

CIS 508 Final Group Project

# EMPLOYEE ATTRITION PREDICTION

Mangesh Patil

Minjeong Kim

Shruthi Sunil

Alok Kumar Krishnan

# AGENDA

## KEY TOPICS DISCUSSED IN THIS PRESENTATION

- 1. Problem & Solution
- 2. Conceptual Model
- 3. Data Mining Solution
- 4. Data
- 5. Data Exploration
- 6. Data Preprocessing
- 7. Modeling
- 8. Evaluation Metrics
- 9. Evaluation
- 10. Final Model Selection
- 11. Deployment





## PROBLEM

Employee turnover is a major challenge for many organizations.

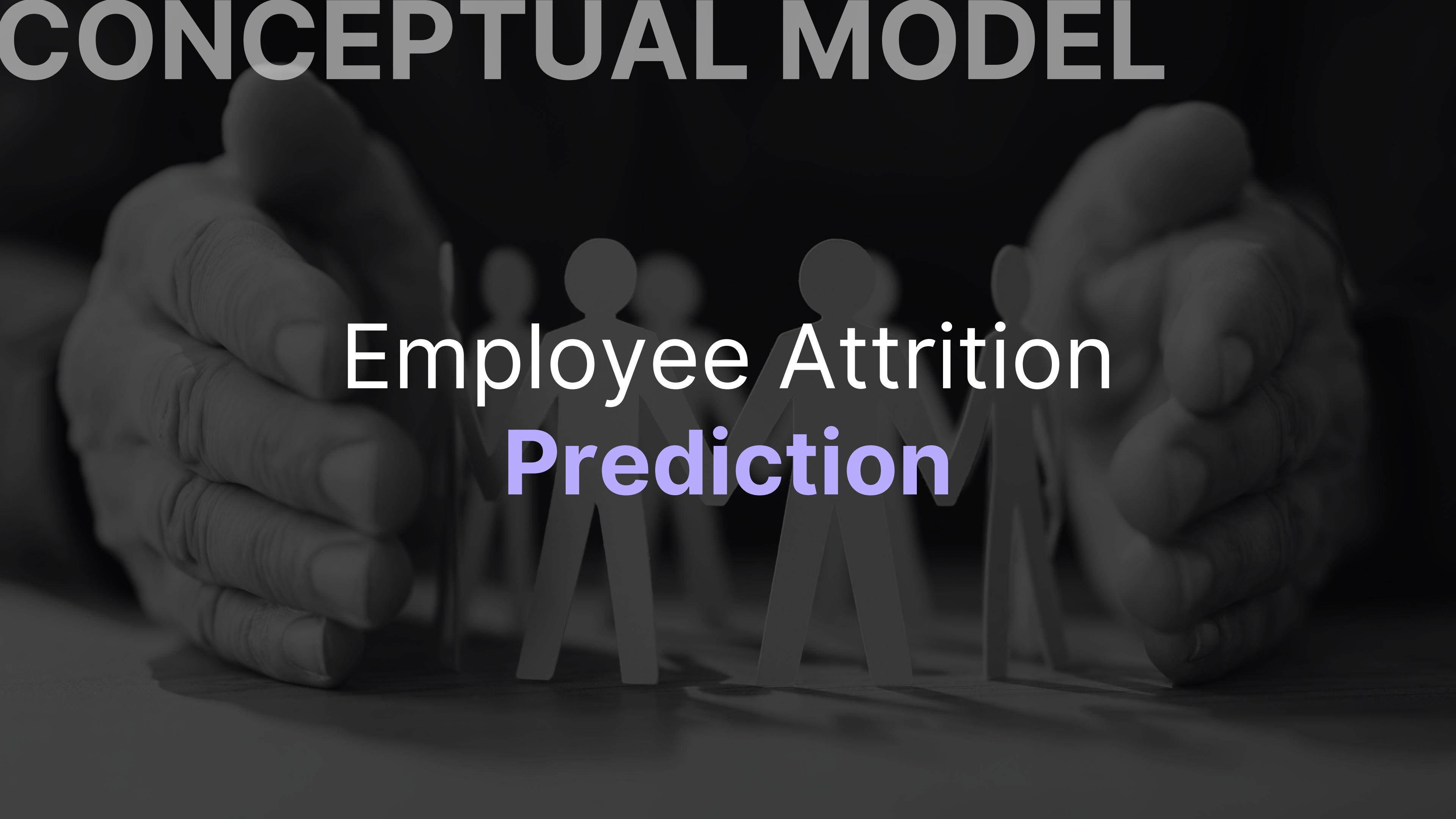
High attrition rates can lead to increased recruitment costs, loss of institutional knowledge, and decreased morale.



## SOLUTION

By predicting and understanding employee attrition, companies can devise and implement strategies to retain their valuable talent.

# CONCEPTUAL MODEL



Employee Attrition  
Prediction

# DATA MINING SOLUTION



## OBJECTIVE

---

**Predict** whether a given employees  
will leave the company

## APPROACH

---

**Classification** using **supervised** learning

# DATA

k

Data source  
found in Kaggle

## DATASET

Provides a detailed examination of employee attrition and its contributing factors, providing a robust set of quantitative and qualitative data

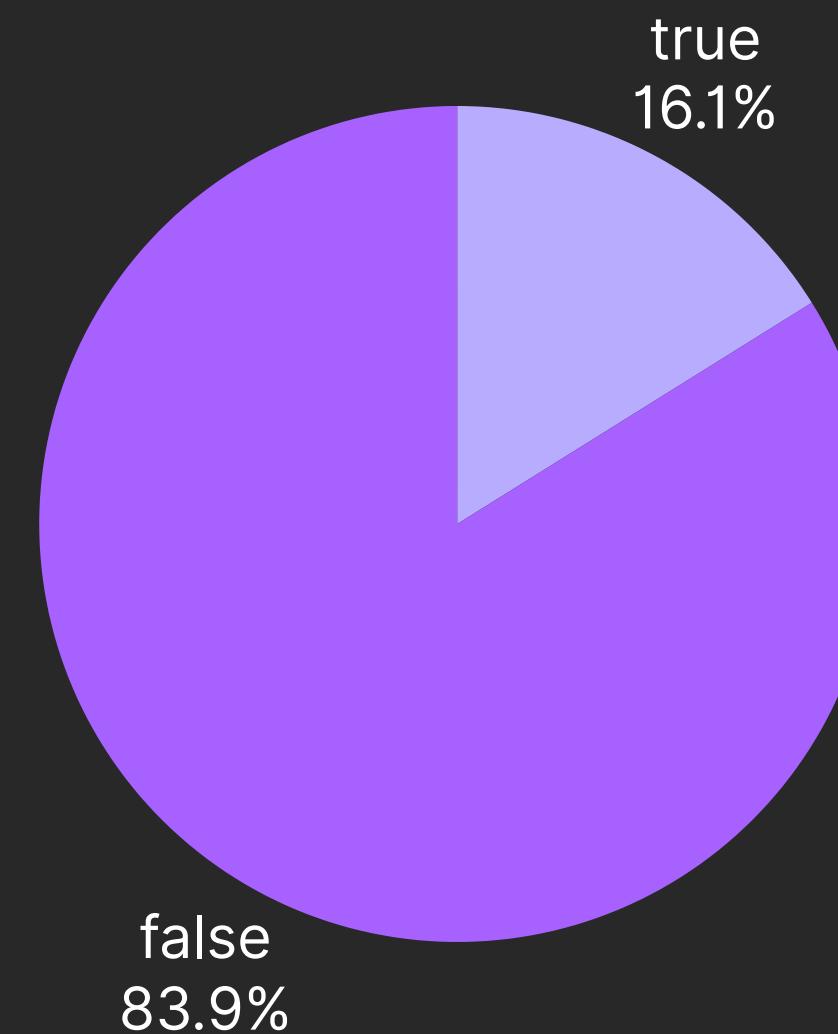
Columns: **35** / Rows: **1470**

Age	Attrition	Business Travel	DailyRate	Department	Distance From Home	Education	...	Worklife Balance	YearsAt Company
41	Yes	Travel_Rarely	1102	Sales	1	2	...	1	6
49	No	Travel_Frequently	279	Research & Development	8	1	...	3	10
37	Yes	Travel_Rarely	1373	Research & Development	2	2	...	3	0
33	No	Travel_Frequently	1392	Research & Development	3	4	...	3	8
27	No	Travel_Rarely	591	Research & Development	2	1	...	3	2
...	...	...	...	...	...	...	...	...	...

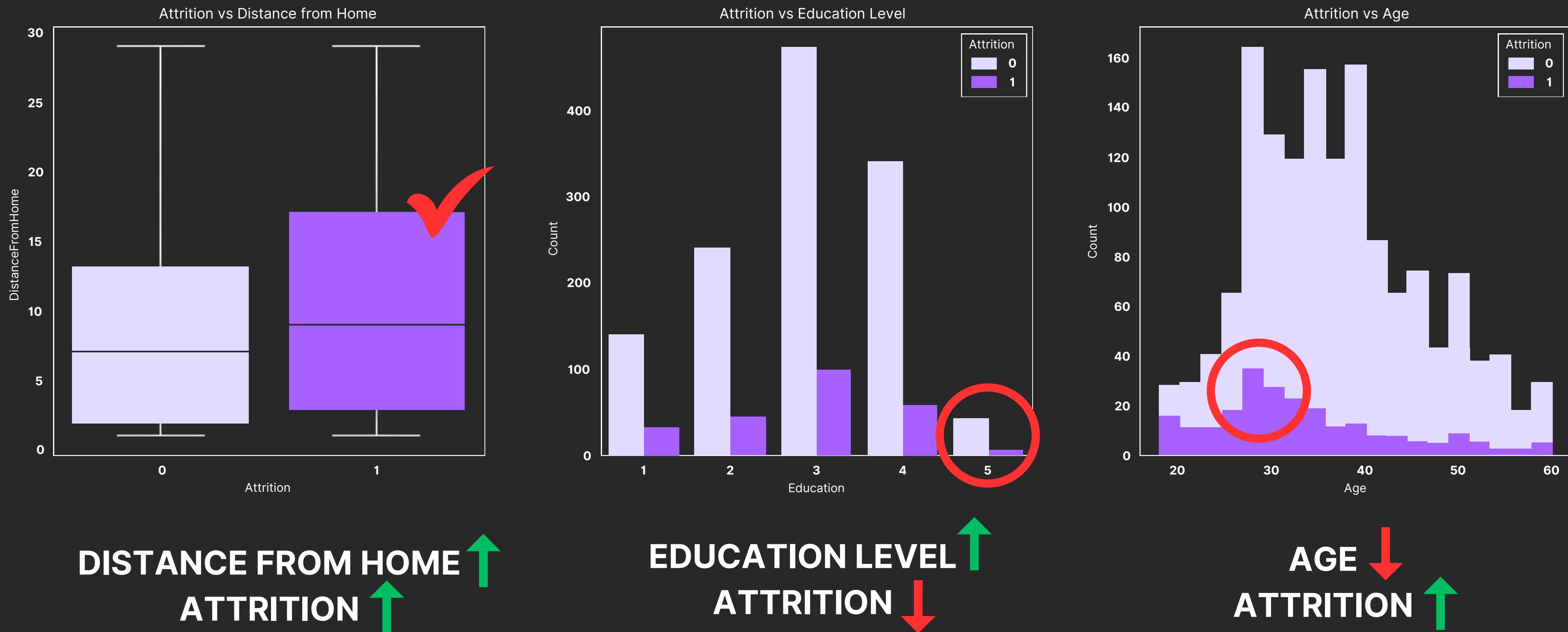
## TARGET VARIABLE

**Attrition:** Yes/No (binary)

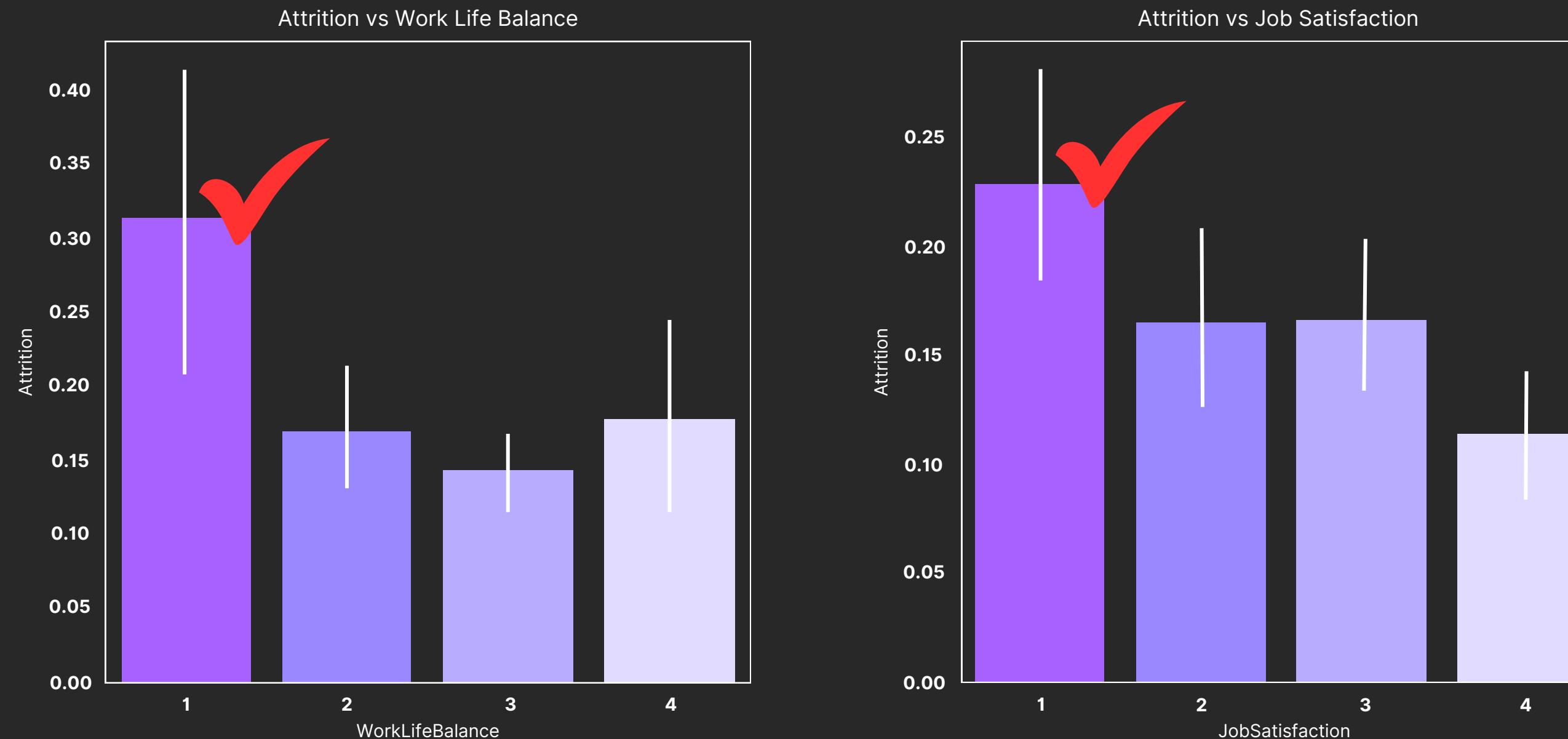
Whether or not the employee has left the organization (categorical)



# DATA EXPLORATION



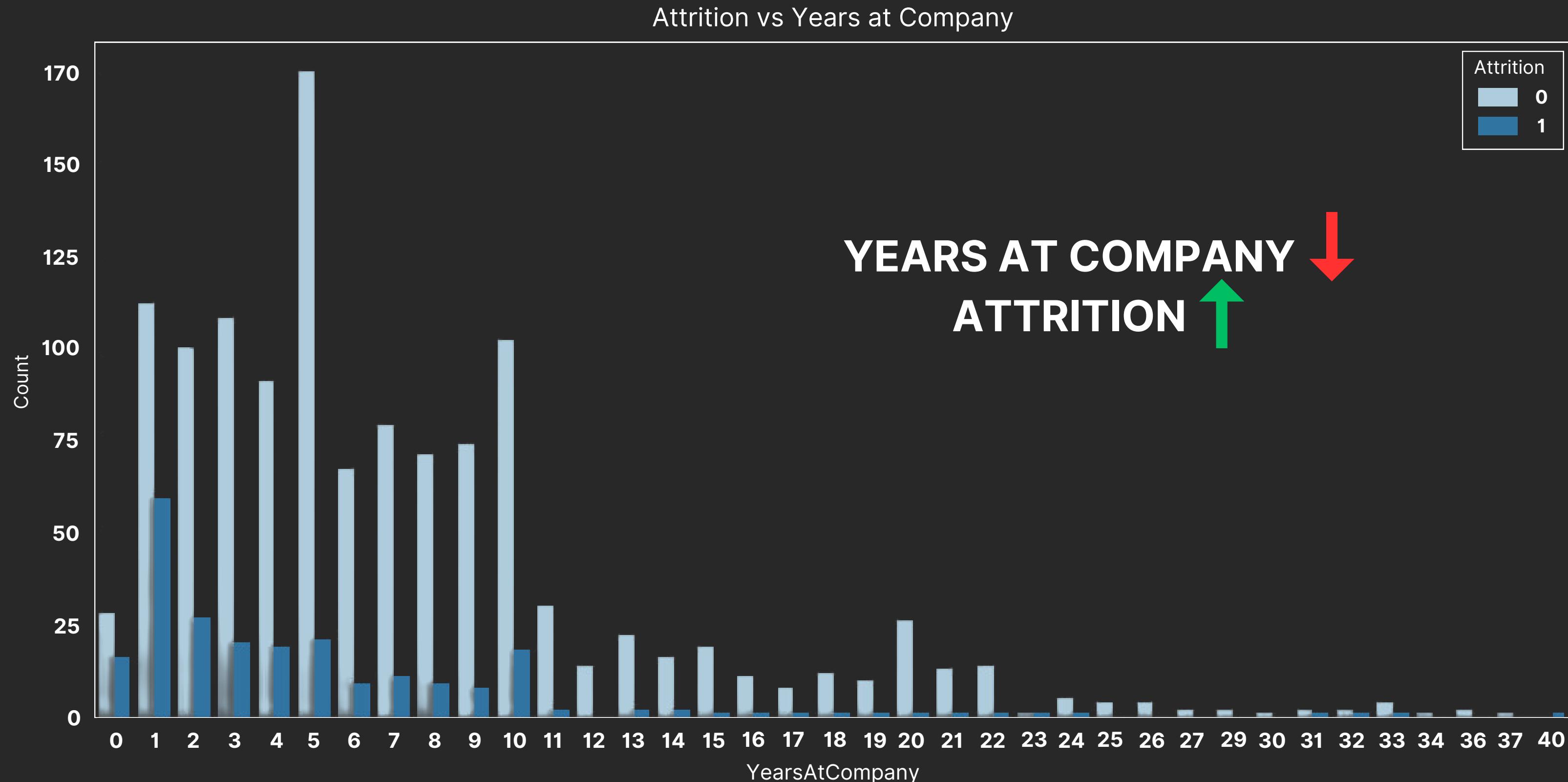
# DATA EXPLORATION



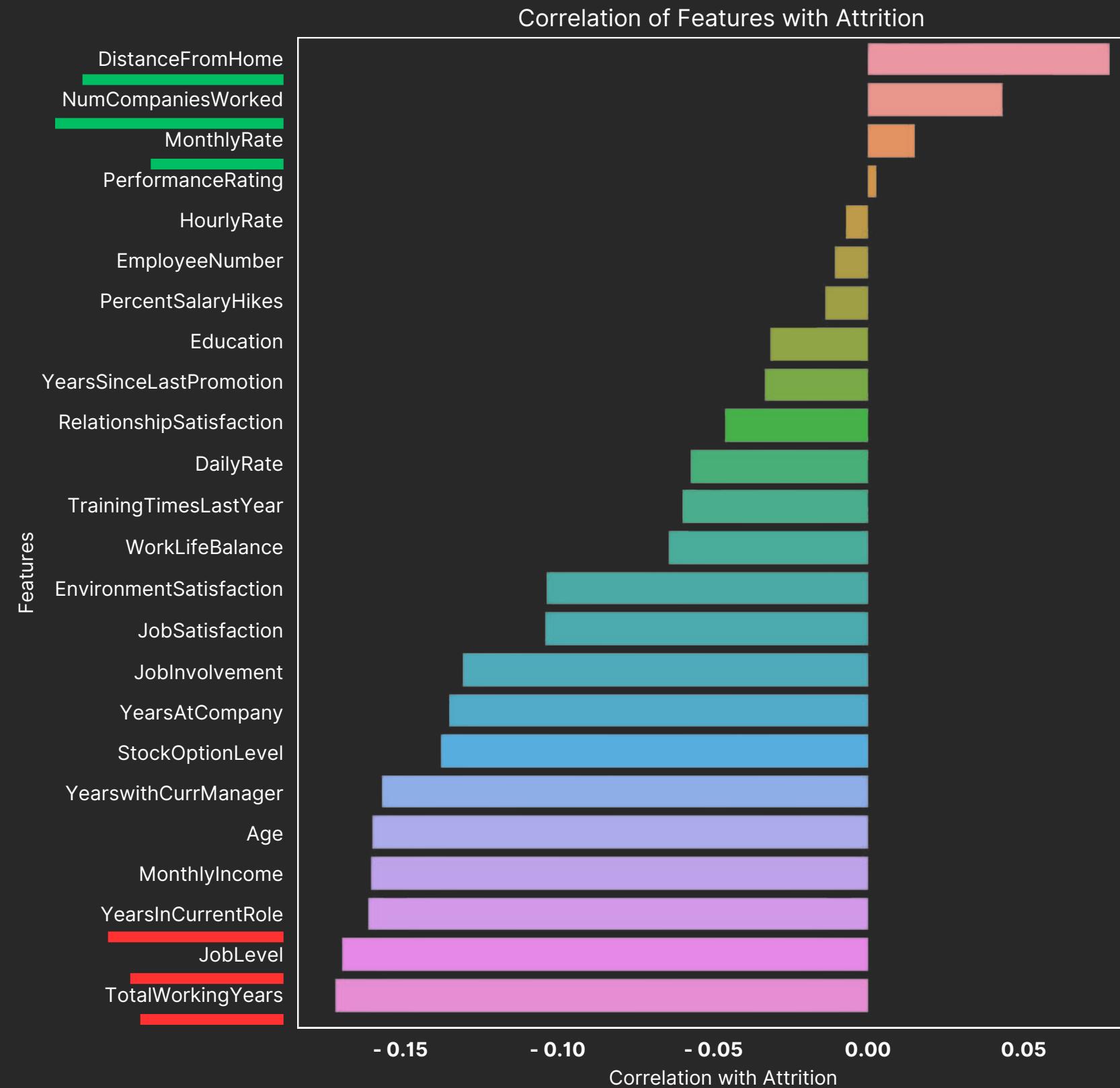
WORK-LIFE BALANCE ↓  
ATTRITION ↑

JOB SATISFACTION ↓  
ATTRITION ↑

# DATA EXPLORATION



# DATA EXPLORATION



## 1. Top 3 Highest **Positive** Correlations:

- DistanceFromHome
- NumCompaniesWorked
- MonthlyRate

## 2. Top 3 Highest **Negative** Correlations:

- TotalWorkingYears
- JobLevel
- YearsInCurrentRole

# DATA PREPROCESSING

0

## ORIGINAL DATA

Age	Attrition	MartialStatus
41	Yes	Single
37	No	Married
33	Yes	Single

1

## DATA ENCODING

Age	Attrition	MaritalStatus _Married	MaritalStatus _Single
41	1	0	1
37	0	1	0
33	1	0	1

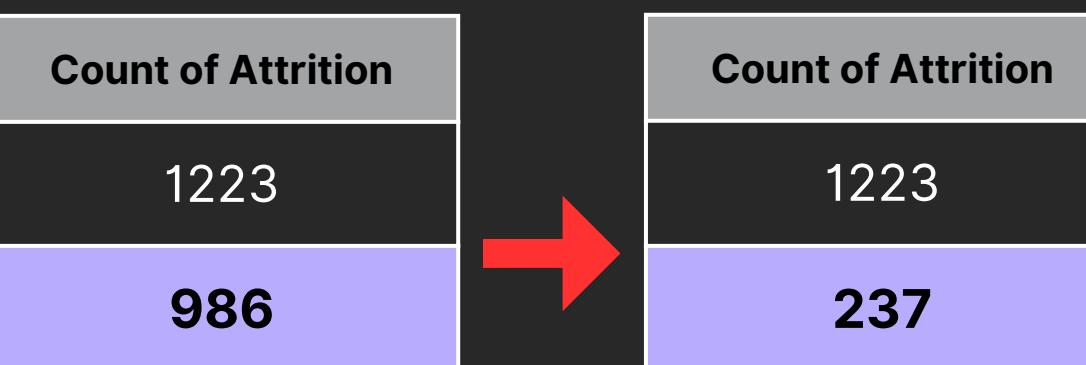
2

## FEATURE SCALING

Age	Attrition	MaritalStatus _Married	MaritalStatus _Single
0.446350	1	0	1
1.322365	0	1	0
0.008343	1	0	1

3

## DATA BALANCING



4

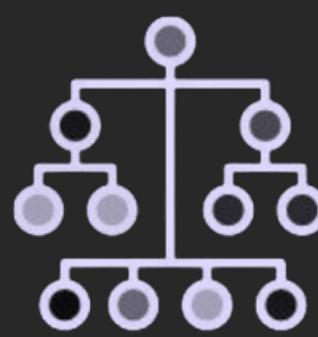
## DATA PARTITIONING

$X_{train}, X_{test}, y_{train}, y_{test}$   
 $test\_size=0.2, random\_state=42$

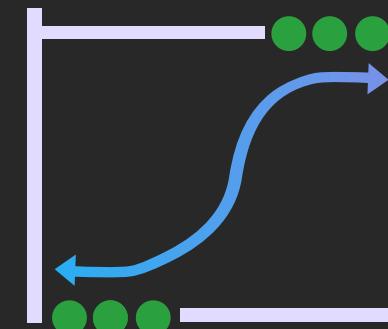
# MODELING

# 17

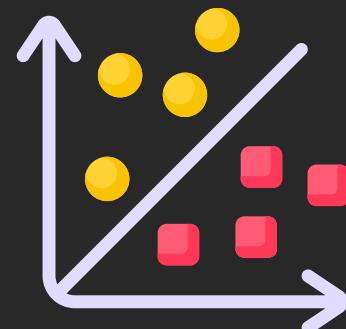
SUPERVISED LEARNING MODEL



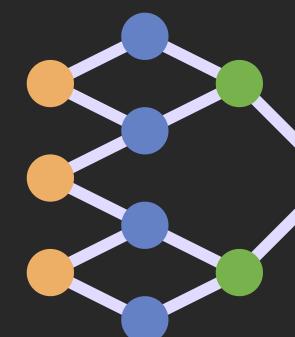
DECISION  
TREE \*



LOGISTIC  
REGRESSION \*



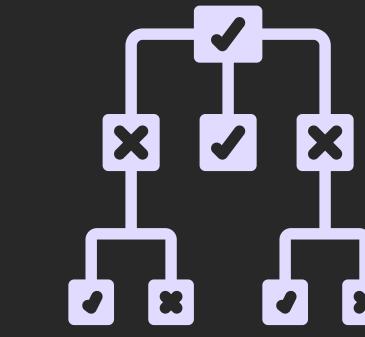
SVM \*



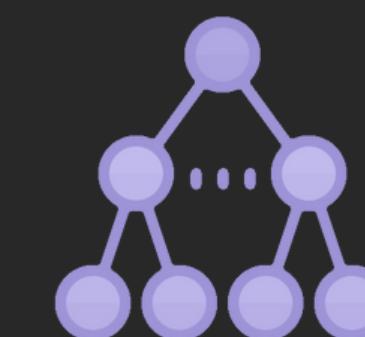
NEURAL  
NETWORKS \*



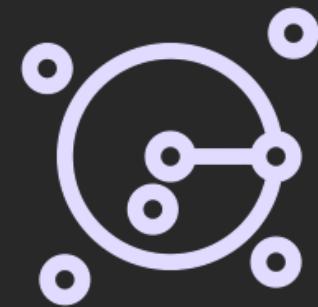
NAIVE  
BAYES



RANDOM  
FOREST \*,  
ENSEMBLE  
MODEL



GRADIENT  
BOOSTING \*,  
XGBOOSTING \*



KNN \*

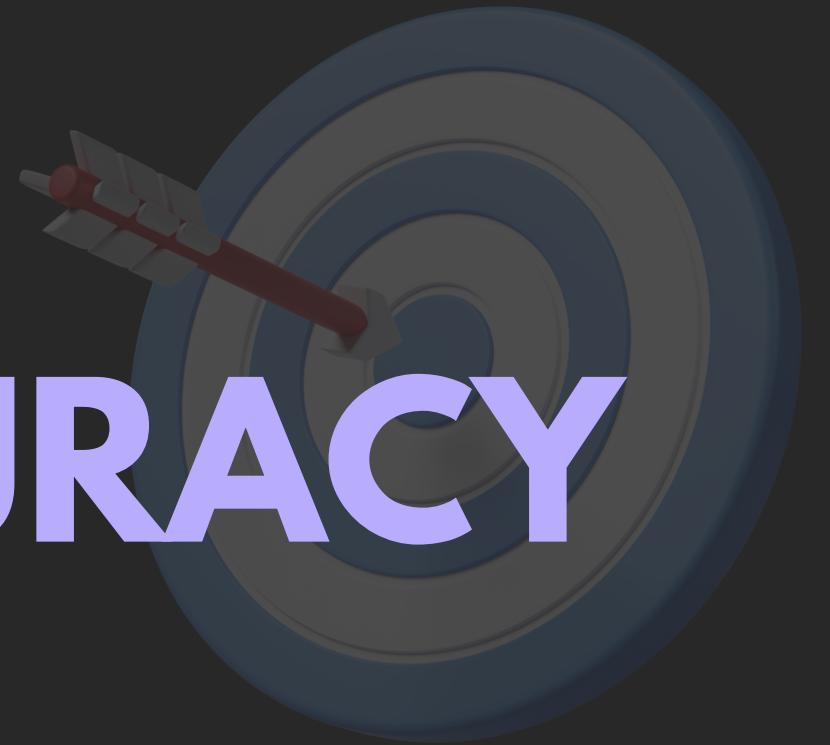
\*: includes hyper-parameter tuned model

# EVALUATION METRICS

## ACCURACY

### Performance

- Measures correct predictions of **both** staying and leaving employees
- **General performance** of our attrition model



## F1-SCORE

### Balance

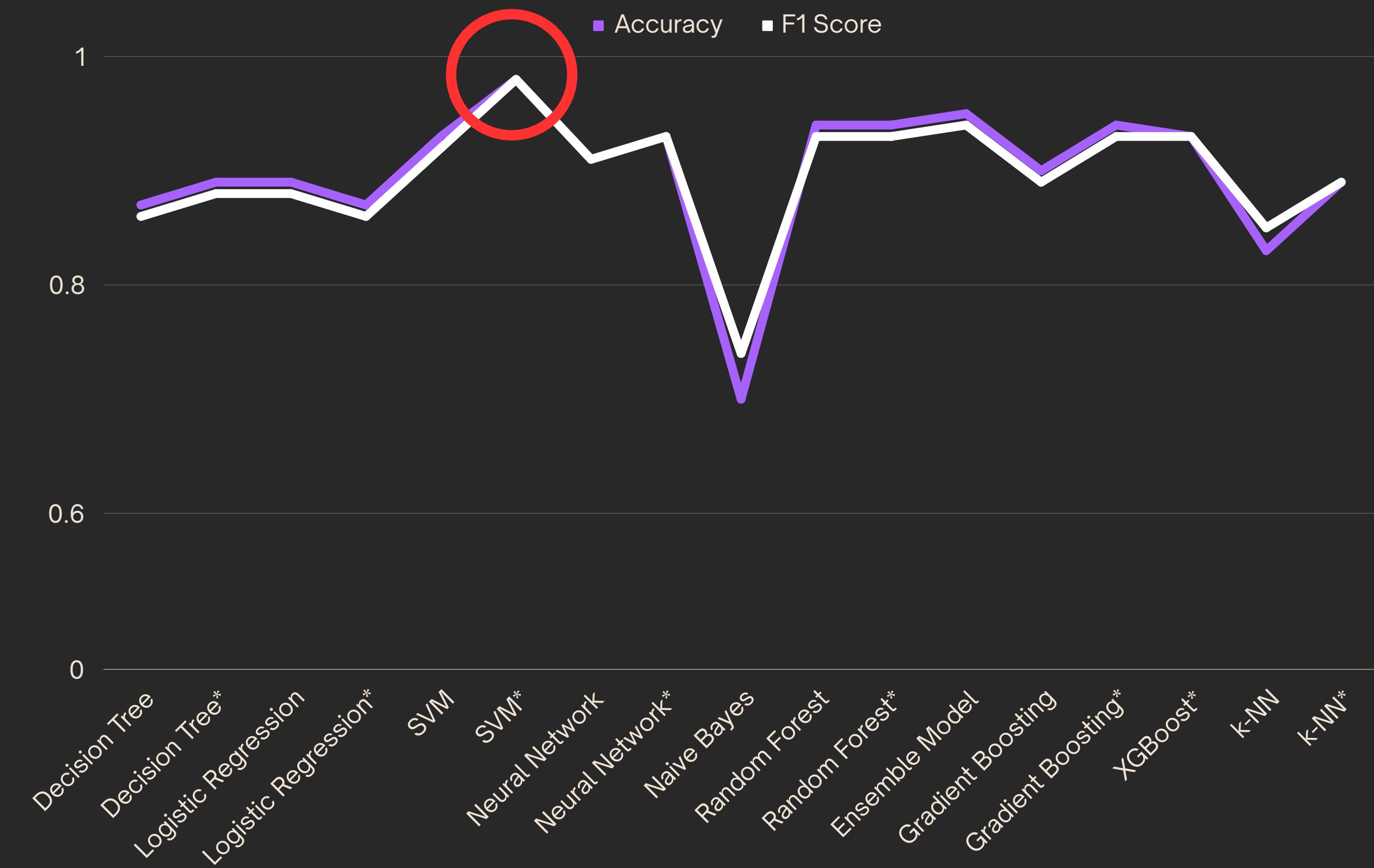
- Harmonic mean
- **precision** (avoiding false attrition alarms) & **recall** (catching true cases of attrition)



### Prediction

- Focuses on the accuracy of predicting the **minority class** (employees likely to leave)

# EVALUATION



# SVM

with hyperparameter tuned

\*: indicates hyper-parameter tuned model

# FINAL MODEL

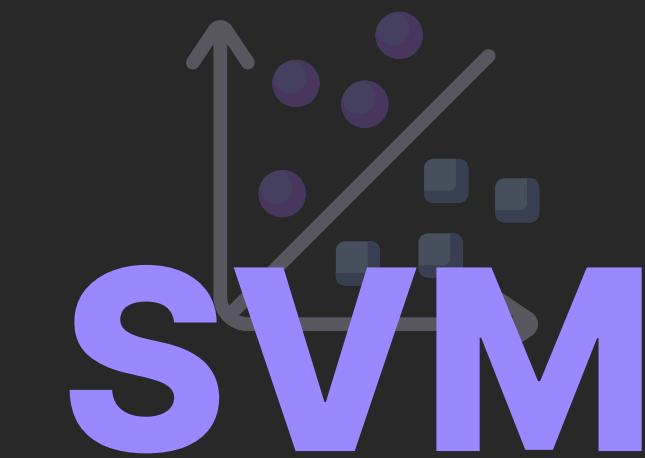
BEST PARAMETERS: {'C': 10, 'gamma': 0.1, 'kernel': 'rbf'}

CLASSIFICATION REPORT:

	Precision	Recall	f1-score	support
0	0.97	0.98	0.98	235
1	0.98	0.97	<u>0.98</u>	209
accuracy			<u>0.98</u>	444
macro avg	0.98	0.98	0.98	444
weighted avg	0.98	0.98	0.98	444

CONFUSION MATRIX :

	Positive	Negative
Positive	231	4
Negative	6	203



with hyperparameter tuned

OUTPUT:

	Attrition
0	1
1	0
2	1
3	0
4	0
5	0
6	0

# DEPLOYMENT PLAN



## MODEL DEVELOPMENT

- Process employee data and finalize the XGBoost model
- Export model for deployment



## INTEGRATION & LAUNCH

- Choose cloud or on-premises platform
- Set up a secure API for model access
- Implement authentication and encryption



## MONITORING & IMPACT

- Ensure system scalability and load management
- Implement monitoring, logging, and CI/CD for updates
- Conduct staging tests and provide API documentation

# BUSINESS IMPACT



## COST EFFICIENCY

By predicting attrition, companies can  
**reduce hiring and training costs**



## TALENT RETENTION

By understanding the reasons behind attrition,  
HR can implement **strategies to retain  
valuable employees**



## ENHANCED EMPLOYEE EXPERIENCE

By addressing potential reasons for attrition,  
the **overall employee experience**  
can be improved



**ANY QUESTIONS?**



**THANK YOU FOR  
LISTENING!**