



TalentaHub



Rakamin

Probabilistic Turnover Risk Scoring with Tiered Classification

FINAL PROJECT REPORT - RAKAMIN DATA SCIENCE BATCH 55
July - August

Mentor : Dedi Irwanto & Rahmat Faizal

[Dataset](#)

[Python Code](#)

[Github](#)

Our Team



Sukma Indah
PM & Data Engineer



Fitria Shinta A.
Business Analyst



Naufal Iffa M. R.
Data Engineer



Ferry Irwanto
Data Scientist

Outline

01

Overview
&
Problem Statement

02

EDA
&
Preparation

03

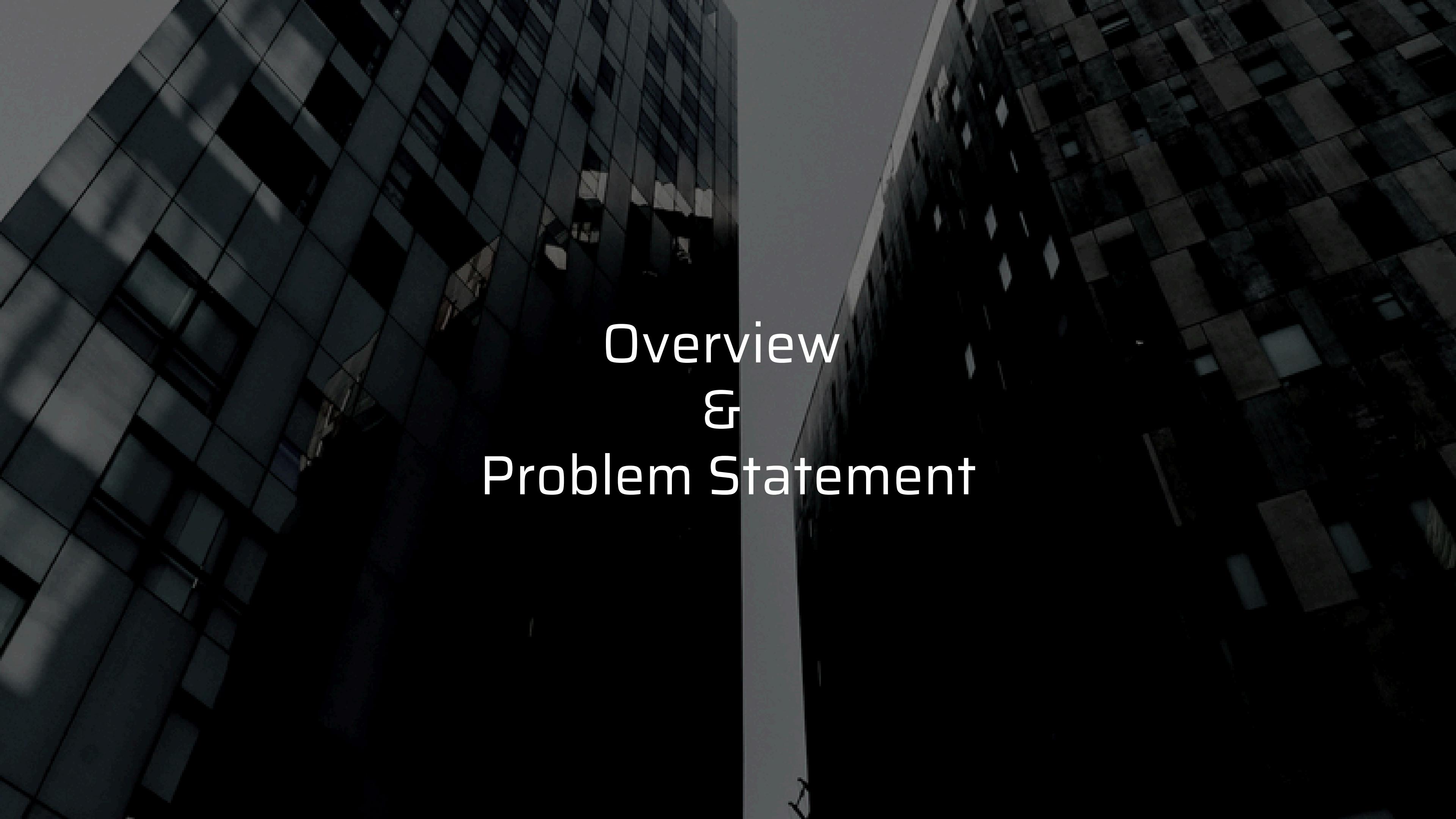
Model Development
&
Experimentation

04

Model Evaluation

05

Bussiness Impact
&
Recommendation



Overview & Problem Statement



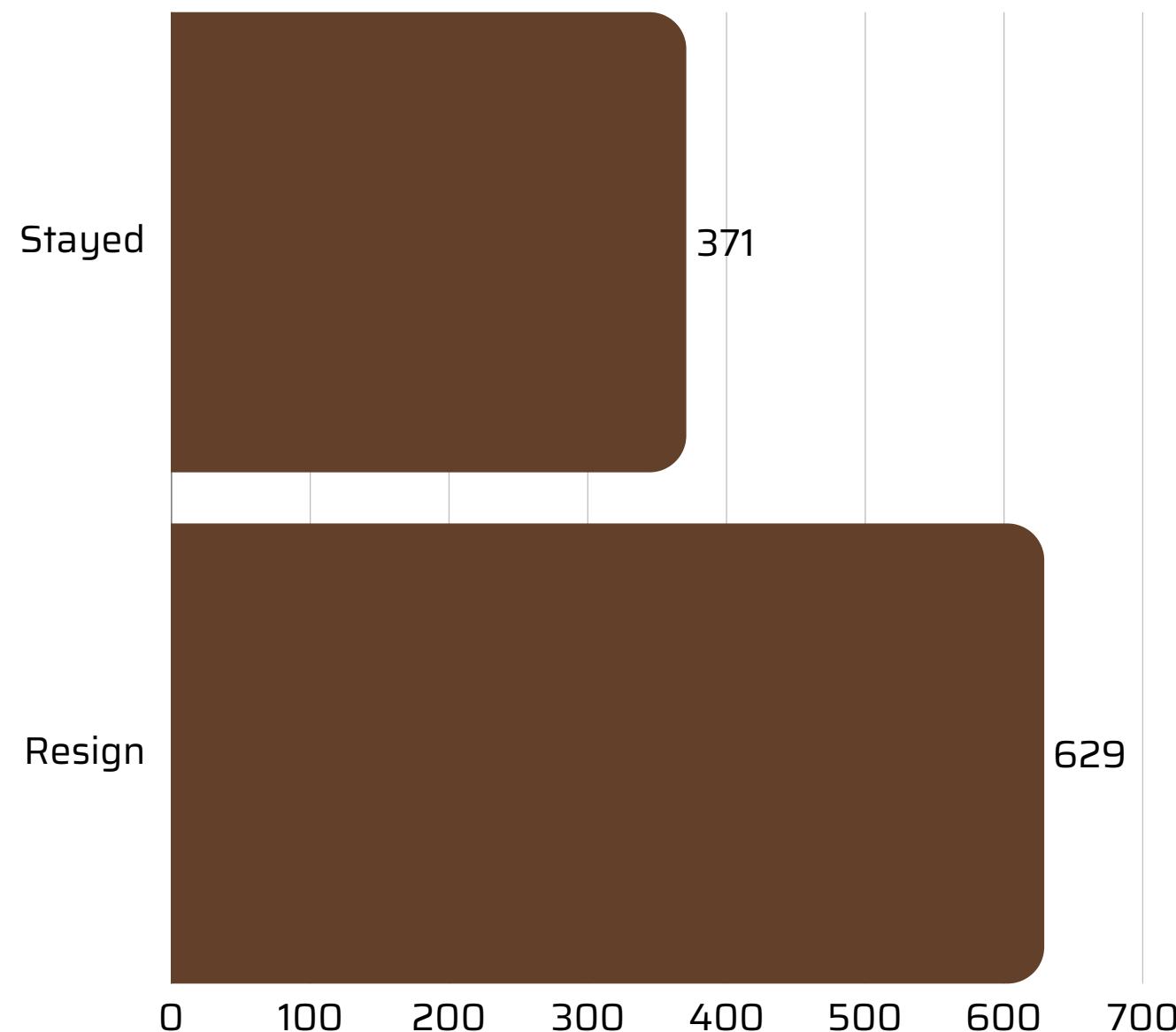
Overview

Talentahub is an Outsourcing Services company that provides digital services for recruitment and talent management, especially in the sales sector.

Last year, the company experienced an **increase in turnover rate**, which is the rate at which the company changes employees and replaces them with new employees.^[1]

[1] <https://sundayinsurance.co.id/blog/turnover-adalah>

Problem Statement



In 2024, **62.9%** of our talents **resigned**, significantly higher than the industry average for B2B sales organizations, which stands at **around 30%.**^[2]

TURNOVER GAP

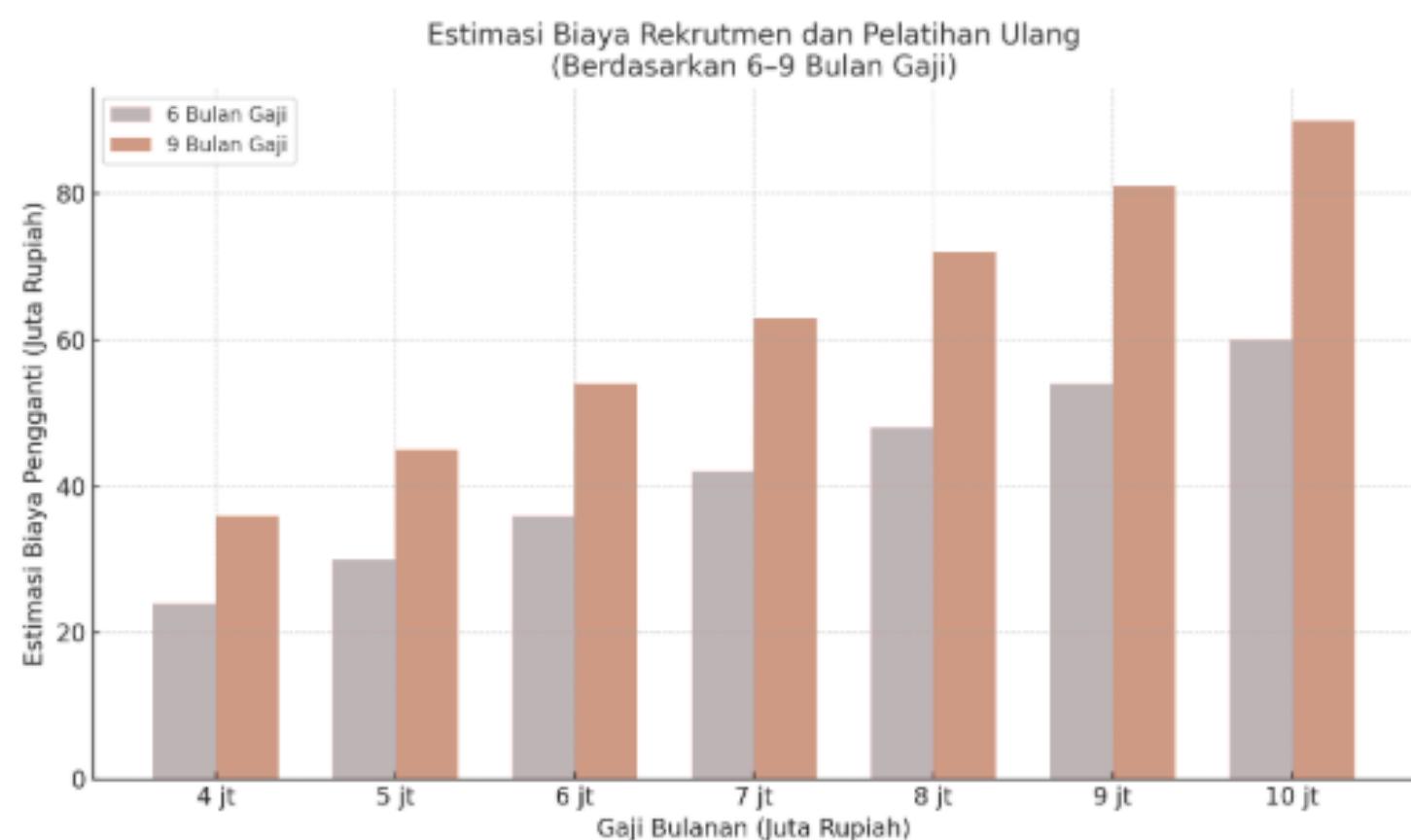
32.9%

This significant gap requires

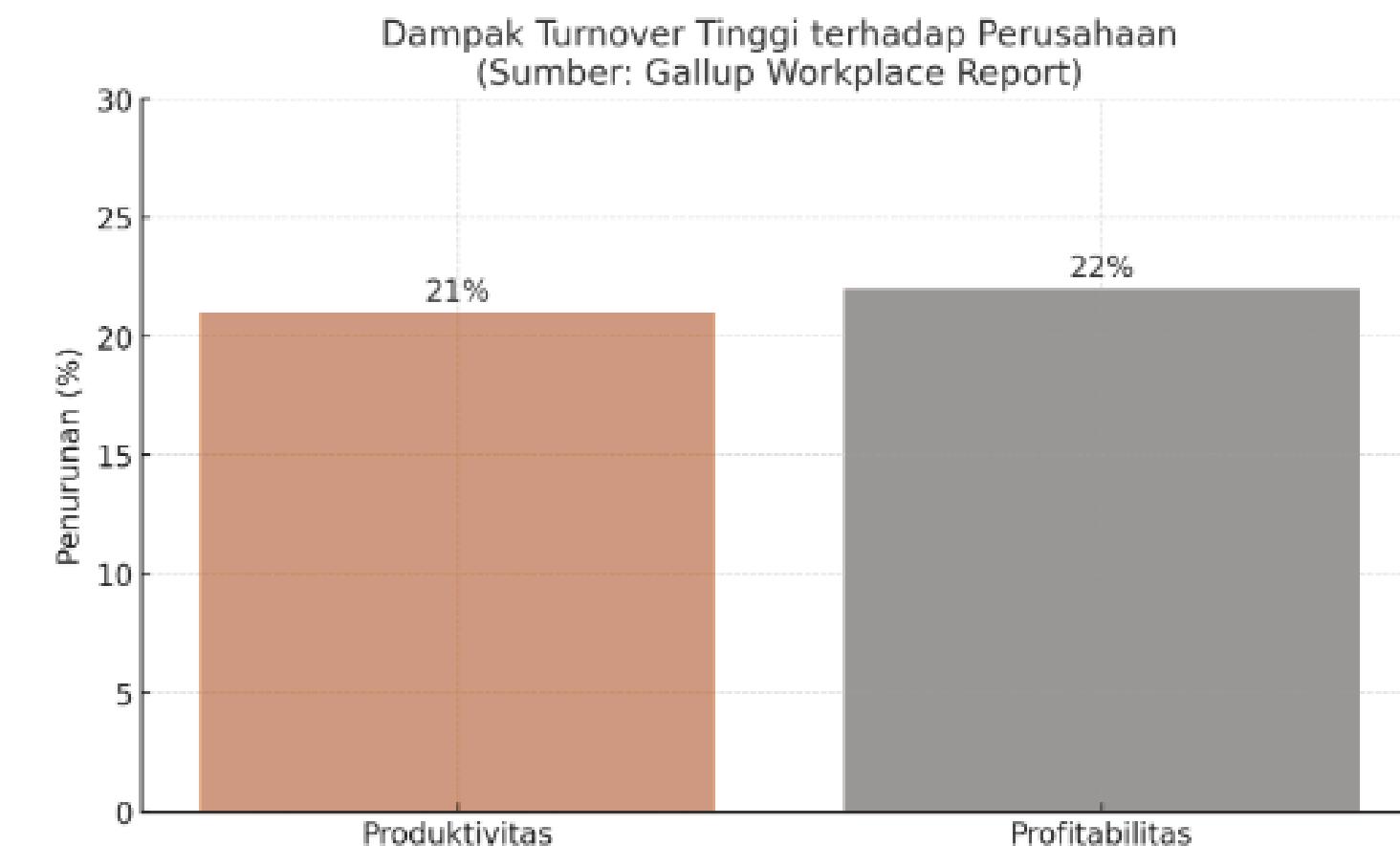
Proactive Talent Retention Strategies

[2] <https://www.xactlycorp.com/blog/motivation/sales-turnover-statistics>

Problem Statement



The cost of replacing an employee is equivalent to 6 to 9 months of that employee's total salary.^[3]



Companies with high employee turnover can experience a decrease in **productivity** of up to 21% and a drop in **profitability** of up to 22%.^[4]

So, the best solution is

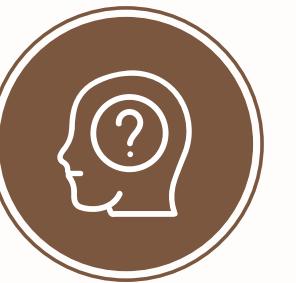
Retain Employees

Goals

Build a system capable of profiling employee risk based on their likelihood of resigning from the company. With this risk profile, we can:



Identify employees with high potential to resign.



Analyze the factors that influence employee decisions to leave.



Improve retention strategies

Business Metrics

Using ML models to reduce employee turnover rates **to equal or less than 30%** by performing Probabilistic Turnover Risk Scoring with Tiered Classification.

Allocate targeted costs to each employee turnover segmentation

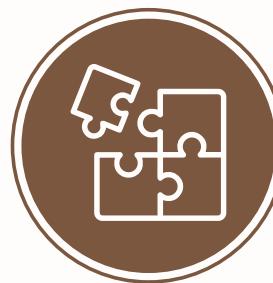
HIGH (60%)

MEDIUM (30%)

LOW (10%)

Based on Risk-Based Prioritization

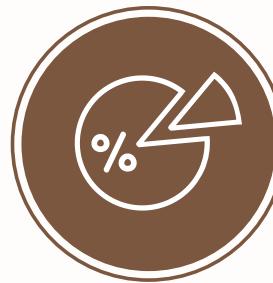
Objective



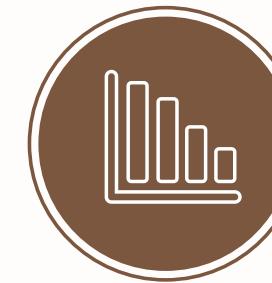
Predict employee resignation with **$\geq 52\%$ recall** using historical HR data to enable early retention efforts.



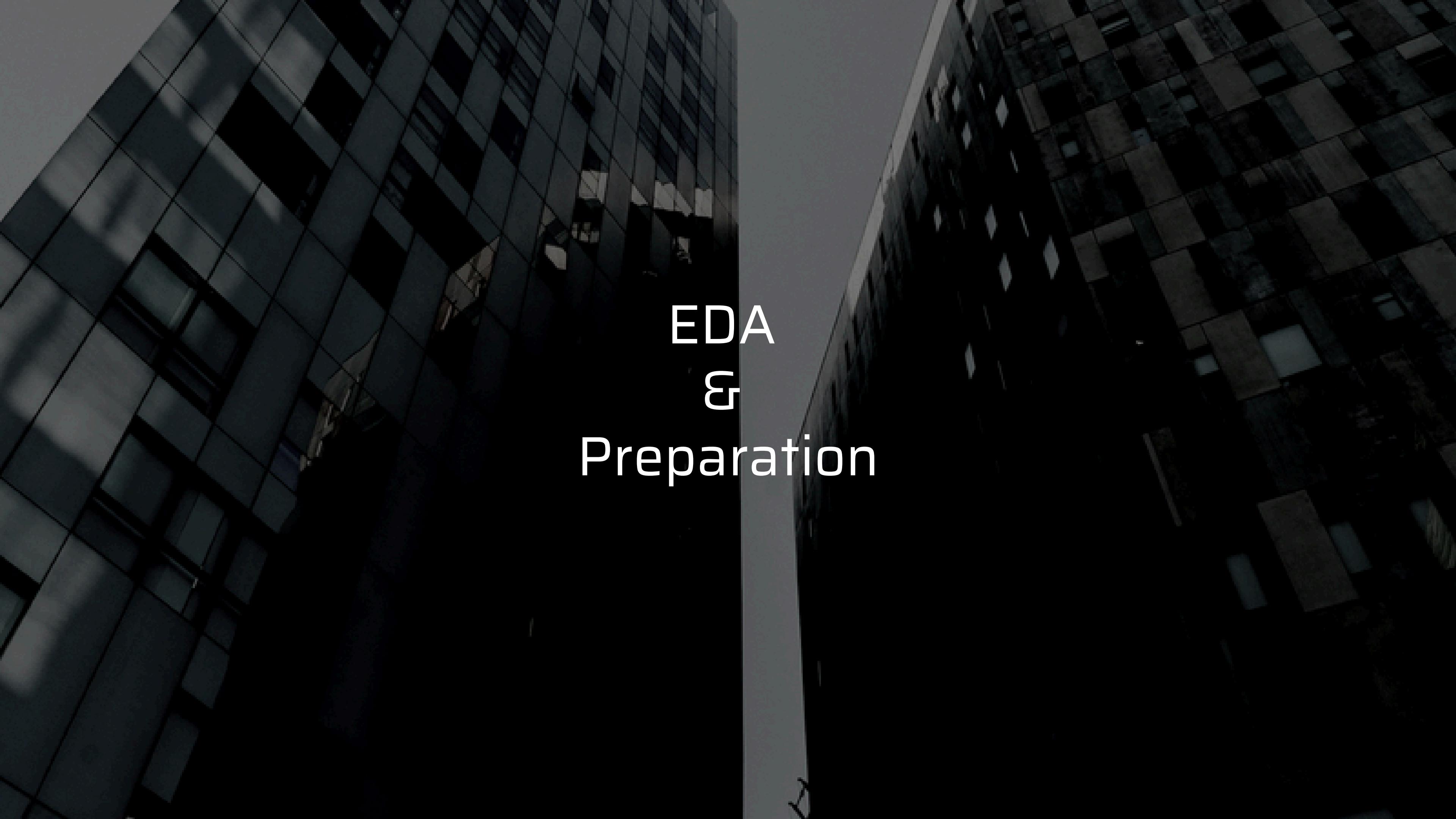
Provides employee risk profiles that can be used as a basis for strategic decision making by HR management.



Conduct risk segmentation to classify employees into several risk levels (**low, medium, high**).



Create Web Apps to help HR teams understand risk prediction and segmentation results.



EDA & Preparation

Dataset Information

Numerical



employee_id



monthly_target



manager_support_score



salary



overtime_hours_per_week



experience_year



working_hours_per_week



target_achievement



distance_to_office_km



job_satisfaction



company_tenure_years



age

The Dataset
consists of
**1000 rows and
19 columns.**

Categorical



gender



education



marital_status



churn_period

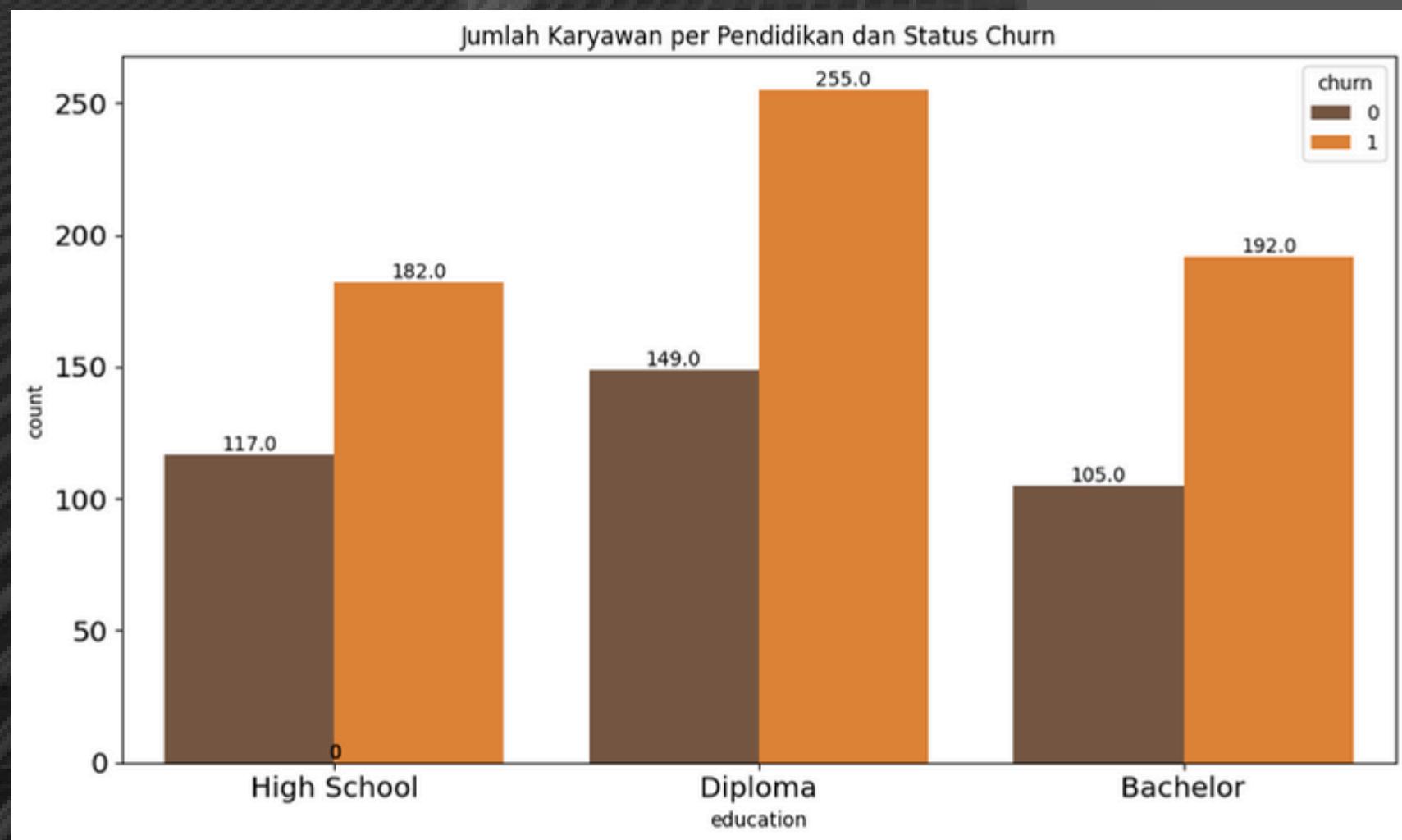


work_location

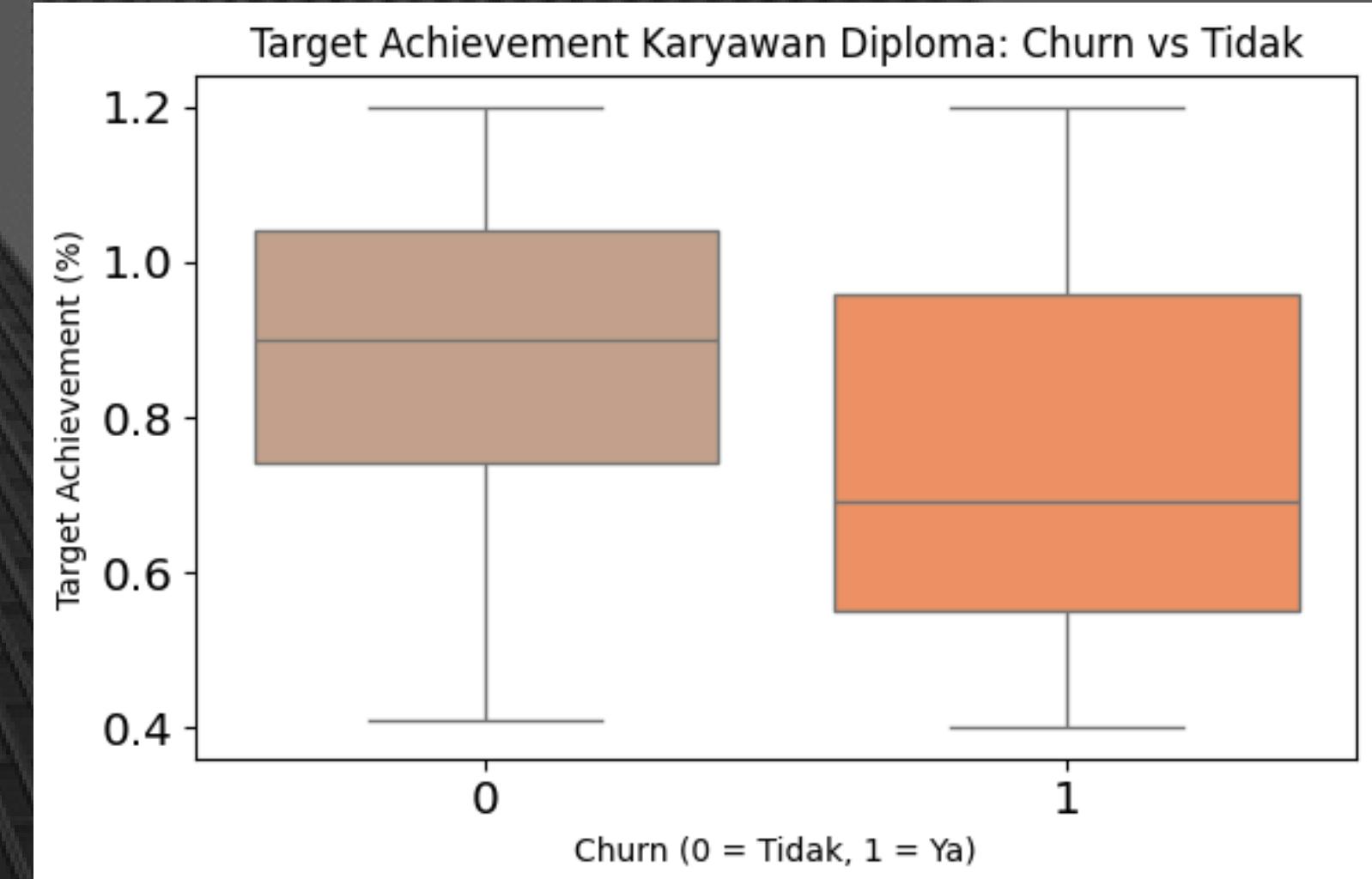


churn
target

Data Insight

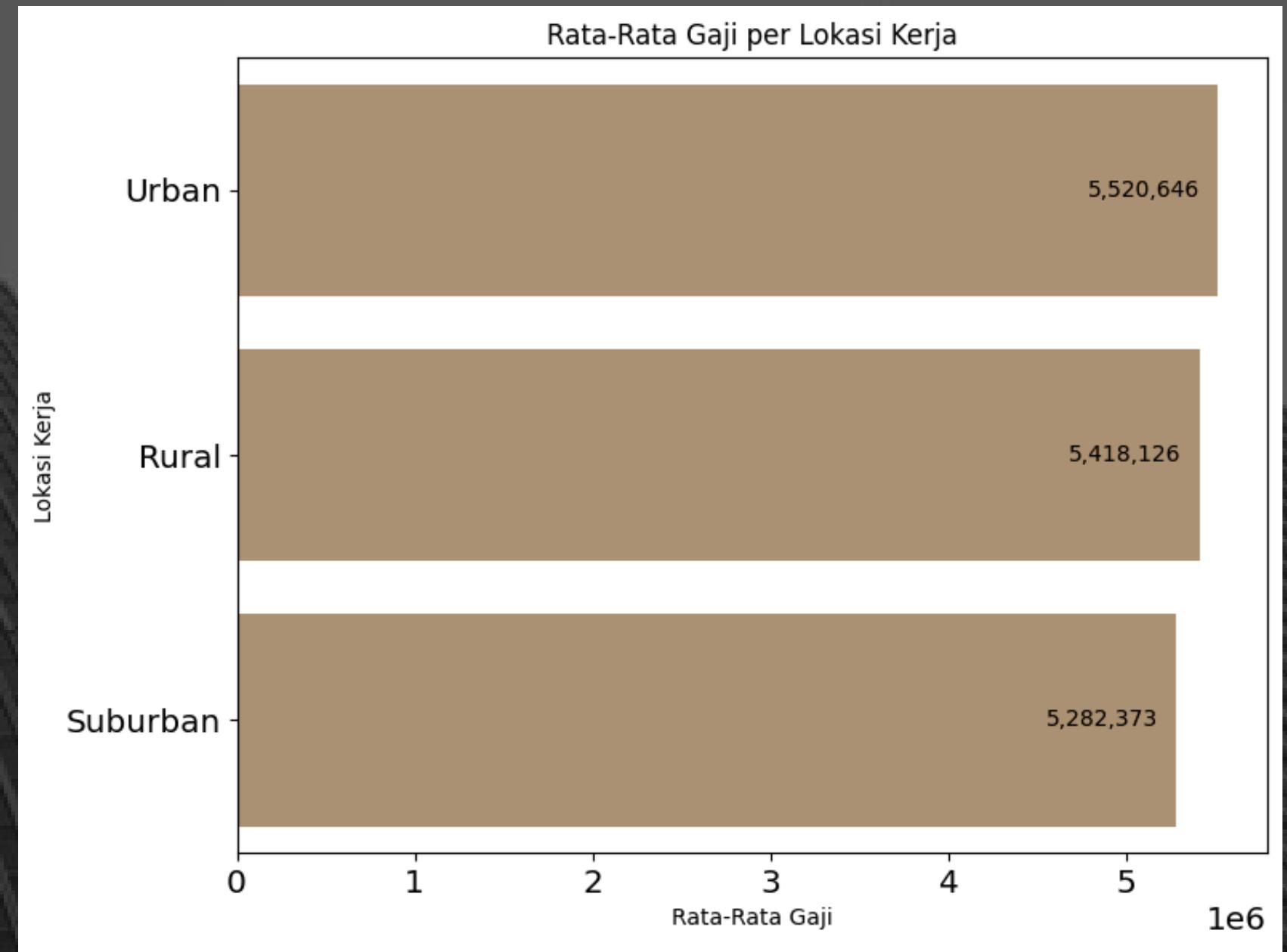
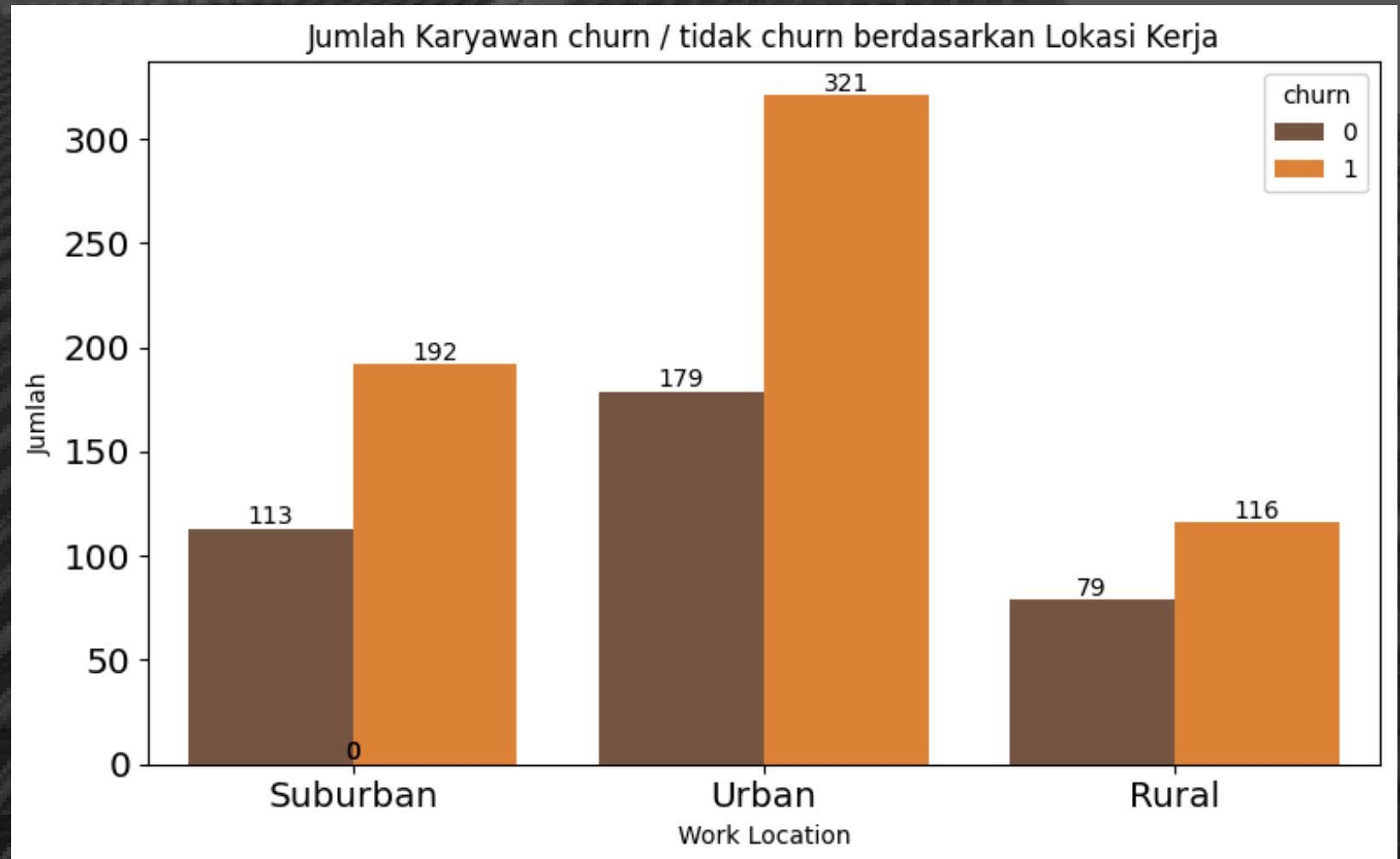


Employees with diplomas show higher turnover rates



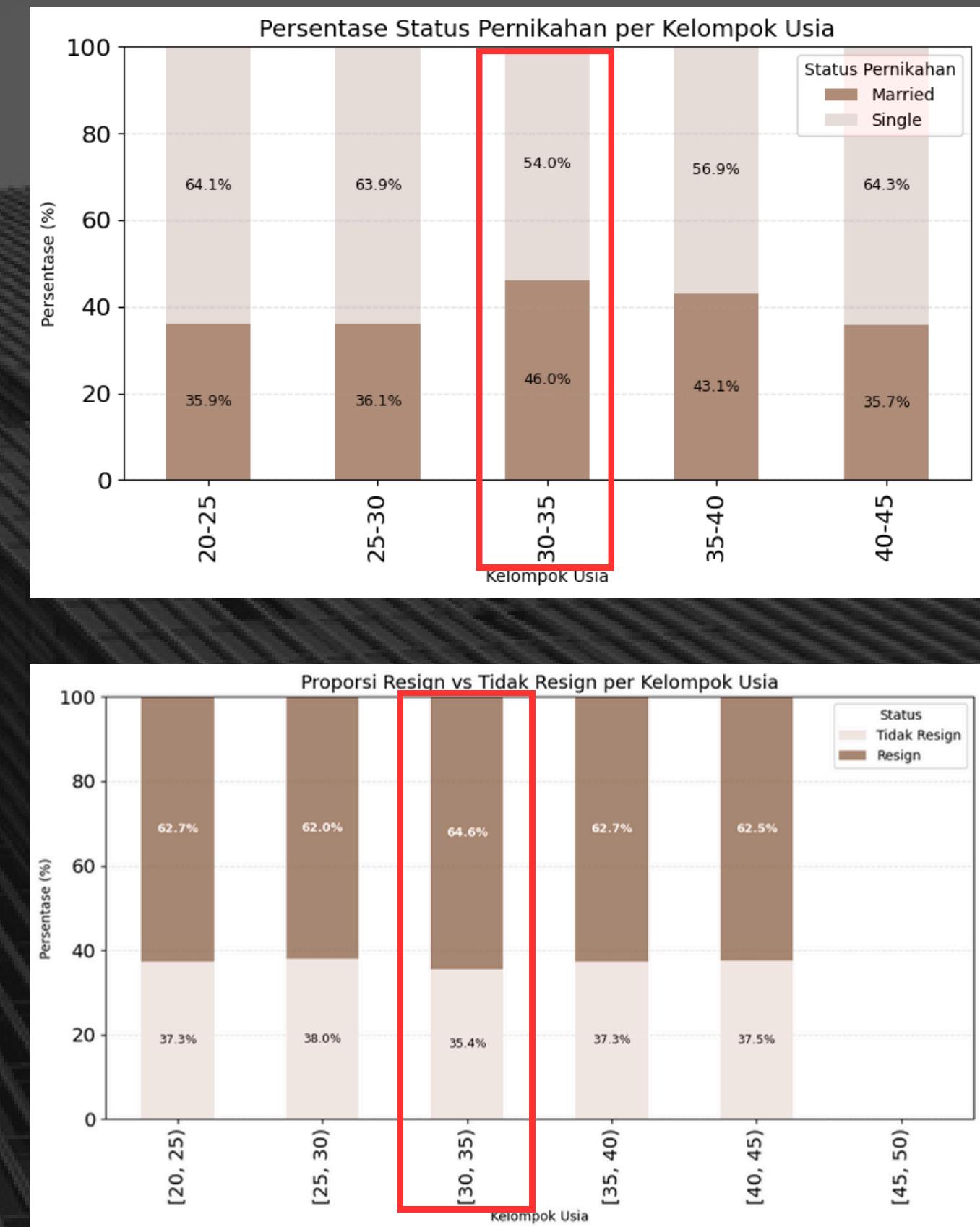
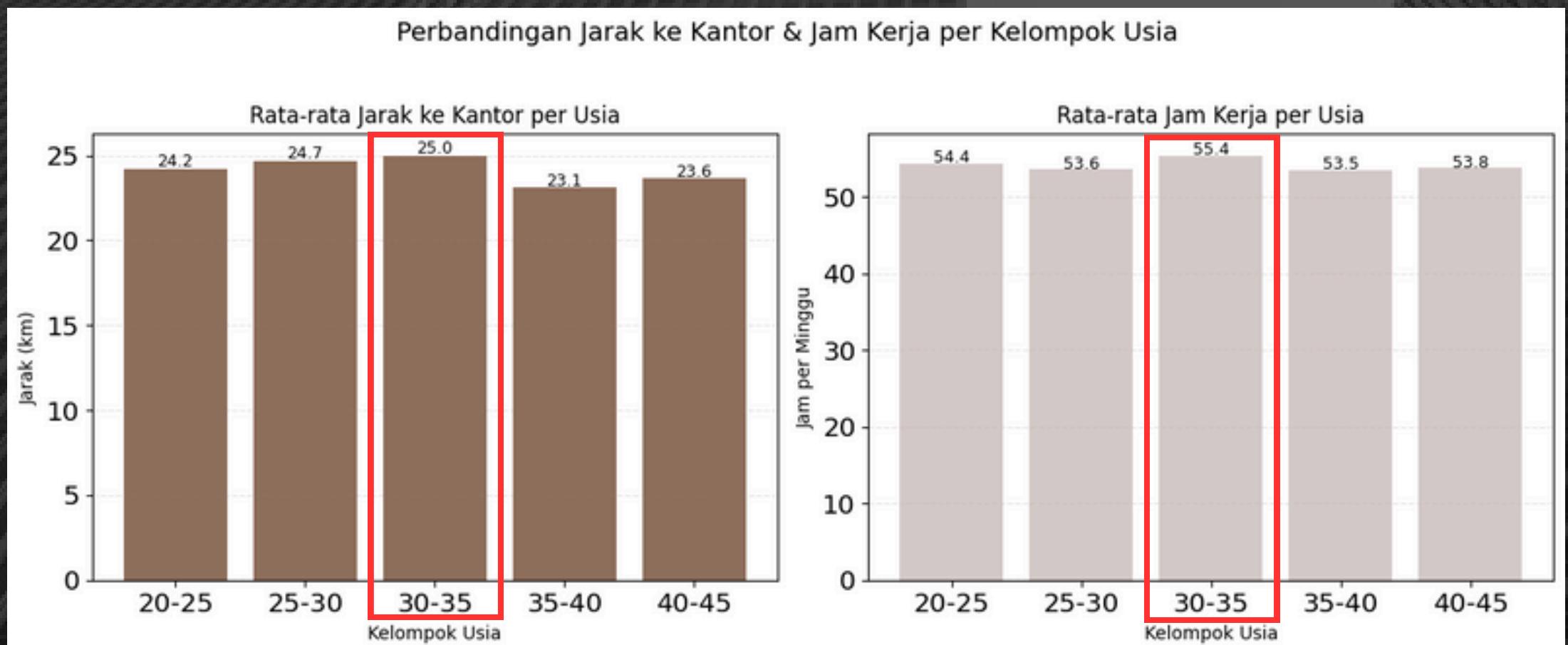
Employees who resign tend to have lower target achievement scores

Data Insight



Employees who live in urban areas tend to have high turnover. This may be due to the equal distribution of salaries across locations, while the cost of living varies (higher in urban areas)

Data Insight

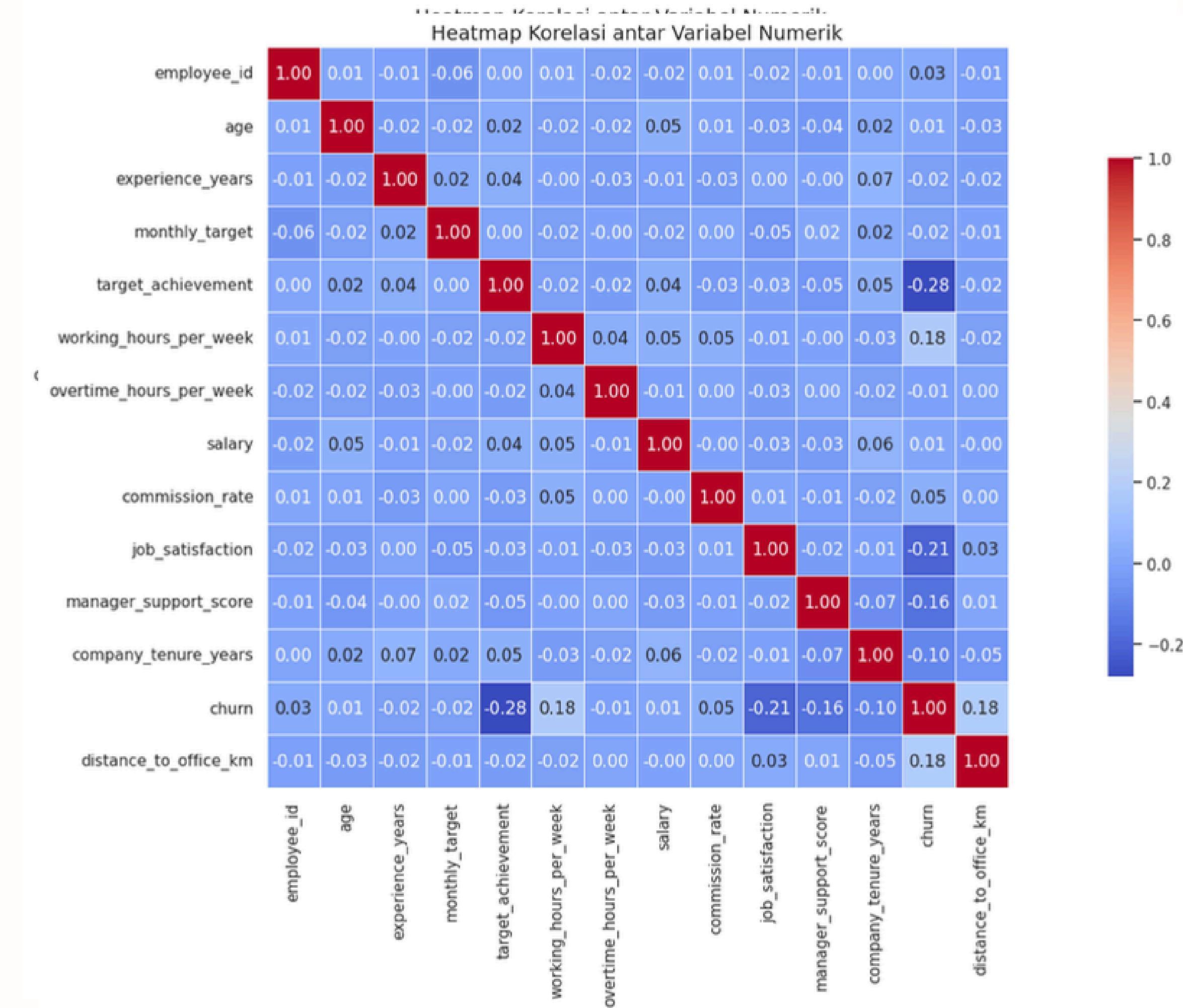


Employees aged 30-35, married, with long working hours per week and living far from home tend to have high churn rates due to burnout and work-life imbalance.

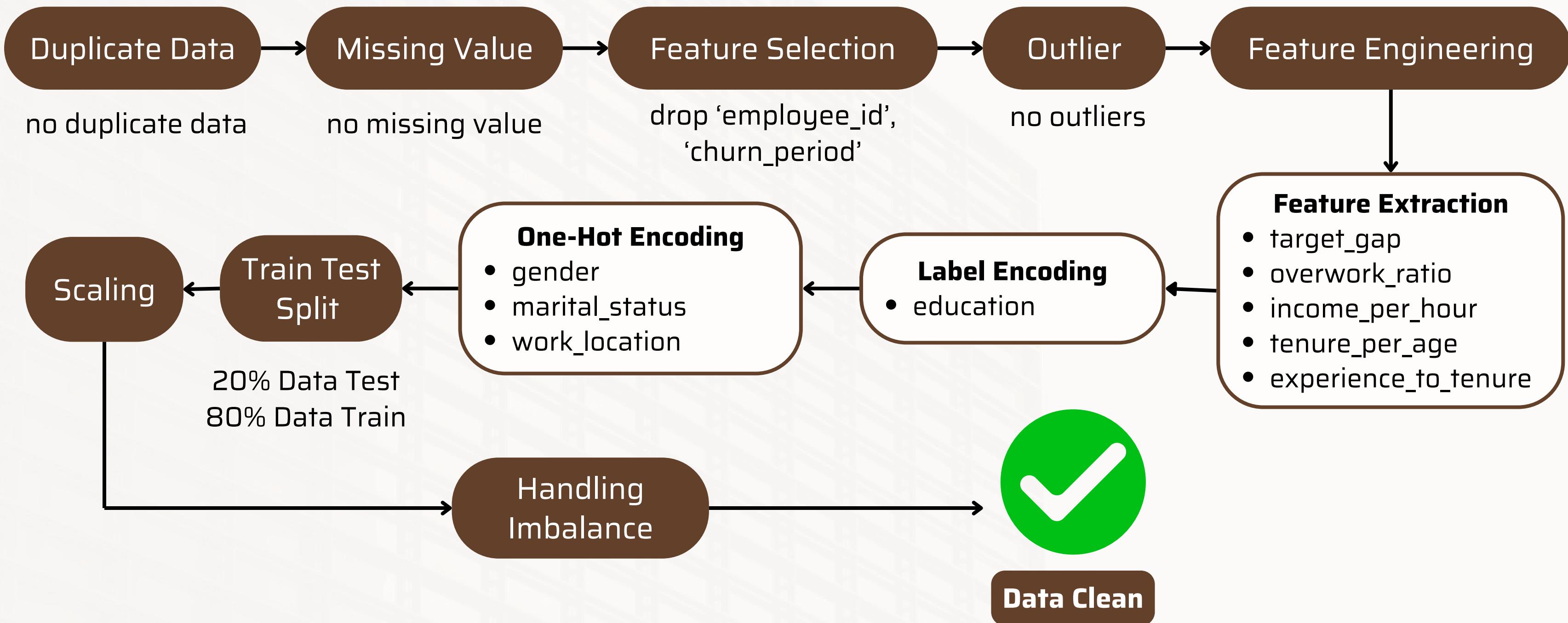
Multivariate Analysis

The heatmap shows that all correlation values are below 0.7, indicating **no multicollinearity** between features.

This suggests the variables are not strongly related and can be used together in modeling.



Data Pre-Processing





Model Development & Experimentation

Model Experiments & Optimization

Best results: Random Forest

- It is robust against overfitting
- It involves many complex features
- It generally performs well

Model	Train Accuracy	Test Accuracy	Train Precision	Test Precision	Train Recall	Test Recall	Train F1	Test F1	Train ROC-AUC	Test ROC-AUC
Logistic Regression	0,63	0,63	0,63	0,63	1	1	0,77	0,77	0,5	0,5
Decision Tree	0,63	0,63	0,63	0,77	1	1	0,77	0,77	0,5	0,5
Random Forest	0,63	0,63	0,63	0,63	0,99	0,97	0,77	0,77	0,43	0,49
SVM	0,63	0,63	0,63	0,63	1	1	0,77	0,77	0,5	0,5
KNN	0,37	0,37	0	0	0	0	0	0	0,5	0,5

Random forest has a low ROC-AUC indicating the need for further Hyperparameter Tuning.

- Recall was chosen as the main success metric to detect as many at-risk employees as possible.
- Precision, accuracy, and ROC-AUC were also considered to ensure model balance and avoid overfitting.



Model Evaluation

Model Evaluation

	Predicted: Stay (0)	Predicted: Resign (1)
Actual: Stay (0)	56 (True Negative)	18 (False Positive)
Actual: Resign (1)	36 (False Negative)	90 (True Positive)

	Precision	Recall	F1-Score	Support
Class 0 (Stay)	0,61	0,76	0,67	74
Class 1 (Churn)	0,83	0,71	0,77	126
Accuracy			0,73	200
Macro Avg	0,72	0,74	0,72	200
Weighted Avg	0,75	0,73	0,73	200
ROC-AUC Score			0,83	

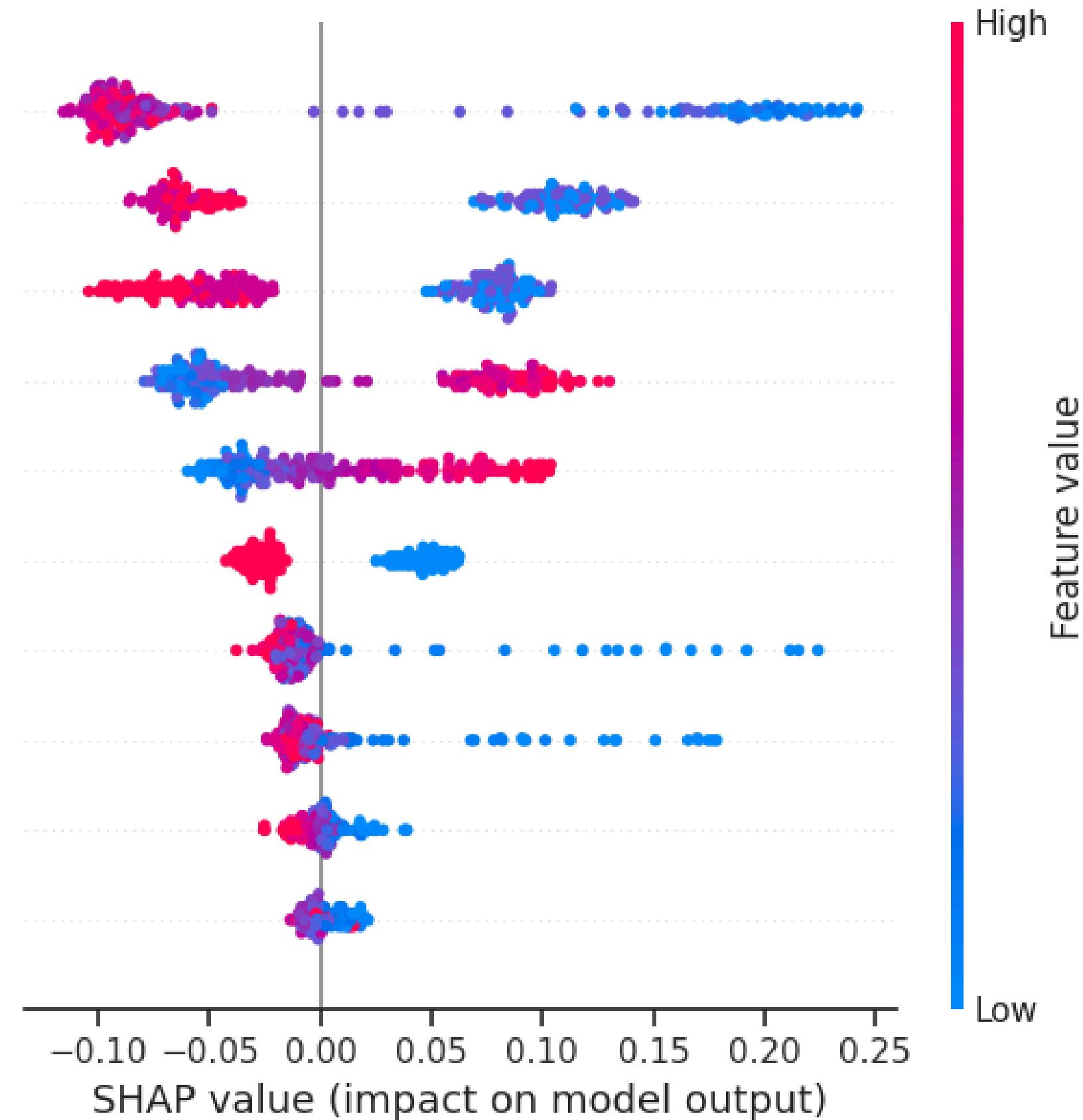
High recall ensures that most true resignations are identified (low false negatives), allowing HR to intervene early and prevent as much potential turnover as possible.

Feature Importance

The SHAP summary plot shows that **low target achievement, low job satisfaction, and low manager support** score are the most influential features contributing to a higher likelihood of churn.

There are 10 best features used for prediction deployment

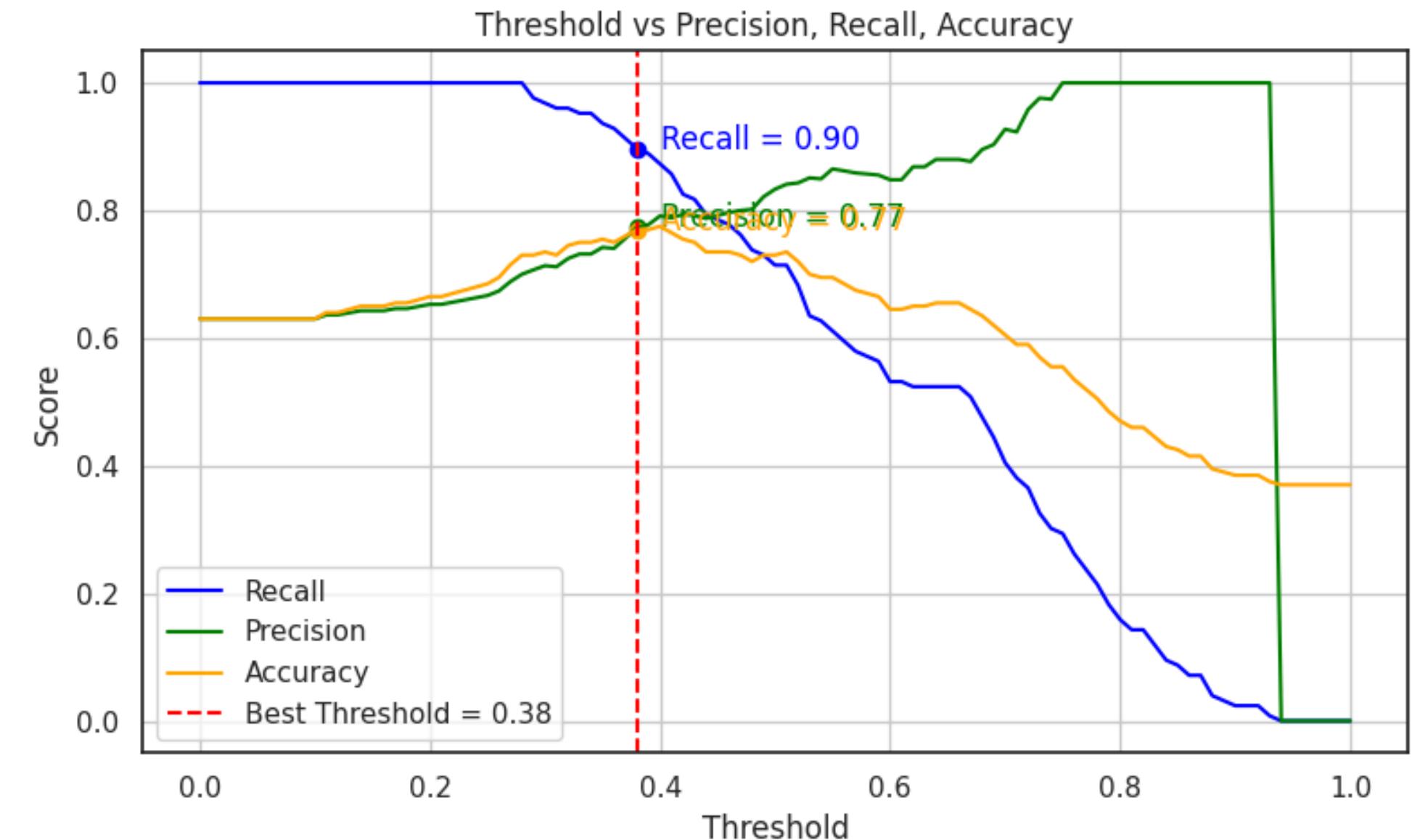
target_achievement
job_satisfaction
manager_support_score
distance_to_office_km
working_hours_per_week
marital_status_Single
tenure_per_age
company_tenure_years
income_per_hour
experience_to_tenure



Threshold Optimization

Threshold Optimization helps in minimizing false negatives, making it ideal for resign prevention strategies.

Best Threshold: 0,38



Segmentation

Turnover segmentation is divided into **3 categories** based on churn probability thresholds

Category	Probability Range	Description
High	> 69%	Very high likelihood of turnover (churn)
Medium	38% – 69%	Moderate risk of churn, needs attention
Low	< 38%	Low likelihood of churn, relatively safe



Bussiness Impact & Recommendation

Business Metrics Evaluation

Total Turnover **629**

Identified **447**

Recall **0.71**

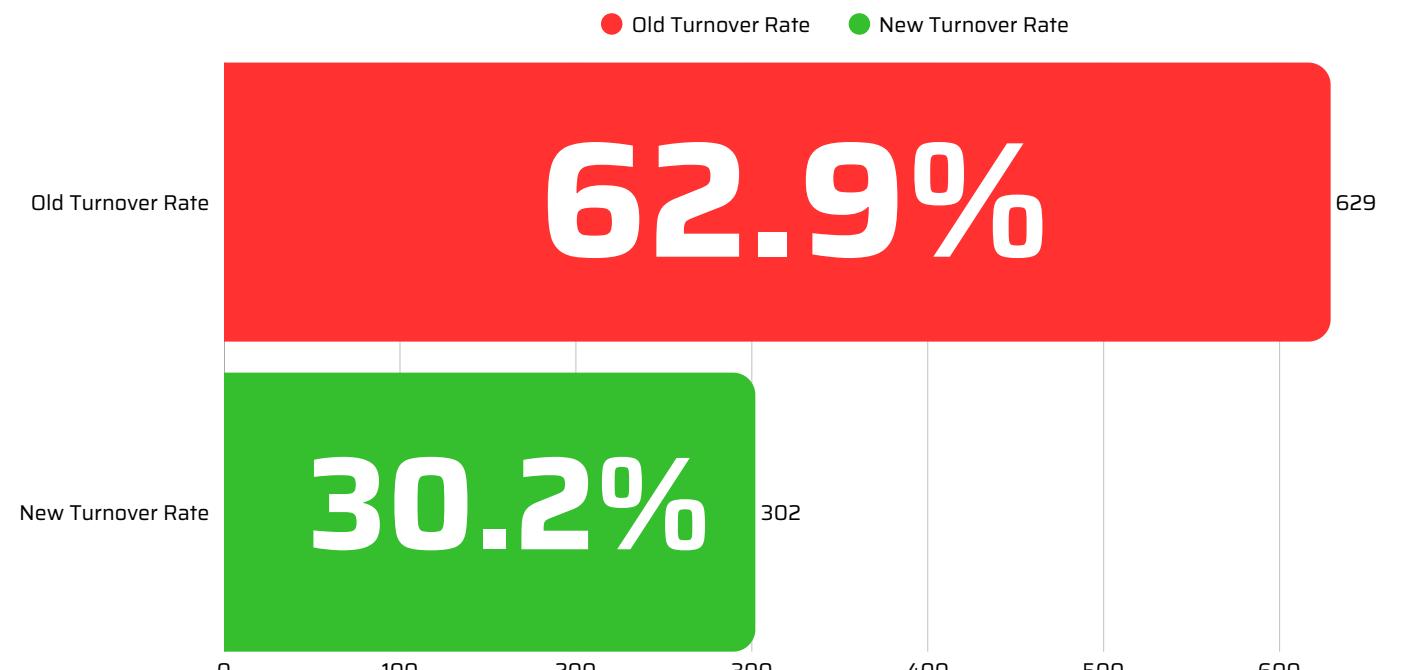
Accuracy **0.73**

Analysis of Prevented Resignation

The model is able to identify 71% of employees at risk of resignation, with an accuracy of 73%, this means approximately 327 employees can potentially be retained through early intervention.

Metric	Value	Explanation
Predicted Resignations (TP)	447	71% of 629
Prevented Resignations	327	73% of 447
Remaining Resignations	302	629 - 327

As a result, the turnover rate drops from 62.9% to 30.2%, indicating a significant reduction of 32.7% in employee resignations.



**MORE THAN
2X
TURNOVER RATE
DECREASE**

Business Impact

Risk Level	Probability of Resignation	Cost Components	Estimated Cost / Employee	Explanation
● High	> 69%	<ul style="list-style-type: none">• Recruitment cost for new hire• Onboarding training• Loss of productivity• Organizational knowledge loss (IP loss)• Team disengagement	60%	Based on Risk-Based Prioritization , resources should be focused first on areas with the highest risk.
● Medium	38% – 69%	<ul style="list-style-type: none">• Recruitment cost (moderate resignation risk)• Partial retraining• Additional supervision	30%	May not resign yet, but potential loss if risks are not mitigated properly.
● Low	< 38%	<p>Passive retention costs such as:</p> <ul style="list-style-type: none">• Satisfaction & wellness surveys• Check-ins & engagement initiatives	10%	Preventive costs for maintaining engagement and well-being.

Business Recommendation

Insight

Actionable Insight

Educational background affects target achievement

Employees with consistently low target achievement tend to have a higher likelihood of resigning

- Improve competencies through targeted training such as workshops, certifications, and e-learning.
- Implement mentoring or a buddy system to pair lower-educated employees with more senior or skilled mentors.
- Re-evaluate targets and KPIs to match job expectations with employee educational levels.

Salary and work location mismatch

Employees resign due to unfair salary distribution that does not align with living costs in their work location.

- Re-evaluate salary schemes based on location by adjusting for local cost-of-living indexes.
- Provide location-specific allowances, such as transport or meal subsidies.
- Focus on non-monetary compensation, such as work shuttles, daycare, or WFH/WFO flexibility.

Work-life imbalance among married employees

Lack of time for family causes resignations.

- Offer flexible working arrangements, such as flexi hours or hybrid models, especially for employees with family commitments.
- Review and reduce excessive working hours by auditing departments with frequent overtime or extreme weekly hours (>50 hours/week).
- Assign work locations based on employee domicile, prioritizing proximity to home, particularly for married employees with young children.

Employee behavior and organizational conditions change over time

- Continuous Learning and Model Fine-Tuning: Regularly update predictive models using the latest data and feedback from HR interventions.

Recommendation

Adding supporting features, such as:



Reason for Resignation

Useful to understand why employees leave, whether it's their own choice or not



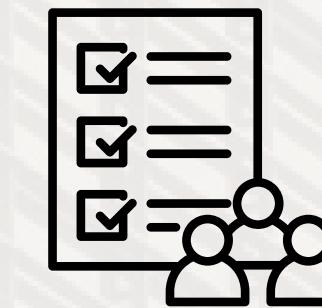
Promotion Amount

Helps understand if a lack of promotions or small promotion increases is linked to why employees resign



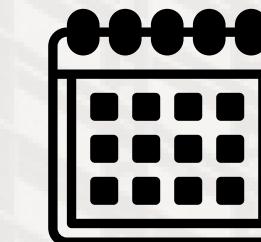
Position

Helps identify which job roles have higher resignation rates, so the company can focus on improving those roles



Employee Status (permanent / contract)

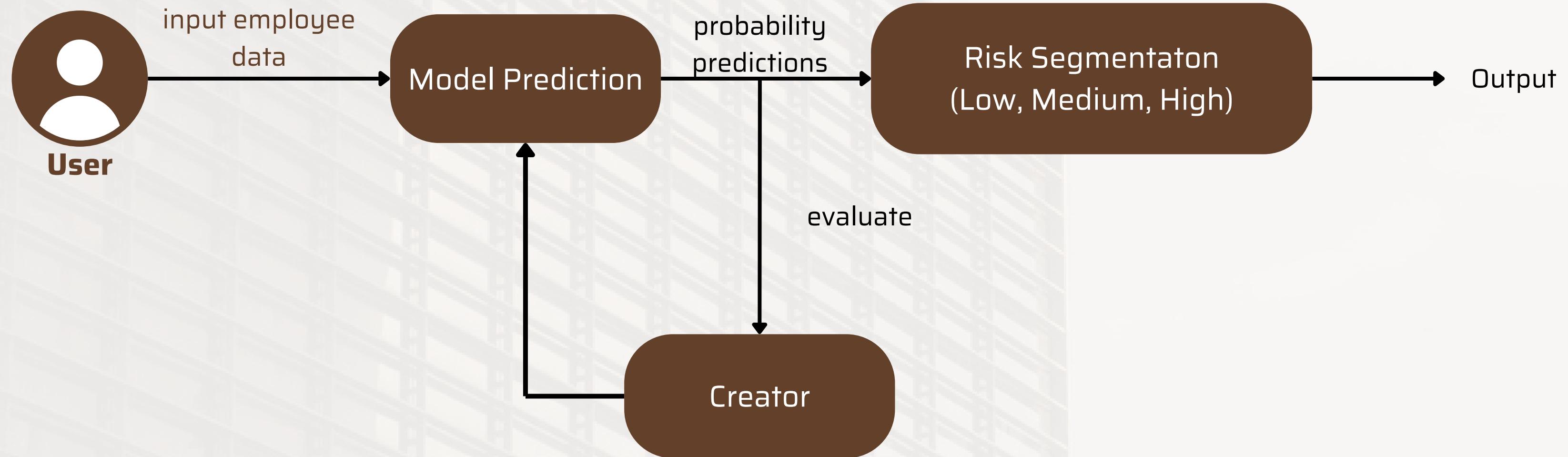
Shows if contract employees are more likely to resign compared to permanent staff



Entry Date and Resignation Date

Helps calculate how long an employee worked before resigning — useful to see if early turnover is a problem

Deployment



The model is deployed using Streamlit to predict employee turnover rates and generate risk-based segmentation

Deployment

Navigation

Prediction

Previous Predictions

Churn Prediction

Target Achievement (%)
26

Job Satisfaction (1-5)
3.70

Manager Support Score (1-5)
2.53

Distance to Office (km)
20

Marital Status
 Single
 Married

Working Hours per Week
40

Company Tenure (years)
1,50

Commission Rate
0,10

Age
27

Predict

Churn Probability: 44%

Risk Group: Medium

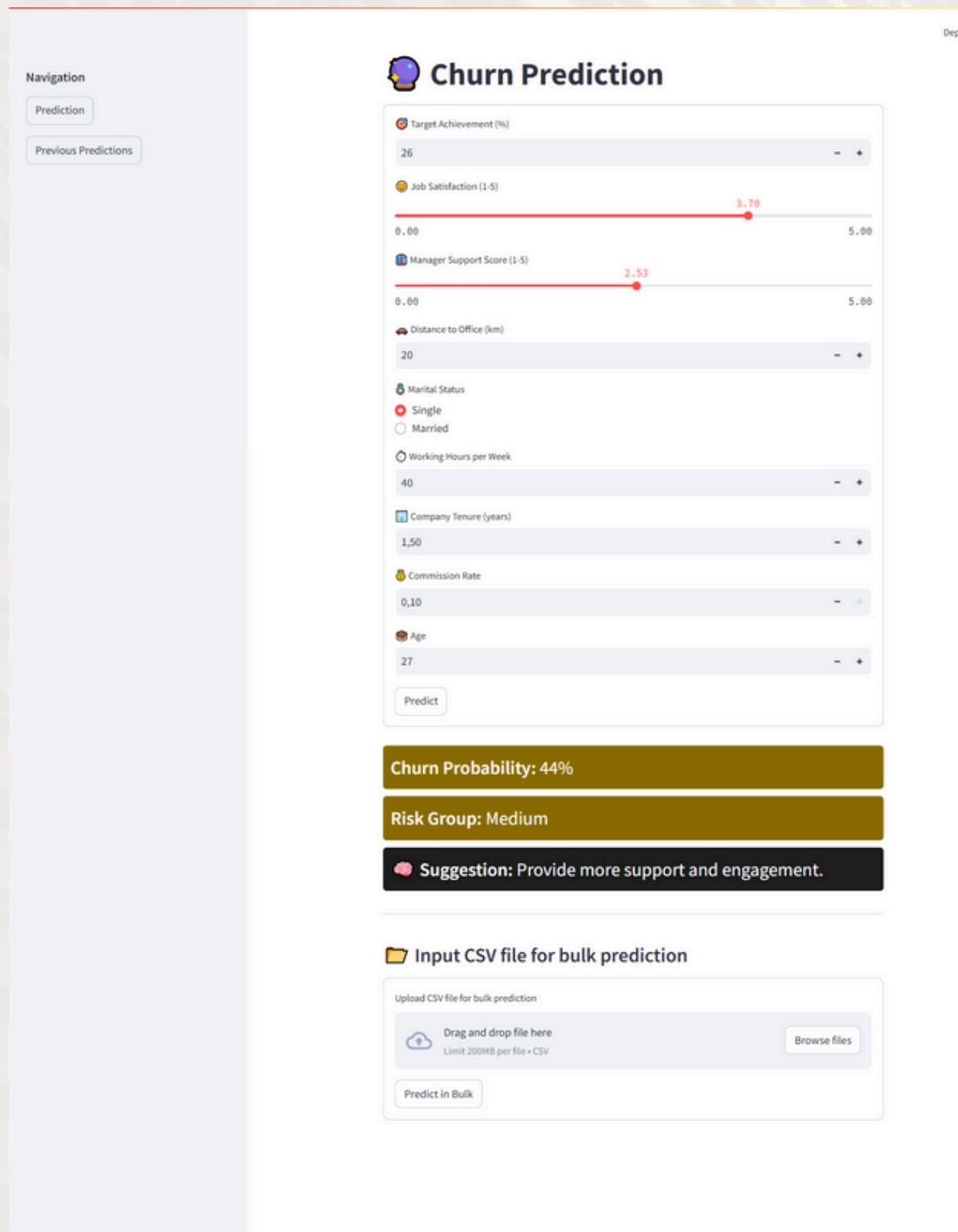
Suggestion: Provide more support and engagement.

Input CSV file for bulk prediction

Upload CSV file for bulk prediction
Drag and drop file here
Limit 200MB per file • CSV

Browse files

Predict in Bulk



[Link Deployment](#)



TalentaHub



Rakamin



Thank You

Gerente General