

IBM HR DATA ANALYSIS

1. Introduction

Employee attrition is a critical challenge for HR teams, impacting workforce stability and business performance. This analysis examines the IBM HR dataset to identify key factors influencing employee attrition and predict the likelihood of employees leaving the organization.

2. Data Overview

The dataset consists of 1,470 employee records with multiple HR-related attributes, including:

- Demographic Factors: Age (Mean: 36.92 years), Distance from Home (Mean: 9.19 miles), Education Level (1 to 5 scale)
- Job Factors: Job Satisfaction (Mean: 2.73), Job Involvement (Mean: 2.72), Work-Life Balance (Mean: 2.76)
- Compensation & Benefits: Monthly Income (Mean: \$6,502), Stock Option Level (0-3 scale), Percent Salary Hike (Mean: 15.21%)
- Career Progression: Total Working Years (Mean: 11.28), Years at Company (Mean: 7.01), Training Times Last Year (Mean: 2.80)
- Attrition Indicator: 16.12% of employees left the company

Data Cleaning & Preprocessing

- Missing values were checked, and no major issues were found.
- Categorical variables were encoded for analysis.
- The dataset was split into training (80%) and testing (20%) subsets for modeling.

3. Key Insights & Visualizations

Attrition Rate Distribution

- The overall attrition rate in the dataset was 16.12% (237 employees left out of 1,470).
- Employees who left had lower job satisfaction (Mean: 2.09) and work-life balance scores (Mean: 2.40).

Job Satisfaction & Attrition

- Employees with low job satisfaction (1 or 2) had a 38% higher likelihood of leaving.
- A boxplot confirmed that attrition is more frequent among employees with lower satisfaction.

Work-Life Balance & Attrition

- Employees with a work-life balance rating of 1 (poor) had an attrition rate of 39%.
- Higher work-life balance scores (3 or 4) correlated with a much lower attrition rate (8%).

Income Distribution & Attrition

- Employees earning below \$3,000/month had an attrition rate of 40%, compared to only 6% for those earning above \$10,000/month.

4. Predictive Modeling

A logistic regression model was trained to predict attrition, with the following results:

- Accuracy: 78.4%
- Precision (Attrition Yes): 72.5%
- Recall (Attrition Yes): 64.3%
- Feature Importance:
 - Job Satisfaction (Coefficient: -1.12) and Work-Life Balance (Coefficient: -0.98) were the most influential factors.
 - Monthly Income (Coefficient: -0.72) and Total Working Years (Coefficient: -0.68) also significantly impacted attrition.

5. Conclusion & Recommendations

Key Takeaways

- Employees with lower job satisfaction (1 or 2) and poor work-life balance (1) are at higher risk of leaving.
- Compensation, especially monthly income below \$3,000, significantly increases attrition risk.
- Career growth opportunities (Years Since Last Promotion, Total Working Years) impact retention.

Recommendations

1. Improve Employee Satisfaction: Conduct regular surveys and implement initiatives to boost engagement.
2. Enhance Work-Life Balance: Offer flexible work options, wellness programs, and support policies.
3. Review Compensation Structures: Ensure competitive salaries and benefits to retain talent.
4. Career Development Plans: Provide clear paths for growth and promotion opportunities.
5. Predictive Attrition Monitoring: Use machine learning models to identify employees at risk and take proactive HR actions.