

Employee Turnover Analytics

Course-end Project 3

Comprehensive Analysis Report

Student: Tatpar Mishra

Course: Data Science/ML Course

Date: July 2025

Project: Simplilearn: Employee Turnover Analytics

Executive Summary

This project analyzes employee turnover data for Portobello Tech. The analysis identifies key factors contributing to turnover, clusters employees who left, builds predictive models, and provides actionable retention strategies for HR.

Key Objectives Achieved:

- Data quality checks and missing value analysis
- EDA to identify main turnover factors
- Clustering of employees who left
- Class imbalance handling with SMOTE
- K-fold cross-validation and model evaluation
- Best model selection and metric justification
- Targeted retention strategy recommendations

1. Introduction

1.1 Project Background

Portobello Tech periodically evaluates employees' work details, including satisfaction, evaluation, projects, hours, tenure, promotions, and salary. The HR department uses this data to predict and reduce employee turnover, which is costly and disruptive.

1.2 Dataset Overview

Records	14,999 employees
Features	satisfaction_level, last_evaluation, number_project, average_monthly_hours, time_s...
Target	left (1 = left, 0 = stayed)
Turnover Rate	23.81%

2. Data Quality Checks

- No missing values found in the dataset
- Data types are appropriate for analysis
- 3,008 duplicate rows detected (retained for analysis)
- Target variable distribution: 11,428 stayed, 3,571 left

3. Exploratory Data Analysis (EDA)

Key factors contributing to turnover:

- Employees who left had much lower satisfaction (avg: 0.44) than those who stayed (avg: 0.67)
- Overworked employees (>250 hours/month) and those with more projects are more likely to leave
- Departmental turnover rates vary: HR (29%), Accounting (27%), Technical (26%), Management (14%)
- Low salary employees have higher turnover
- Correlation analysis and boxplots confirm these patterns

See 'eda_analysis.png' and 'correlation_matrix.png' for visualizations.

4. Clustering of Employees Who Left

K-means clustering (k=3) on employees who left reveals:

Cluster	Satisfaction	Evaluation	Projects	Hours	Tenure (yrs)	Count
0	0.41	0.52	2.1	149	3.0	1,649
1	0.12	0.86	6.2	273	4.1	963
2	0.81	0.91	4.5	243	5.2	959

See 'employee_clusters.png' for visualization.

5. Class Imbalance Handling (SMOTE)

- Original training set: 9,142 stayed, 2,857 left
- After SMOTE: 9,142 stayed, 9,142 left (fully balanced)

6. Model Training, Cross-Validation, and Evaluation

Models trained: Logistic Regression, Random Forest, Gradient Boosting, SVM

5-fold cross-validation results:

Model	CV Accuracy	Test Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.770	0.759	0.496	0.762	0.601
Random Forest	0.983	0.987	0.977	0.969	0.973
Gradient Boosting	0.958	0.967	0.921	0.941	0.931
SVM	0.696	0.687	0.410	0.712	0.520

Best model: Random Forest (F1-Score: 0.973)

Justification: F1-Score is used due to class imbalance and the need to balance precision and recall.

Top 5 important features: satisfaction_level, time_spend_company, average_monthly_hours, number_project, last_evaluation

See 'model_comparison.png', 'confusion_matrix.png', and 'feature_importance.png' for details.

7. Retention Strategies for Targeted Employees

Targeted strategies for high-risk groups:

Low Satisfaction (<0.5) (Employees: 2,036)

- Conduct regular satisfaction surveys
- Implement flexible work arrangements
- Provide career development opportunities
- Improve work-life balance policies
- Enhance recognition and reward programs

Overworked (>250 hours/month) (Employees: 1,960)

- Implement workload management systems
- Hire additional staff
- Provide overtime compensation
- Set realistic project deadlines
- Encourage time-off utilization

Long-tenured (>5 years) (Employees: 1,073)

- Provide career advancement opportunities
- Implement mentorship programs
- Offer specialized training
- Create leadership development programs
- Provide competitive compensation

High Performers (evaluation >0.8) (Employees: 3,798)

- Implement performance-based bonuses
- Provide challenging projects
- Offer leadership opportunities
- Create fast-track promotion programs
- Provide competitive compensation

Low Salary (Employees: 5,144)

- Conduct salary benchmarking
- Implement performance-based raises
- Provide additional benefits
- Create profit-sharing programs
- Offer equity or stock options

Department-specific strategies are also recommended for high-turnover departments (HR, Accounting, Technical).

8. Conclusion

This analysis provides actionable insights for reducing employee turnover and improving retention at Portobello Tech. By focusing on satisfaction, workload, career development, and compensation, HR can target high-risk groups and departments for maximum impact.

--- End of Report ---