



FINAL PROJECT CODING STUDIO COHORT 1

Employee Attrition Analysis

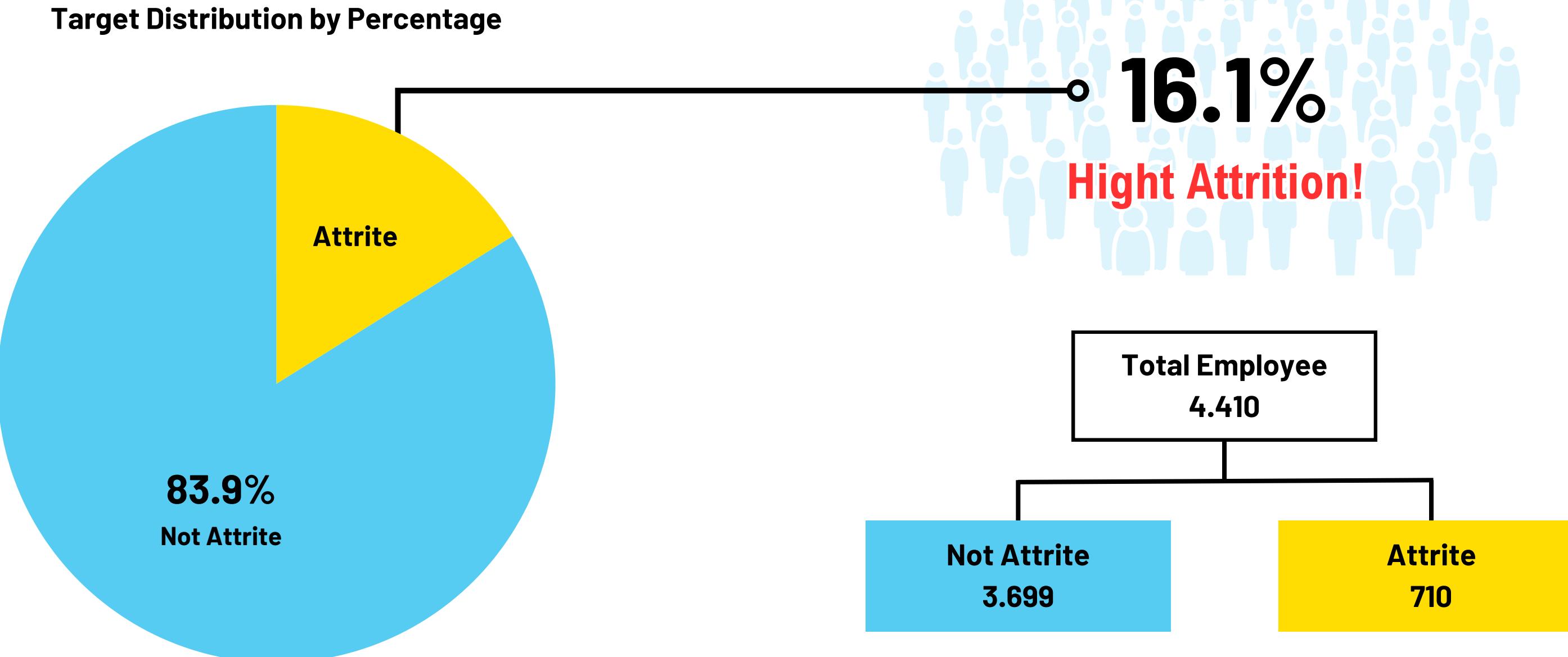
By: Shania Widianingrum Puspitasari

Sourcecode





The Main Problem



What is a good attrition rate for a company?

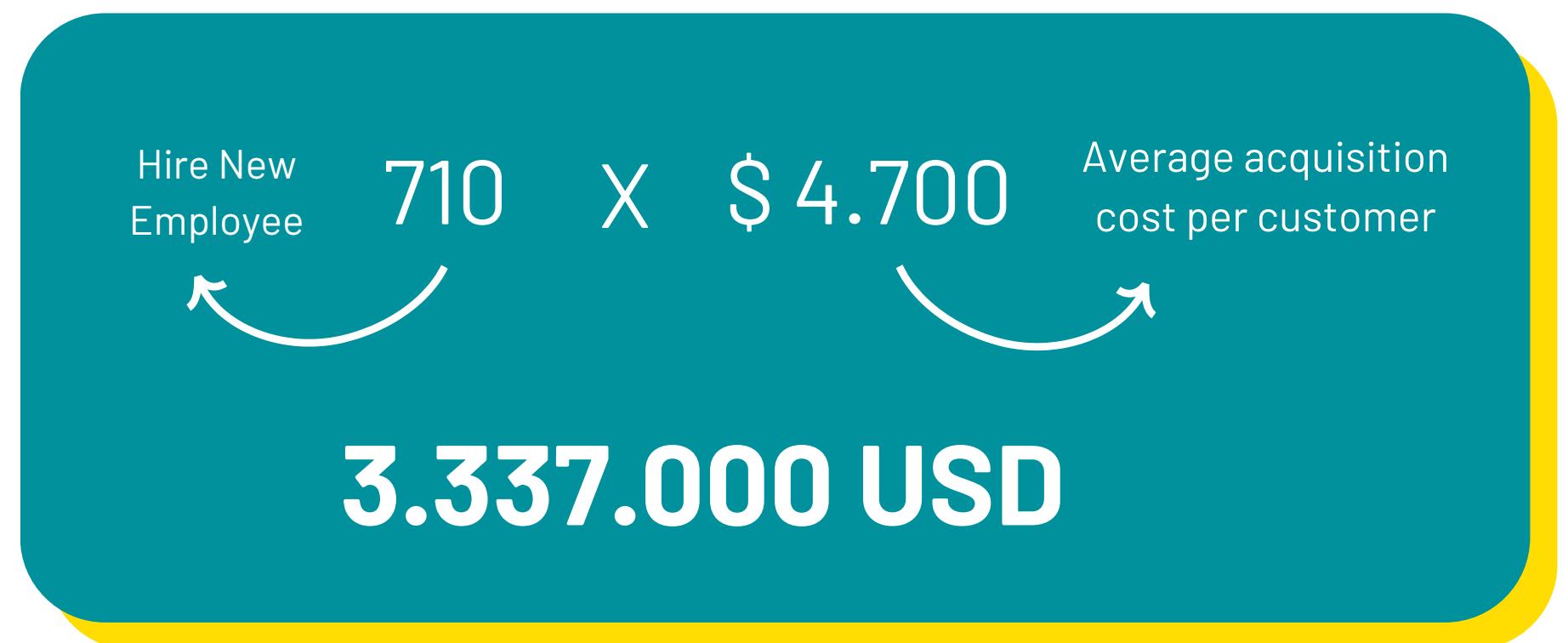
According to experts, healthy organizations have an attrition rate of 10% or less

Source: www.bamboohr.com



The Main Problem

Reqrutment Cost Loss



HIGHT RATE = HIGHT COST

The high rate of employee attrition causes spending to be wasted.

According to data from the Society for Human Resource Management (SHRM), the average cost per hire is nearly \$4,700. **(Snigda Patel, 2024)**

Source: www.businessnewsdaily.com





Problem Summary

Background

The company wants to know **why employee to attrite** and which variables need to be addressed immediately.

Goals

Decrease the attrition rate from **16.1%** to at least **under 10%** in order to reduce cost

Objectives

Create a model machine learning to predict **potential causes of employee attrition**

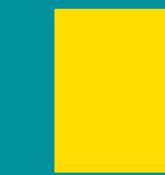
Business Metrics

Attrition Rate (%)

Exploratory Data Analysis



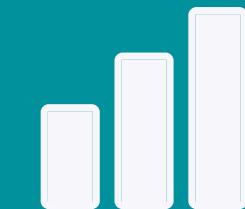
About Dataset



Churn by Feature



About Dataset



The dataset consist of **4.410 rows** and **29 columns**



It have **missing value** and **no duplicated** data

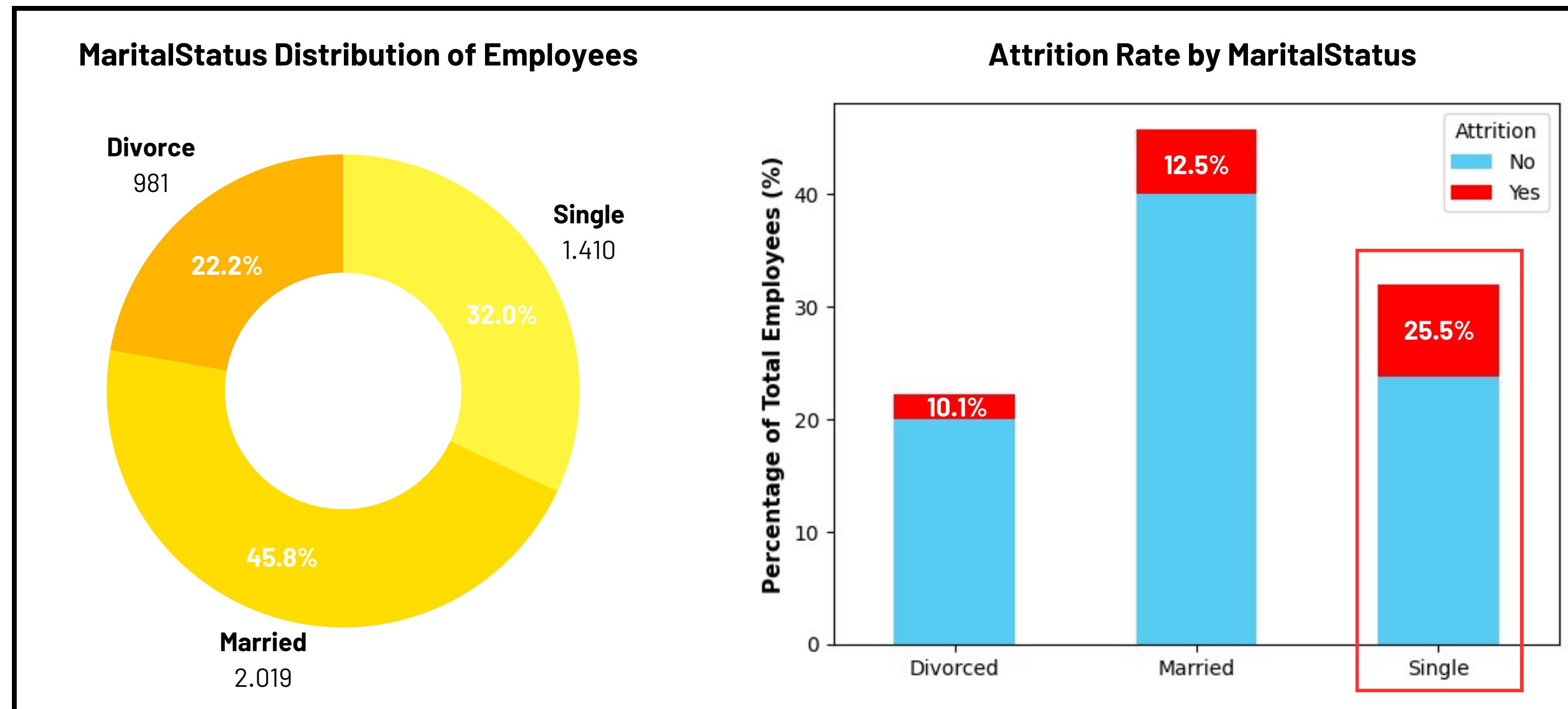


Target variabel is the **Attrition** columns



Attrition by MaritalStatus

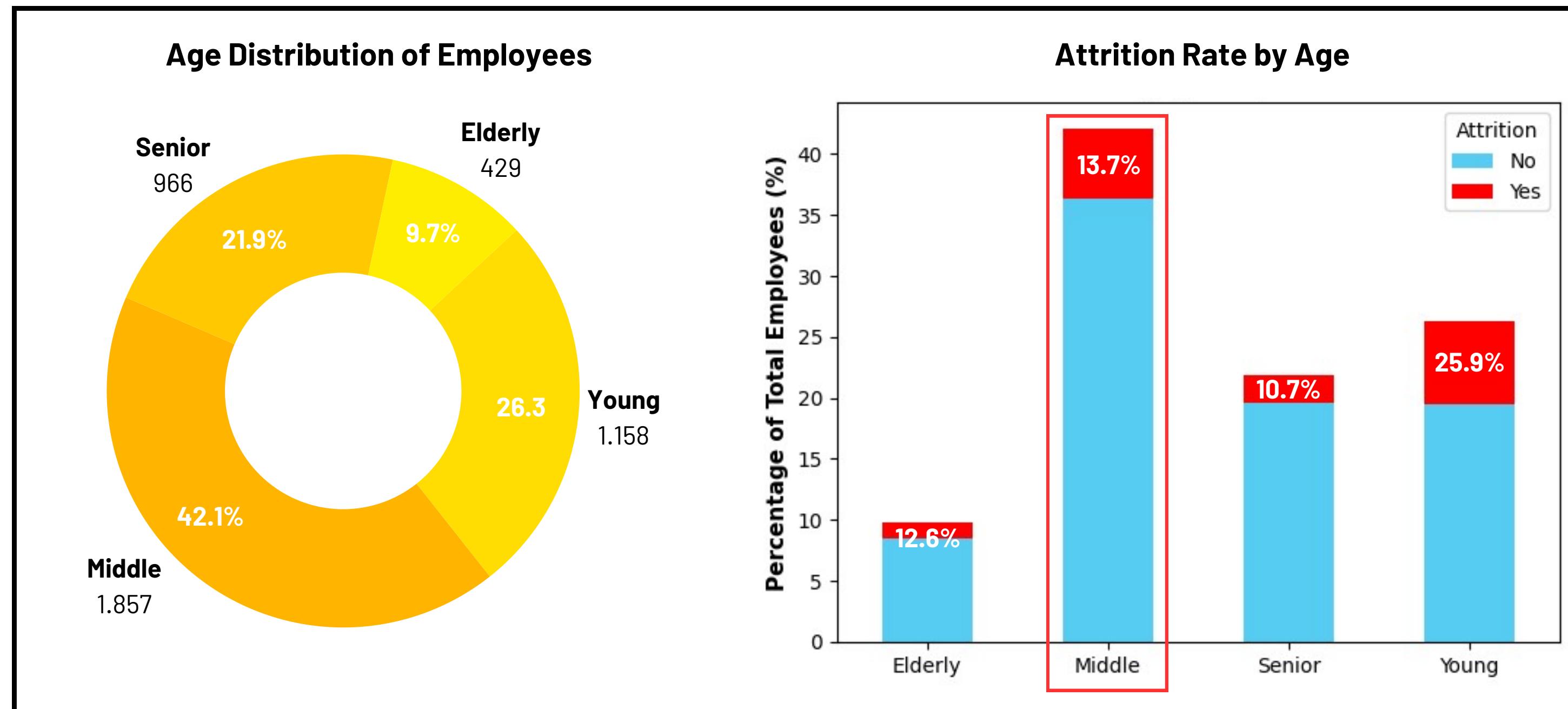
The **majority** of employee are **Married**, the lowest are Divorce, and the **highest attrition rate** is **Single..**





Attrition by Age

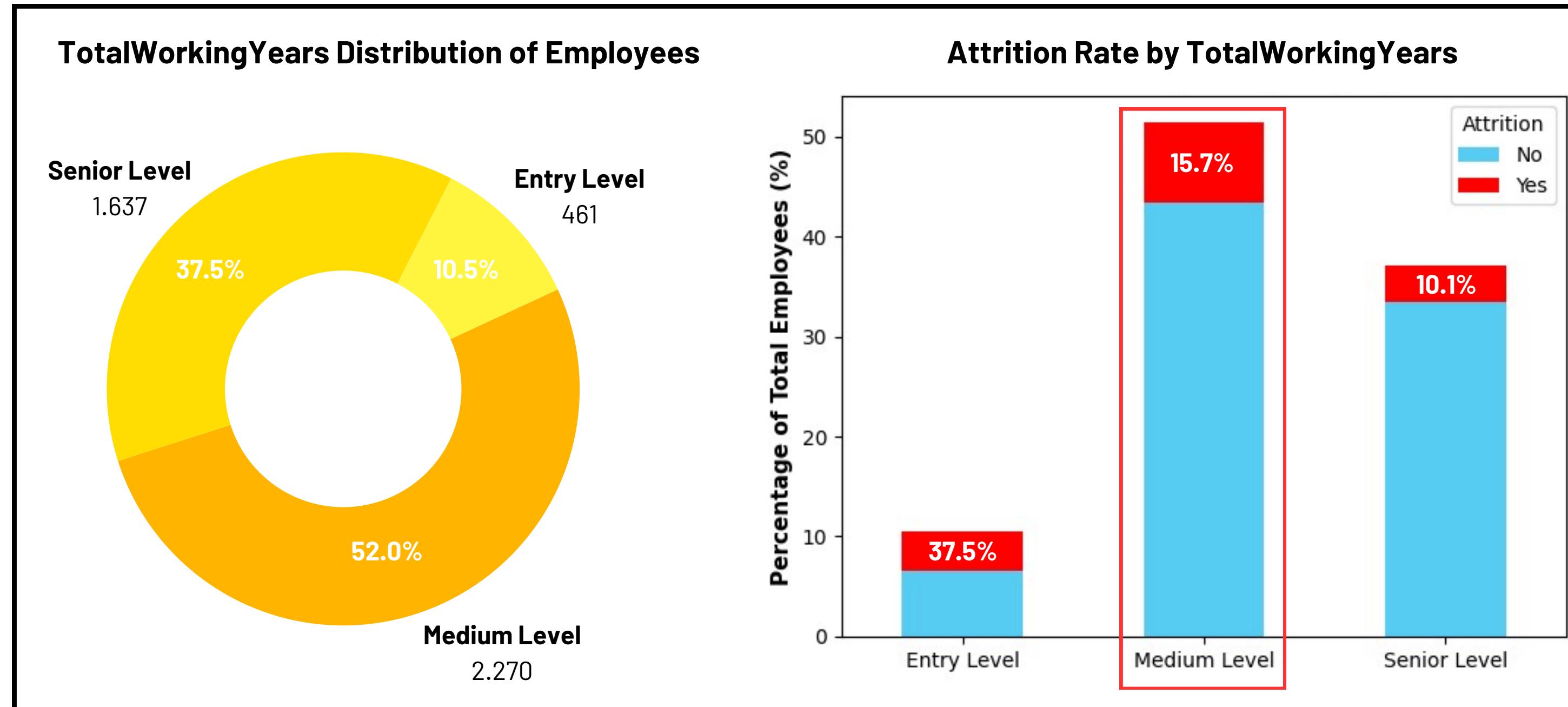
The **majority** of employee are **Middle**, the lowest are **Elderly**, and the **highest attrition rate** is **Middle**.





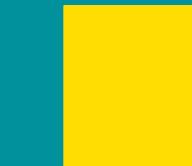
Attrition by TotalWorkingYears

The **majority** of employee are **Medium Level**, the lowest are Entry Level, and the **highest attrition rate** is **Medium Level**.





Data Preprocessing



Preprocessing Flow



Preprocessing Flow



START

Missing Values Duplicated Data Feature Extraction Feature Selection

Handling missing value:

- NumCompaniesWorked
- TotalWorkingYears
- EnvironmentSatisfaction
- JobSatisfaction
- WorkLifeBalance

No duplicated data

No new feature to
be extracted

- Mutual information
- Classification
- Chi Square Test
- Anova

Drop columns:

- EmployeeCount
- EmployeeID
- Over18
- StandardHours
- EducationField_Medical

Handle Outlier

Handling outlier
using **logarithmic**
transformation

Data Train
80%

Train Test Split

Data Test
20%

Feature Engineering

One hot encoding

- BusinessTravel
- Department
- EducationField
- JobRole
- MaritalStatus
- Gender

Feature Transformation

Transform features
using **StandardScaler**

Imbalance Data

Handling imbalance
data using **SMOTE**

Data Train Final
80%

Modelling



FINISH

Machine Learning Modelling



Model Evaluation



Feature Importance

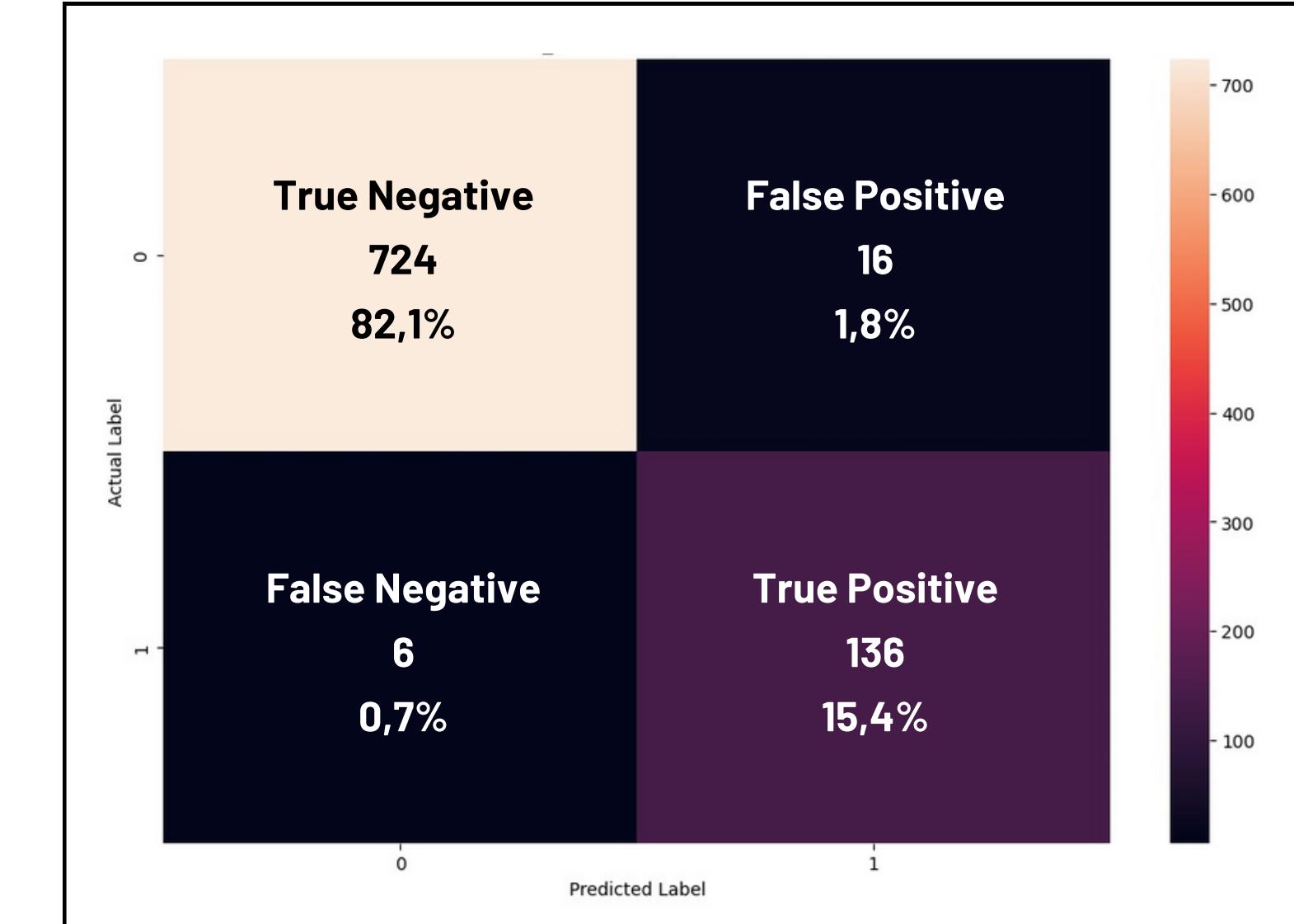
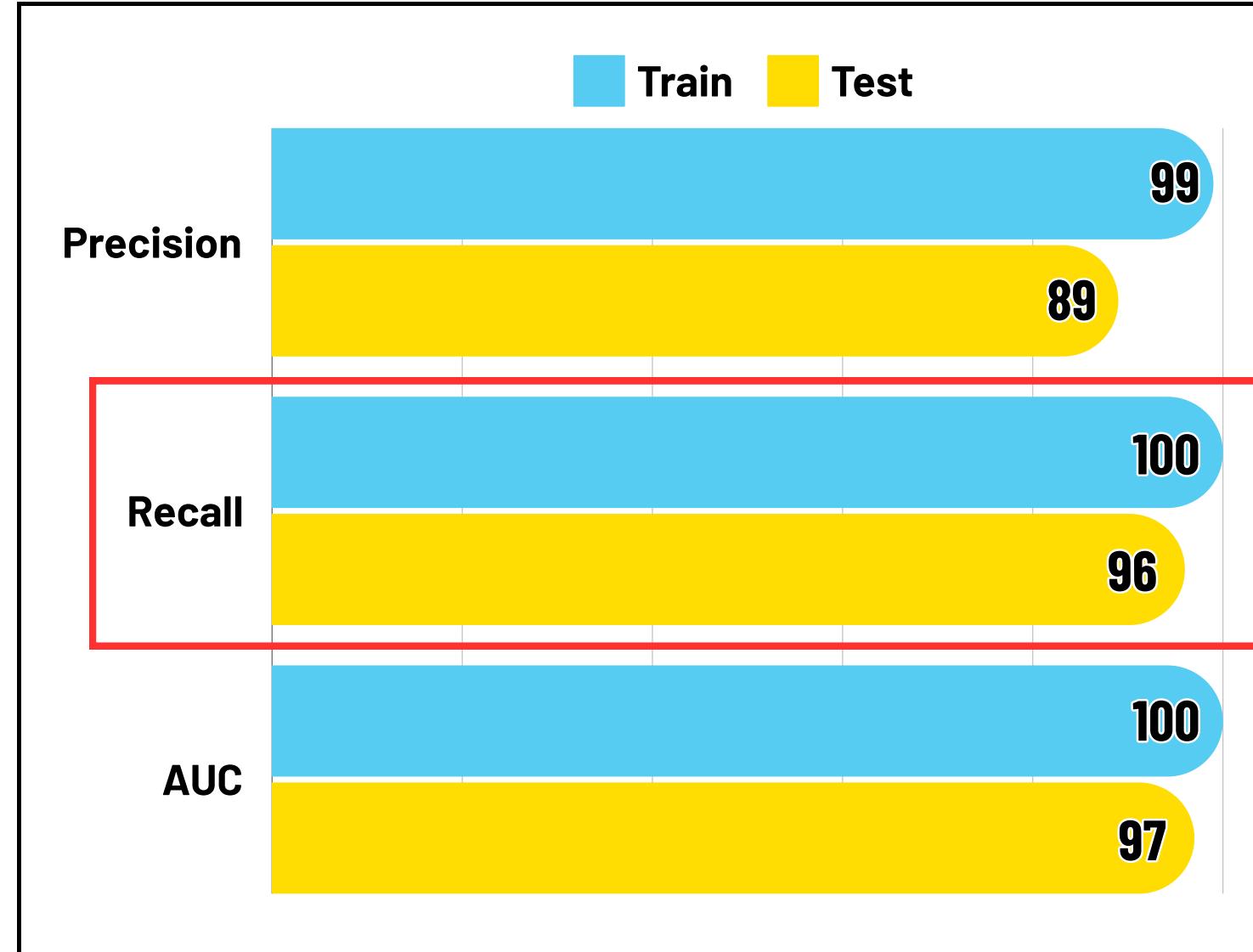


Model Evaluation

MODELS	PRECISION (TRAIN)	PRECISION (TEST)	RECALL (TRAIN)	RECALL (TEST)	AUC (TRAIN)	AUC (TEST)
LOGISTIC REGRESSION	0.39	0.33	0.60	0.51	0.79	0.73
DECISION TREE	0.96	0.88	1.00	0.94	1.00	0.97
ADABOOST	0.30	0.27	0.65	0.60	0.76	0.68
RANDOM FOREST	0.24	0.22	0.61	0.56	0.68	0.63
XGBOOST	0.99	0.89	1.00	0.96	1.00	0.97



Model Evaluation

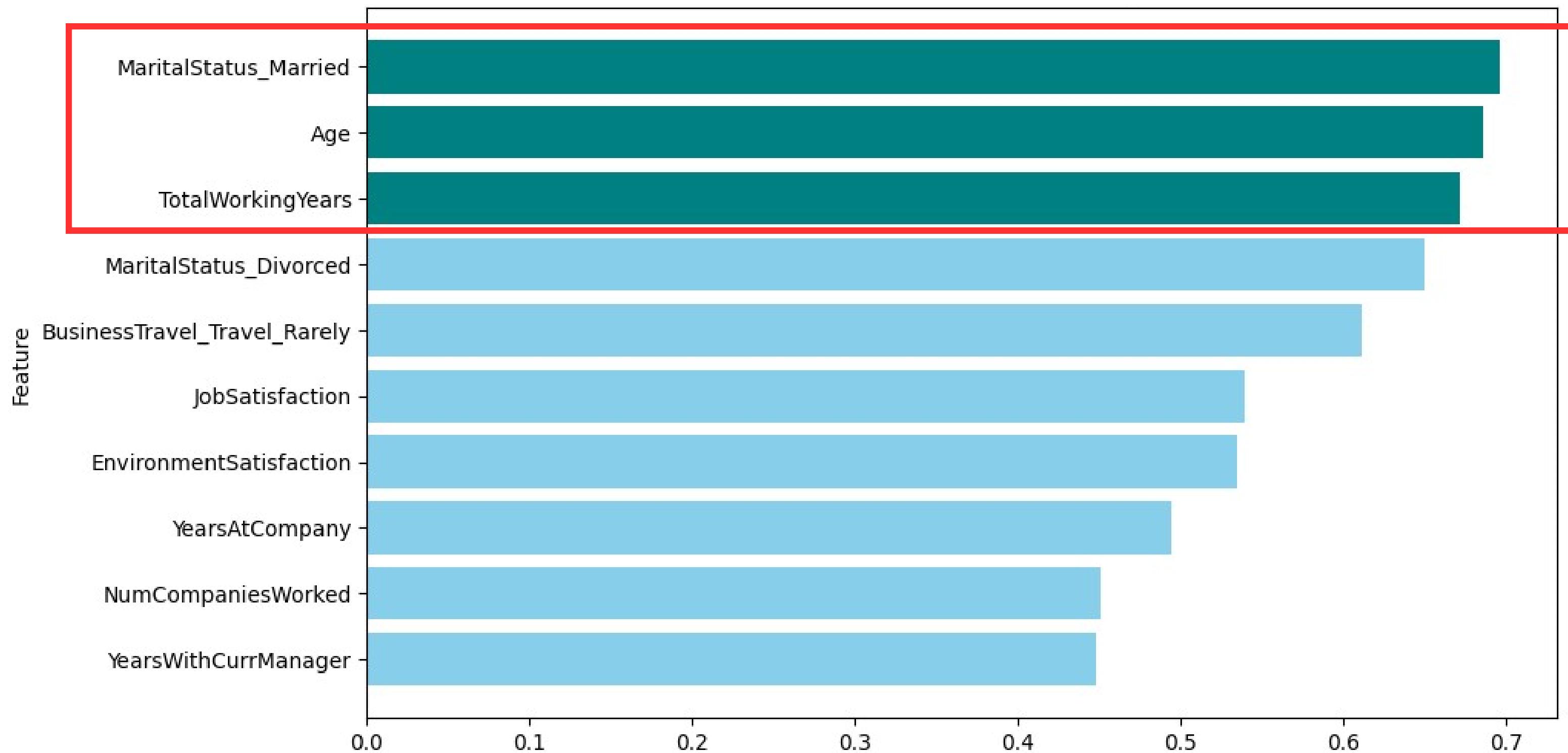


Recall **measures the ratio** between the number of employee that are **predicted** to **attrite** who actually left the company, and the number of employee that are **predicted** to **stay** but **actually** left the company.

Maximizing recall means minimizing the number of employee who are **incorrectly** predicted to not attrite.



Feature Importance





Business Recommendation



Recommendation



Business Impact



Recomendation for Marital Status



- Mengembangkan **program manfaat karyawan** dengan menawarkan manfaat tambahan seperti asuransi kesehatan keluarga, cuti tambahan untuk peristiwa keluarga, atau dukungan pendidikan untuk pasangan atau anak-anak karyawan.
- Mengembangkan **program dukungan keluarga** yang mencakup konseling keluarga, bantuan perencanaan keuangan, atau fleksibilitas waktu kerja untuk membantu karyawan menyeimbangkan antara pekerjaan dan kehidupan pribadi mereka



Recomendation for Age



- Mengembangkan program pelatihan yang dirancang khusus untuk generasi tertentu, mentoring lintas-generasi, atau kesempatan pembelajaran online yang fleksibel
- Mengembangkan konseling kesehatan mental, program kebugaran fisik, atau seminar kesehatan untuk masalah yang relevan dengan usia.
- Menawarkan opsi tabungan pensiun tambahan atau asuransi kesehatan dengan cakupan tambahan untuk karyawan yang mendekati pensiun.



Recomendation for Total Working Years



- Mengembangkan pelatihan tambahan, kursus pengembangan keterampilan, atau program mentoring untuk membantu karyawan meningkatkan kompetensi mereka dan memajukan karir.
- Merancang program pengelolaan pengetahuan internal atau forum kolaborasi untuk memfasilitasi pertukaran pengetahuan dan pengalaman antara karyawan.
- Mengembangkan program retensi khusus yang menawarkan insentif atau keuntungan tambahan kepada karyawan berpengalaman untuk mempertahankan mereka dalam jangka panjang.



Business Impact

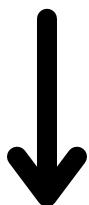
136 = number of employee we correctly predict to attrite (True Positive)

6 = number of employee we misclassified as people who stay (False Negative)

882 = total employee in data test

Assumption 50%

$$\text{Attrition Rate} = (0.5(\text{TP}) + \text{FN}) / \text{Total}$$



Before Model

16.1%



710

Decrease of

7.7%



340

48% ↓

After Model

8.4%

370



Business Impact

Assumption average **recruitment cost** is \$4,700

Assumption **marketing cost** is \$100

Before model

710 x 4,700 = 3.337.000 USD

After model

370 x 4,700 = 1.739.000 USD

Retention Cost

152 x 2,350 = 357.200 USD

1.240.800 USD



Saving Money



Thank You
For Your Attention

