# Employee Retention: Actionable Insights

-RIA SINGH

Business Objective

A mid-sized technology company wants to improve its understanding of employee retention to foster a loyal and committed workforce. While the organization has traditionally focused on addressing turnover, it recognizes the value of proactively identifying employees likely to stay and understanding the factors contributing to their loyalty.

Strategy

- Import the data

- Perform data cleaning by removing duplicates, null values and outliers

- Perform train test split

- Perform EDA To identify how different features impact employee retention

- Apply feature encoding and scaling

- Use RFE to select most important features

- Build logistic regression model

- Determine optimal threshold to build final model

- Apply the final model to validation data

- Calculate metrics like accuracy, precision and recall
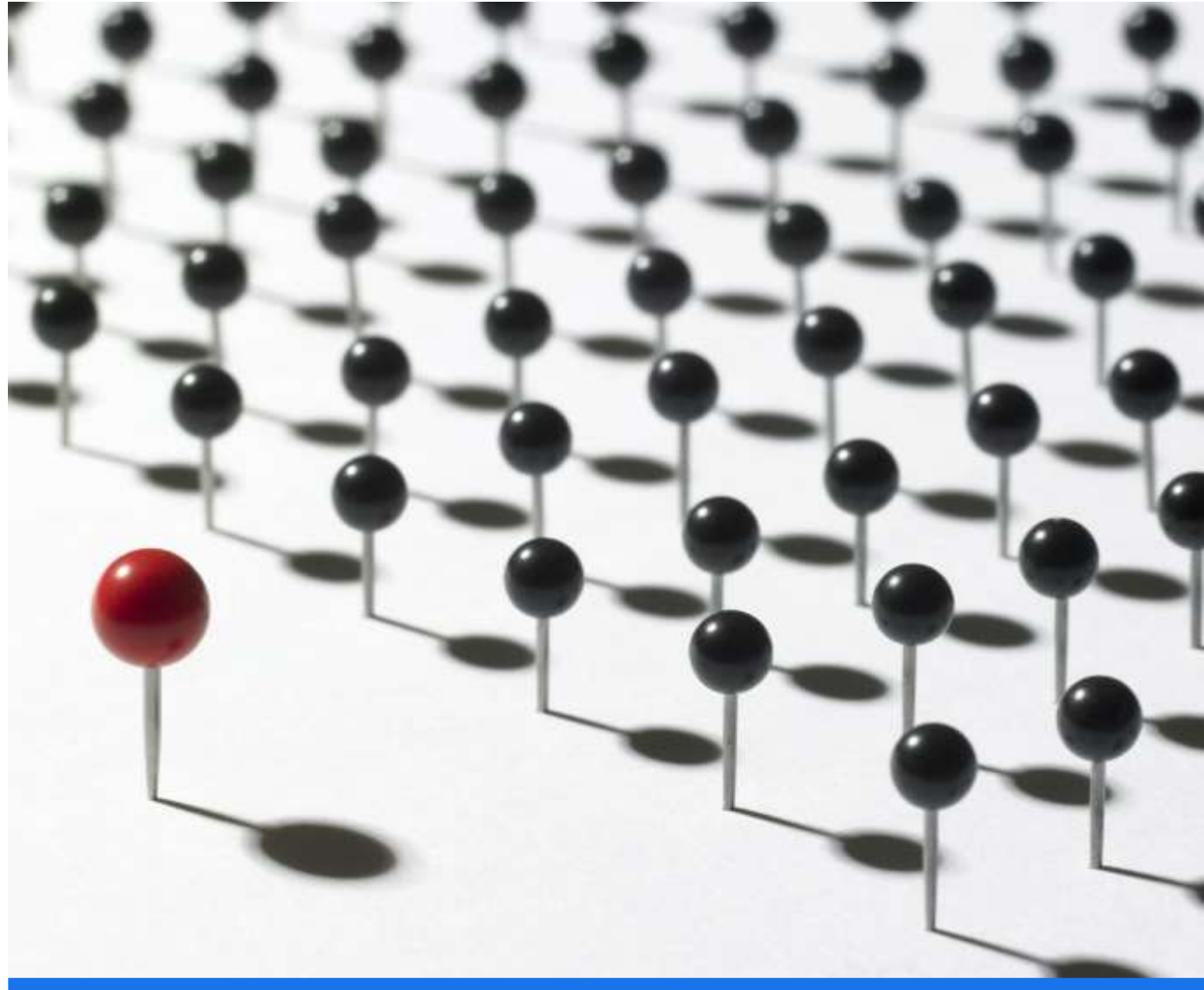
Exploratory Data Analysis (Train Data)
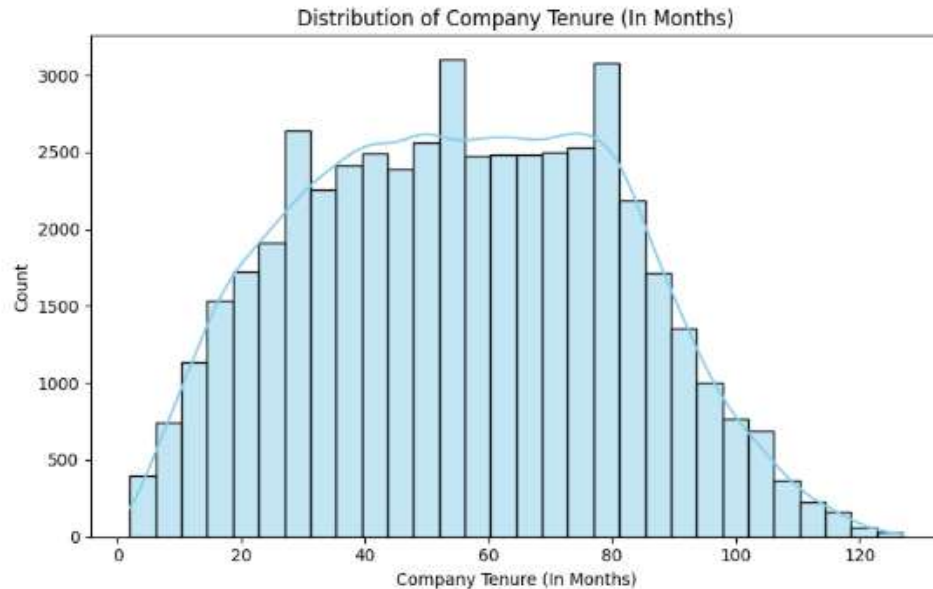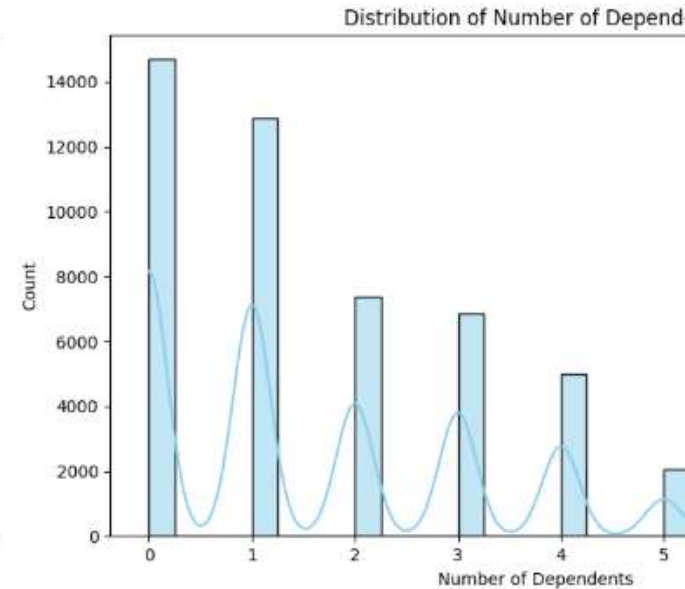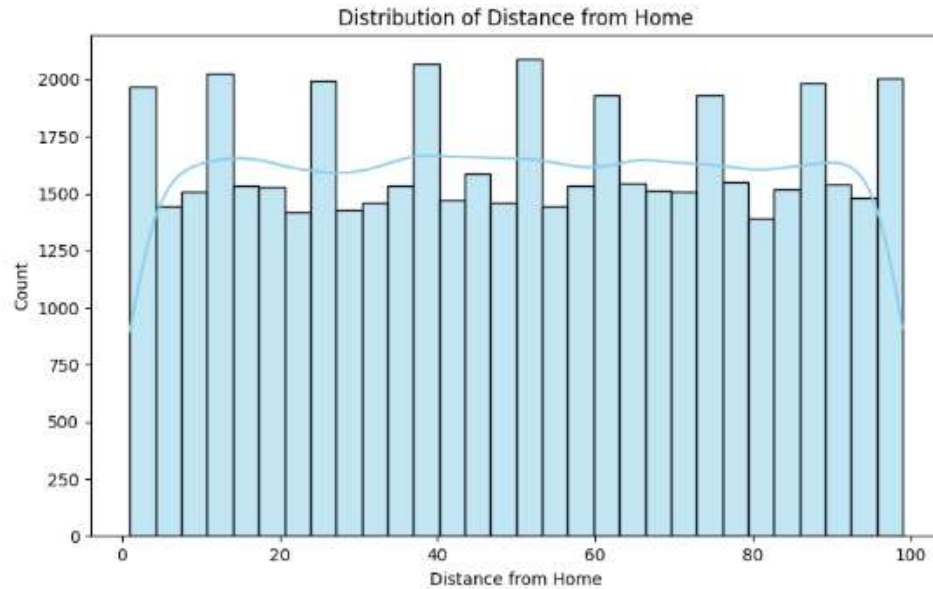
Univariate Analysis
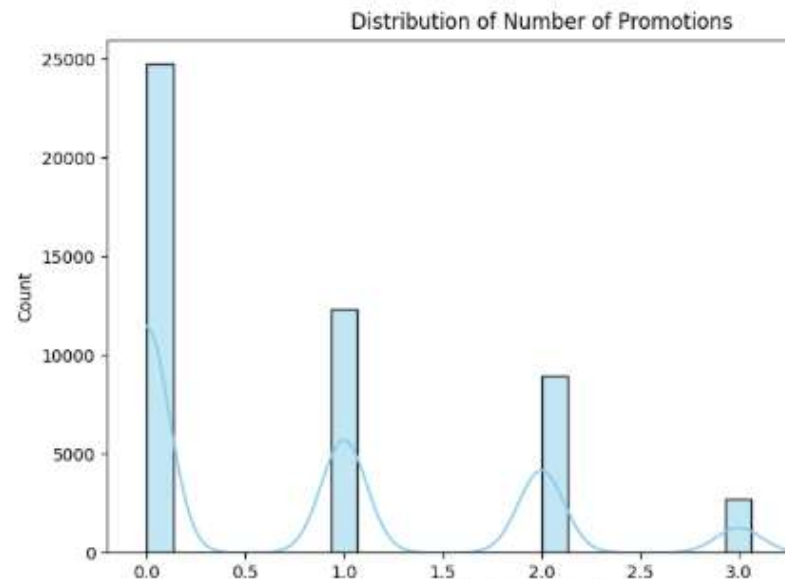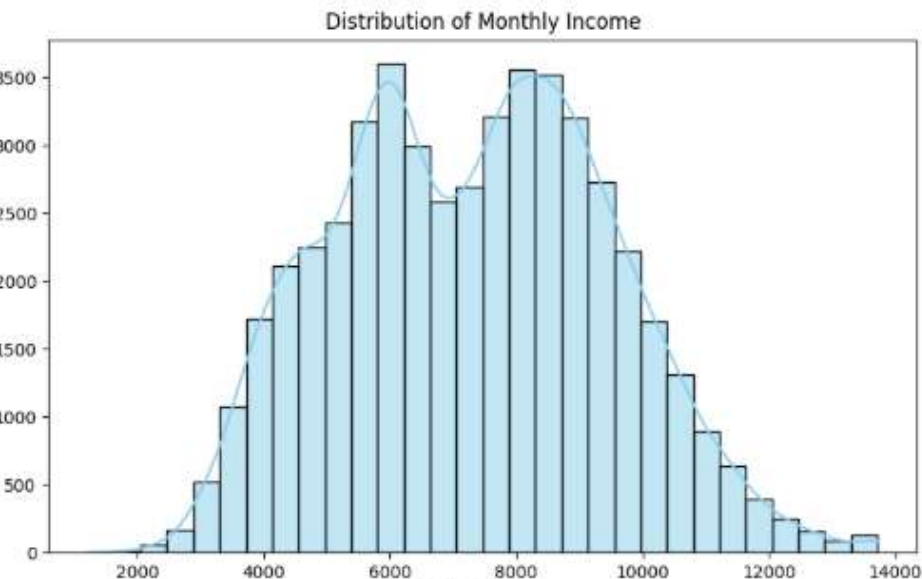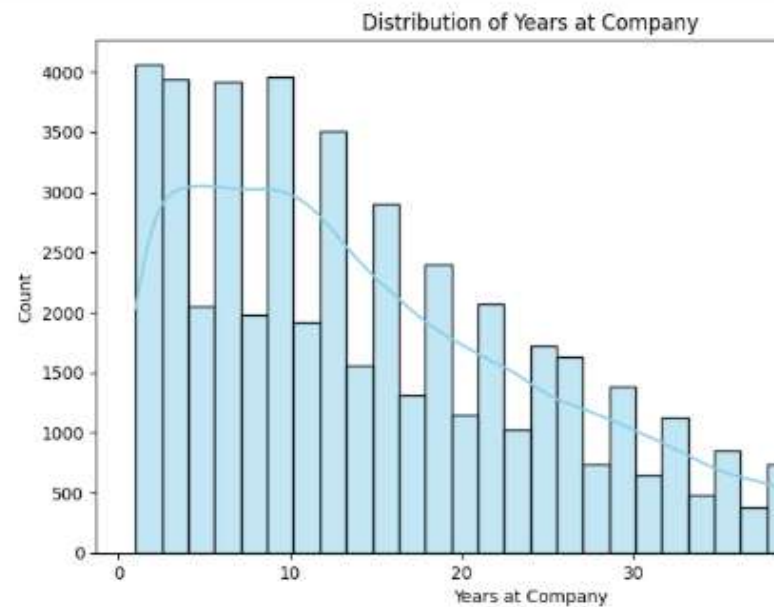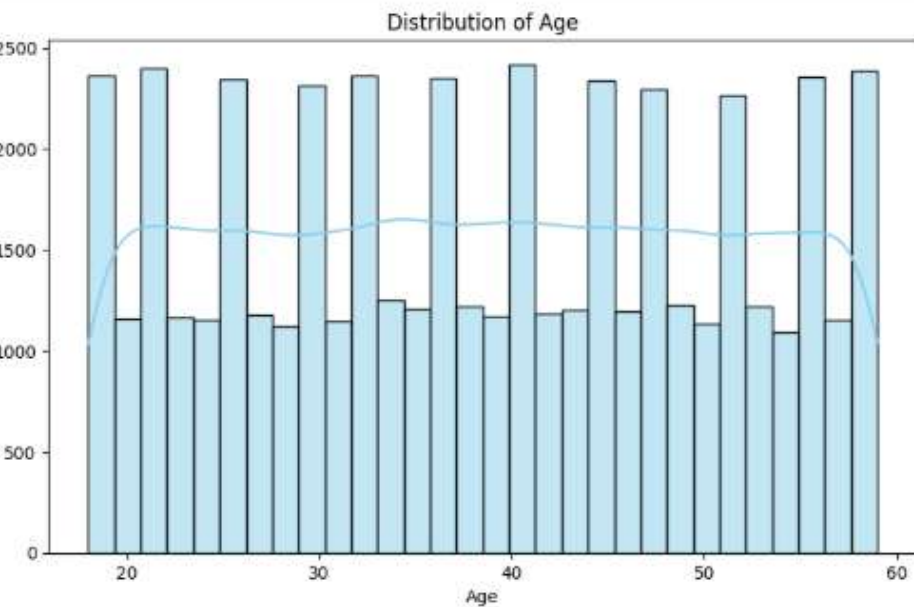
- The age distribution is relatively even across the data.

- Majority of the employees have been with the company for 0-10 years.

- Most employees fall within the income range of 6000-8000, indicating high concentration around mid-level salaries.

- Majority of employees have zero promotions, which may suggest limited career progression.

Distribution of Distance from Home

Distribution of Number of Dependents

Distribution of Company Tenure (In Months)

- The distances from home is spread evenly for the employees.

- Most employees have either 0 or 1 dependents, showing they have fewer familial responsibilities.

- Company tenure peaks around 60 and 80 months, suggesting that significant number of employees stay in the company for 5-6.65 years.
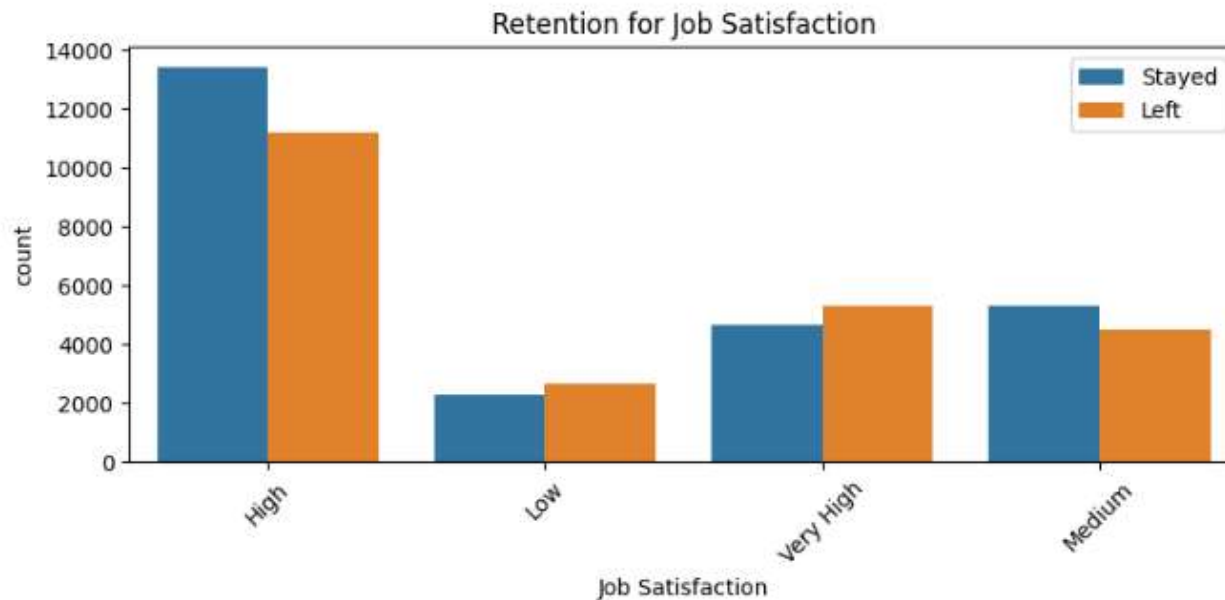
Bivariate Analysis

- Male employees show high retention rate compared to female employees
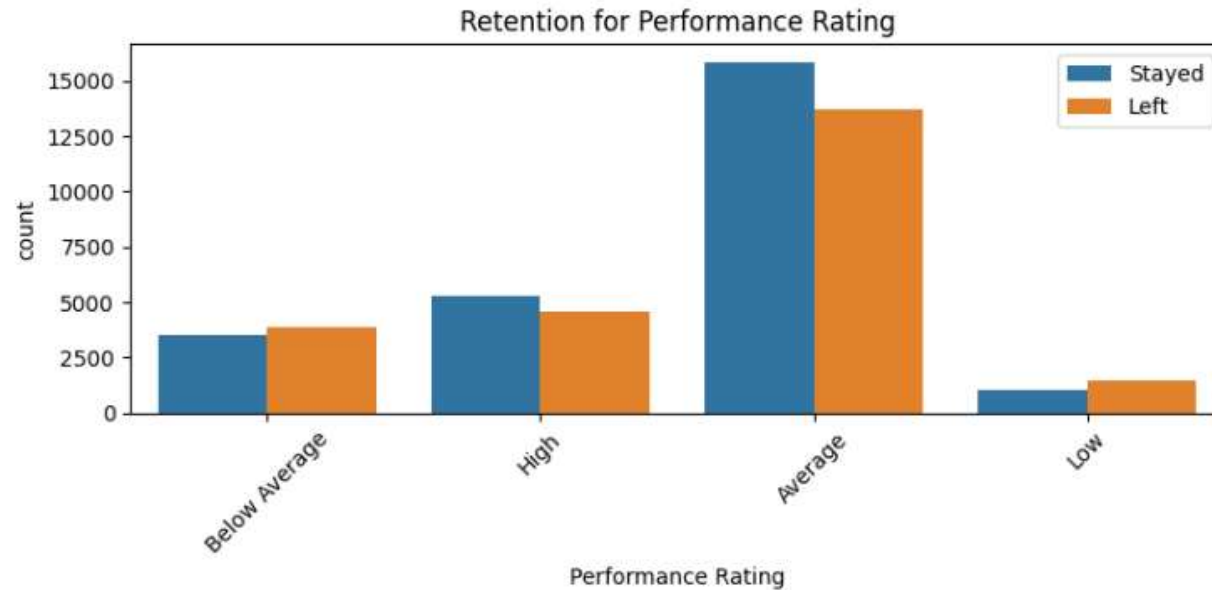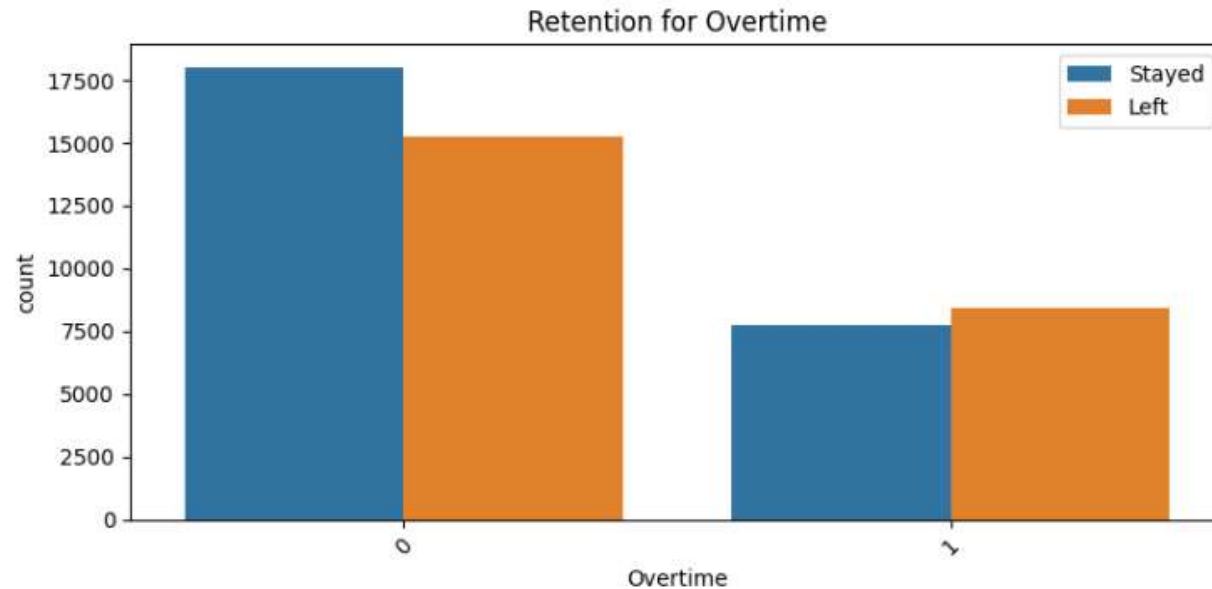
- Finance and technology sectors have high retention rate

Retention for Work-Life Balance

Retention for Job Satisfaction
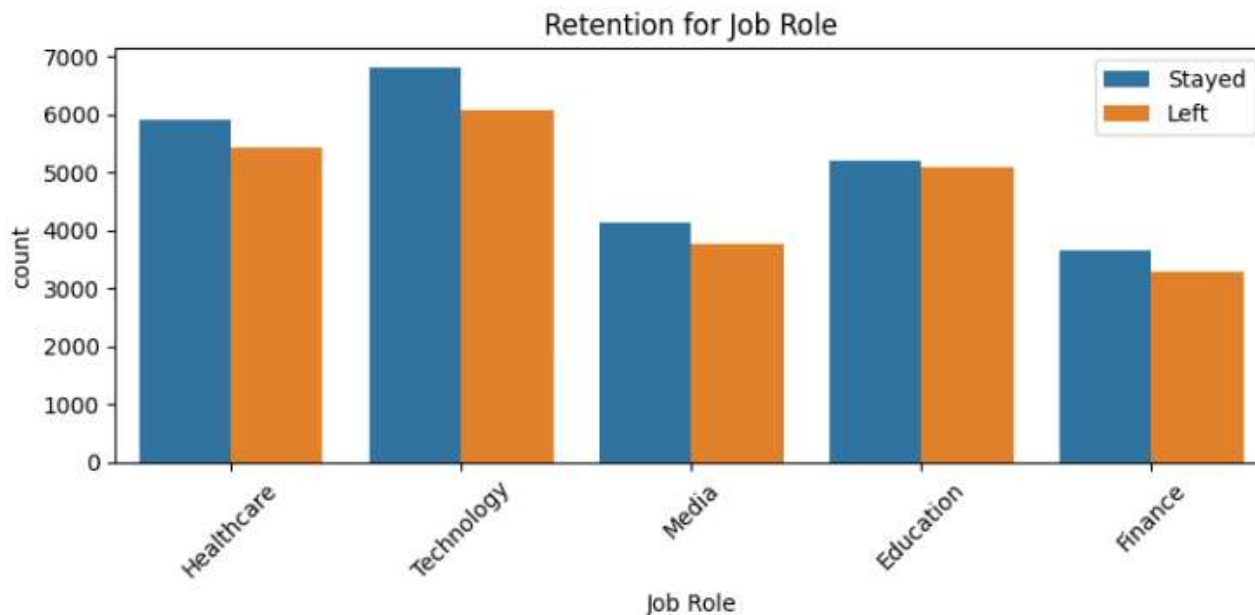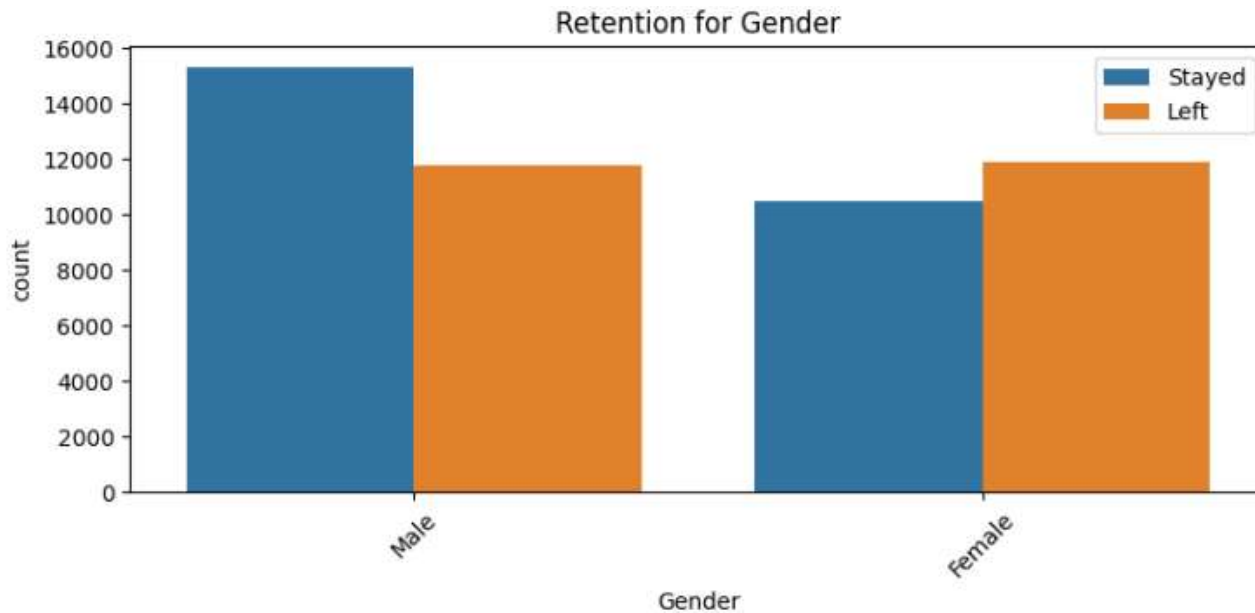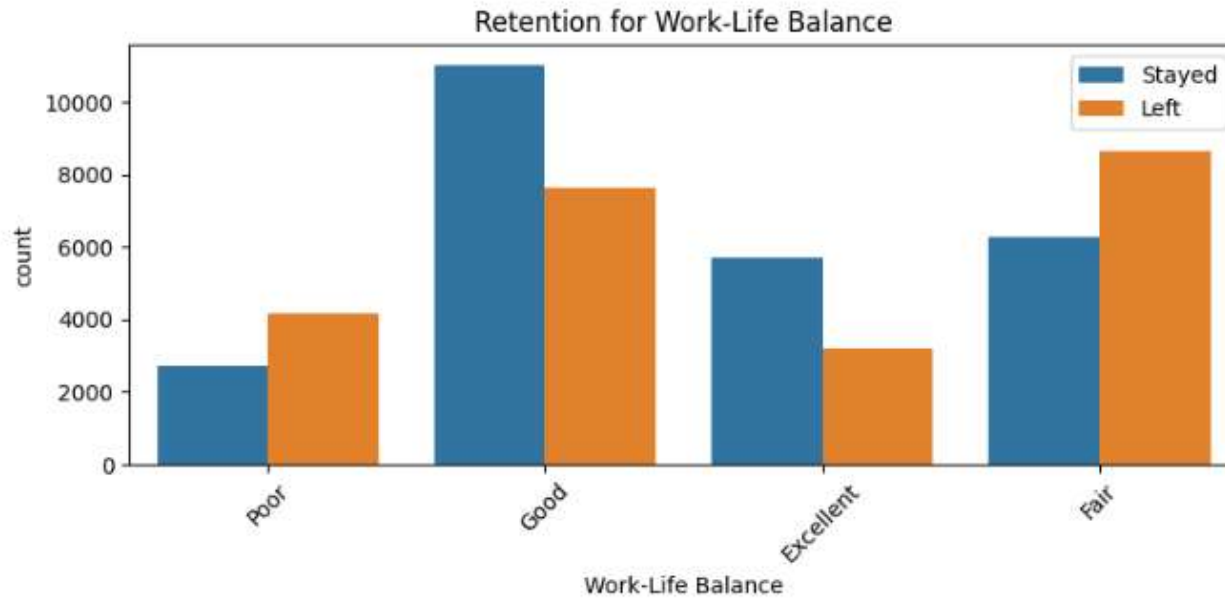
- Employees with the good work-life balance tend to stay

- Employees with high satisfaction tend to stay the most
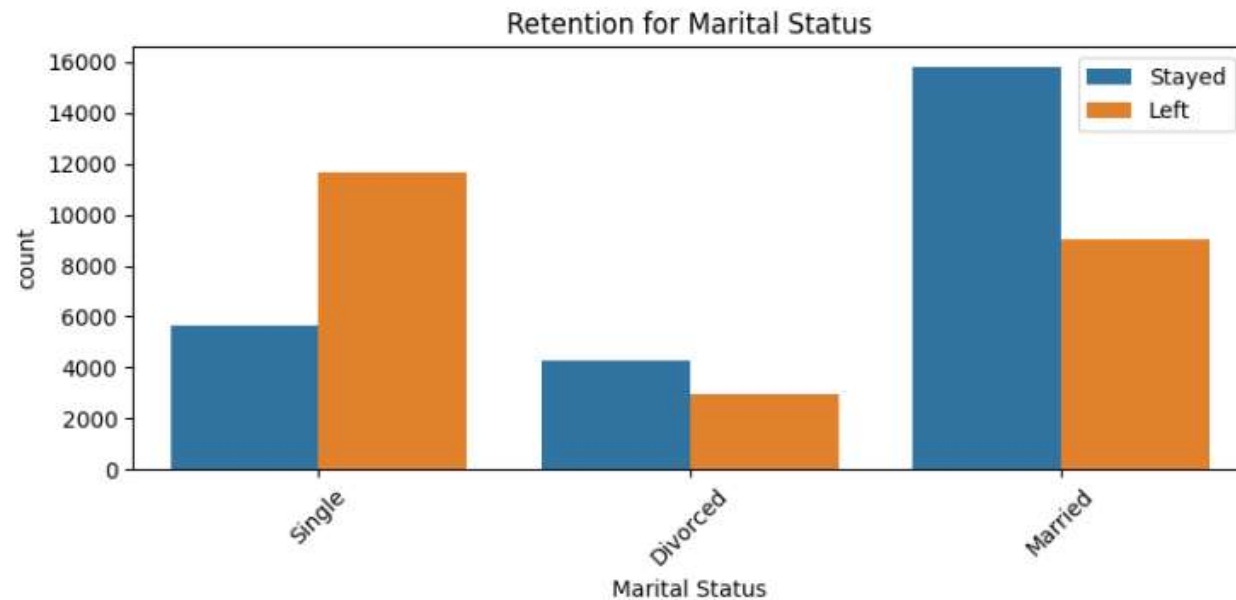
Retention for Performance Rating

Retention for Overtime

- Retention is more for average and high-performance ratings

- Higher retention is observed among the employees who don't work overtime
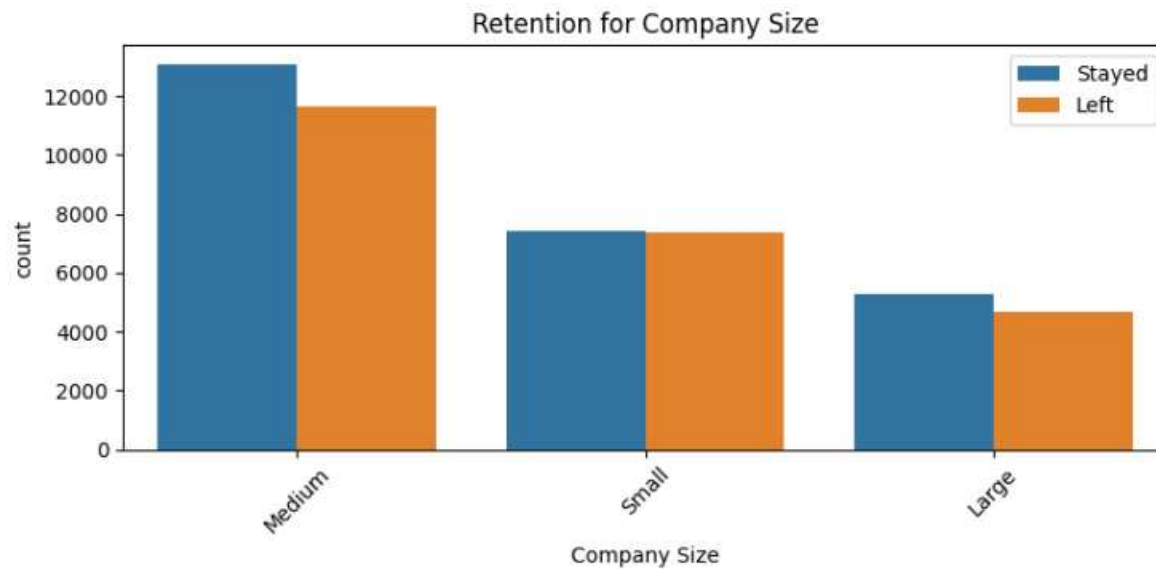
Retention for Education Level

Retention for Marital Status
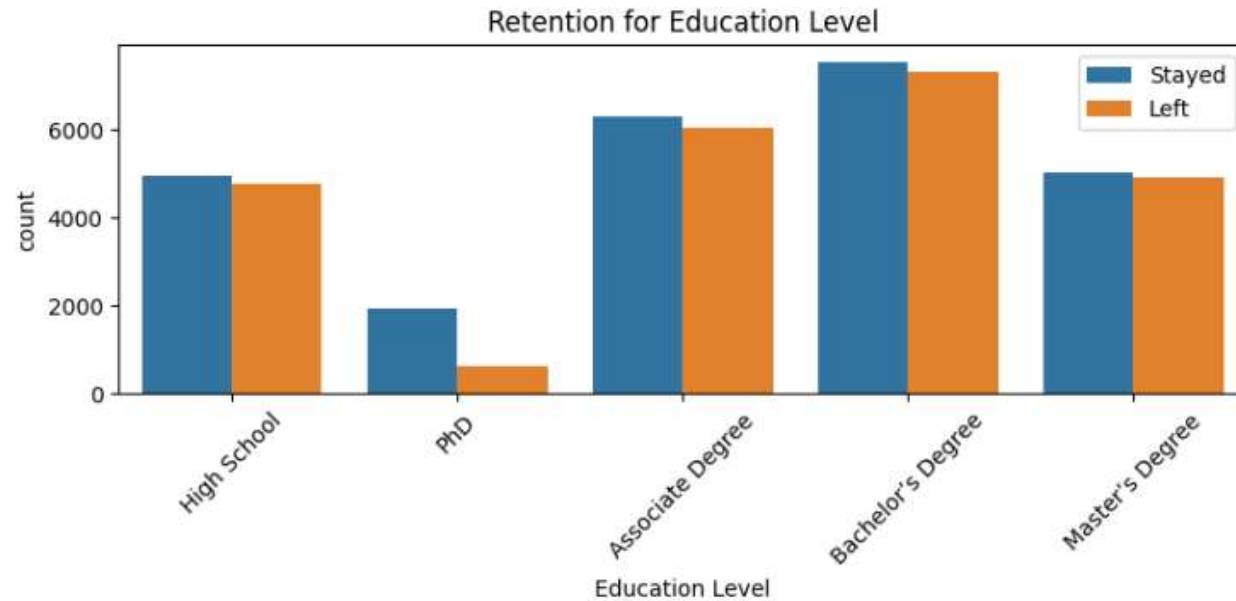
- Employees with PHD are least likely to leave, showing strong retention rate

- Married employees are most likely to stay
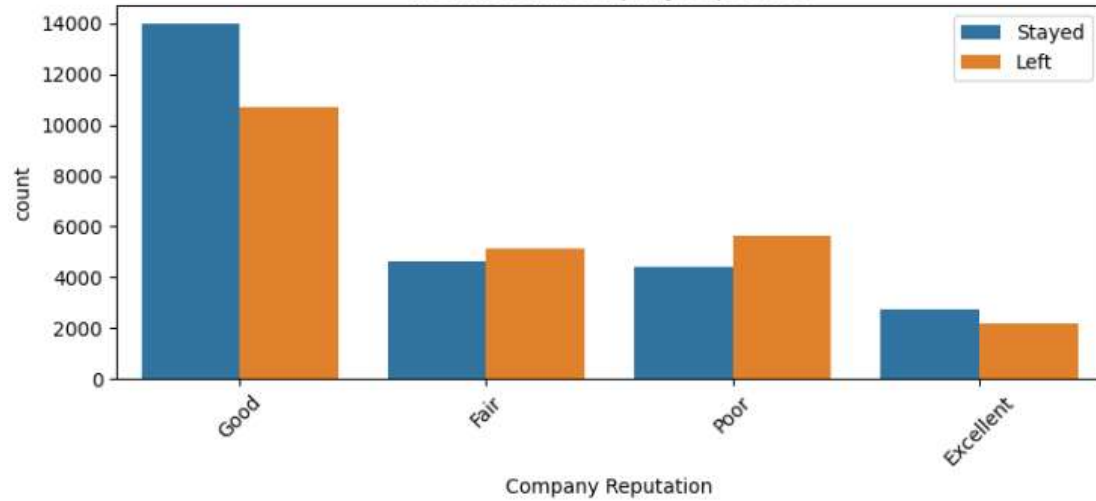
Retention for Job Level

Retention for Company Size
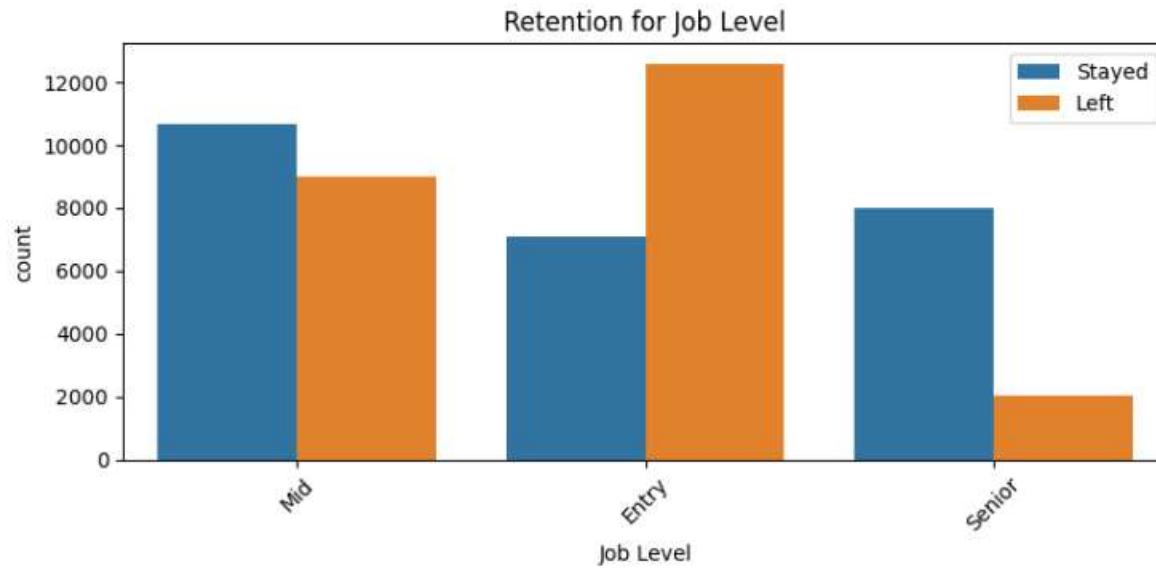
- Senior employees show high retention

- High retention is observed among medium sized companies

Retention for Company Reputation

Retention for Remote Work

- Companies with good reputation are more likely to retain employees
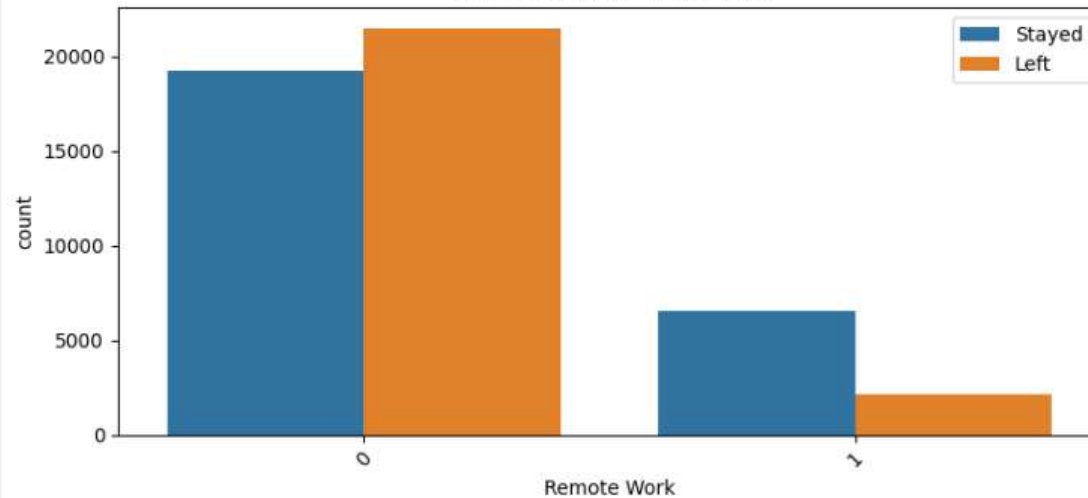
- Employees with remote work show high retention

Model Building and Evaluation

___

Receiver operating characteristic example
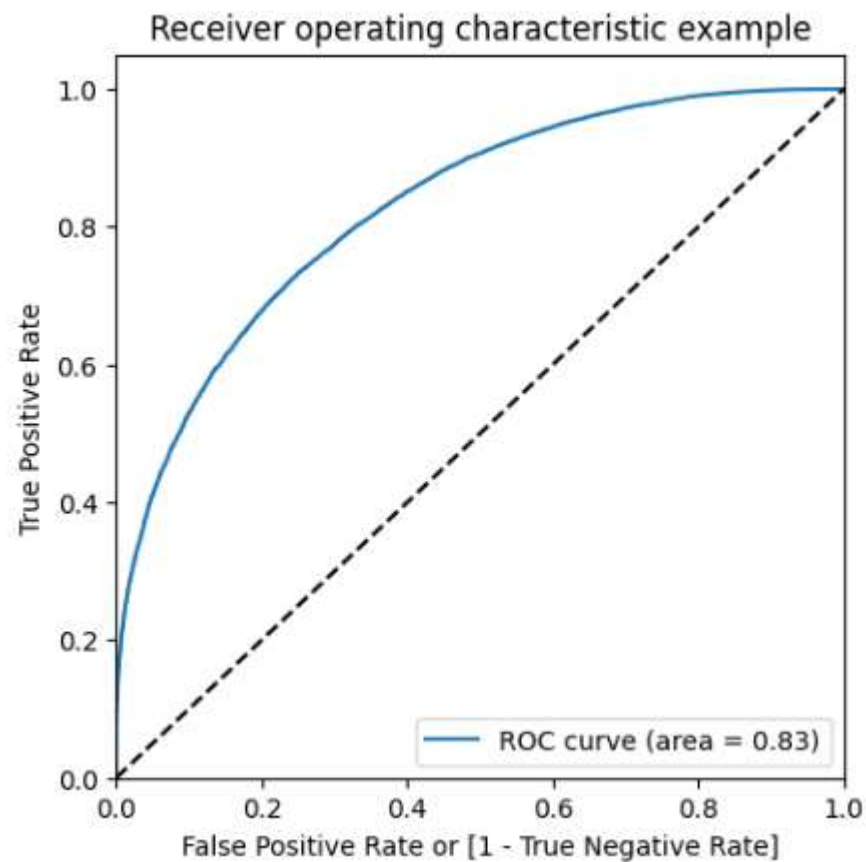
Area under the ROC curve is 0.83
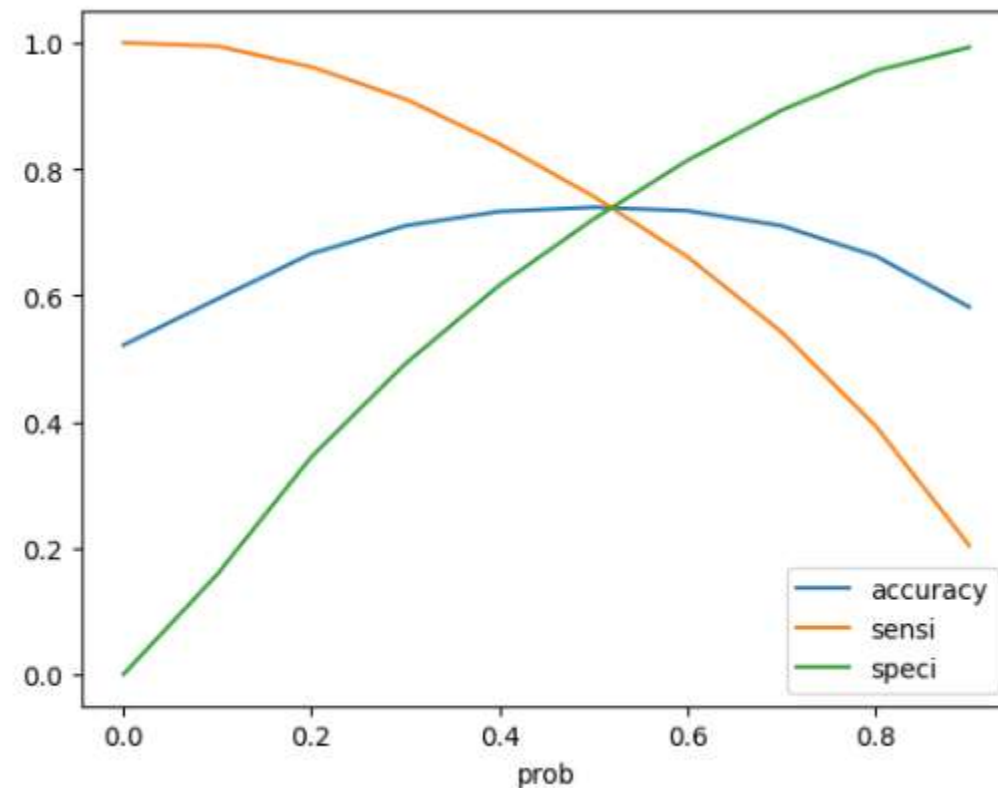
The optimal threshold is 0.54

Training Data Metrics

Accuracy 74.04	**Precision** 76.62
Recall 72.26	**Specificity** 75.99
Sensitivity 72.26	

Validation Data Metrics

Accuracy 73.74	**Precision** 76.46
Recall 71.66	**Specificity** 76.02
Sensitivity 71.66	

Conclusion

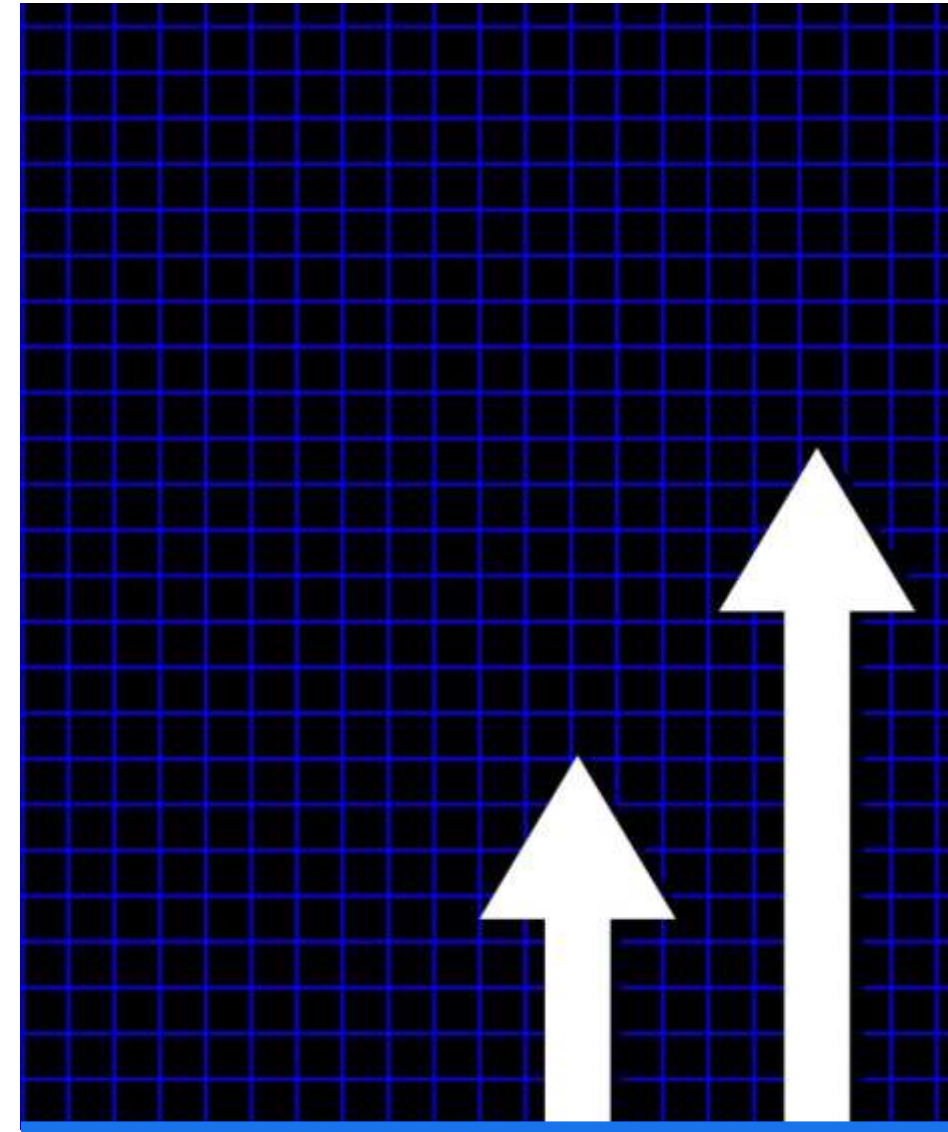- The accuracy of the model on training data is 74.04% and validation data is 73.74% which indicates that model generalizes well to unseen data

- The recall on training data is 72.26% and validation data is 71.66% which means model is predicting large number of the actual positive cases consistently

Senior employees are most likely to stay so the company shall offer more leadership and growth opportunities.

Company should offer more remote/hybrid options as remote employees are more loyal.

Good work-life balance is a major factor in employee retention so the company should invest in flexible hours and wellness benefits.

Employees who have been promoted and who have dependents stay longer. Company should offer clear promotion pathways and family friendly policies.

High performers tend to stay longer. Company can incentivize high performance and offer perks and recognitions.

Married employees are more loyal. Company can offer spousal insurance and benefits to reward them.

THE END