

PYTHAGORS

Proudly Present



ATTRITION RATE PREDICTION

Pharmaceutical Company Attrition Problem to
be solved.

[Explore More](#)

29/01/2022



Faisal Adam Y

Leader of Pythagoras



Jehezkiel Natan

Member of Pythagoras



Anoga Rizky

Member of Pythagoras



Esa Tauran

Member of Pythagoras



Danienta A.

Member of Pythagoras

Table of Content

01

Project Background

- Problem Background
- Goal & Objectives

02

Insight after Data Preparation Stage

Show important insight from EDA

04

Modelling

A/B Testing Algorithm
Modelling (3 Dataset & 6 Algorithm Model)

03

Data Preparation

- Dataset Explanation (Feature & Target)
- Data Cleansing
- Feature Engineering

05

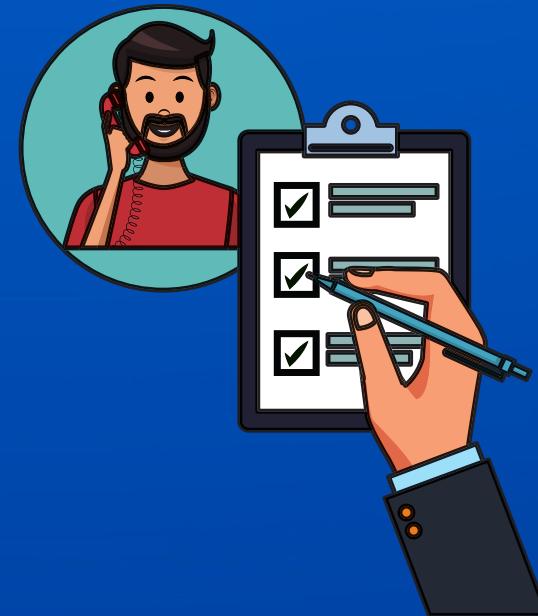
Simulation and Recommendation

- Simulation to Project Goal
- Recommendation to Stakeholders

Our Job Role

A Brief Story About Our Client Problem

As one of the biggest Pharmaceutical company, PT. Goras have attrition problem to manage their human resource. In order to help their problem, as Human Resource Consultant, Phytogoras have responsibility to give them solution.





Project Background

Pitch Deck By

Pythagoras Inc.

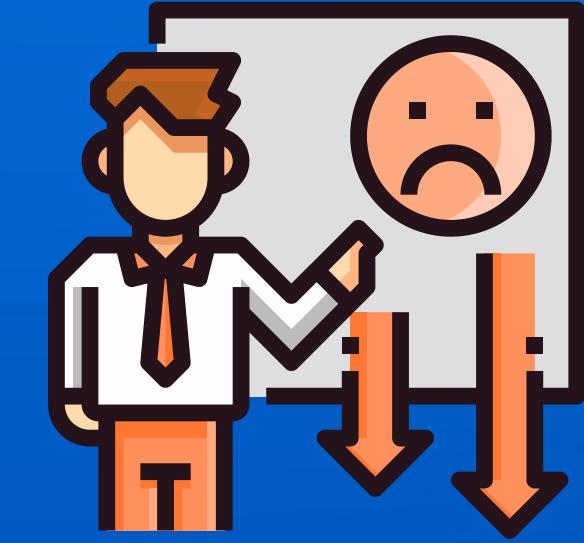
Rakamin Academy

Problem Statement



Employees choose to leave:
16.4%

Average rate in Indonesia:
6.3%



**Brings disadvantage for
PT.Goras**

We Are Here To Bring You The Solutions

Goals



Reduced the number of Attrition Rate by
2-3%

Objectives

- Get the insight and prediction from indicators about why employees choose to leave with Machine Learning
- Give Recommendation to PT. Goras about reducing the attrition rate

Business Metrics

- Attrition Rate

Preview Dataset

1470 data of employees

Employee Profile

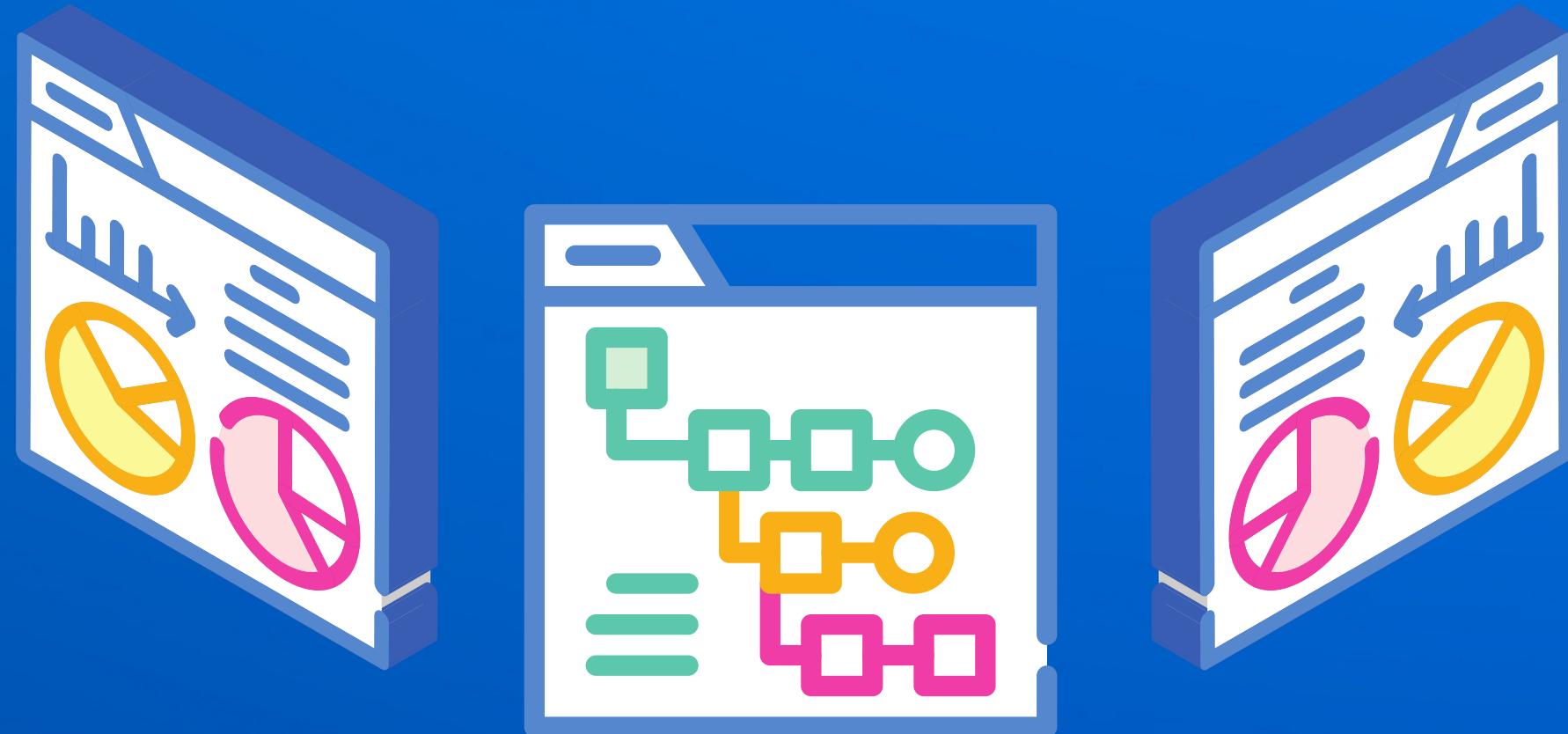
Age
Gender
Marital Status
Education
...
etc.

Employee Status

Job Role
Job Level
Department
Years at Company
...
etc.

Employee Performance

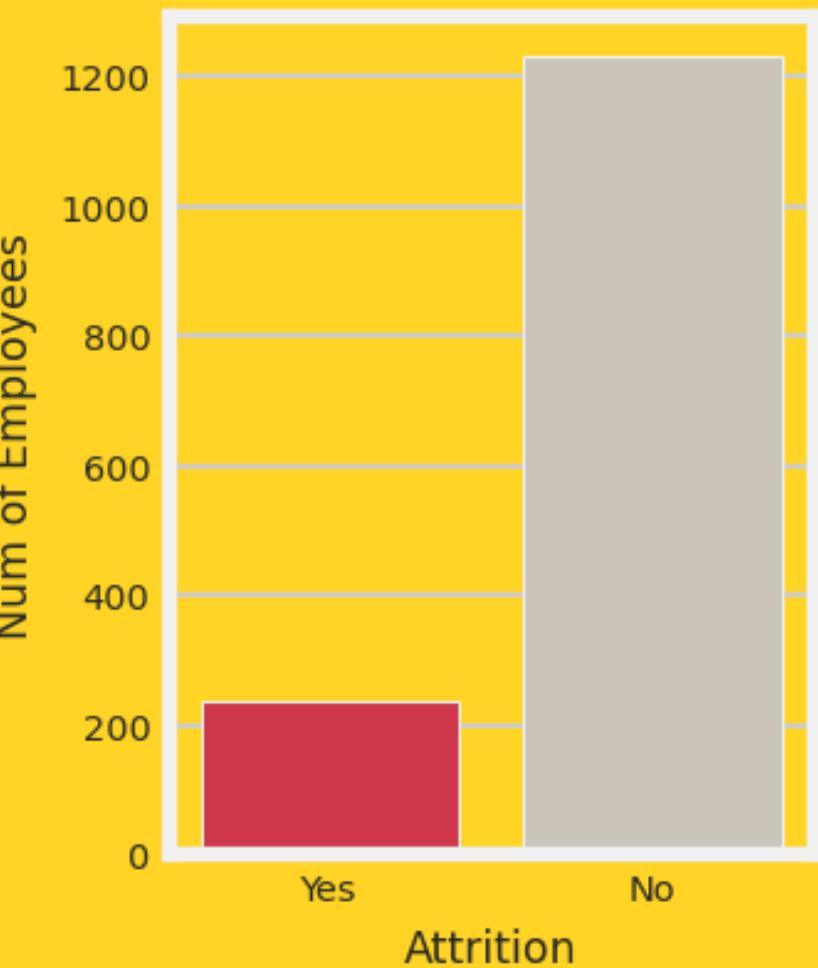
Performance Rating
Job Involvement
Job Satisfaction
Overtime
...
etc.



Exploratory Data Analysis

Attrition Rate

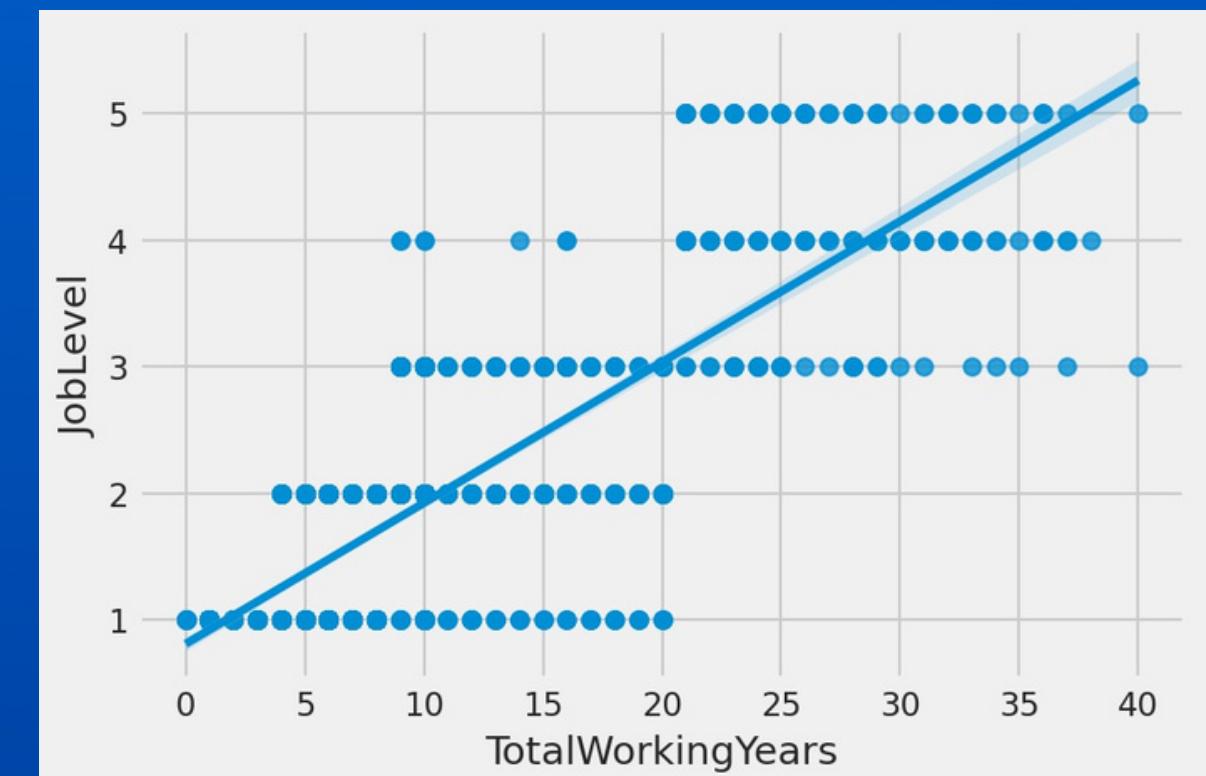
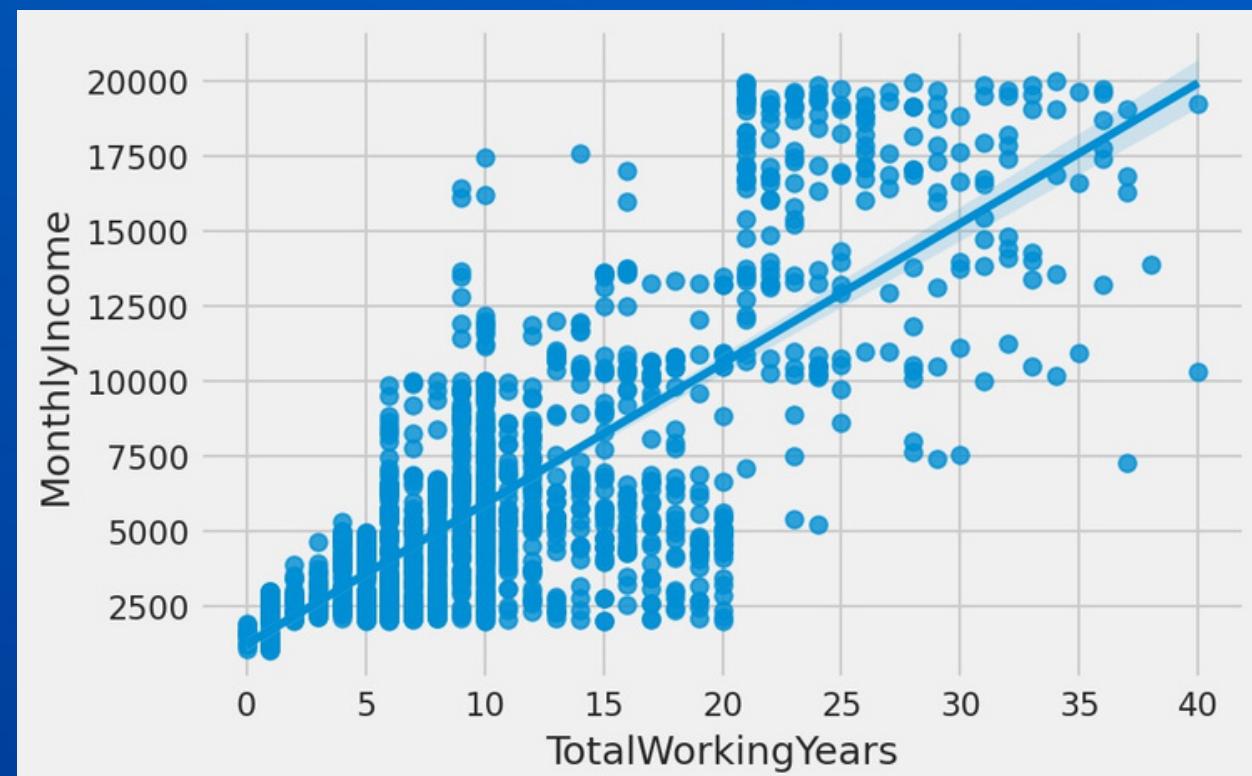
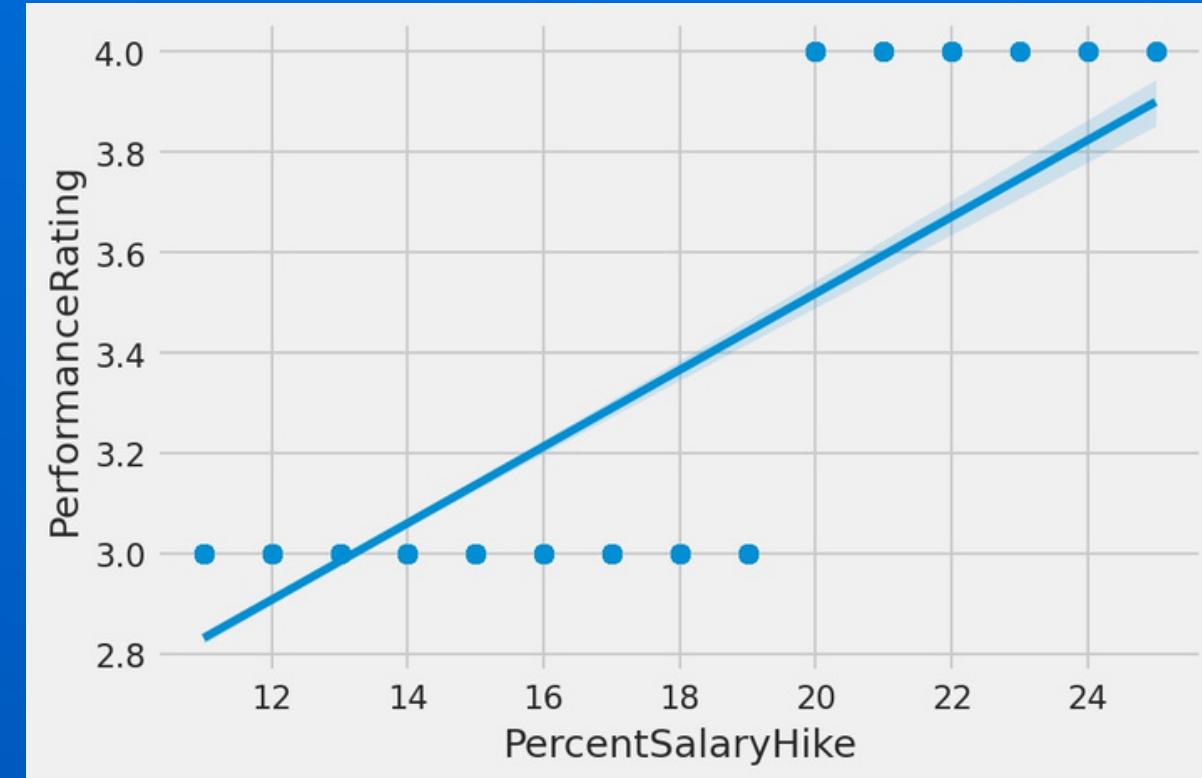
The Attrition Rate of PT Goras is 16.4%



Formula =
Attrition Rate % = $\frac{\text{No. of employee Resigned}}{\text{Total employees}}$

Multivariate Analysis

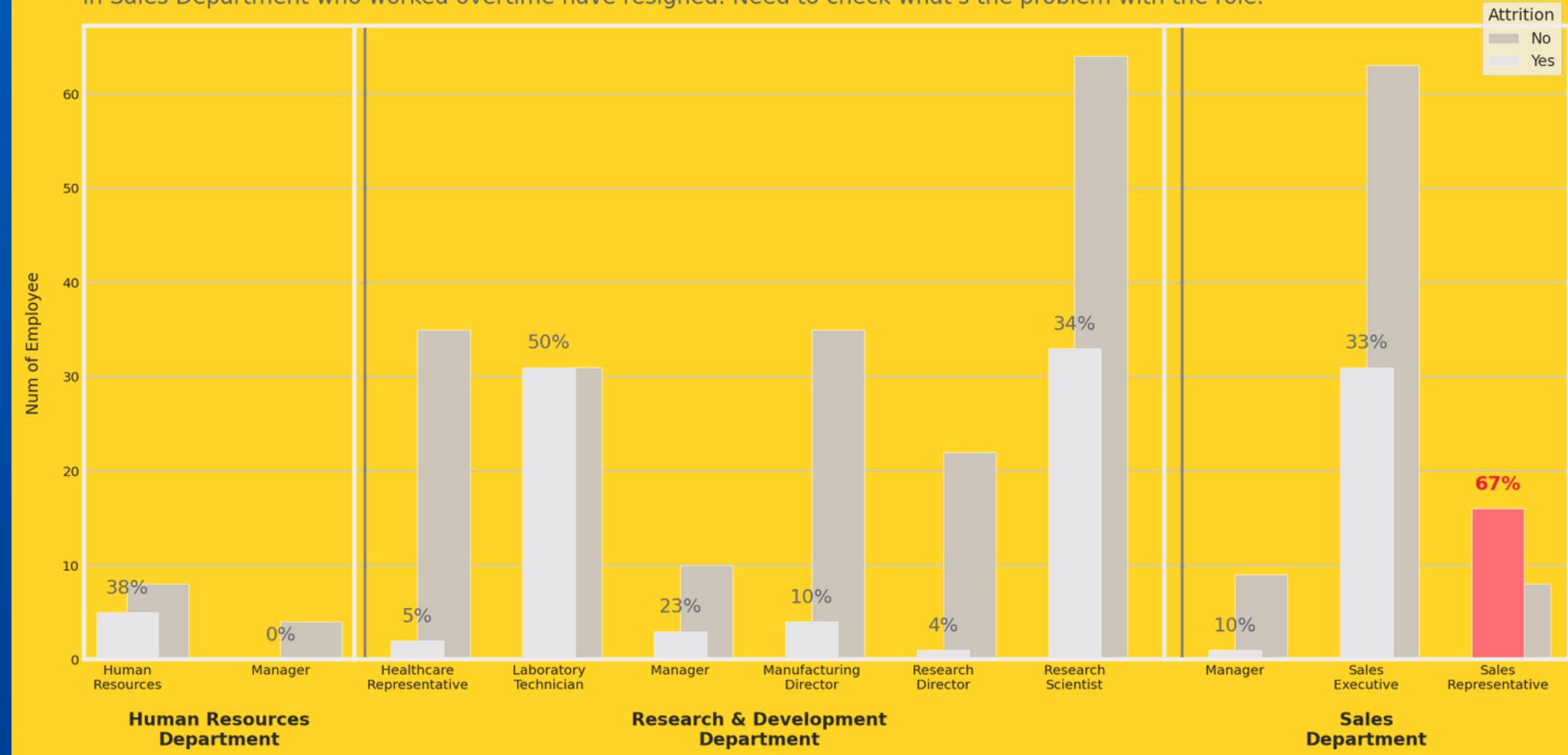
- Salary has a great influence on the employee's performance
- The longer employees work in PT. Goras, the more pay they will get
- Most employees who are in the higher position have worked for PT Goras for a long time



Data Invention of Exploratory (1)

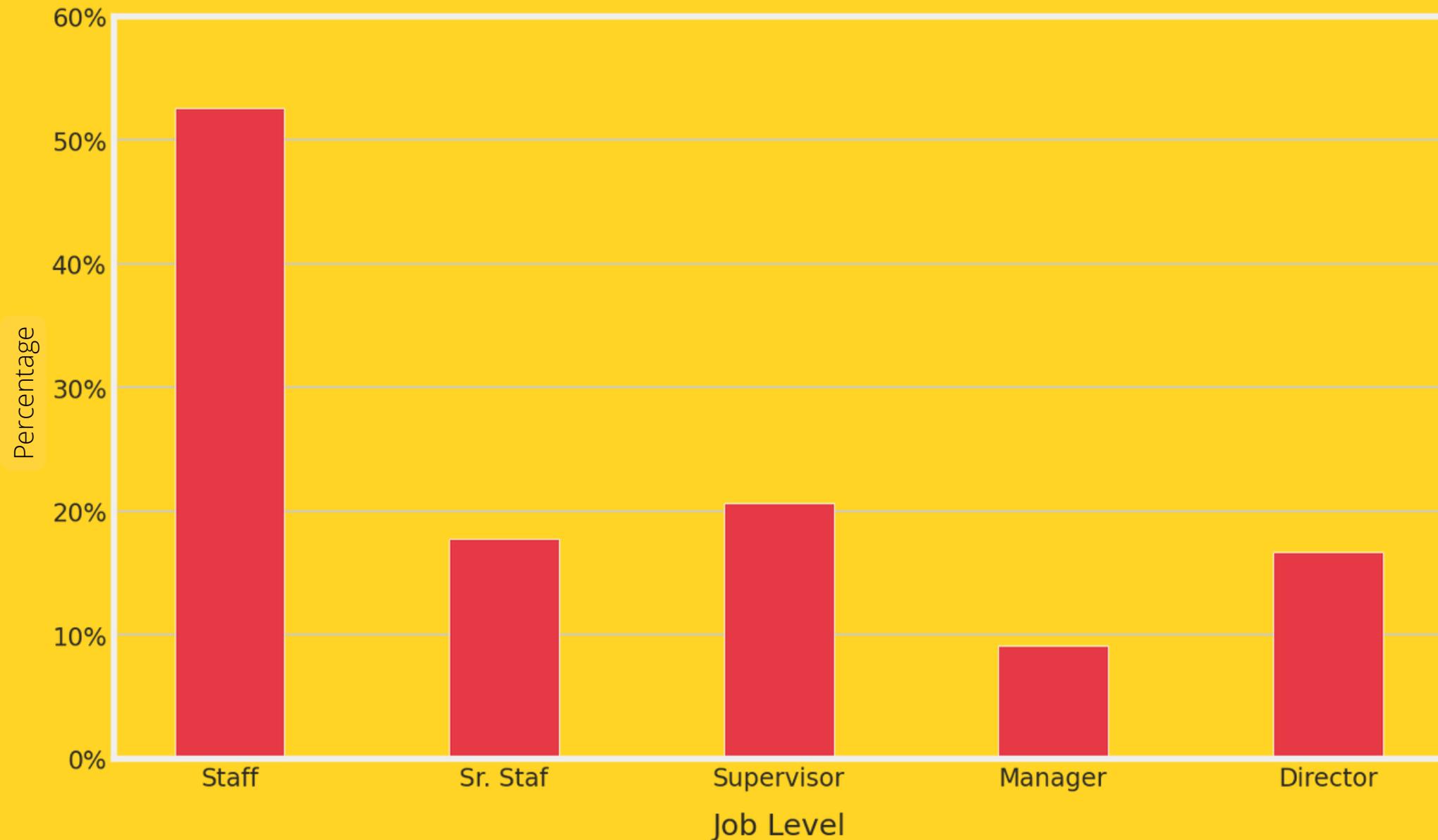
Most of the Sales Representative in Sales Department who Worked Overtime Have Resigned

The attrition rate of employees working overtime in each role in each department is below or equal 50%. But two thirds of the Sales Representatives in Sales Department who worked overtime have resigned. Need to check what's the problem with the role.



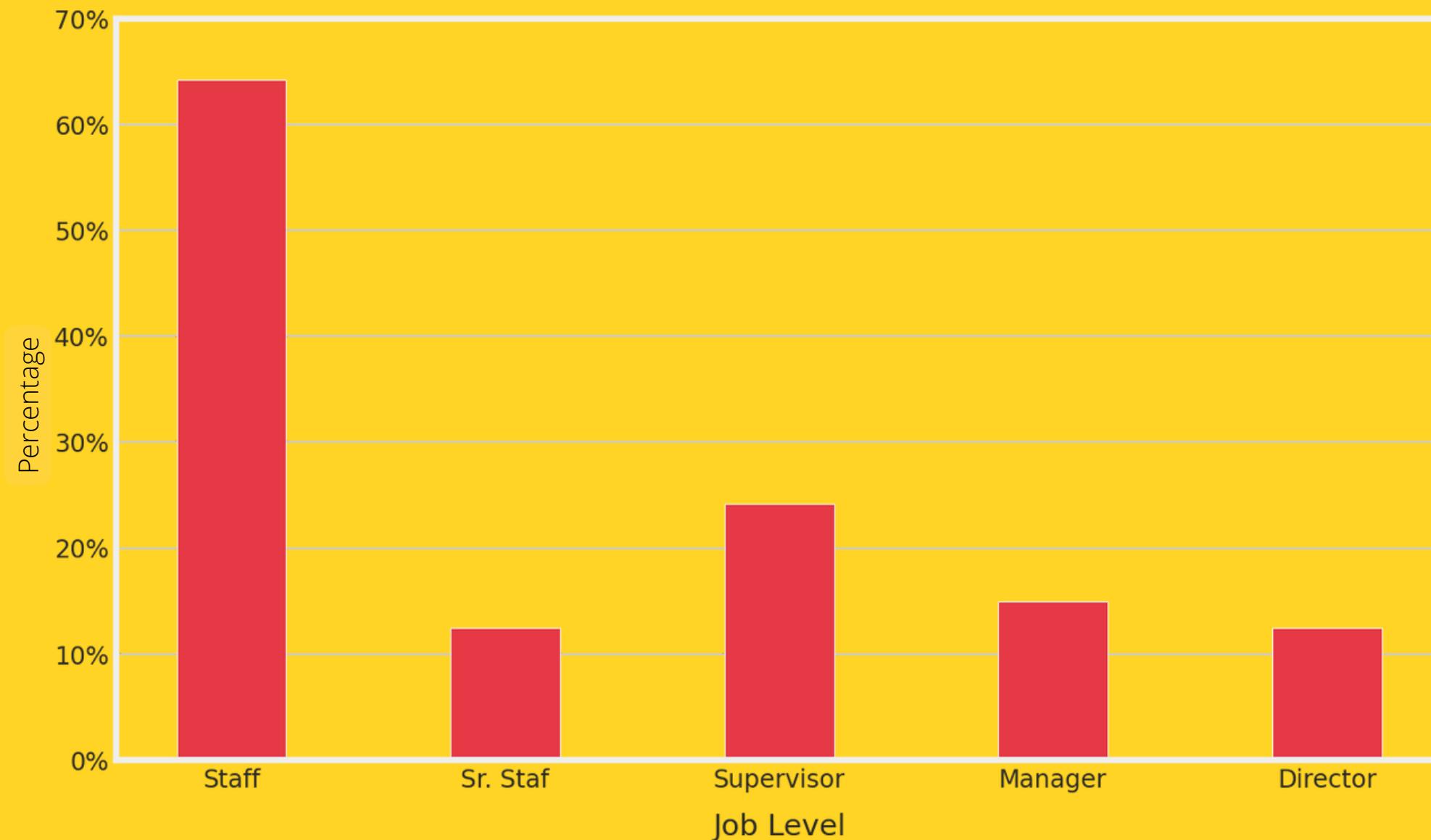
Data Invention of Exploratory (2)

Attrition Rate by Job Level of Overtime Employees



Data Invention of Exploratory (3)

***Attrition Rate by Job Level of Employees
who Work Overtime and Below the Median Monthly Income***



EDA Result



01

· 67% of sales representative in sales department who worked overtime have resigned, followed by Laboratory Technician

02

· Employees who are in the age group of 18 – 30 years old, and are in job level 1 have the highest attrition rate compared to other groups

03

· Overtime is the number one key factor in sales and R&D Department

04

· In HR Department, the distance from home is more important to them

Why do we need Machine Learning Model?



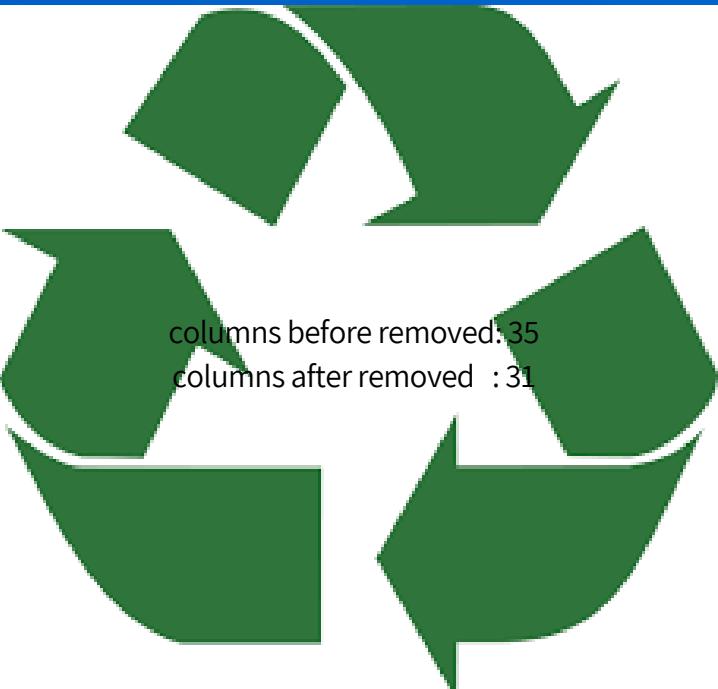
Data Preparation

Pitch Deck By

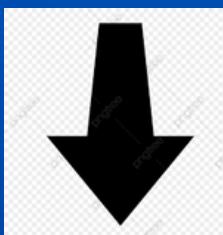
Pythagoras Inc.

Rakamin Academy

Why do We need Preparation?



**Data
Exploration**

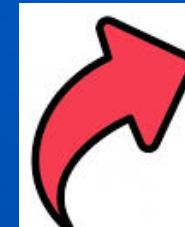


Modelling

Handle Single and Unique Data

EmployeeCount	1
Over18	1
StandardHours	1
EmployeeNumber	1470

columns before removed: 35
columns after removed : 31



```
df_clean.drop(['EmployeeCount', 'Over18', 'StandardHours', 'EmployeeNumber'])
```

Handle Missing Data & Duplicated Data

Age	0
Attrition	0
BusinessTravel	0
DailyRate	0
Department	0
DistanceFromHome	0
Education	0
EducationField	0
EnvironmentSatisfaction	0
Gender	0
HourlyRate	0
JobInvolvement	0
JobLevel	0
JobRole	0
JobSatisfaction	0
MaritalStatus	0
MonthlyIncome	0
MonthlyRate	0

NumCompaniesWorked	0
Overtime	0
PercentSalaryHike	0
PerformanceRating	0
RelationshipSatisfaction	0
StockOptionLevel	0
TotalWorkingYears	0
TrainingTimesLastYear	0
WorkLifeBalance	0
YearsAtCompany	0
YearsInCurrentRole	0
YearsSinceLastPromotion	0
YearsWithCurrManager	0



`df_clean.isna().sum()`

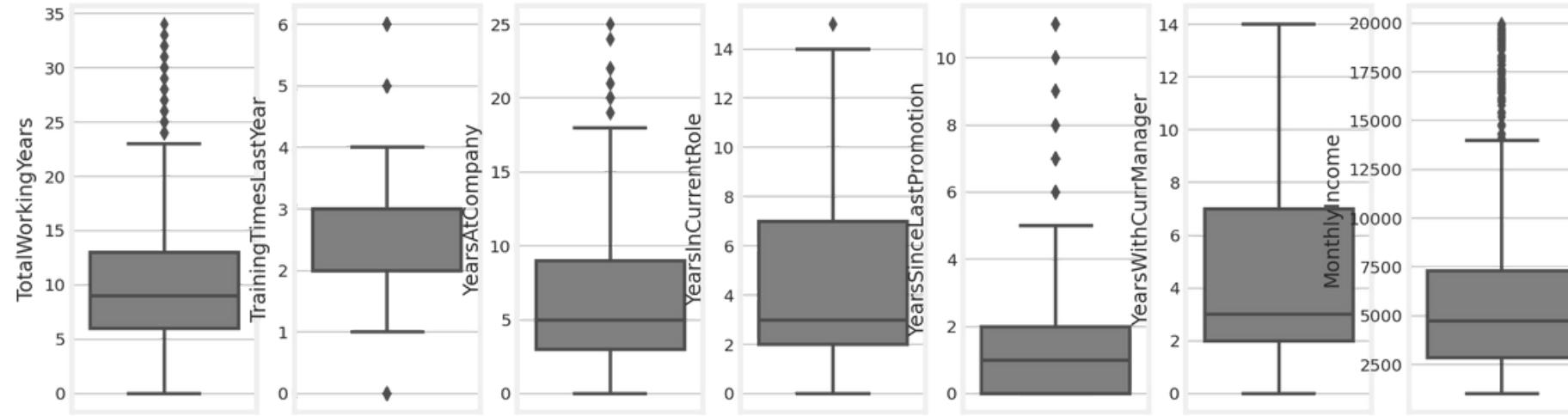
`df_clean.duplicated().sum()`

= 0 (Zero)

Fortunately,
We don't have missing and
duplicated data

Remove Outliers

Outliers After Cleansing by Z-Score



Total row before filtering the outliers:

1470

Total row after filtering the outliers:

1387

5,6% data removed

Total row before filtering the outliers:

1470

Total row after filtering the outliers:

1024

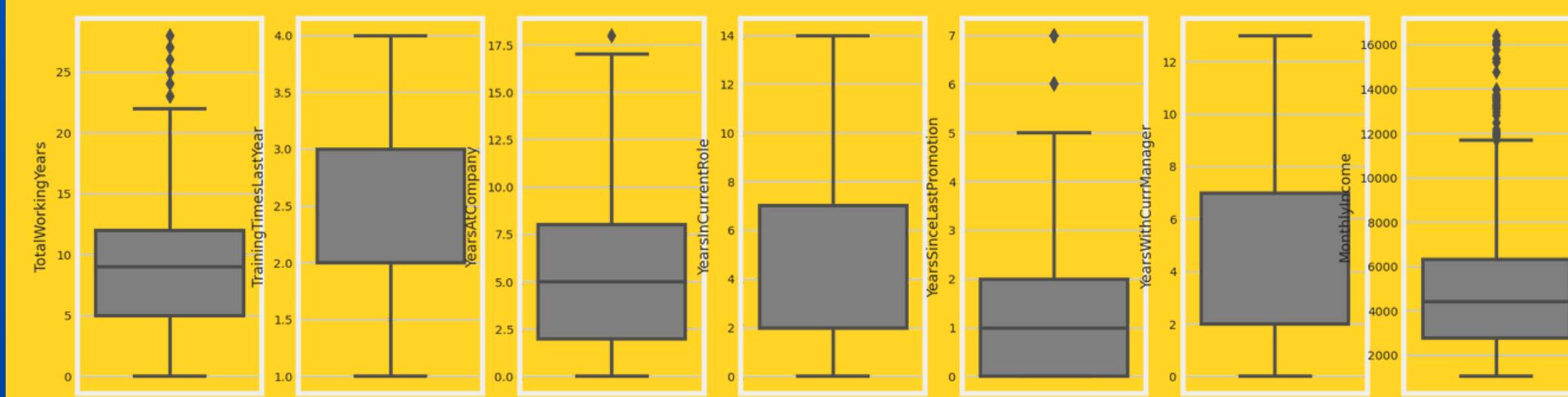
30,3% data removed

Pitch Deck By

Pythagoras Inc.

Rakamin Academy

Outliers After Cleansing by IQR



Feature Engineering

Create New Features

GroupAge	BelowMedIncome	GroupAge_Overtime	JobLevel_Overtime	JobLevel_BelowMedIncome_Overtime
41-50	0	0	0	0
41-50	1	0	0	0
31-40	1	0	1	1
31-40	0	0	1	0
18-30	0	0	0	0

Group Age
Below MedIncome
GroupAge Overtime
JobLevel Overtime
JobLevel BelowedIncome_Overtime

Feature Encoding

label columns : for Columns contain only 2 unique values (yes or no, male or female ,etc)

one hot columns : for Columns contain levels (1,2,3,4,5), ex: Education

Feature Scaling

Standarization : Age Column (normal distribution)

Normalization : Numerical Columns except Age

Feature Engineering

Split Unseen Data

```
print(df_clean.shape)  
print(df_unseen.shape)  
  
(1248, 84)  
(139, 84)
```

Training Data

Validation Data



Handling Imbalance Class

Using Oversampling SMOTE

Initial Number of Each Class:

Attrition
0 1051
1 197
dtype: int64

Number of Each Class After Oversampling (SMOTE):

Attrition
1 1051

There are 6 Dataset and 6 Algorithm that we use for Modeling:

Dataset 1

**2102 row
84 features**

Dataset 2

**2102 row
20 features**

Dataset 3

**2102 row
28 features**

Validation Dataset 1

**139 row
84 features**

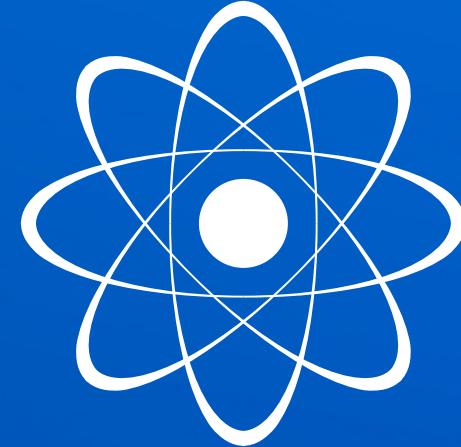
Validation Dataset 2

**139 row
20 features**

Validation Dataset 3

**139 row
28 features**

- Logistic Regression
- K-Nearest Neighbor
- Support Vector Machine
- Decision Tree
- Random Forest
- XGBoost



DATA MODELING

Pitch Deck By

Pythagoras Inc.

Rakamin Academy

Model Evaluation

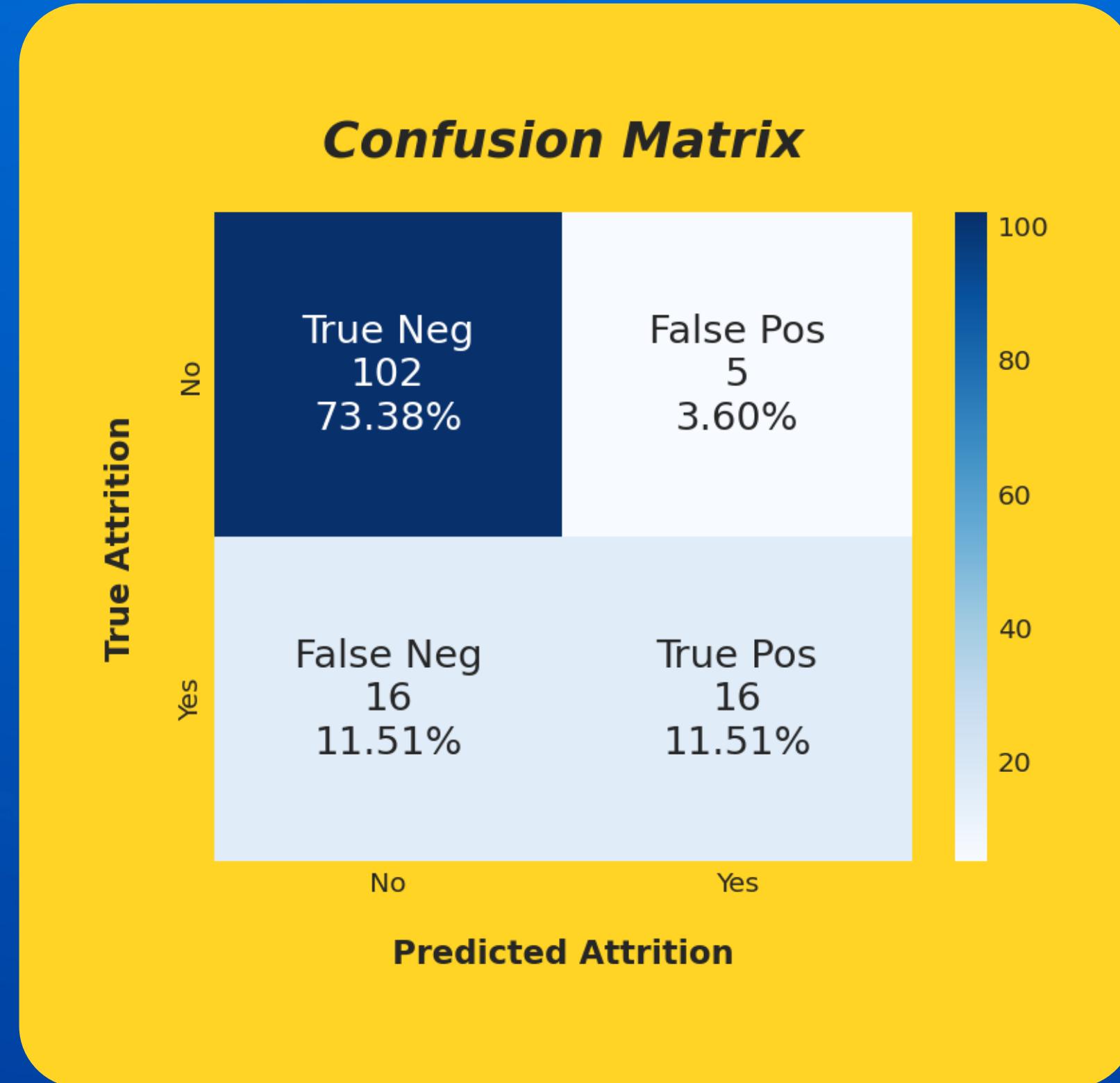
	Accuracy	Precision	Recall	F1	AUC
Logistic Regression (1)	0.848921	0.761905	0.50000	0.603774	0.726636
Support Vector Machine (1)	0.791367	0.548387	0.53125	0.539683	0.700204
Logistic Regression (3)	0.726619	0.434783	0.62500	0.512821	0.691005
Logistic Regression (2)	0.676259	0.385965	0.68750	0.494382	0.680199
XGBoost (1)	0.805755	0.619048	0.40625	0.490566	0.665742

(1) = Dataset 1; (2) = Dataset 2; (3) = Dataset 3

From the comparison table, we found that
Logistic Regression with Dataset 1 has the highest F1 Score.

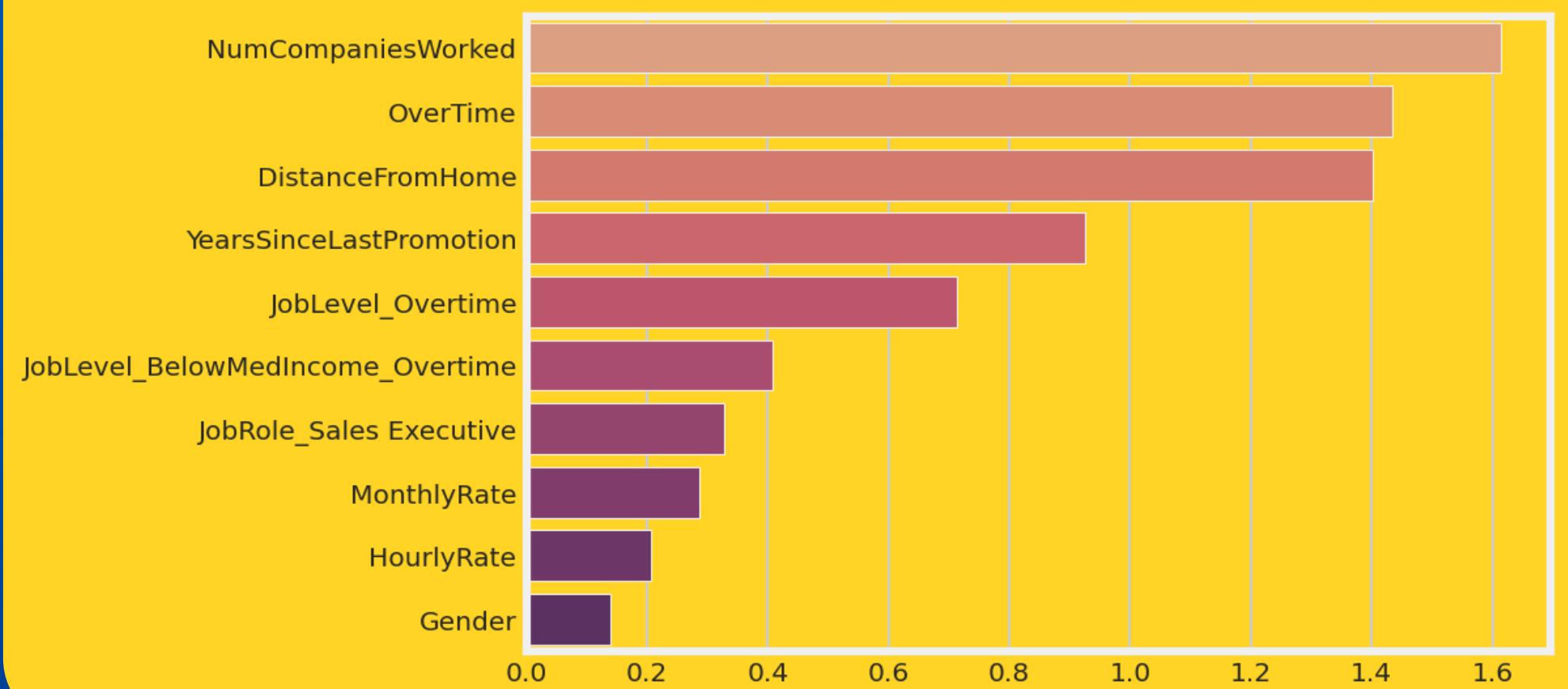
Final Training Model

Doing Training Model using Logistic Regression
algorithm with scenario of Dataset 1
All Features



Coefficients Feature

Top 10 Feature Coefficients



Based on the top list of Coefficient Feature, we decided to use 'OverTime' and 'YearsSinceLastPromotion' as key point of simulation.

RECOMMEND



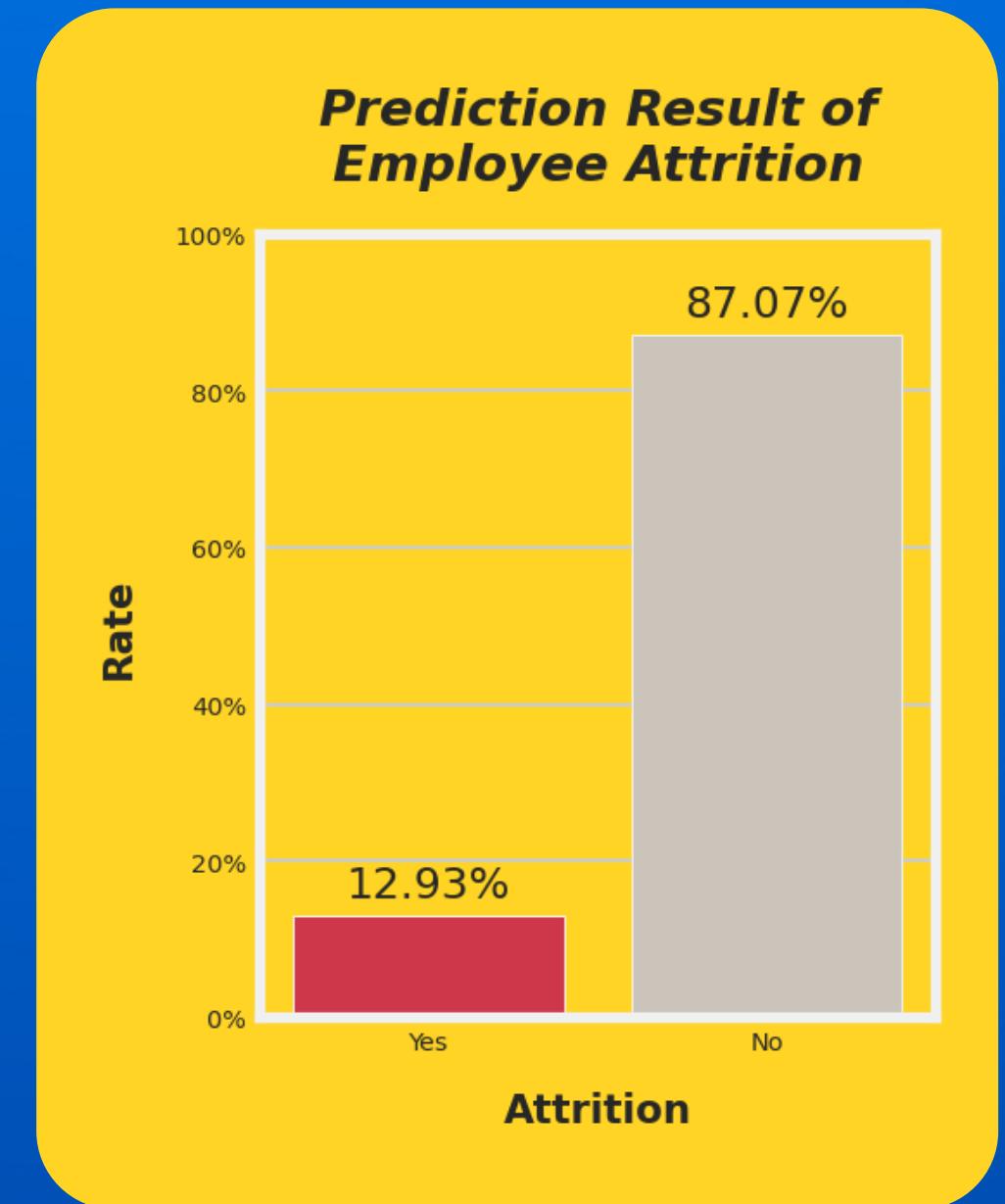
SIMULATION AND OUR BUSINESS RECOMMENDATION



1470
Employees



Prediction of Employee Attrition



190 out of 1470 employees have the potential to leave the company (12.93%)

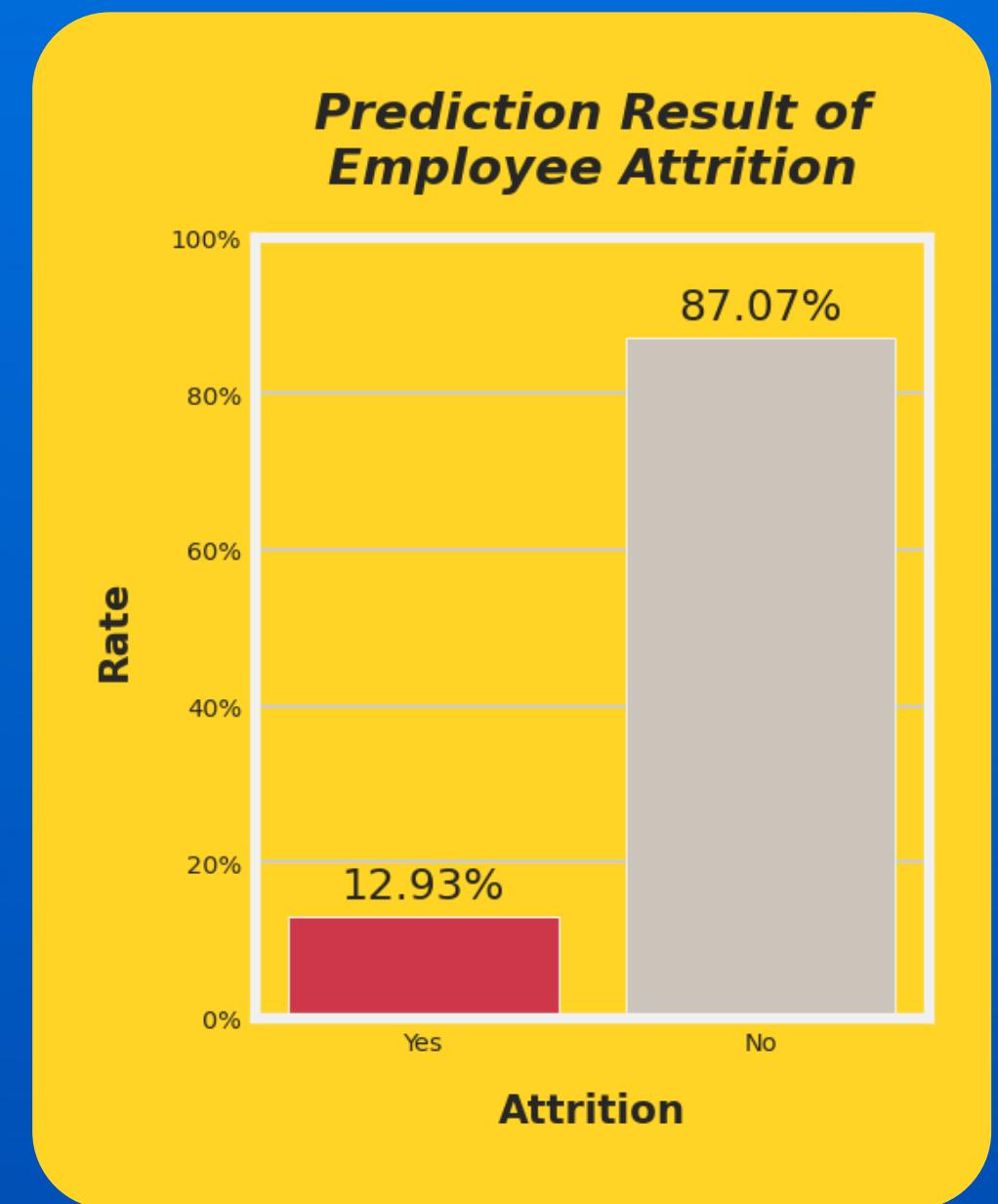
What's Next?



Work Overtime

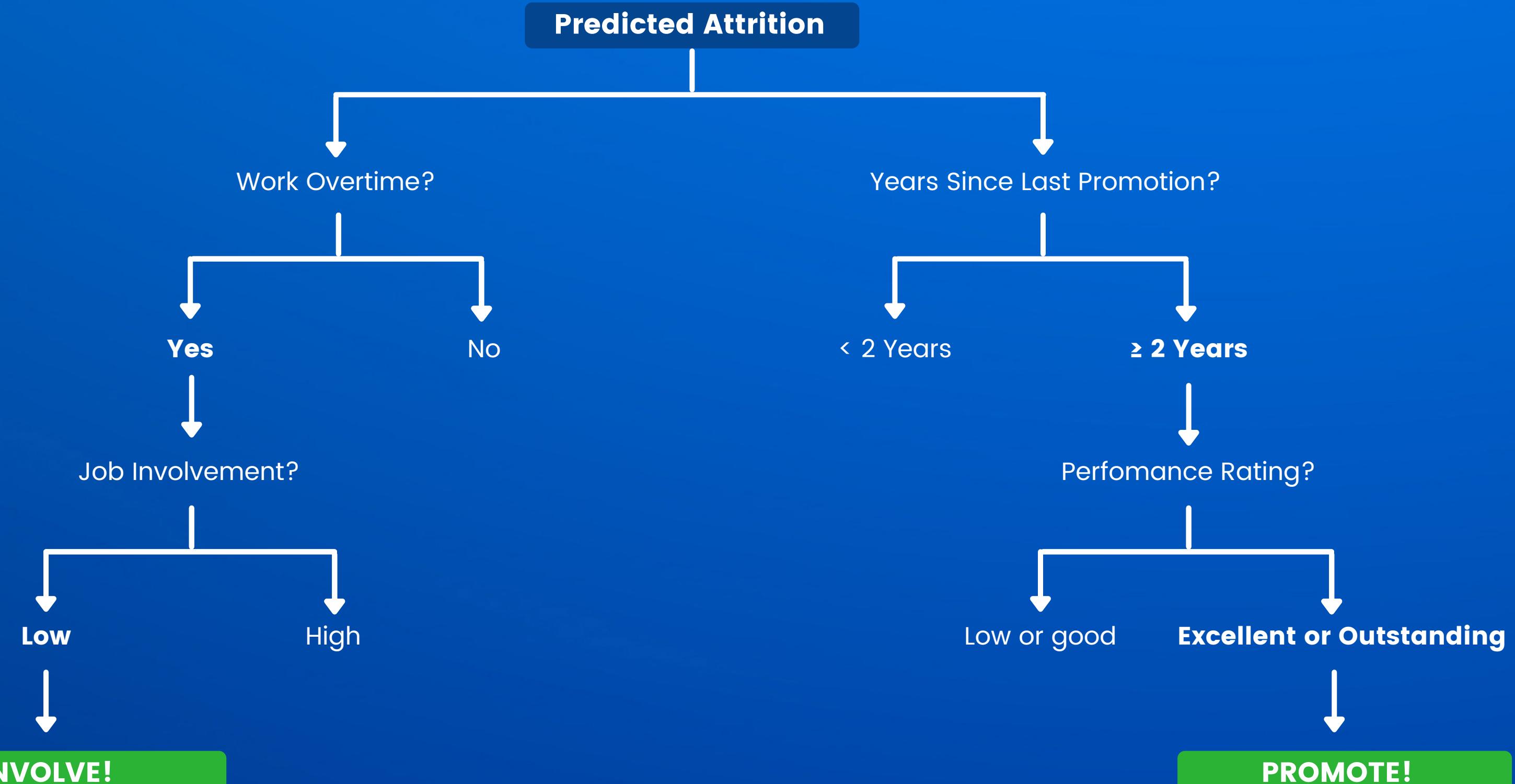


Years Since
Last Promotion



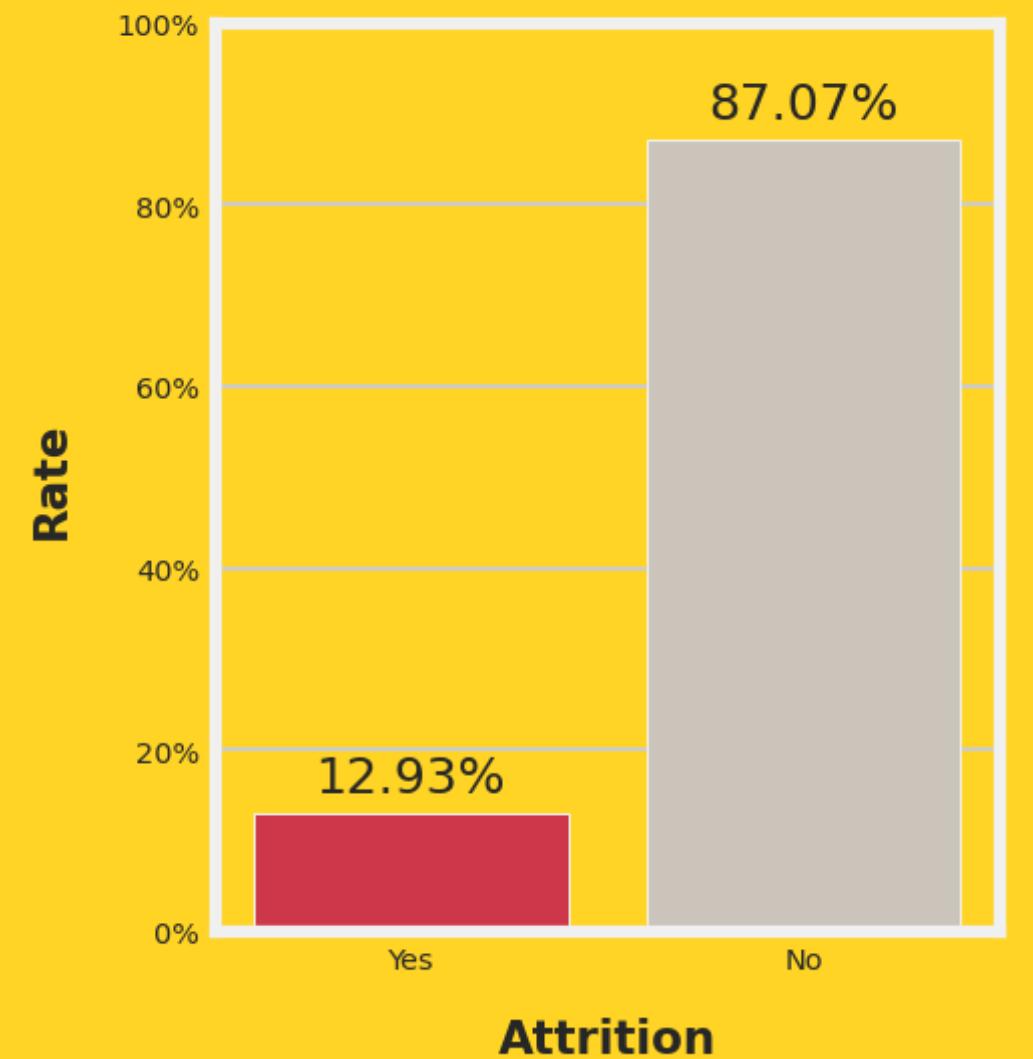
We recommend PT Goras to pay attention to some employee attributes.

Recommendation Flow



If the Scenario is realized...

**Attrition Rate
Before Doing Treatment**

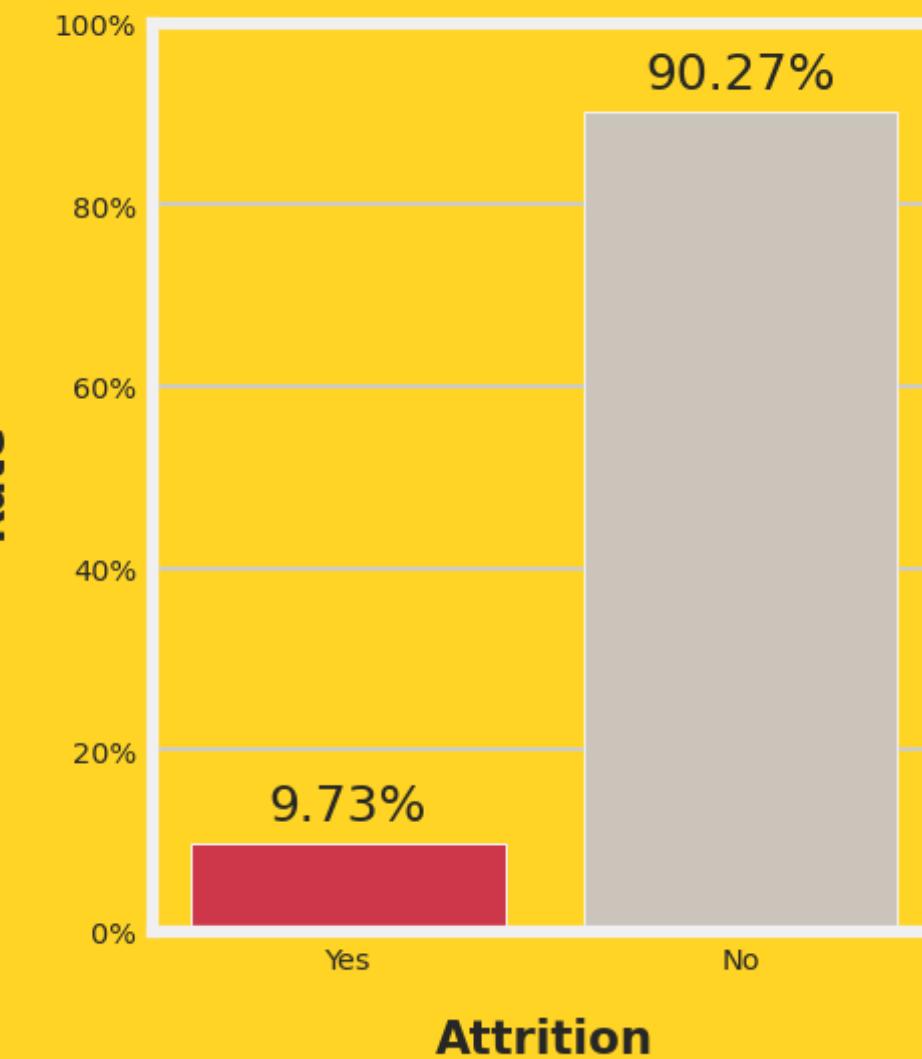


INVOLVE!

PROMOTE!

reduced by 3.2%

**Attrition Rate
After Doing Treatment**



So, We Recommend to...



PROMOTE!

- Pay attention to the performance of your employees who are **entitled to a promotion**

INVOLVE!

- **Improve the job involvement of your employees:**
 - Coaching or mentoring, and give positive feedback
 - Employee involvement programs
 - Open-communication and suggestion boxes
- **Reduce your employee's overtime:**
 - Cross-train your employees
 - Try flexible work schedules



Thank You!

PYTHAGORAS

Because, we're here to help

PT. Goras will improve their business vision by follow our recommendation.

 Address

Rakamin Academy- Pythagoras

 Telephone

Call me anytime

 Website

Rakamin.com