

HR Attrition Analysis Report

August 14, 2024

1 Introduction

The purpose of this report is to analyze HR attrition data to understand the key factors contributing to employee turnover. The analysis includes a correlation analysis, data visualizations, and a K-Nearest Neighbors (KNN) classification model to predict the likelihood of attrition based on various factors.

2 Data Summary

The dataset contains 15000 entries with the following columns: `Satisfaction Level`, `Last Evaluation`, `Number of Projects`, `Average Monthly Hours`, `Experience in Company`, `Work Accident`, `Promotion in Last 5 Years`, `Role`, `Salary`, `Left` (target variable).

- `Satisfaction Level`: Employee satisfaction level.
- `Last Evaluation`: Last evaluation score.
- `Number of Projects`: Number of projects handled.
- `Average Monthly Hours`: Average monthly working hours.
- `Experience in Company`: Number of years of experience in the company.
- `Work Accident`: Binary indicator of whether the employee had a work accident.
- `Promotion in Last 5 Years`: Binary indicator of whether the employee was promoted in the last 5 years.
- `Role`: Employee's role in the company.
- `Salary`: Employee's salary level (Low, Medium, High).
- `Left`: Binary indicator of whether the employee left the company.

3 Correlation Analysis

A correlation analysis was performed to understand the relationships between the variables. The correlation matrix is presented below.

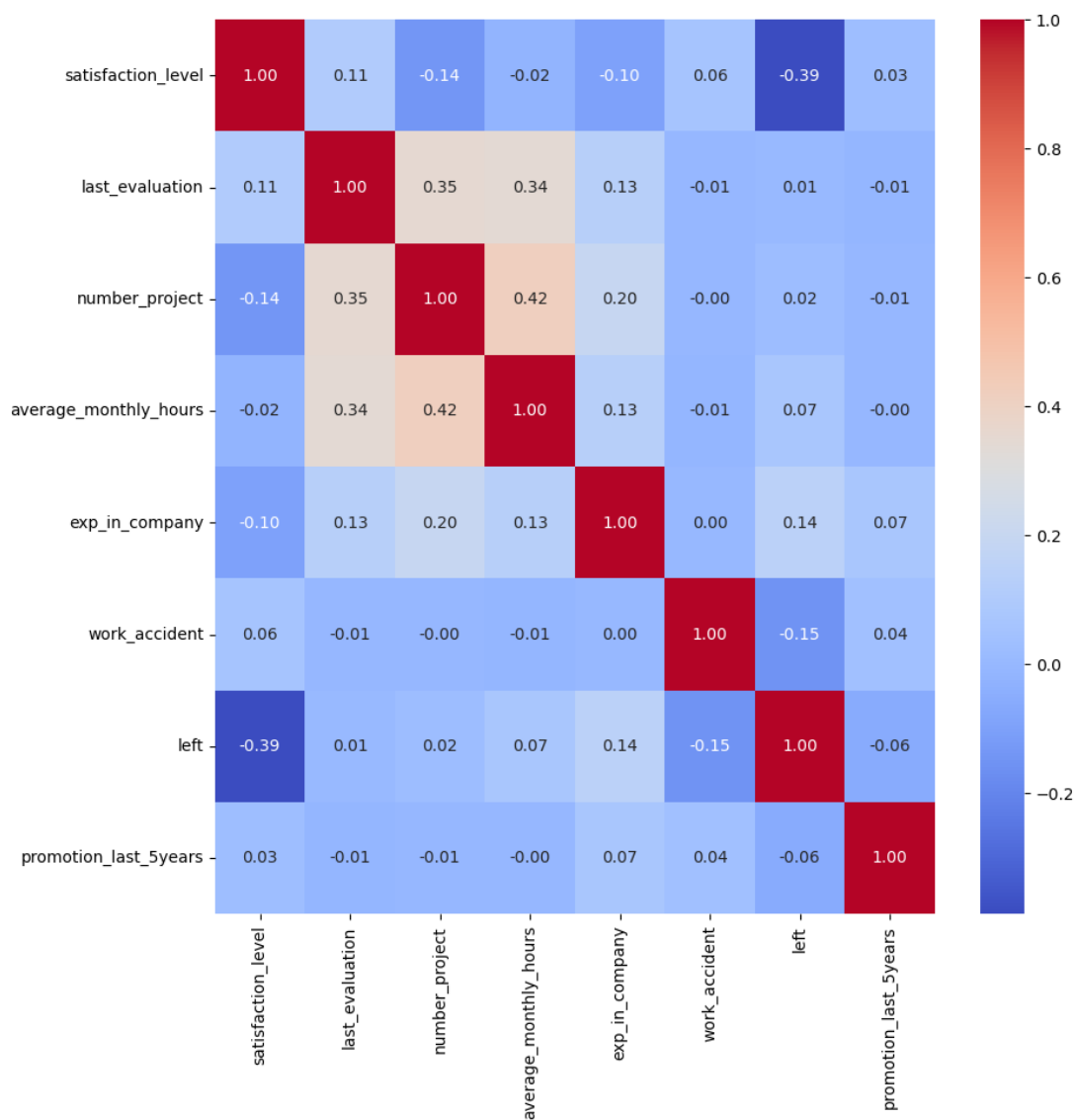


Figure 1: Correlation Heatmap

4 Data Visualization

Several visualizations were created to explore the data.

4.1 Distribution of Roles

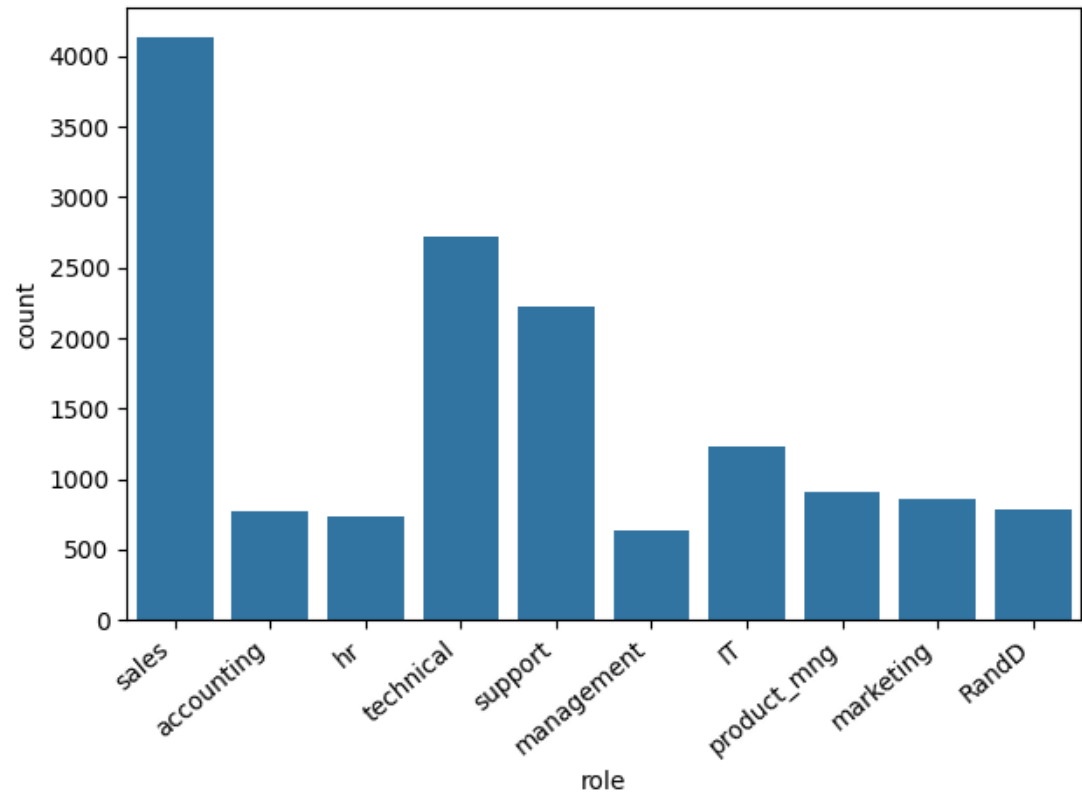


Figure 2: Distribution of Roles in the Company

4.2 Salary Distribution

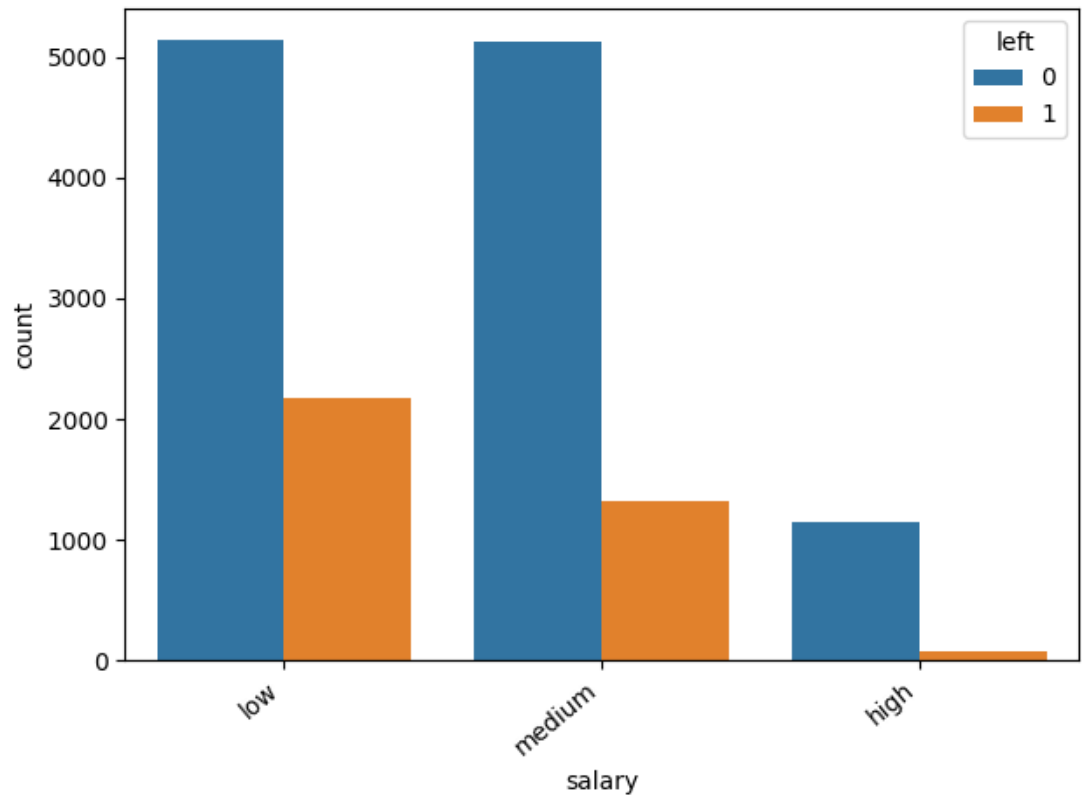


Figure 3: Salary Distribution Across Employees

4.3 Satisfaction Level vs Attrition

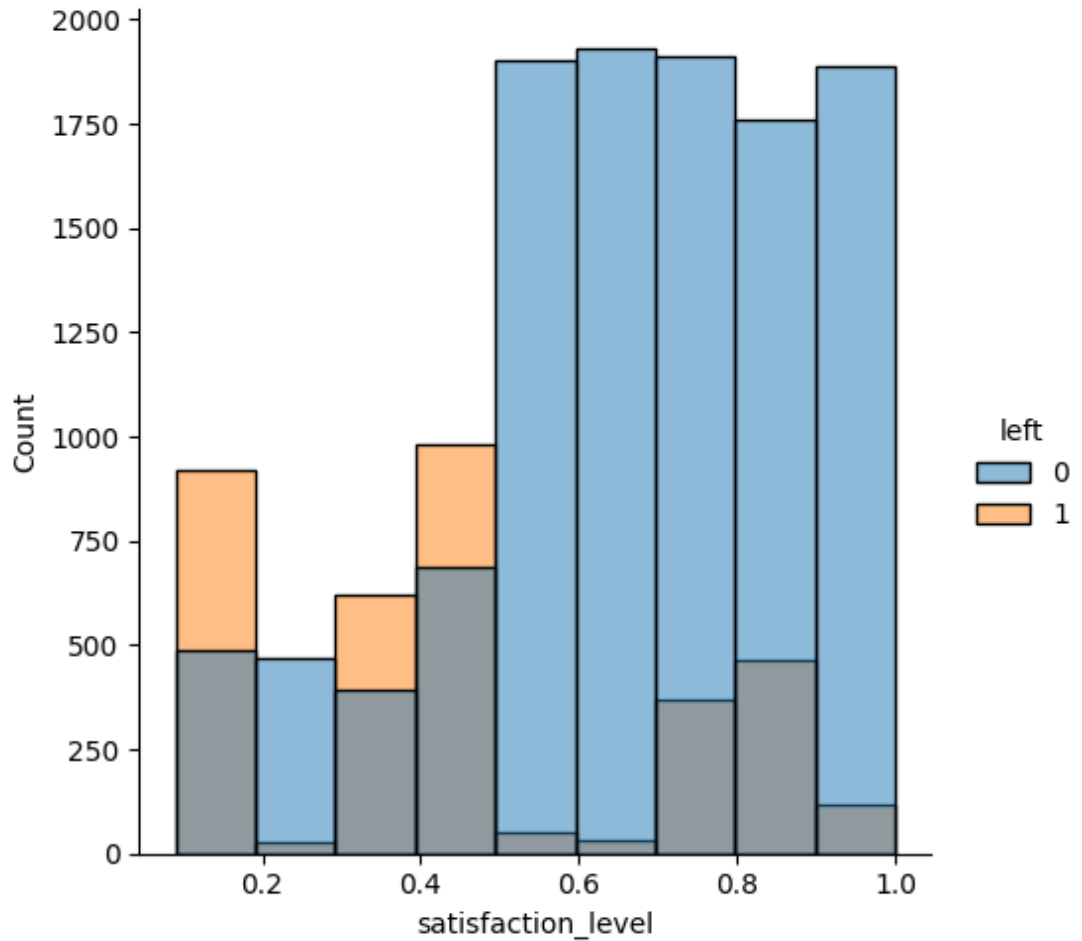


Figure 4: Employee Satisfaction Level vs Attrition

5 K-Nearest Neighbors (KNN) Classification

A K-Nearest Neighbors (KNN) classification model was used to predict employee attrition based on various factors. The KNN algorithm works by finding the 'k' nearest data points in the feature space and making predictions based on the majority class among these neighbors.

5.1 Model Parameters

The KNN model was trained with the following parameters:

- **Number of Neighbors (k):** [INSERT K VALUE]
- **Distance Metric:** Euclidean distance
- **Weight Function:** [Uniform/Distance]

5.2 Model Performance

The performance of the KNN model was evaluated using accuracy, precision, recall, and F1-score. The confusion matrix is presented below.

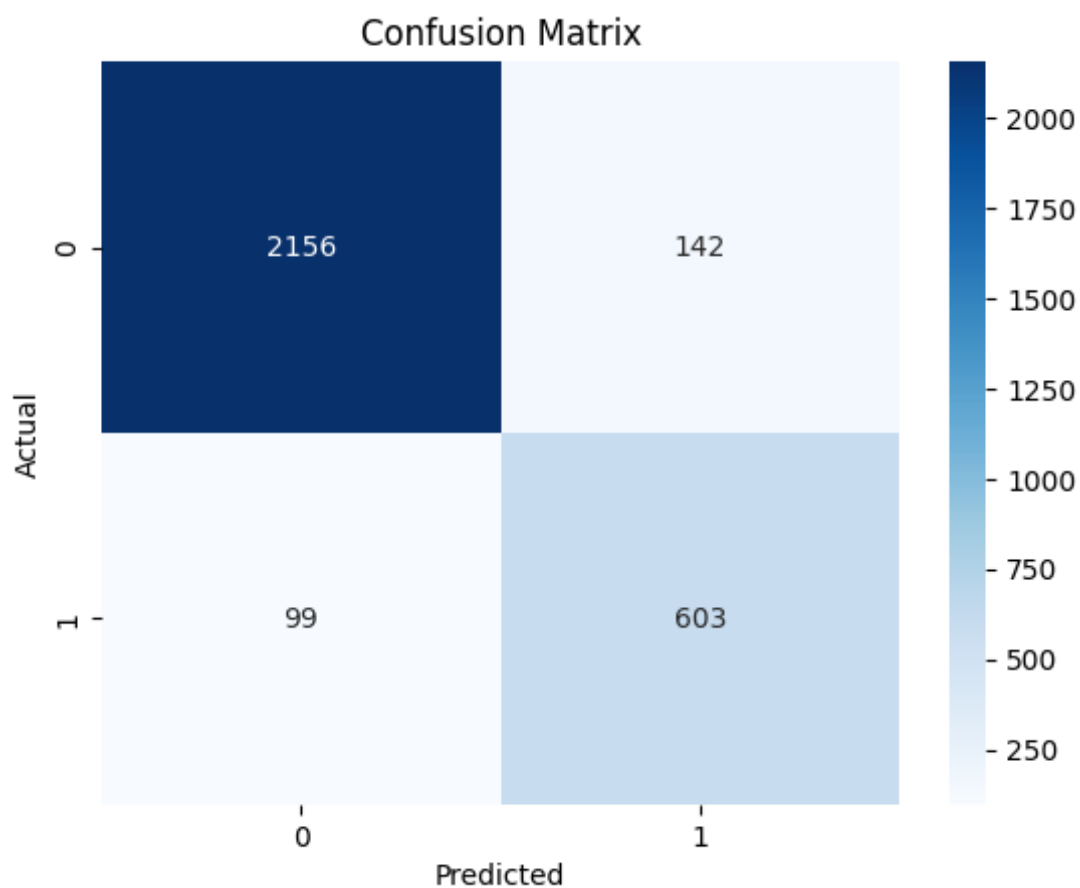


Figure 5: Confusion Matrix for KNN Model

- **Accuracy:** 0.9197
- **Precision:** 0.8094
- **Recall:** 0.8590

- **F1-Score:** 0.8334

6 Conclusion

The analysis reveals that factors such as satisfaction level, last evaluation, number of projects, and salary level significantly impact employee attrition. The KNN model provides a reliable method for predicting employee attrition, with an accuracy of 91.97%. These findings suggest that HR departments can use KNN to proactively identify employees at risk of leaving and take appropriate measures to improve retention.