{WorkLifeBalance}) + \beta_6(\text{YearsAtCompany}) + \beta_7(\text{DistanceFromHome}) + \dots + \varepsilon_i$$

The model integrates various factors, outputting attrition probability and risk classification. Validation focuses on accuracy and minimizing false negatives.



Key Features

Demographic Features

- Age: Employee age ranging from 18 to 60 years, with most employees in the 30-40 age range
- Distance From Home: Commute distance (1-29 units), affecting work-life balance
- Education: Education level on a scale of 1-5, indicating qualification levels
- Education Field: Academic background including Life Sciences (41%), Medical (32%), and other fields

Employment Features

- Department: Three main departments - Research & Development (65%), Sales (30%), and Human Resources (4%)
- Business Travel: Travel frequency with most employees traveling rarely (71%), some frequently (19%), and others with no travel (10%)
- Daily Rate: Daily compensation ranging from \$102 to \$1,499, showing wide salary variation
- Employee Number: Unique identifier for each employee (1-2068)

Target Variable

- Attrition: Binary outcome indicating whether an employee left the company (Yes/No)

Distribution Insights

- Age Distribution: Fairly normal distribution with peak in 30s
- Department Concentration: Heavy concentration in R&D, suggesting a research-focused organization
- Travel Patterns: Most employees have minimal travel requirements
- Salary Range: Wide compensation range indicating diverse roles and seniority levels

Data Quality

- Completeness: Dataset appears complete with no missing values
- Consistency: Categorical variables show consistent formatting
- Employee Count: Constant value of 1 for all records (likely a data artifact)

Analytical Potential /Predictive Modeling

This dataset is well-suited for:

- Classification models to predict employee attrition
- Survival analysis to understand time-to-leave patterns
- Clustering analysis to identify employee segments