# DATA-DRIVEN HIRING

## PREDICTIVE MODELING FOR ATTRITION RISK


**PRESENTED BY**
MA. FERNANDA MALDONADO VALLES | IAN NIKOLAY DOSHNER GALLAND | ANDREA PONCE DE LEÓN | JUAN CARLOS SÁNCHEZ GUTIÉRREZ | SEBASTIÁN ZENTENO BARREIRO

# BUSINESS MODEL

“One Business Network” is a consulting company that is constantly recruiting new personal from all over the world.  Its current goal is to modify the recruitment process in order to evaluate if a new prospect could, by taking into account numerous factors, resign or not.

# OUR OBJECTIVES

1

Identify the different factors
that drive desertion

2

Build a model that predicts if an employee will
quit or not

# DATA

Attrition Age Business Travel Daily Rate Department Distance From Home Education Education Field Environment Satisfaction Gender Hourly Rate Job Involvement Job Level Job Role Job Satisfaction Marital Status Monthly Income Monthly Rate Num. Companies Worked Over Time Percent Salary Hike Performance Rating Relationship Satisfaction Stock Option Level Total Working Years Training Times Last Year Work Life Balance Years At Company Years In Current Role Years Since Last Promotion YearsWithCurrManager

# DATA ANALYSIS

Average age of employees

37

Average number of years until promotion

2

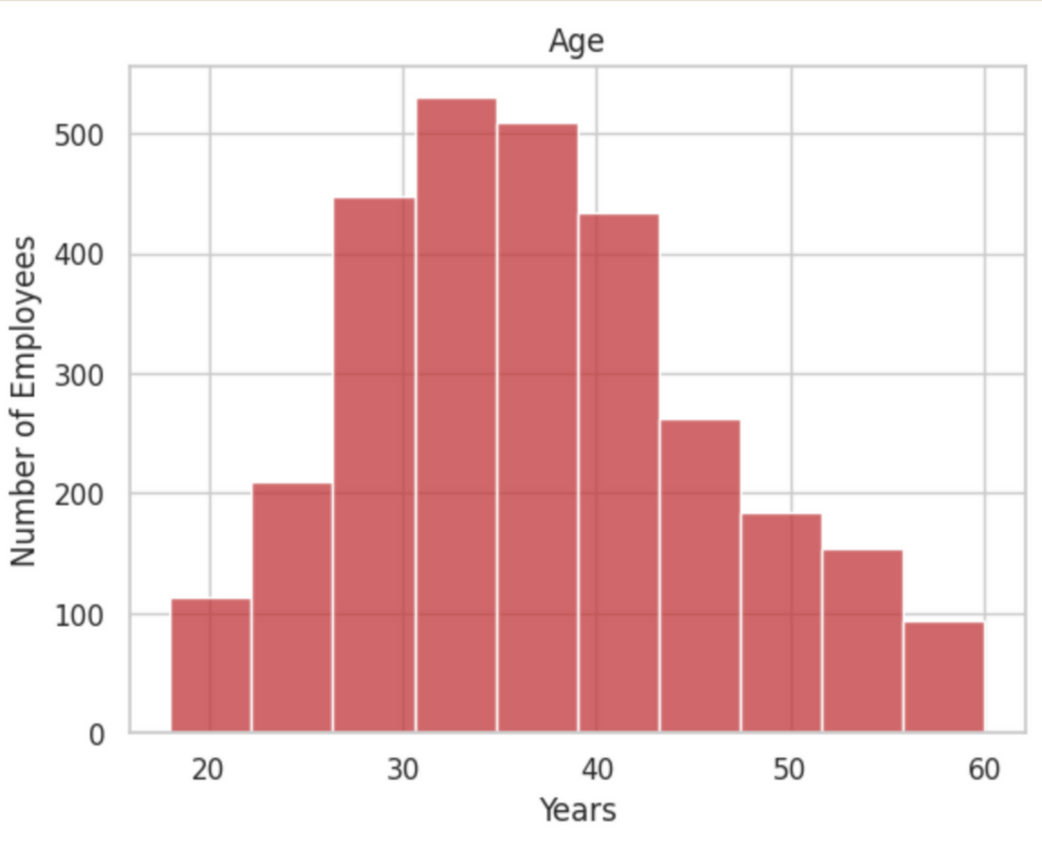
Average pay raise of an employee

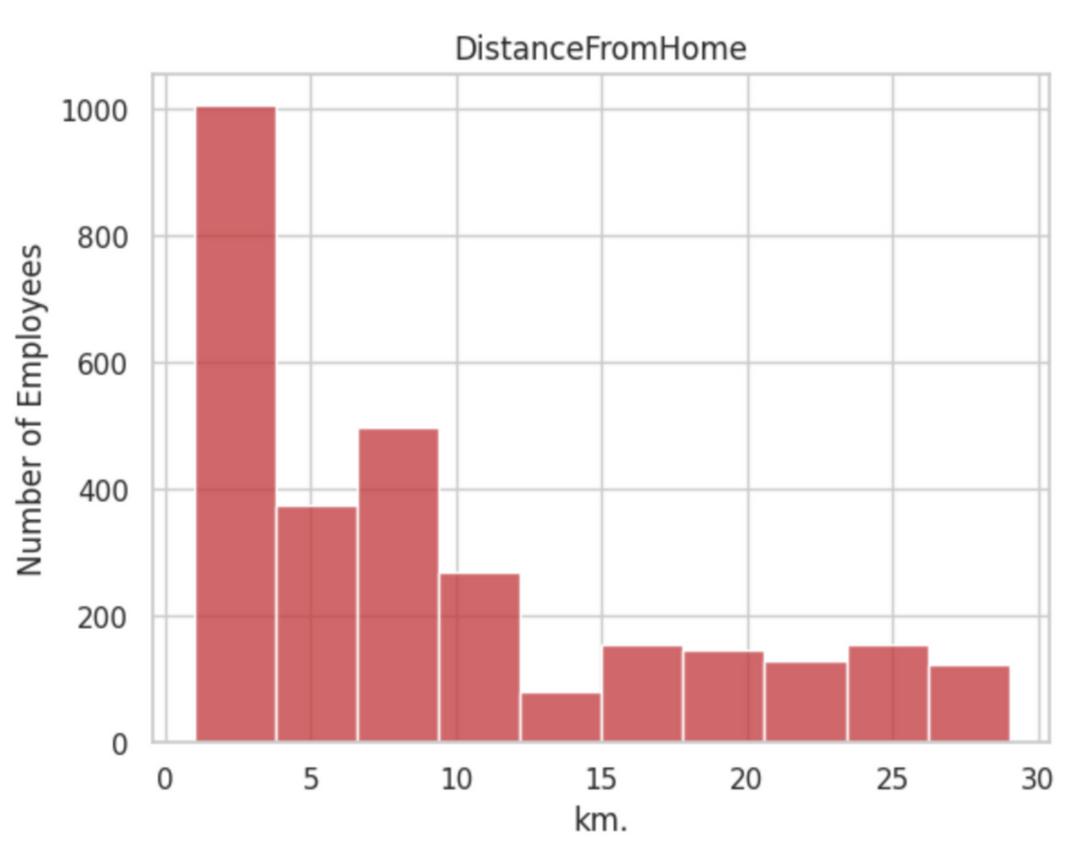
15%

Average monthly income per employee

$6,500

# UNDERSTANDING THE DATA

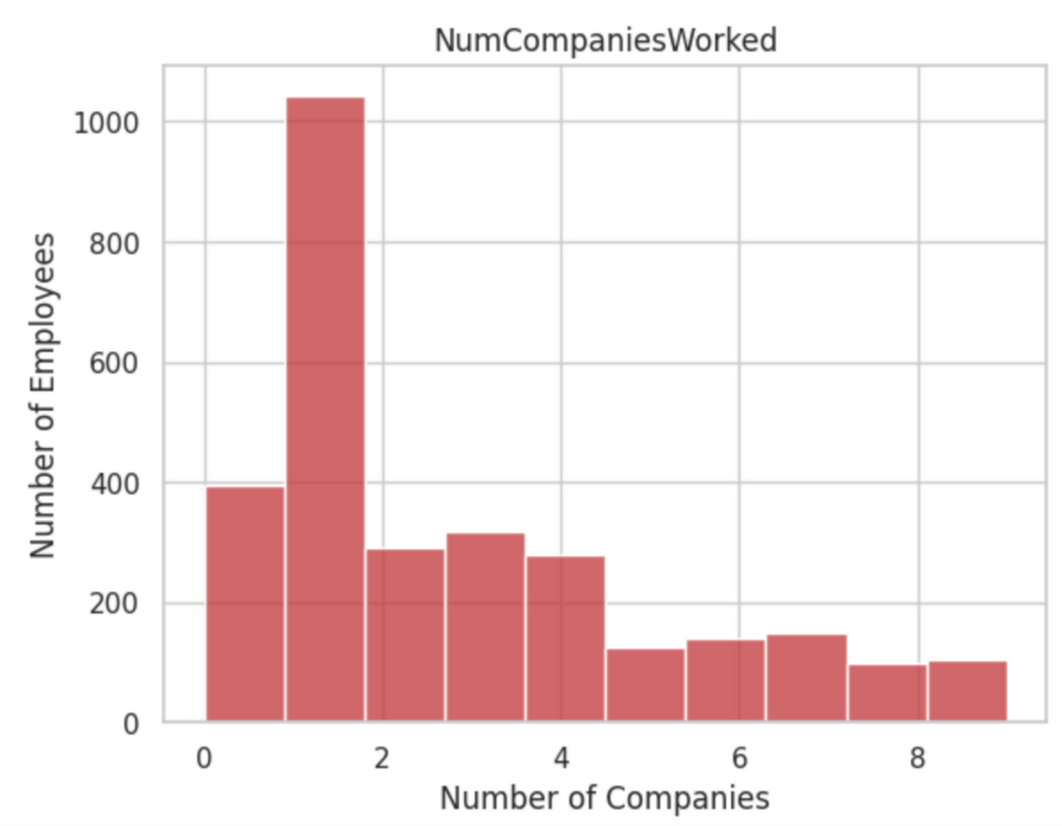


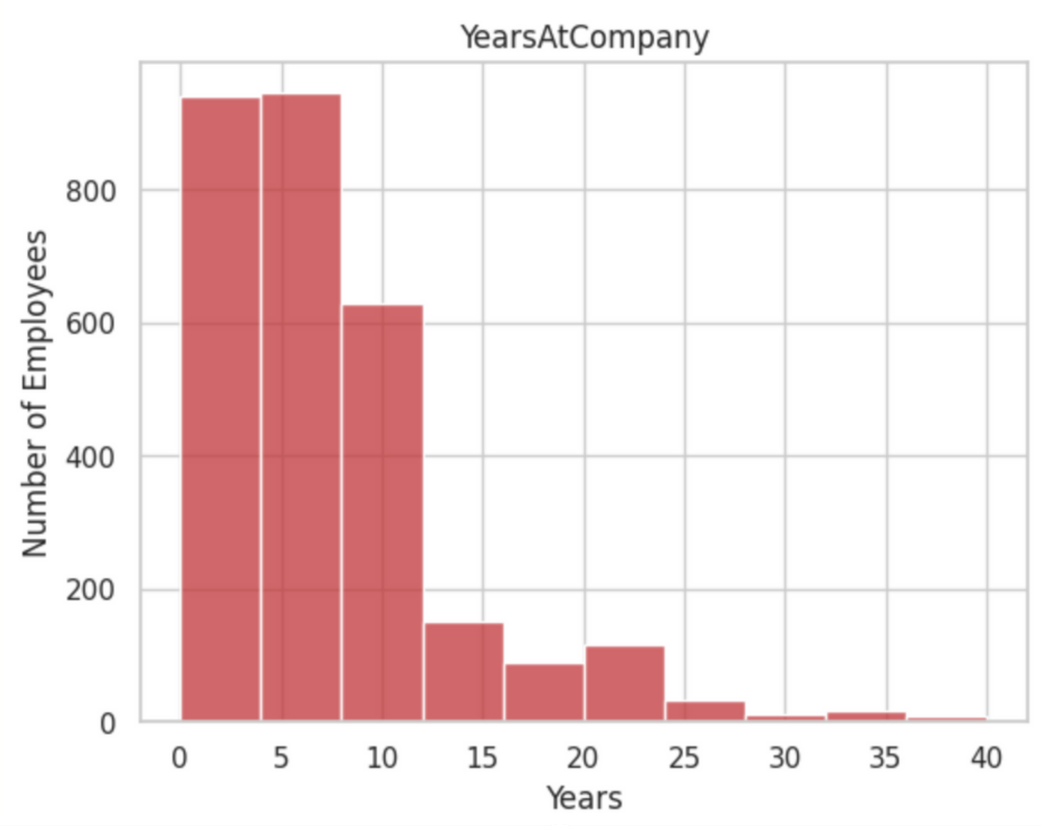




## INSIGHTS

- **The majority of employees are in the age range of 25 to 50**
- **Most of them live close to their workplace, less than 10 km**

# UNDERSTANDING THE DATA

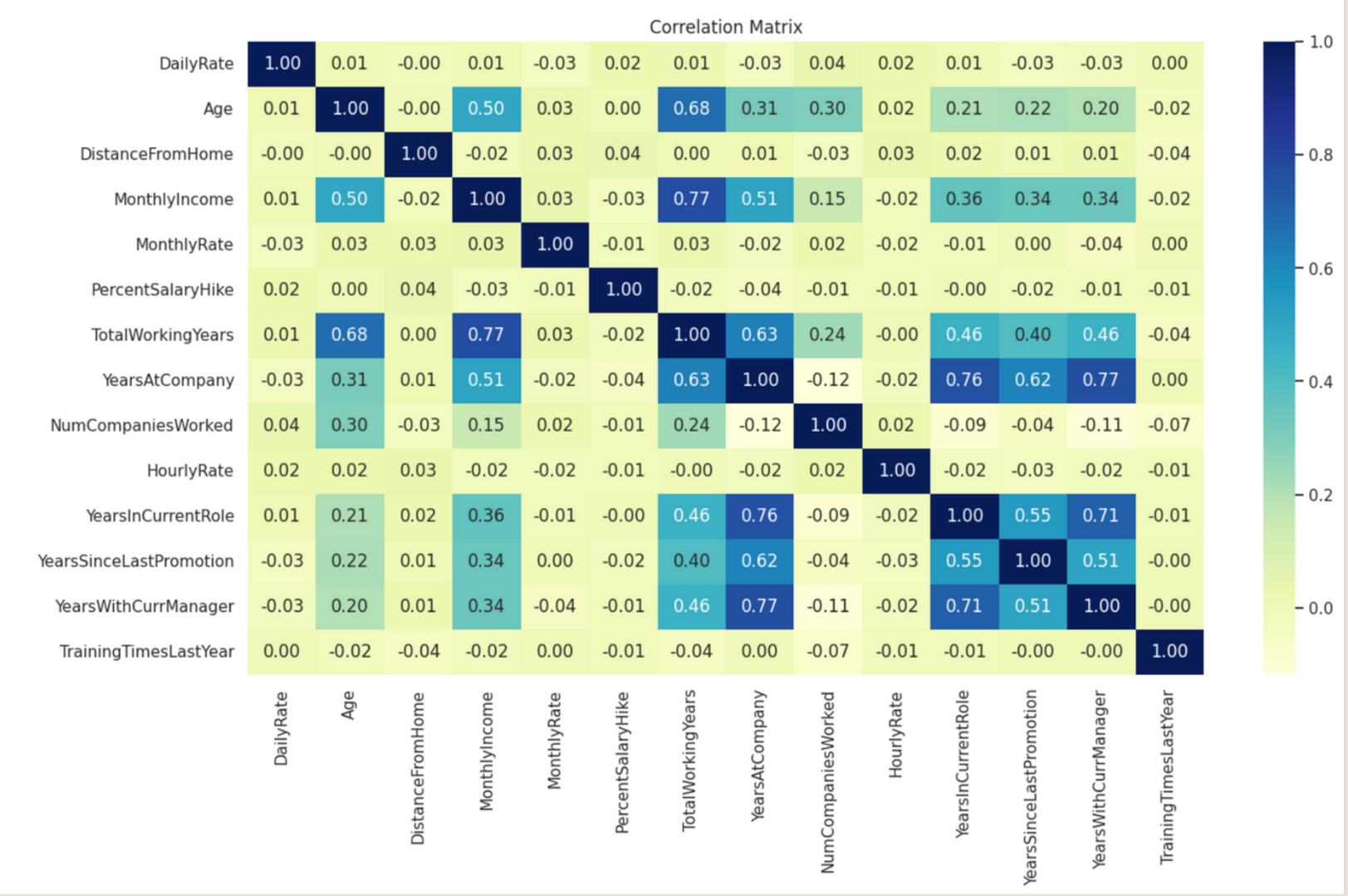



## INSIGHTS

- **The majority has only worked in 1 company**
- **Big number of employees who have been at the company for a decade**

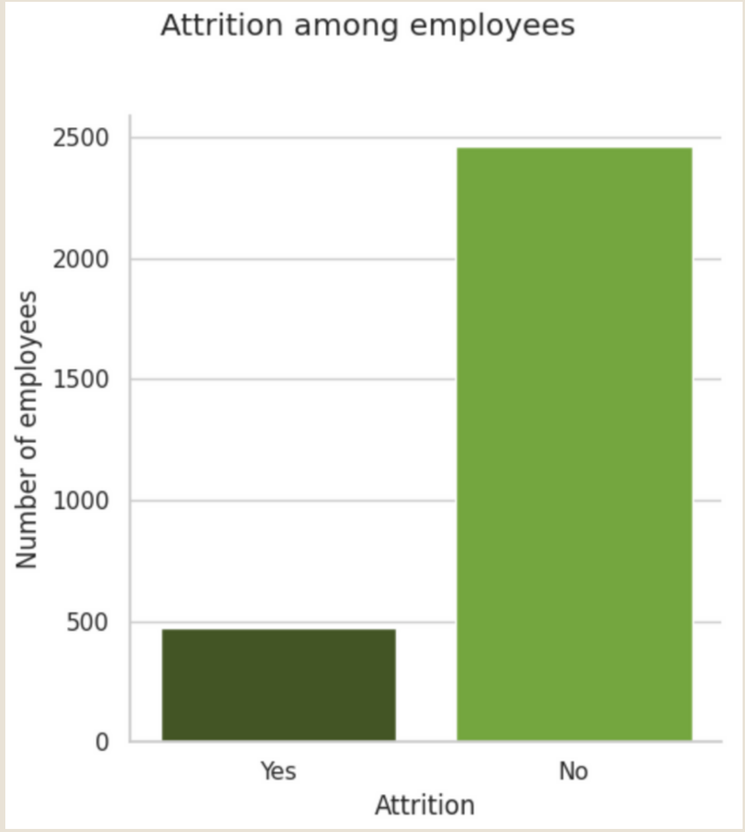
# CORRELATION

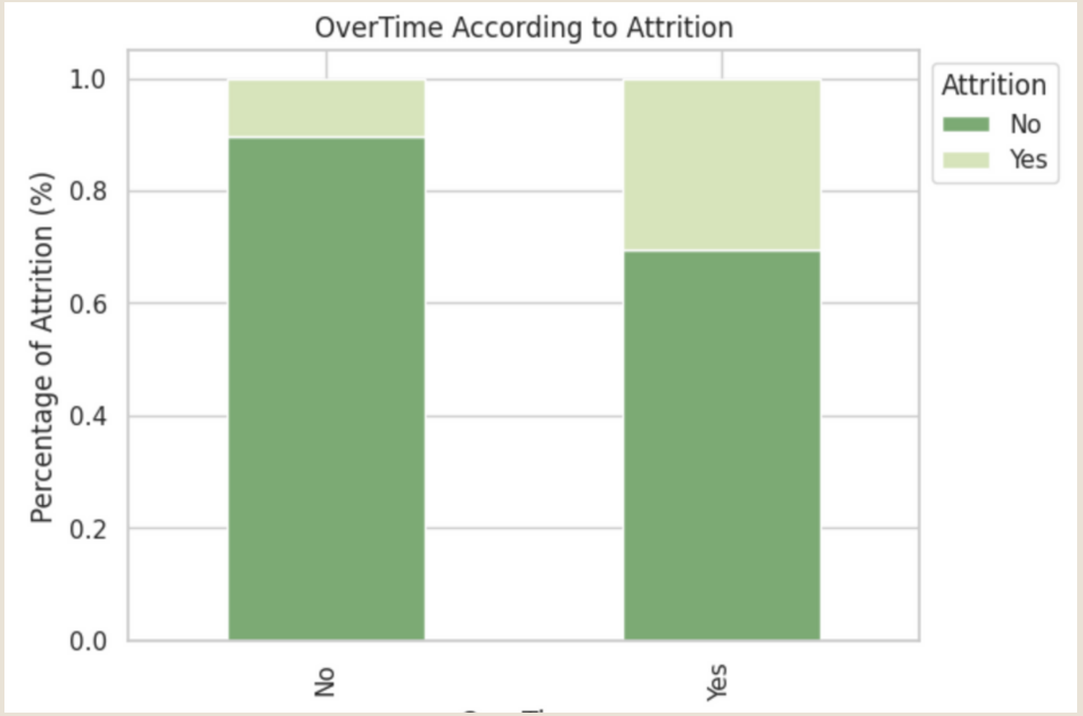



## INSIGHTS

- **Years at company vs Years with current manager**
- **Monthly income vs Total working years**
- **Years at company vs Years in current role**

# RELATION WITH ATTRITION

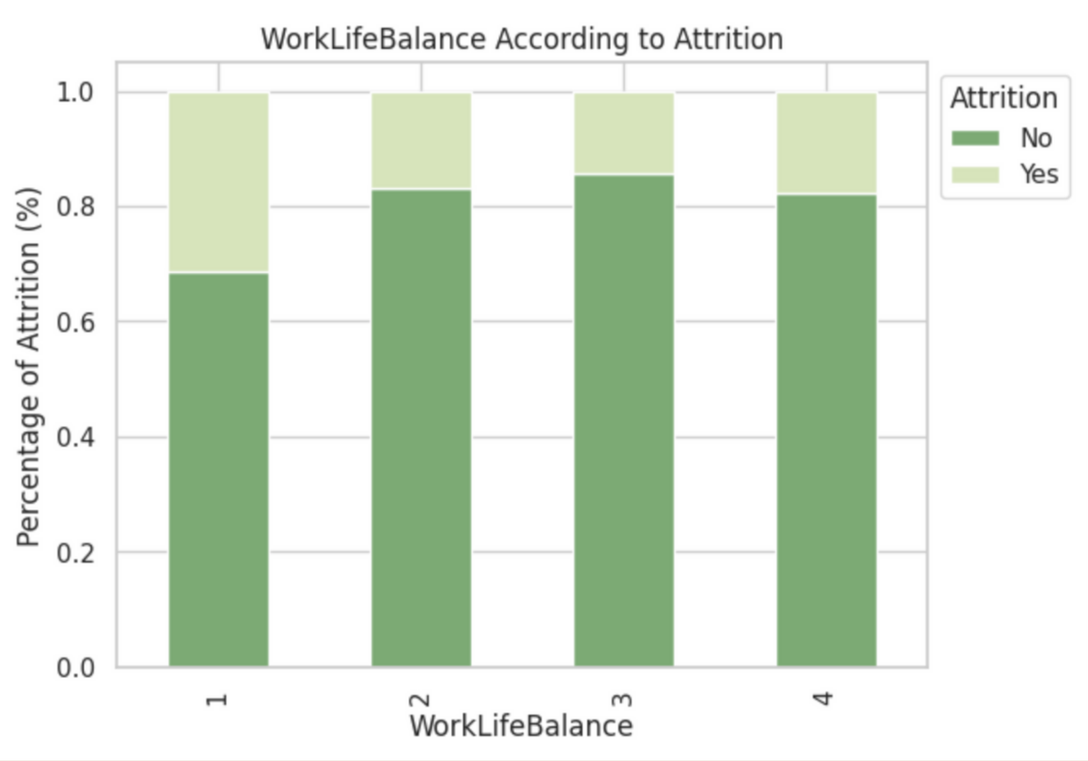


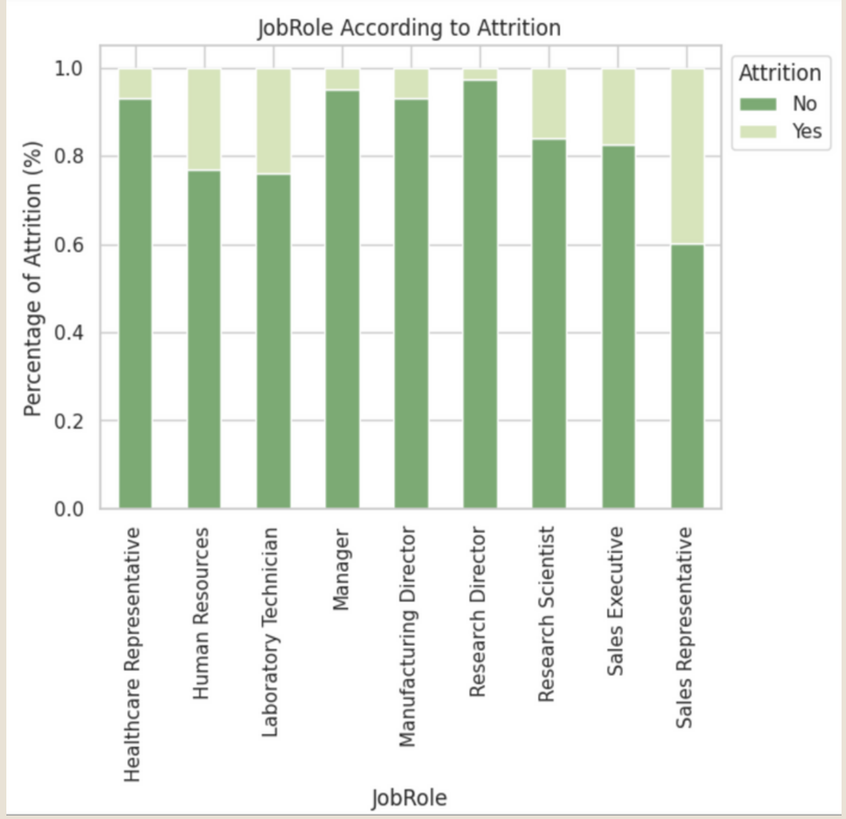




## INSIGHTS

- **Low attrition rate**
- **Employees who work overtime have 30% chance of quitting**

# RELATION WITH ATTRITION

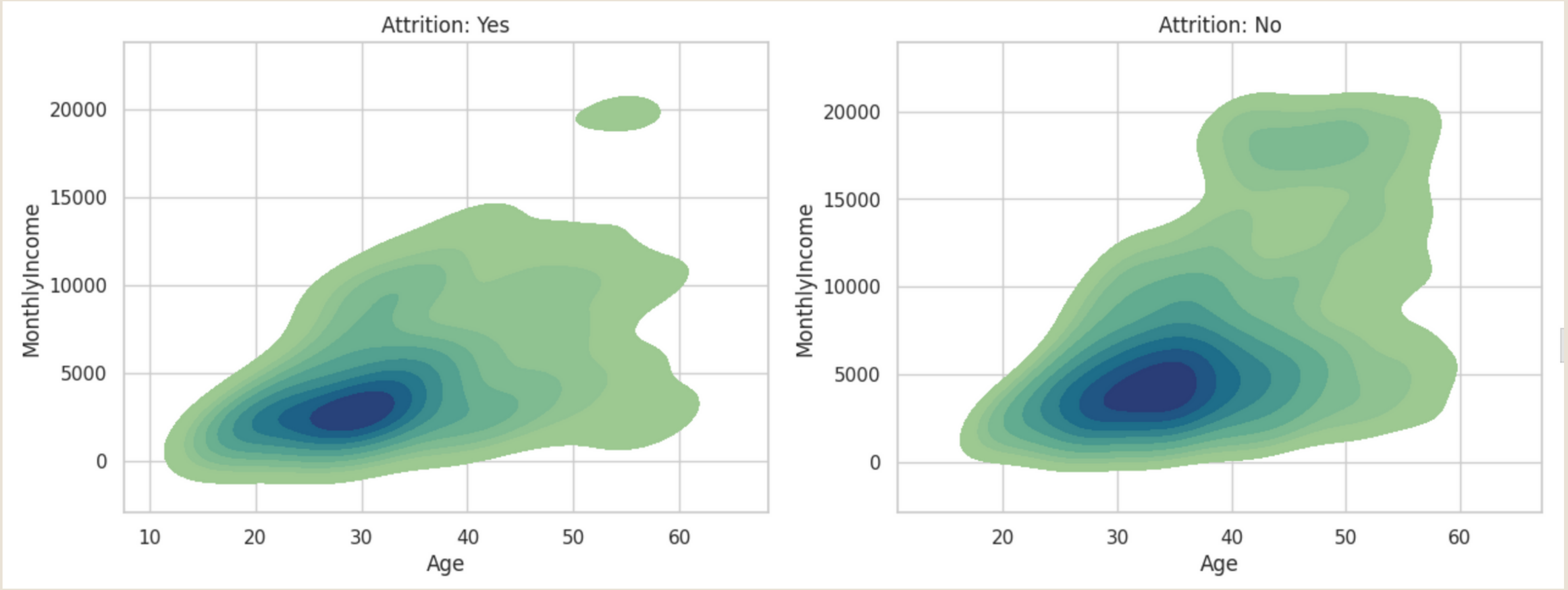



## INSIGHTS

- **Employees with poor work-life balance are more likely to quit, nearly 30%**
- **Sales representatives have the higher attrition rate with approximately 40%**

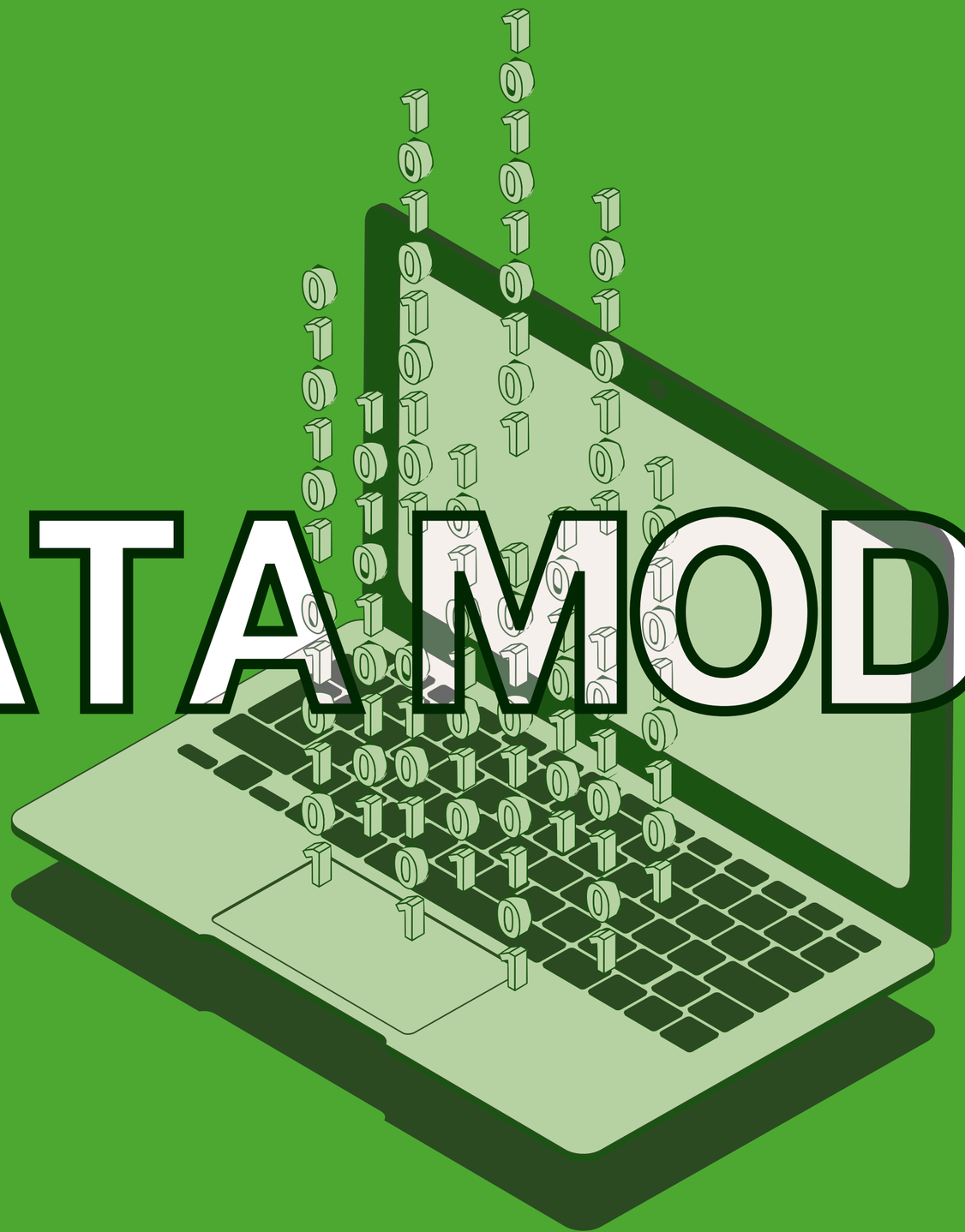
# RELATION WITH ATTRITION




## INSIGHTS

- **Attrition tends to concentrate among lower ages**
- **Monthly income becomes more relevant with age as it relates to attrition**

# DATA MODELS



01. Preparing the Data for Modeling

02. Splitting Data

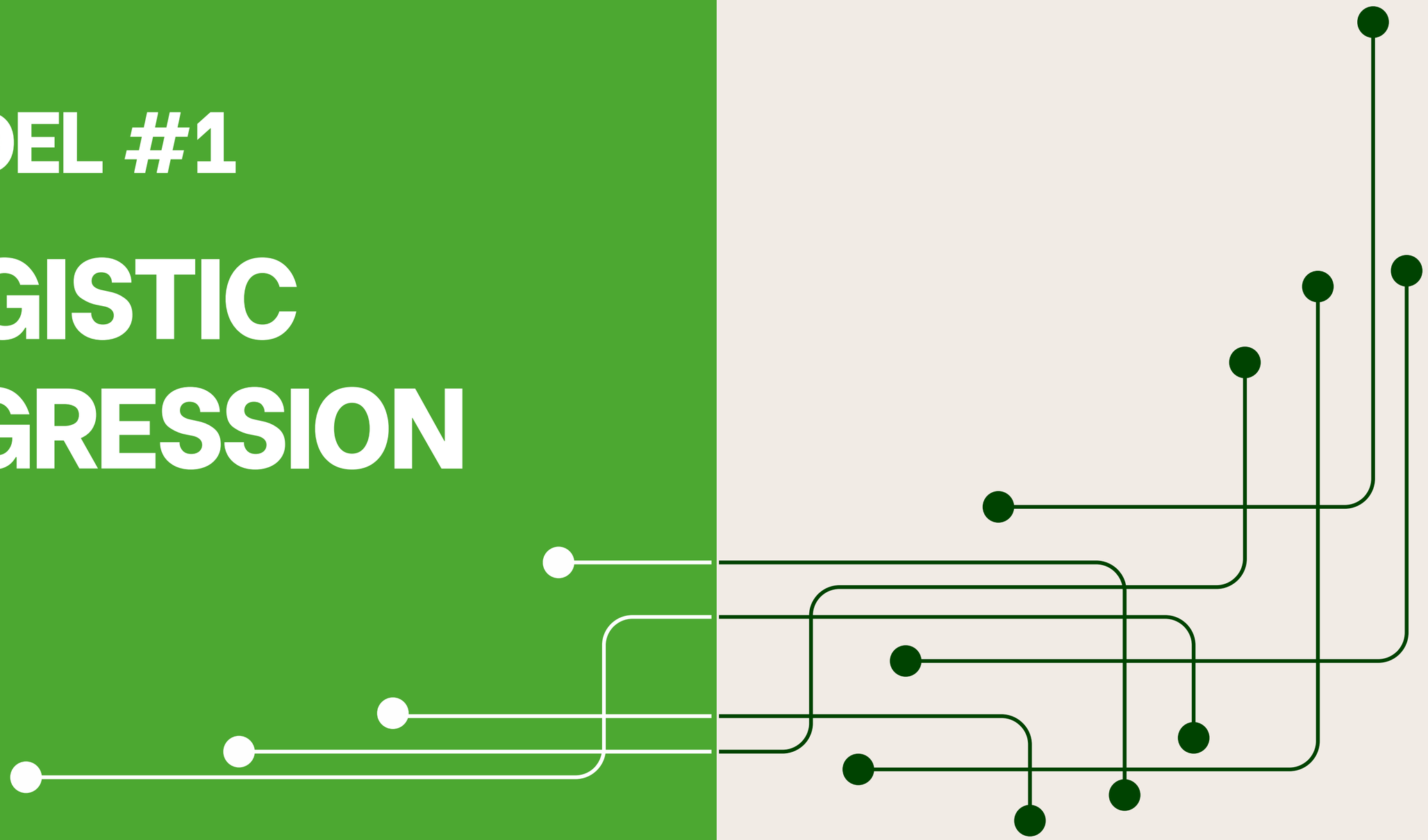
03. Building the Model

04. Fine-Tuning the Model

05. Testing the Model

# MODEL #1

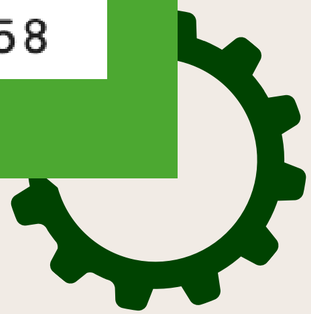
# LOGISTIC REGRESSION

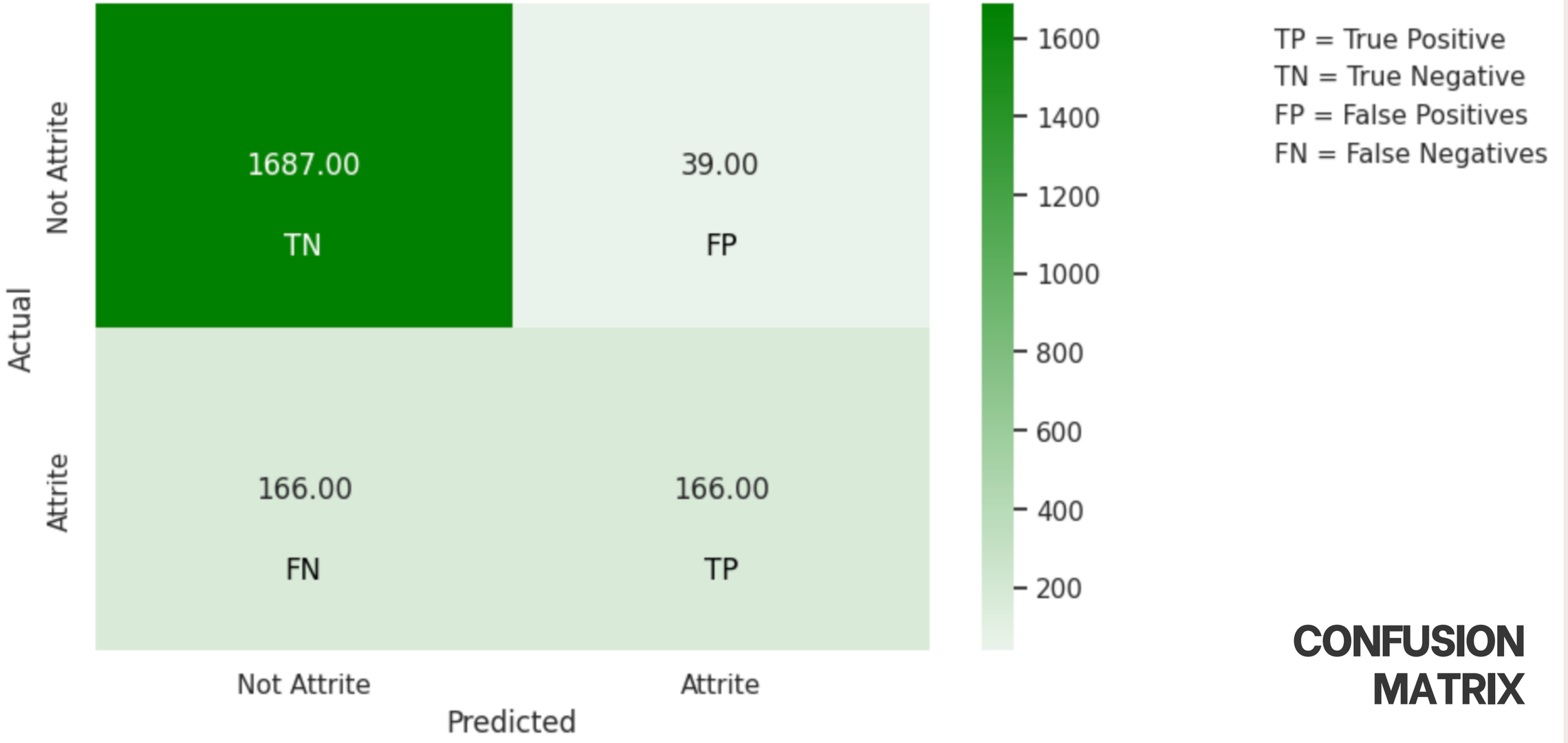


# TRAINING DATA

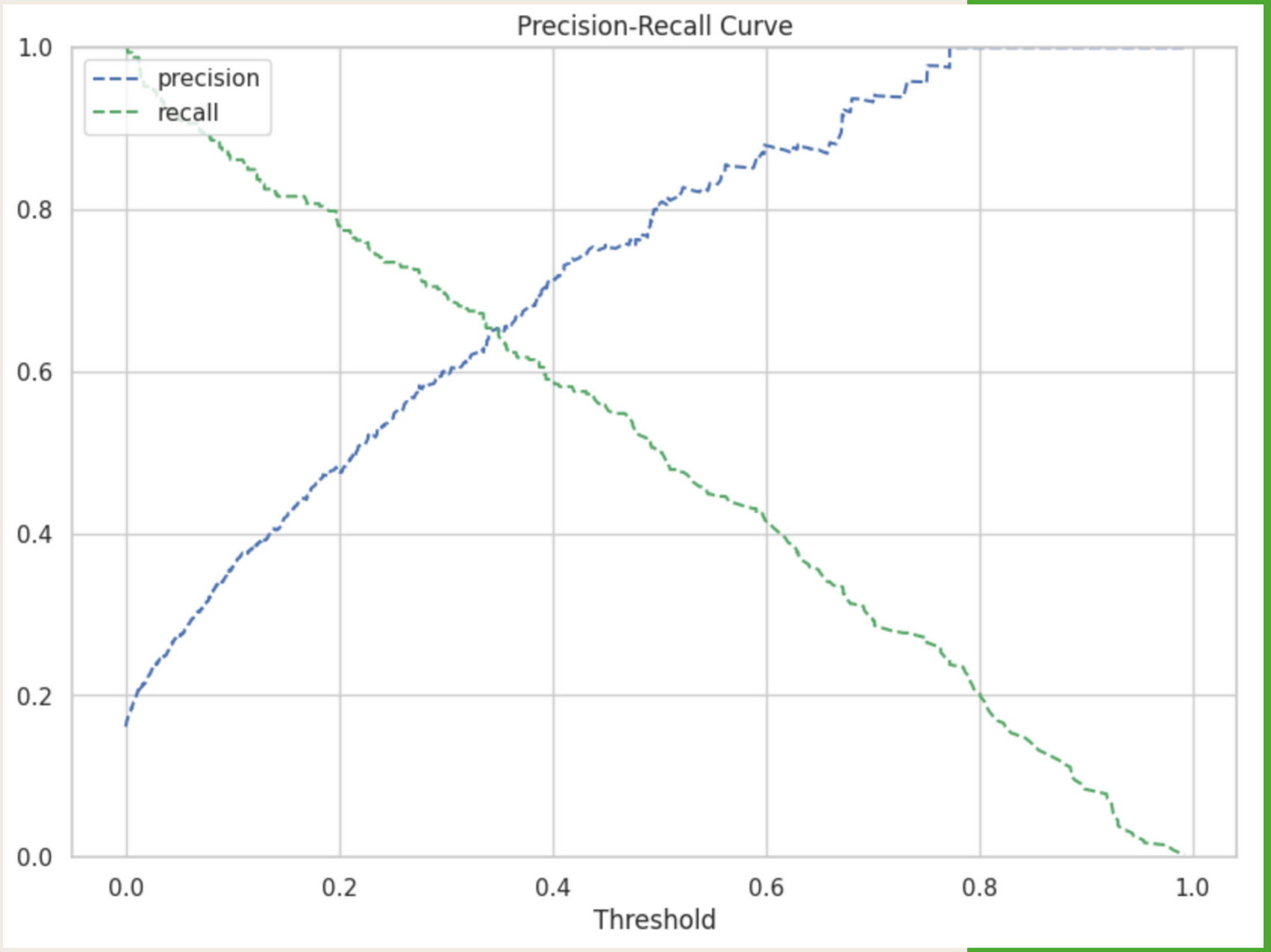
LR

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.91 | 0.98 | 0.94 | 1726 |
| 1 | 0.81 | 0.50 | 0.62 | 332 |
| accuracy | | | 0.90 | 2058 |
| macro avg | 0.86 | 0.74 | 0.78 | 2058 |
| weighted avg | 0.89 | 0.90 | 0.89 | 2058 |



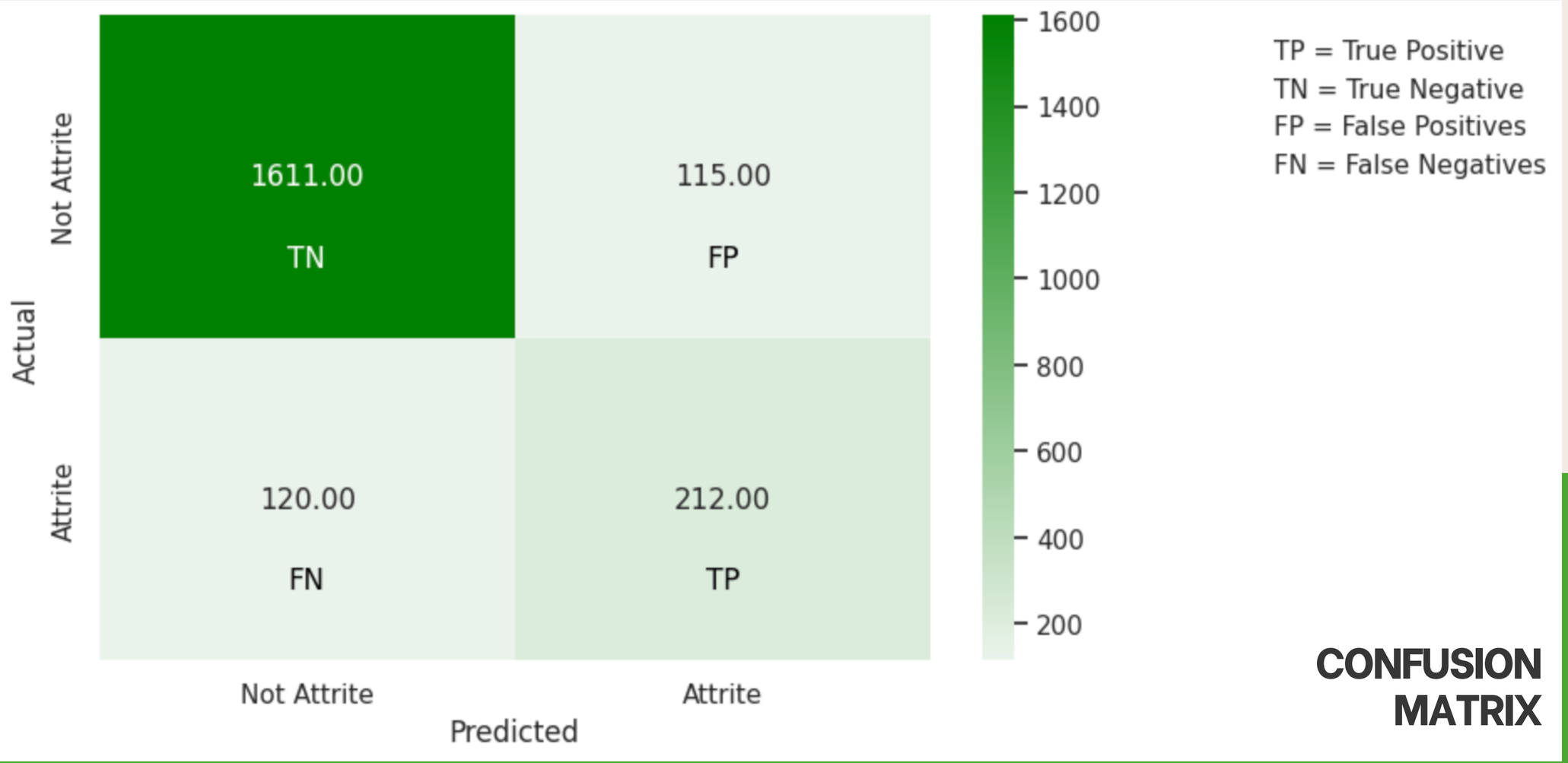



# CONFUSION MATRIX


Precision-Recall Curve
precision
recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Threshold

# NEW TRAINING DATA

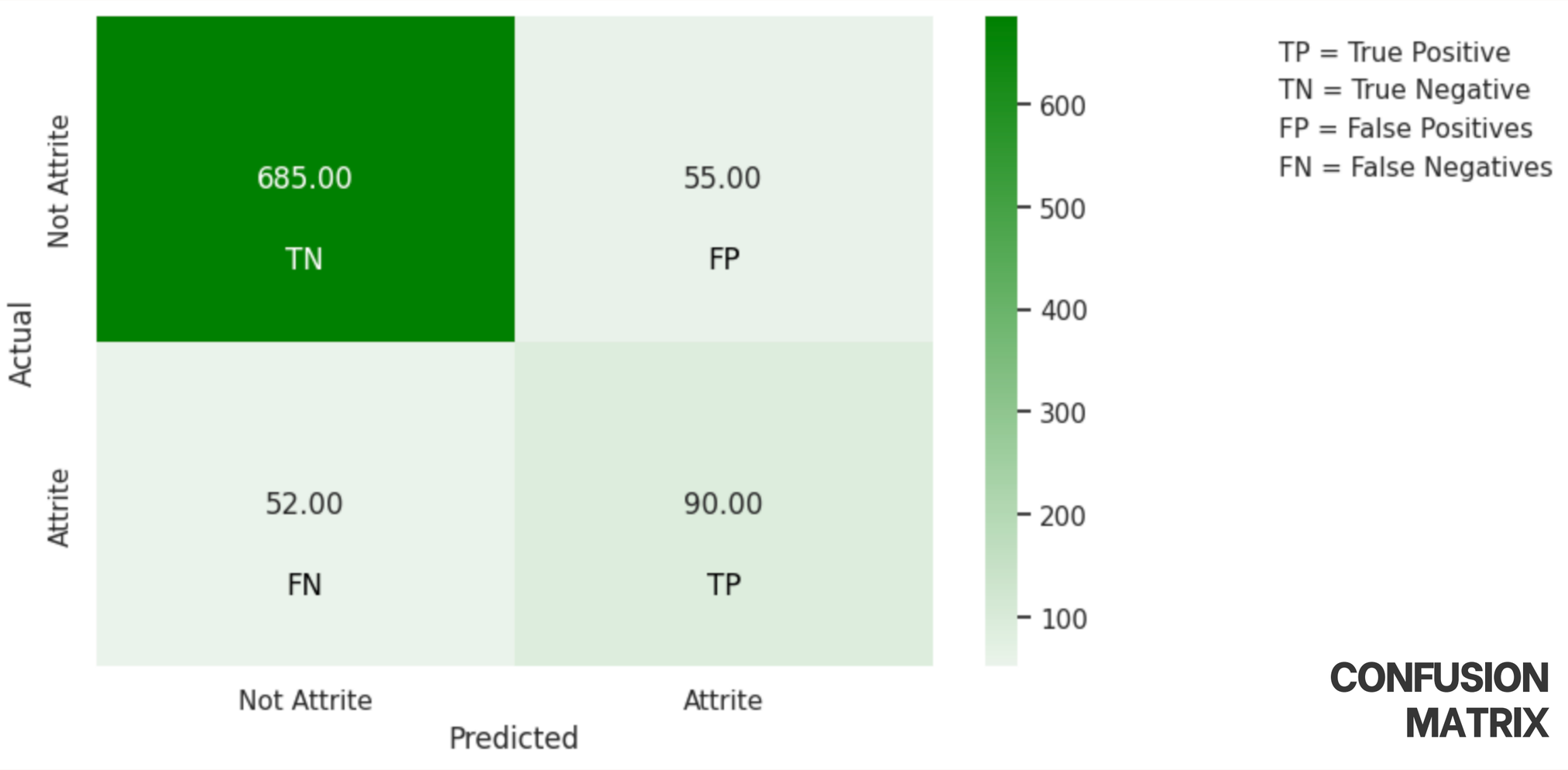
LR




**CONFUSION MATRIX**

# TESTING DATA

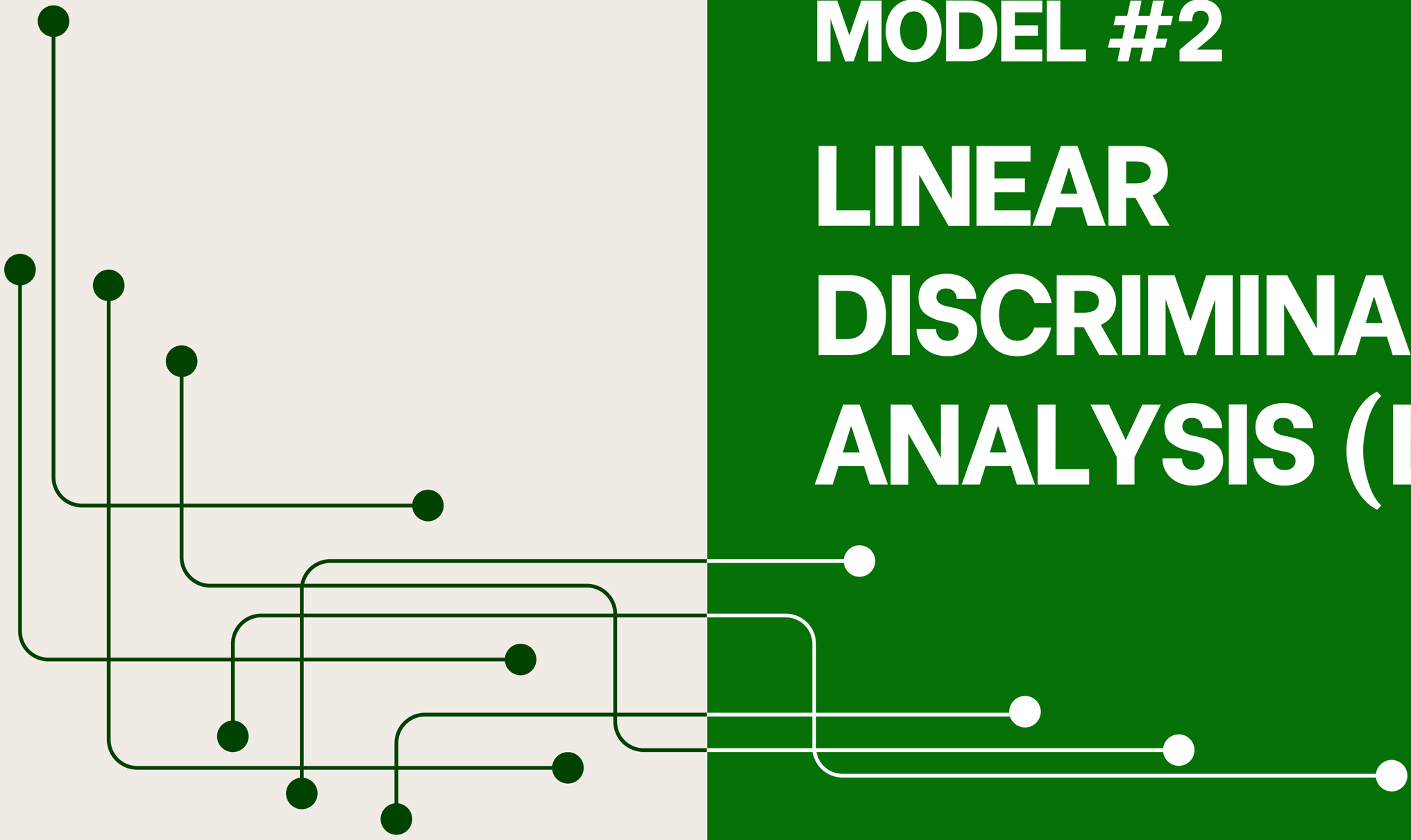
LR




**CONFUSION MATRIX**

# MODEL #2

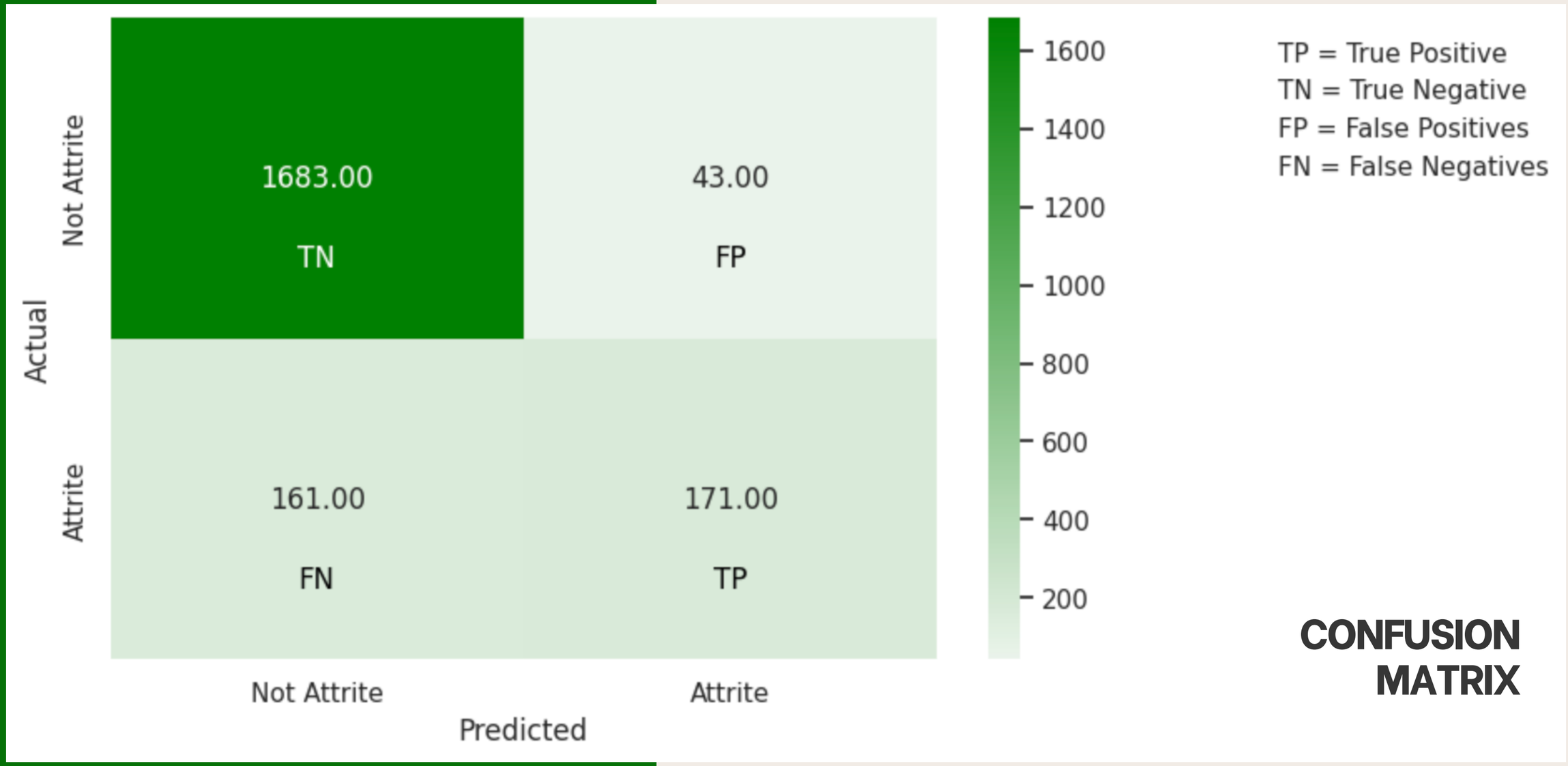
# LINEAR DISCRIMINANT ANALYSIS (LDA)

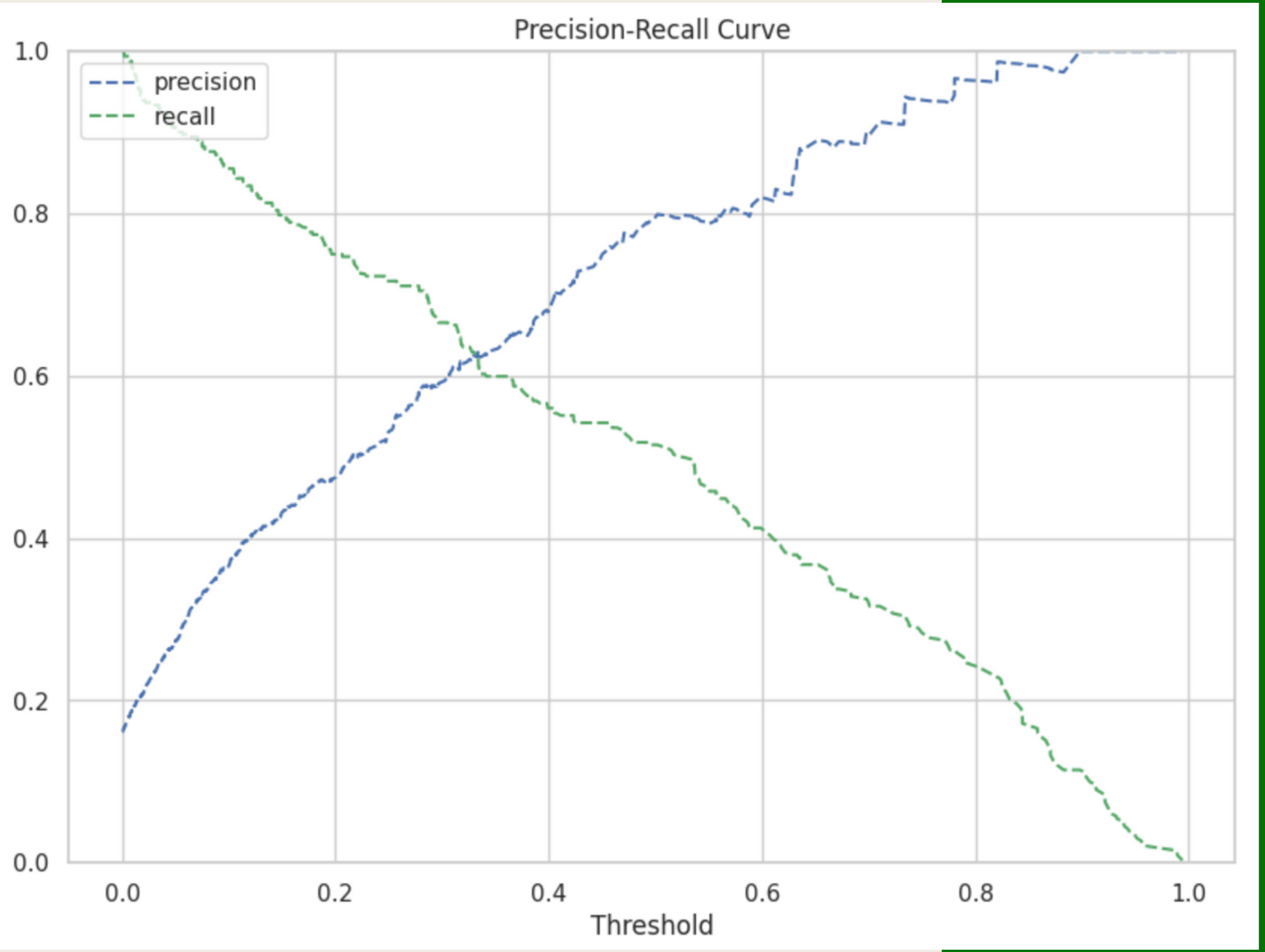


# TRAINING DATA

LDA

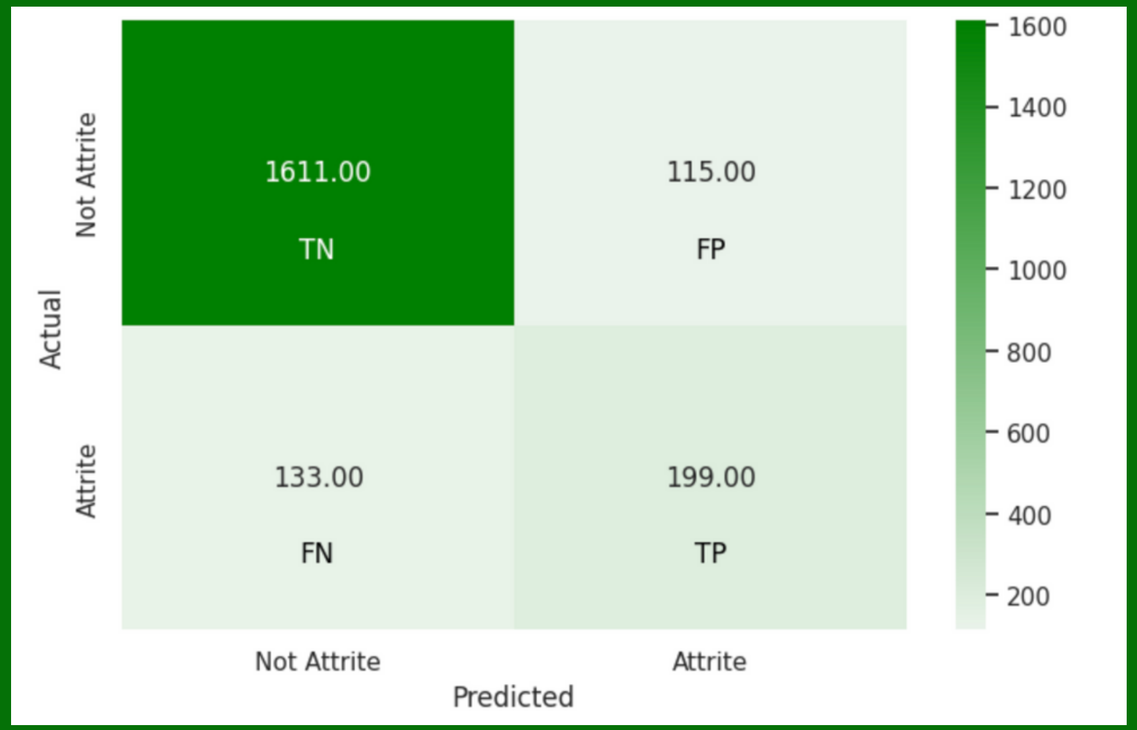
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.91 | 0.98 | 0.94 | 1726 |
| 1 | 0.80 | 0.52 | 0.63 | 332 |
| accuracy | | | 0.90 | 2058 |
| macro avg | 0.86 | 0.75 | 0.78 | 2058 |
| weighted avg | 0.89 | 0.90 | 0.89 | 2058 |


Actual
Not Attrite
Attrite
1683.00
TN
43.00
FP
161.00
FN
171.00
TP
Not Attrite
Attrite
Predicted
1600
1400
1200
1000
800
600
400
200
TP = True Positive
TN = True Negative
FP = False Positives
FN = False Negatives
CONFUSION MATRIX

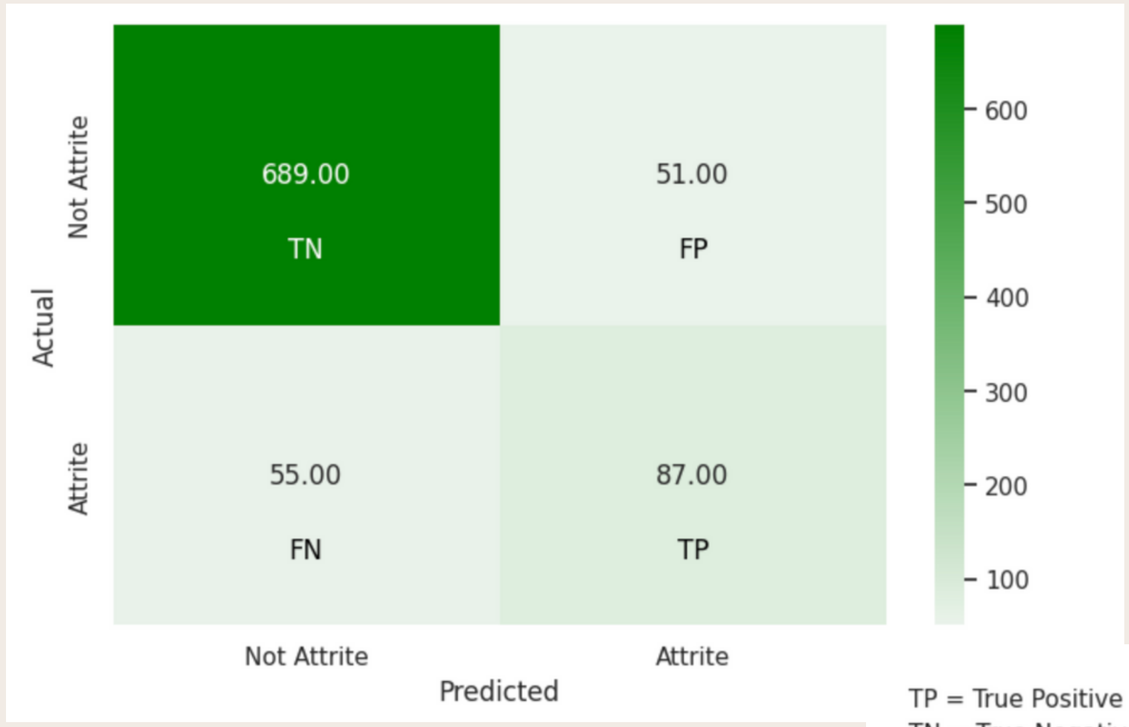

Precision-Recall Curve
precision
recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Threshold

LDA

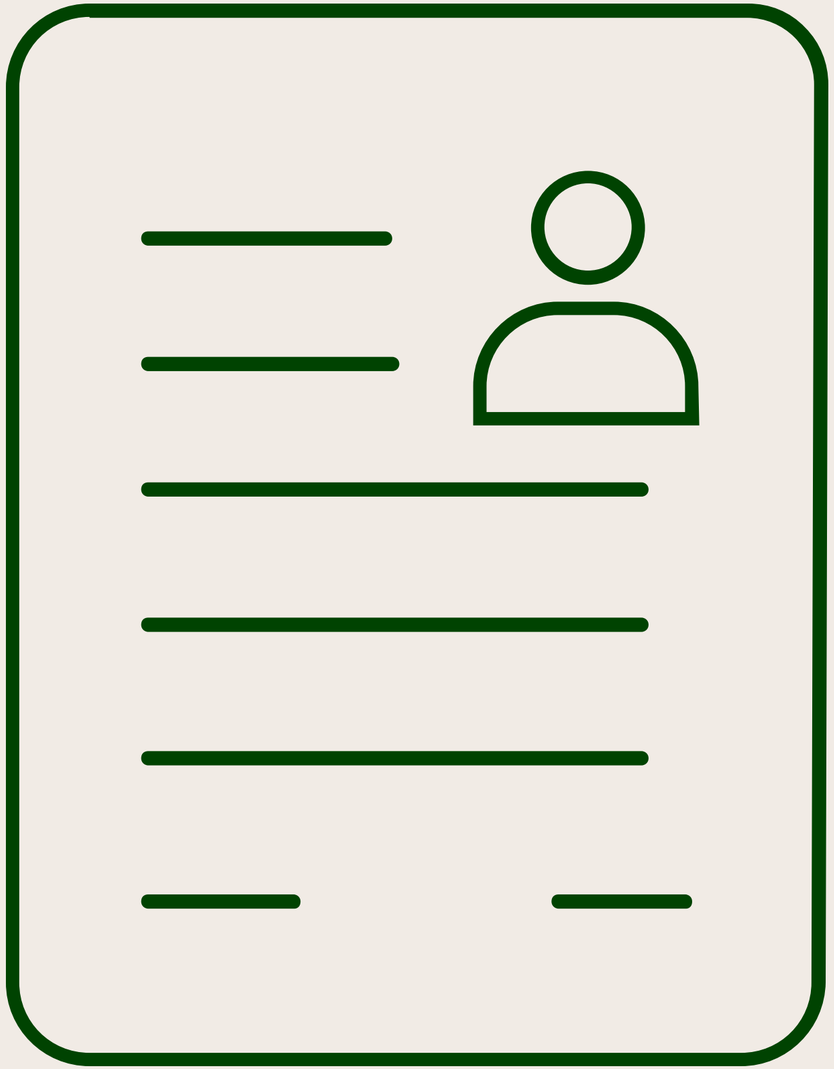
# NEW TRAINING DATA




# TESTING DATA




TP = True Positive
TN = True Negative
FP = False Positives
FN = False Negatives



# INDIVIDUAL PREDICTION

**Model 1: Logistic Regression**

Probability of No Attrition: **98.90%**

Probability of Attrition: **1.10%**

**Model 2: Linear Discriminant Analysis**

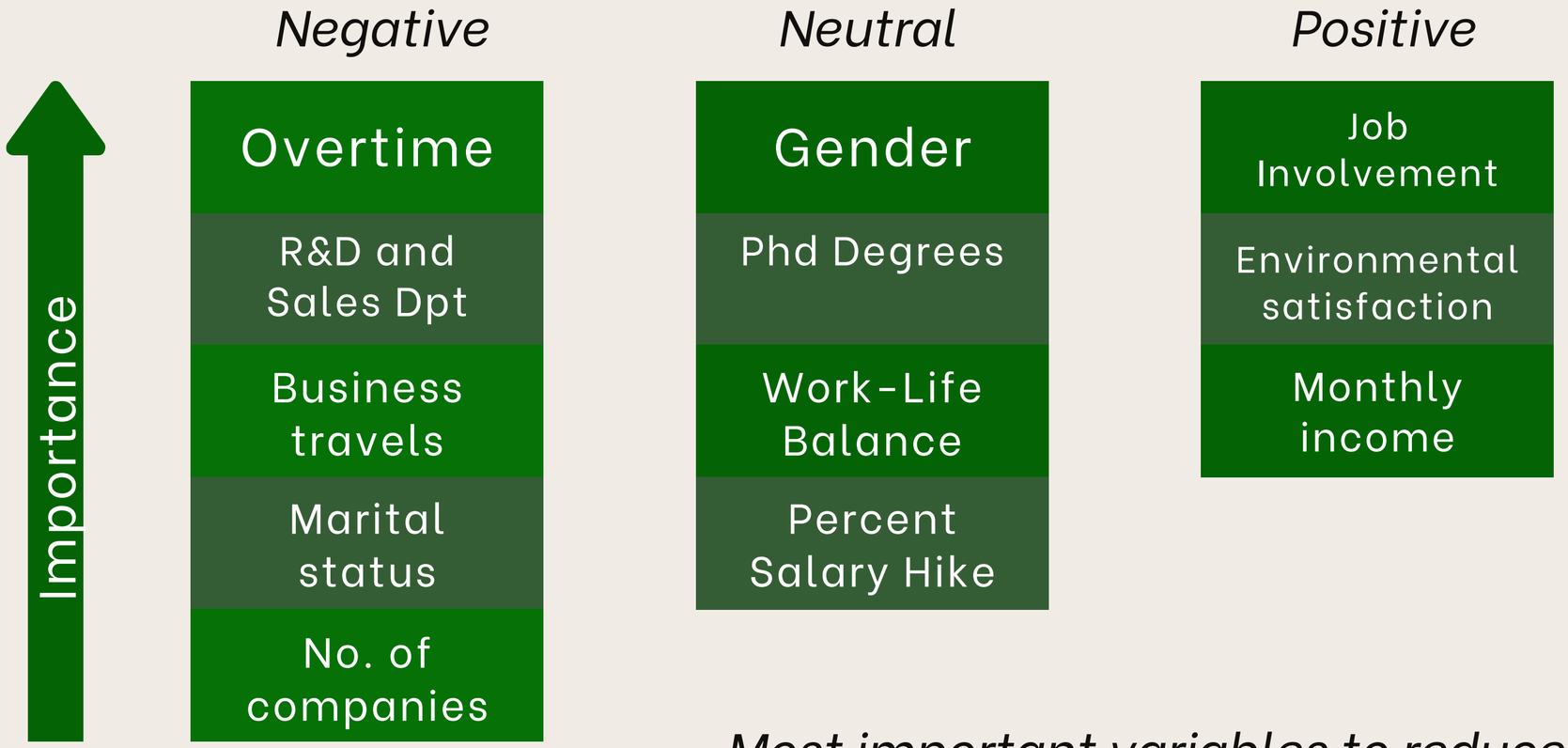
Probability of No Attrition: **99.18%**

Probability of Attrition: **0.82%**

- *Attrition is notably higher for **28 -31 years old employees***
- *Employees **over 37 with a monthly income over $14,000** don't quit as easily as others*

# RESULTS

## & General Guidelines




*Most important variables to reduce attrition*

# Conclusions

1

Attrition variables

Our analysis provides insights in the **influence of each variable** and their relationship between them that provides ways to **control attrition while also reducing costs in an efficient way**

2

LDA & LR Models

Our models are an excellent **tool to predict new talent's attrition** although it shouldn't be regarded as absolute, but as a helpful tool that **provides a scientific insight into hiring decisions.** Our models have **room for improvement** in both accuracy and recall

3

Improvement guidelines

Our results guideline show the **areas where the company should focus** and expect **positive results** while also shows **nonimportant aspects that might be a waste of time & investment**

# THANK YOU !