

# Employee Retention — Holistic Summary Report

Source notebook:

/mnt/data/Predicting\_Employee\_Retention\_Rosy\_Samantaray\_Khushpreet\_Prity\_Raaz.ipynb

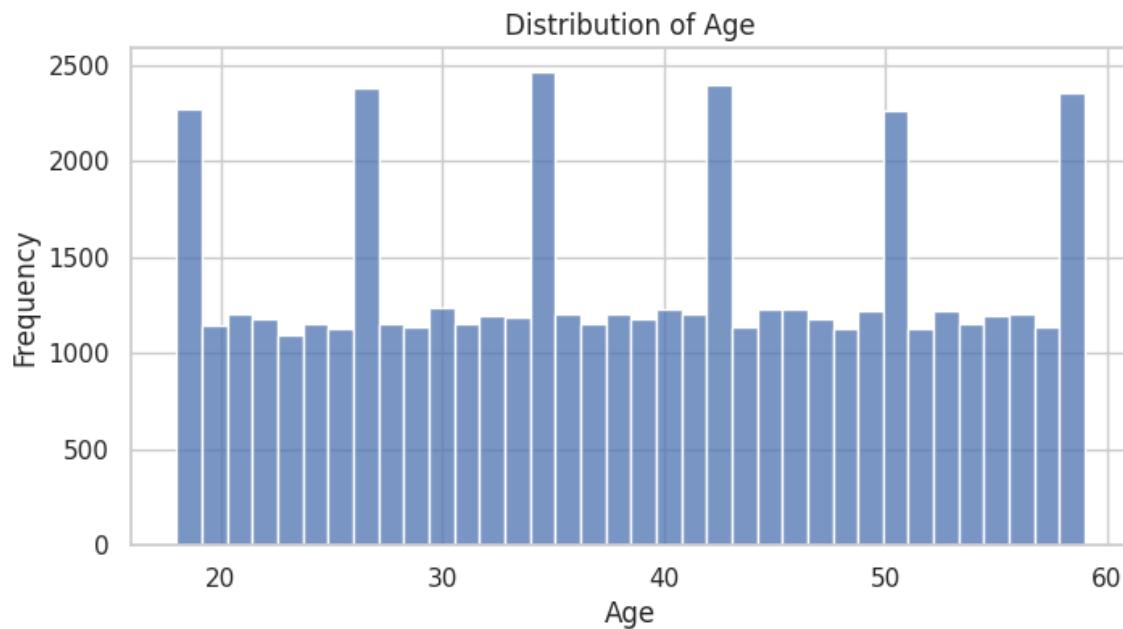
This report summarizes the end-to-end analysis for predicting employee attrition. It includes data understanding, cleaning, exploratory data analysis (EDA), feature engineering, model building, evaluation, and next steps. Key figures and visualizations from the notebook are embedded below.

## Key results (at a glance):

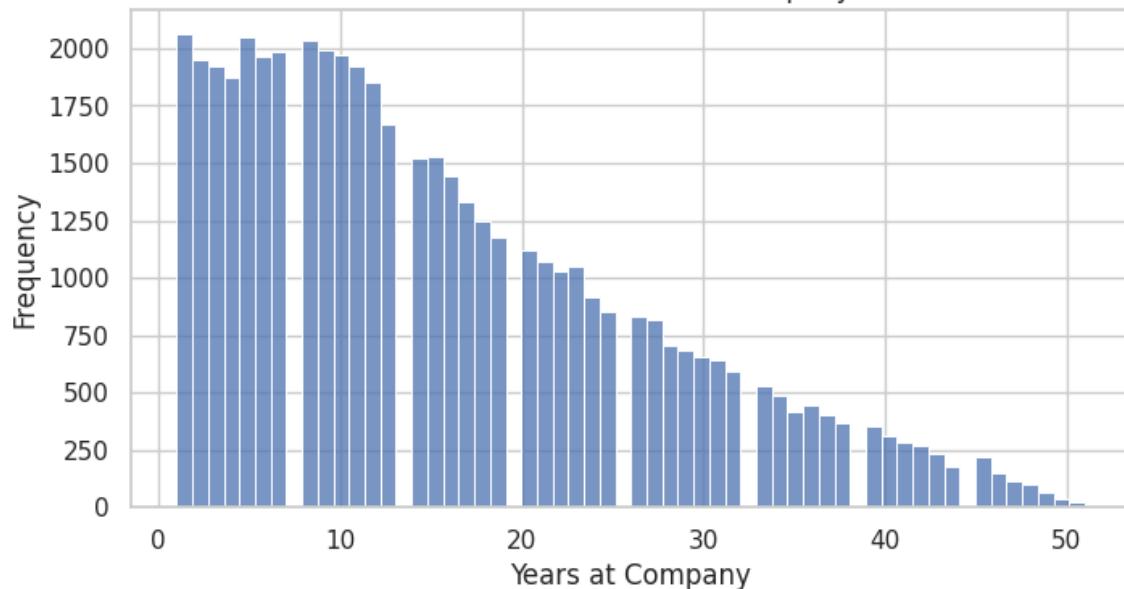
- Auto-detected target: Attrition
- Models trained: Logistic Regression, Random Forest (200 trees)
- Logistic Regression (test): Accuracy = 0.7526, AUC = 0.8425
- Random Forest (test): Accuracy = 0.7500, AUC = 0.8380

## Top feature importances (Random Forest):

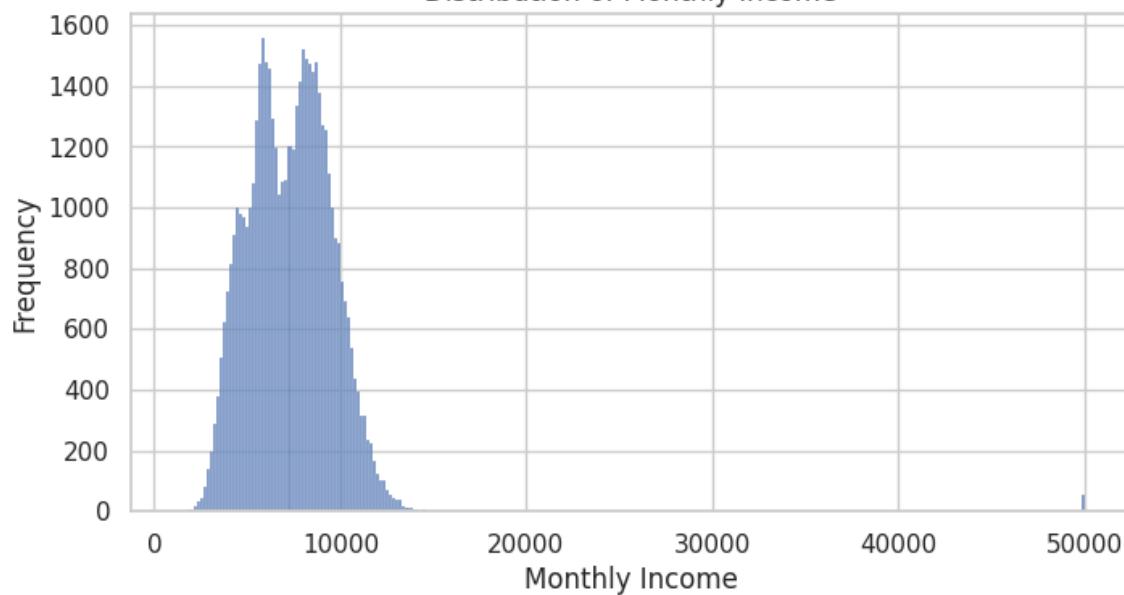
Distance from Home, Employee ID, Monthly Income, Company Tenure, Years at Company, Age, Marital Status\_\_Single, Job Level\_\_Senior, Job Level\_\_Entry, Number of Dependents

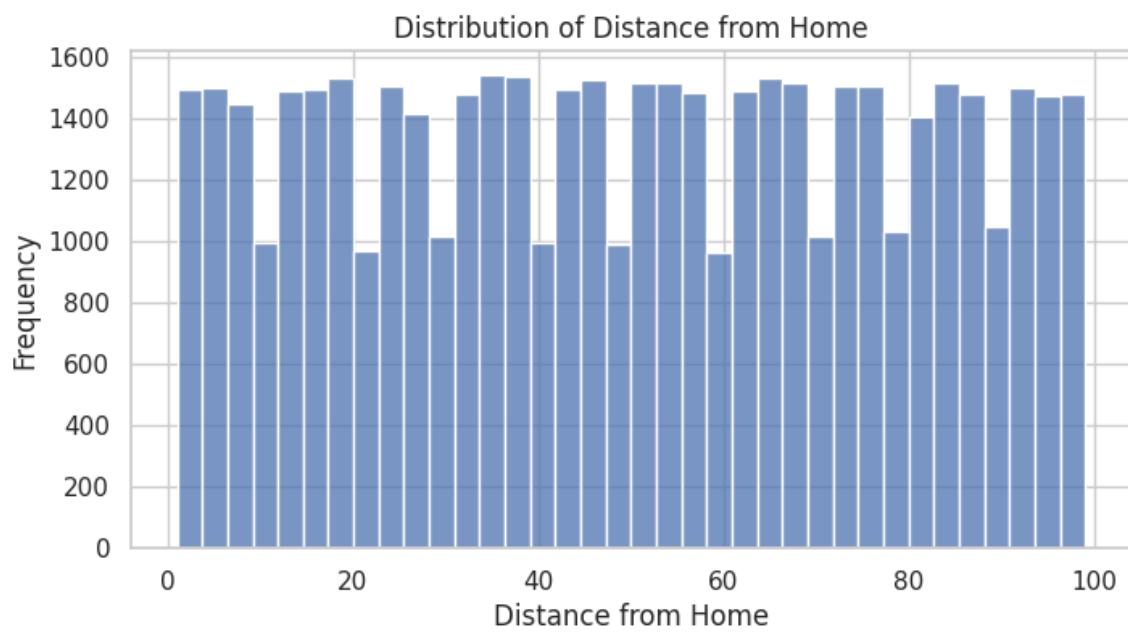
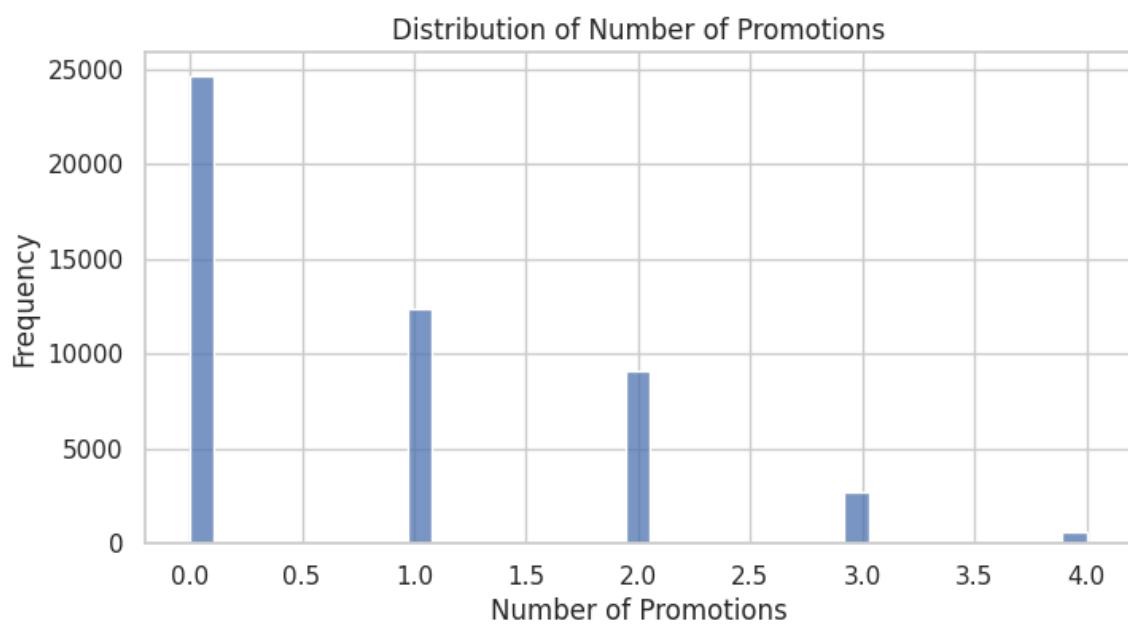


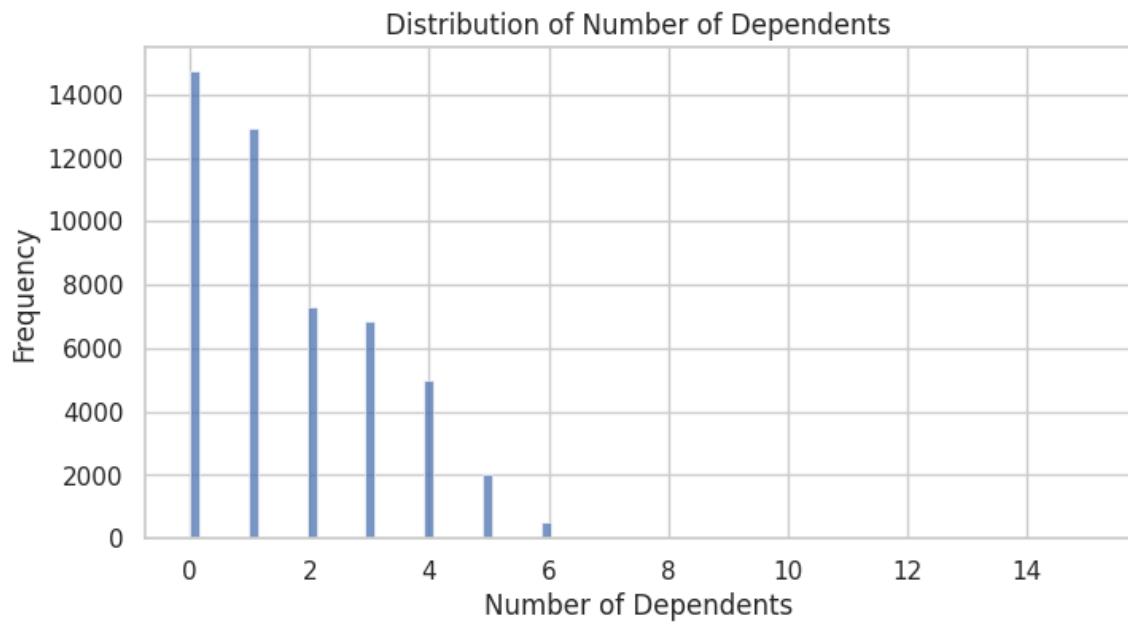
Distribution of Years at Company



Distribution of Monthly Income







## Recommendations & Next Steps

1. Perform hyperparameter tuning (GridSearchCV) for both models.
2. Use cross-validation for more reliable estimates.
3. Explore class imbalance handling (SMOTE or class\_weight) if target is imbalanced.
4. Create interpretable model explanations (SHAP or LIME) for business stakeholders.
5. Deploy the model as a monitoring tool to flag at-risk employees for HR interventions.