Human Resources Attrition

Predicting Employee Attrition

ENTERPRISE DATA SCIENCE BOOTCAMP

Group 9:
Bernardo Madeira
Gonçalo Pereira
Inês Mafra
Matheus Nunes

December 2025

AGENDA

- Introduction
- Employee Profile
- Attrition Prediction Model
- Conclusions

INTRODUCTION

- Business Problem

- Objectives

- Dataset Features

Business Problem

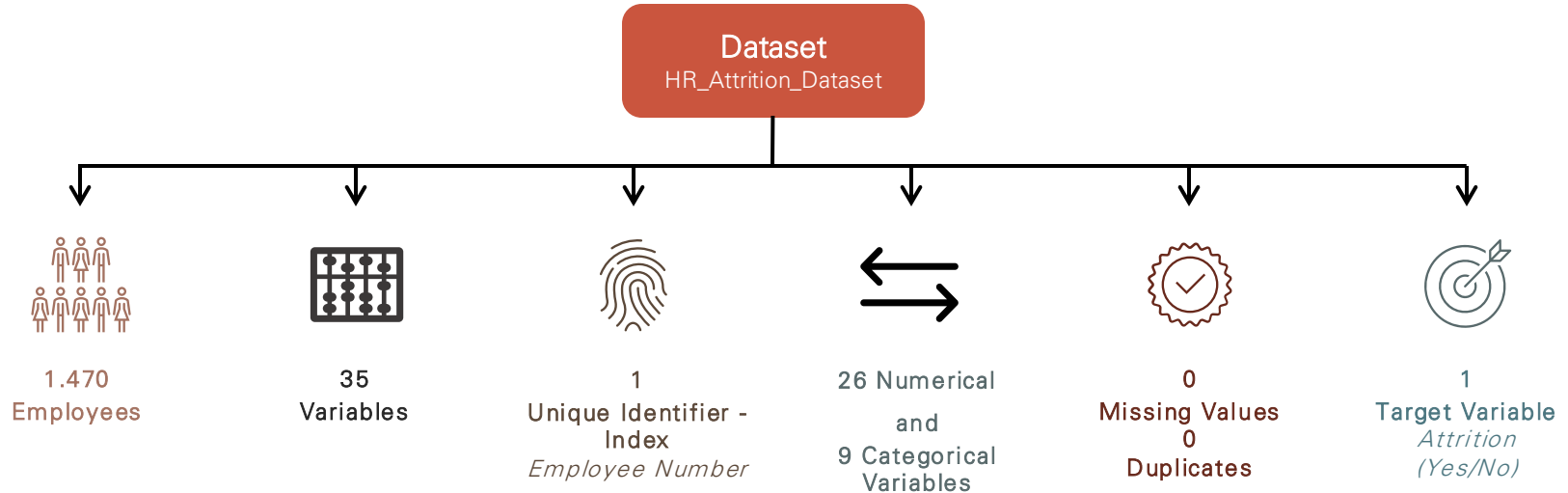**Employee turnover creates significant financial and operational impact:**

- The time needed to recruit and interview candidates;
- Incentives or bonuses tied to new contracts;
- Reduced productivity during the adaptation period;
- Loss of experience and knowledge;
- Maintaining company culture.

A **data-driven approach** can **help anticipate resignations,** reveal the underlying drivers of attrition, and **support HR** in **implementing retention strategies**.

Objectives

- **Identify characteristics** associated with employee resignation;
- **Analyse relationships** between key HR variables and attrition;
- **Build a predictive model** capable of estimating resignation probability;
- **Support HR decision-making** with targeted interventions;
- **Enable continuous monitoring** of attrition patterns;
- **Support workforce planning** by aligning hiring needs with predicted turnover trends.
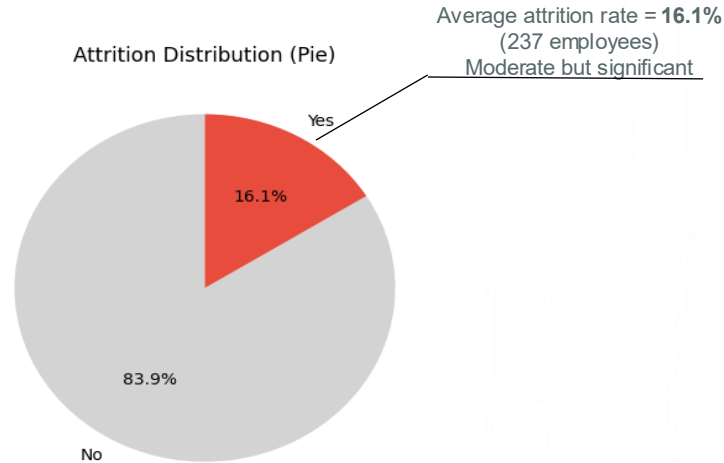
Dataset features

Dataset
HR_Attrition_Dataset

1.470
Employees

35
Variables

1
Unique Identifier -
Index
Employee Number

26 Numerical

and

9 Categorical
Variables

0
Missing Values
0
Duplicates

1
Target Variable
*Attrition
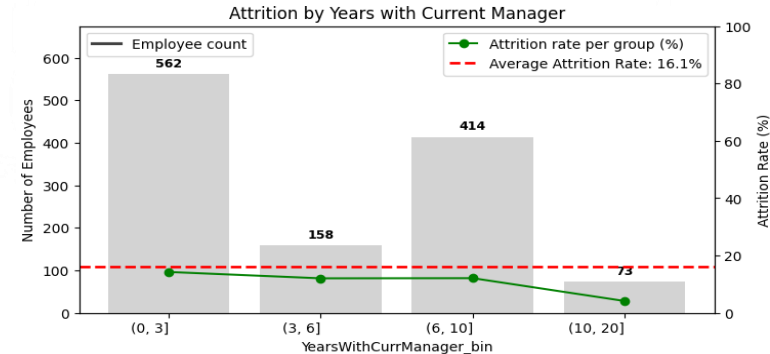(Yes/No)*

EMPLOYEE PROFILE

- Characteristics of Employees Who Resign

Characteristics of employees who resign

Attrition Distribution (Pie)

Average attrition rate = **16.1%**
(237 employees)
Moderate but significant

Yes

16.1%

83.9%

No

- Which variables are related to Attrition?

- Which segments have Attrition rate above the average?

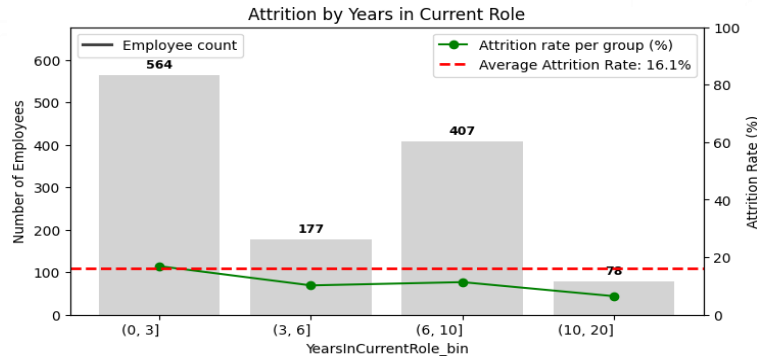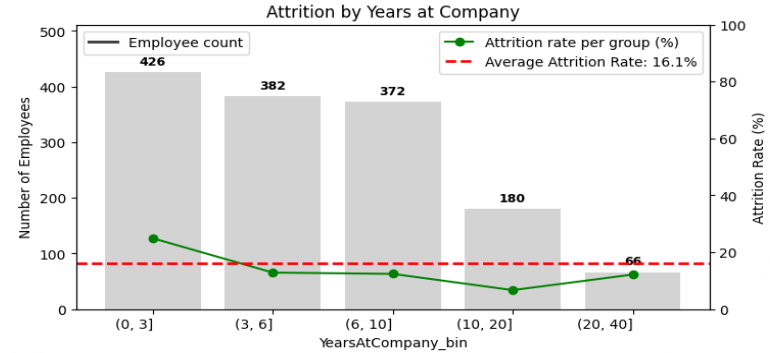- What is the profile of employees who resign?

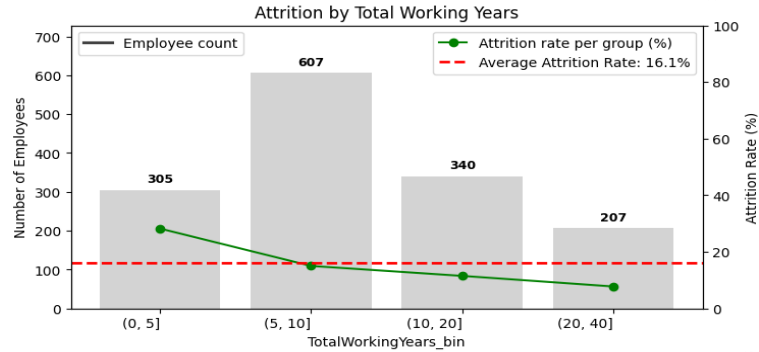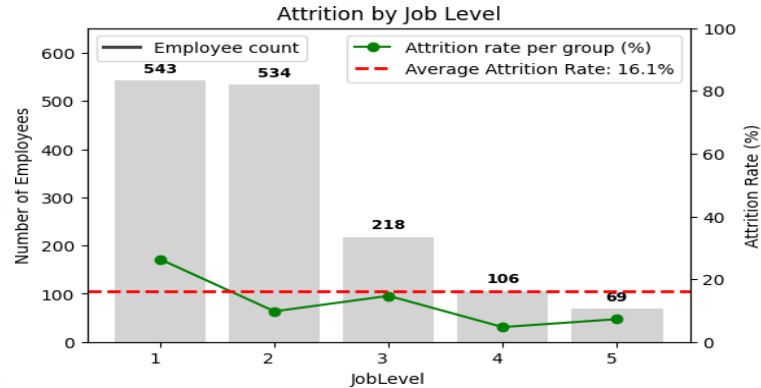# Characteristics of employees who resign

Characteristics of employees who resign

Characteristics of employees who resign

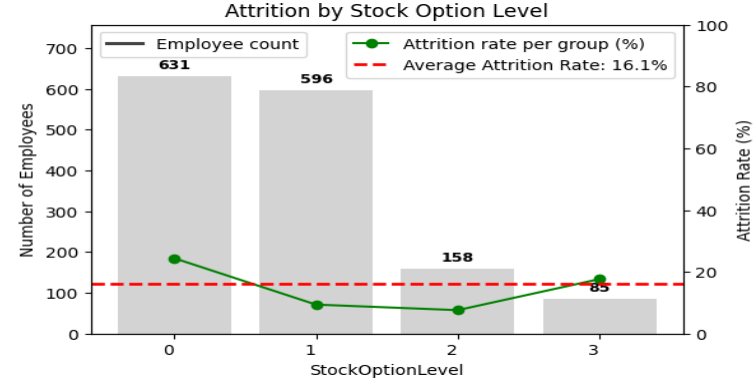# Profile of employees who resign

- Younger employees (under ~35)

- Lower salary ranges (≤ 3000€)

- Low stock option levels (0) or mid-level plateaus (3)

- Early tenure (≤ 3 years) or no recent role changes

- Entry-level roles (Job Level 1)

- Lower job involvement (Levels 1–2)

- Lower job satisfaction (Levels 1–3)

ATTRITION PREDICTION MODEL

- Correlation Analysis

- Predictive Modeling

- Performance Analysis

- Deployment Strategy

- Model Interpretation

- Usefulness of the Model

Correlation Analysis — Top 10 Variables Related to Attrition

- Categorical variables were converted into binary indicators (one-hot encoding)

- Spearman's correlation coefficient

- The top 10 variables have a negative correlation with Attrition

Heatmap — Top 10 Correlations with Attrition

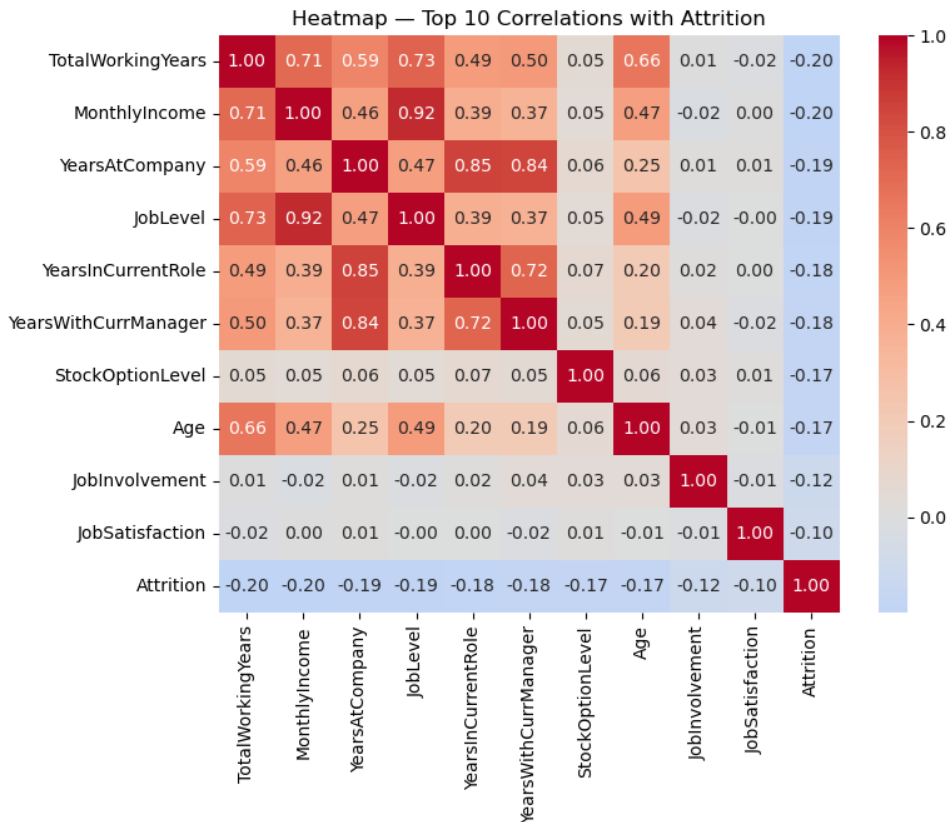|	TotalWorkingYears	MonthlyIncome	YearsAtCompany	JobLevel	YearsInCurrentRole	YearsWithCurrManager	StockOptionLevel	Age	JobInvolvement	JobSatisfaction	Attrition
TotalWorkingYears	1.00	0.71	0.59	0.73	0.49	0.50	0.05	0.66	0.01	-0.02	-0.20
MonthlyIncome	0.71	1.00	0.46	0.92	0.39	0.37	0.05	0.47	-0.02	0.00	-0.20
YearsAtCompany	0.59	0.46	1.00	0.47	0.85	0.84	0.06	0.25	0.01	0.01	-0.19
JobLevel	0.73	0.92	0.47	1.00	0.39	0.37	0.05	0.49	-0.02	-0.00	-0.19
YearsInCurrentRole	0.49	0.39	0.85	0.39	1.00	0.72	0.07	0.20	0.02	0.00	-0.18
YearsWithCurrManager	0.50	0.37	0.84	0.37	0.72	1.00	0.05	0.19	0.04	-0.02	-0.18
StockOptionLevel	0.05	0.05	0.06	0.05	0.07	0.05	1.00	0.06	0.03	0.01	-0.17
Age	0.66	0.47	0.25	0.49	0.20	0.19	0.06	1.00	0.03	-0.01	-0.17
JobInvolvement	0.01	-0.02	0.01	-0.02	0.02	0.04	0.03	0.03	1.00	-0.01	-0.12
JobSatisfaction	-0.02	0.00	0.01	-0.00	0.00	-0.02	0.01	-0.01	-0.01	1.00	-0.10
Attrition	-0.20	-0.20	-0.19	-0.19	-0.18	-0.18	-0.17	-0.17	-0.12	-0.10	1.00

Predictive Modeling

Objective: Predict the probability of resignation for each employee.

Target Variable (label)

- 1 if the employee resigned, 0 if the employee did not resign.

Avg Attrition Rate

- 16,1%

Training Sample

- 70% of data

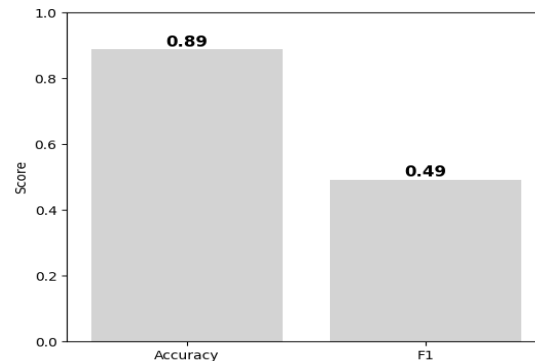**Test Sample**

- 30 % of data

Modeling Approach

- We tested a pletora of models; *KNN, Logistic Regression, Random Forest, High-Gradient Boosting.*

- *Hyper-parametrization of each.*

Gradient Boosting

Best overall result, maintaining high accuracy and achieving the highest F1-Score, showing greater ability to correctly identify attrition cases.
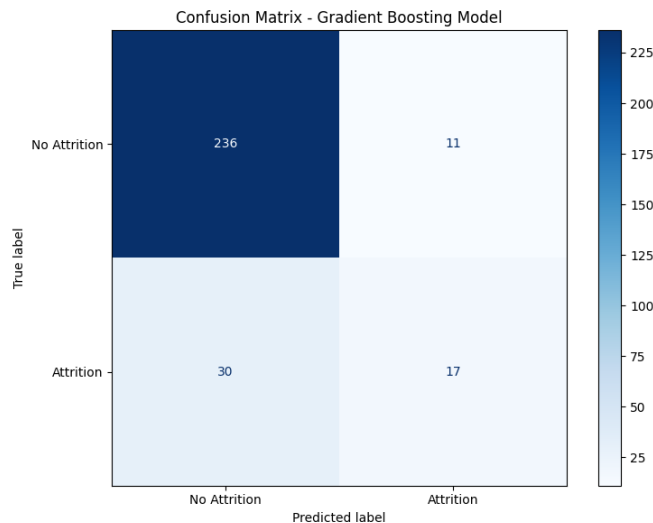
Performance Analysis

According to the Confusion Matrix, the model:

- 236 → True Negatives (Correctly predicted stay) ✔
- 11 → False Positives (Predicted leave, but stayed ✘
- 30 → False Negatives (Left, but predicted stay) ✘
- 17 → True Positives (Correctly predicted leave) ✔

Business Impact

- Stability maintained (low error for stayers);
- Retention risk (many at-risk employees missed);
- Low Recall: poor alert for departures

Confusion Matrix - Gradient Boosting Model

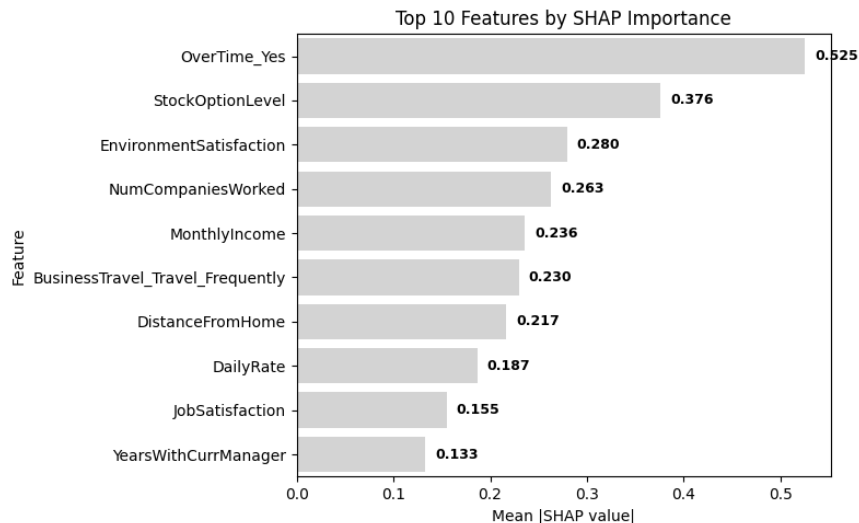# Deployment Strategy

1) Gradient Boosting → Primary model

2) Logistic Regression → Baseline (interpretability)

3) Shadow deploy 2 – 4 weeks

4) Monitor: recall, precision, alert volume, cost

5) Choose final model with HR

6) Retrain quarterly, adjust thresholds

7) Hugging Face Distribution:
   HTTPS://HUGGINGFACE.CO/MAKONEANDTWO/HGB_
   EDSA_GROUP_9_2025/TREE/MAIN

Model interpretation

According to SHAP technique, the TOP 5 most significant variables in differentiating employees are:

- Overtime hours

- Stock options level

- Satisfaction with the work environment

- Number of companies worked at

- Monthly salary

Top 10 Features by SHAP Importance

Feature	Mean |SHAP value|
OverTime_Yes	0.525
StockOptionLevel	0.376
EnvironmentSatisfaction	0.280
NumCompaniesWorked	0.263
MonthlyIncome	0.236
BusinessTravel_Travel_Frequently	0.230
DistanceFromHome	0.217
DailyRate	0.187
JobSatisfaction	0.155
YearsWithCurrManager	0.133

Usefulness of the Model

a) **Targeted HR Interventions**

- **Helps to identify the problems** (mentorship, salary review, career pathing);

- **Avoids blanket spending** — invest only where risk is highest.

b) **Financial Impact**

- **Prevents** resignations and **saves** recruitment, onboarding, and productivity costs;

- Direct **financial gains** by reducing turnover.

c) **Operational Workflow**

- **Weekly dashboard:** flagged employees, alert volume, confirmed attrition;

- **HR reviews** flagged cases and applies tailored retention actions;

- **Feedback loop**: measure precision/recall, adjust threshold regularly;

- Use insights to **justify budget allocation and retention policies.**

CONCLUSIONS

- The Loyalty Levers

Conclusions

The Loyalty Levers

- **Overtime Hours:**

  o Project planning, aligned with resource and time constraints;

  o Flexible schedules and enforce clear limits on working hours.

- **Stock options level:** Consider offering stock options or financial incentives to employees at lower levels.

- **Satisfaction with the work environment:**

  o Create a positive and supportive work culture with open communication and recognition;

  o Improve the physical and psychological work environment by ensuring comfort, safety, and flexibility;

  o Enhance work-life balance policies;

  o Collect regular feedback via engagement surveys.

Conclusions

The Loyalty Levers

- **Number of companies worked at:**

  o Provide clear career growth opportunities and long-term incentives to encourage stability.

  o Strengthen onboarding and engagement programs to build loyalty from the start.

- **Monthly salary:**

  o Strengthen retention strategies for lower-income employees (fair pay reviews, financial support programs);

  o Provide clear career progression paths to help employees move into higher income brackets;

  o Monitor satisfaction among mid-income employees to prevent stagnation and keep engagement for high-income.

Thank You!

Questions?

Hugging Face Results

Employee Attrition Risk Predictor

Provide basic information about an employee to estimate the probability that they will leave the company.

Age			Monthly Income	
	25	↺		1809	↺
18		60	1009		19999

Department	Job Role
Research & Development ▼	Healthcare Representative ▼

Business Travel	OverTime
Travel_Frequently ▼	◉ No ○ Yes

Job Satisfaction (1-4)	Environment Satisfaction (1-4)	Work-Life Balance (1-4)
4 ↺	3 ↺	2 ↺
1 ——— 4	1 ——— 4	1 ——— 4

Years at Company	Distance From Home (km)
3 ↺	24 ↺
0 ——— 40	1 ——— 29

Predict attrition risk

Predicted risk level

Low risk of attrition

Estimated probability of attrition

8.9%

Short explanation

This risk estimate is based on patterns learned from historical employees, considering factors such as overtime, satisfaction, tenure and income. It should support HR decisions, not replace them.