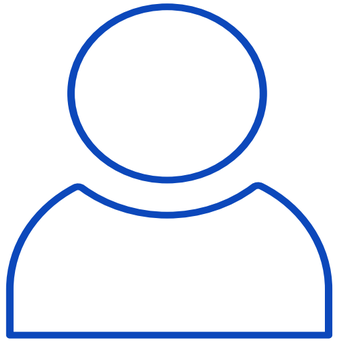
Contents



Situation

IBM is facing significant losses due to high employee turnover.

Task

As consultants, our mission is to develop a machine learning model that predicts whether or not an employee will leave the company based on various factors.

Action

We approached the problem as a binary classification task, with the model outputting either 0 (employee stays) or 1 (employee leaves)

Result

Help IBM take proactive steps to reduce employee departures.

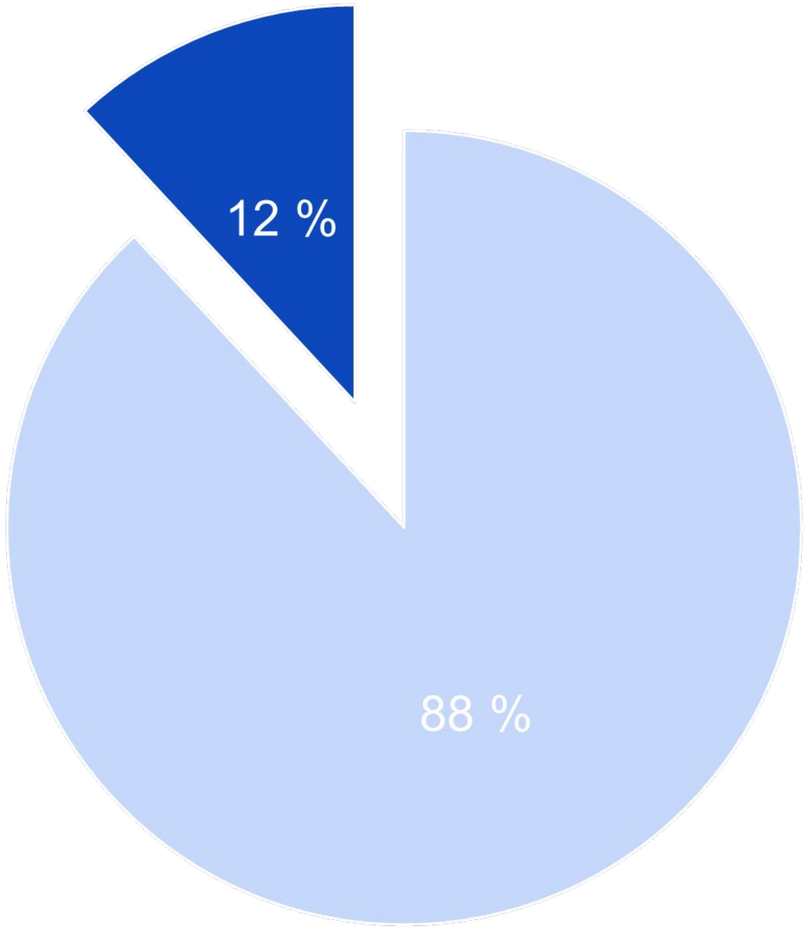
Understanding The Data

- No missing values
- No duplicate values

Corporate Workforce Profile

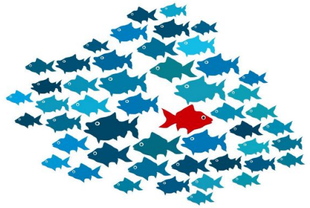
- Average Age: **36 years**
- Gender Ratio: **1.7** Male vs. Female
- Job Satisfaction: **2.8** out of a range of 1 to 4
- Distribution of Employees by Departments :
  - Sales: **28%**
  - Human Resources: **2 %**
  - Research & Development: **70 %**

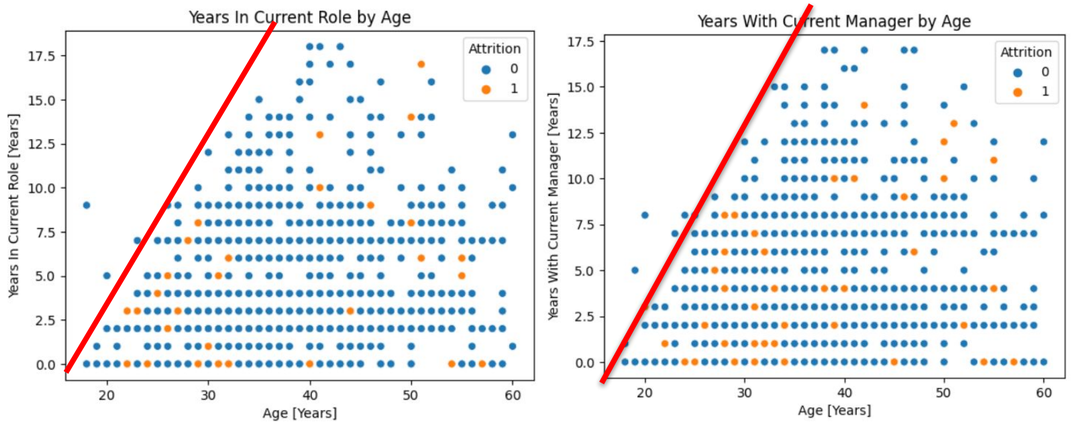
Understanding The Data

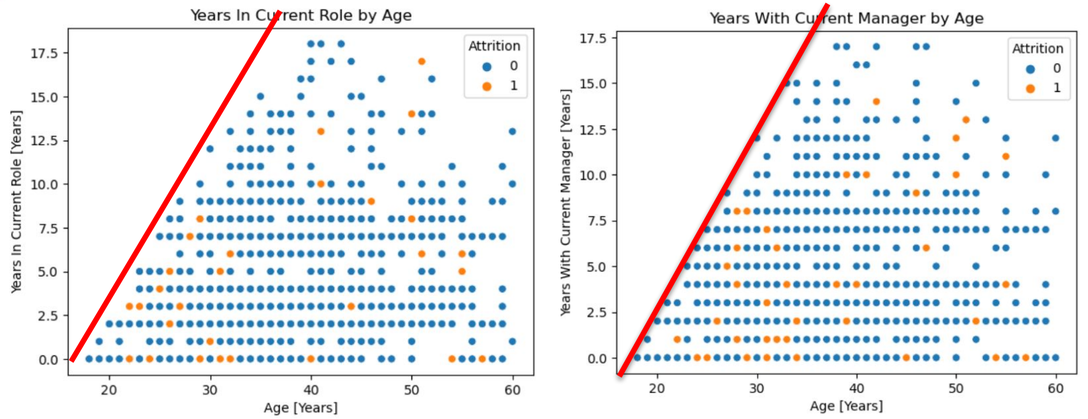


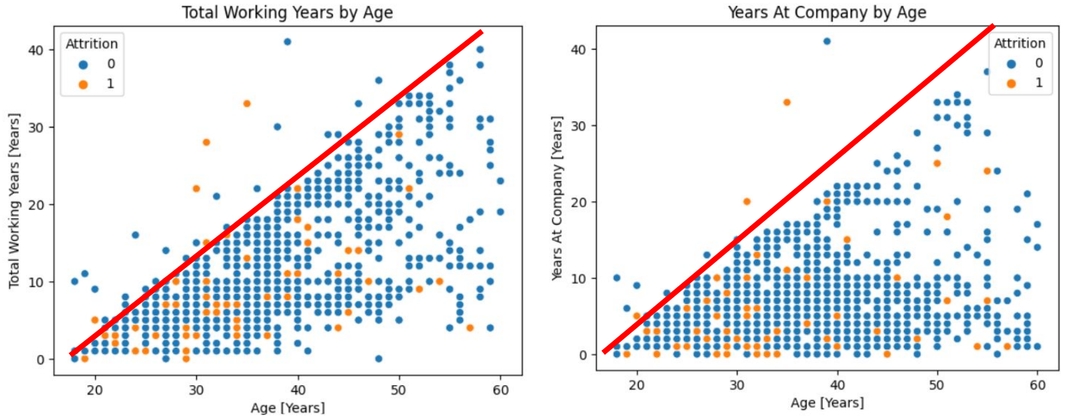
12 % of Employees Leave

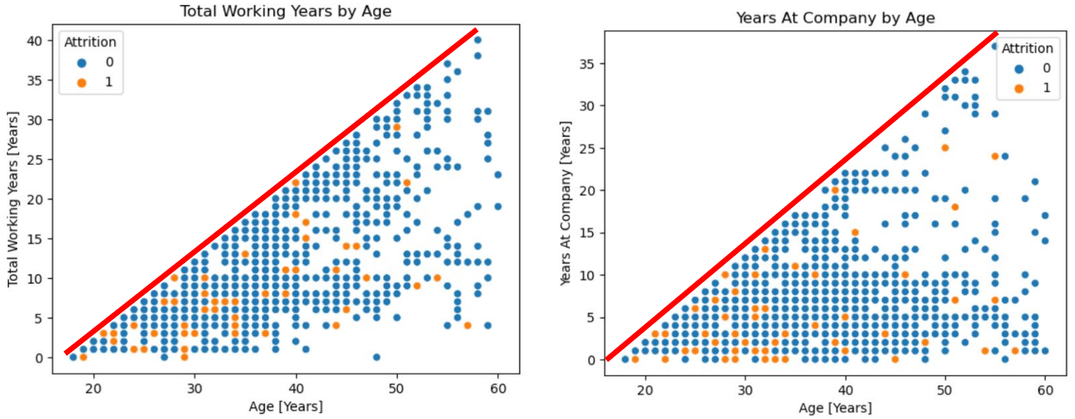
Handling Outliers









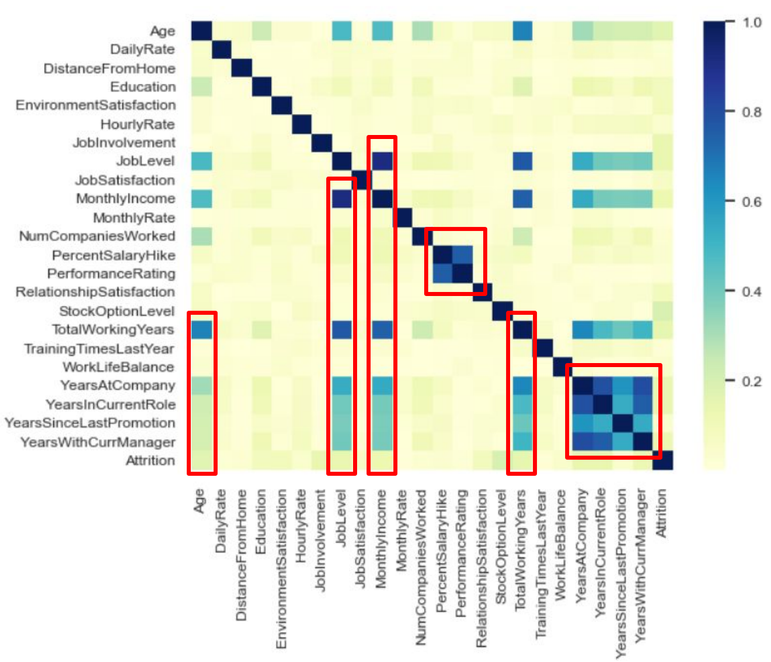


Feature Selection

Drop features whose domain is a single value:

- EmployeeCount
- Over18
- StandardHours

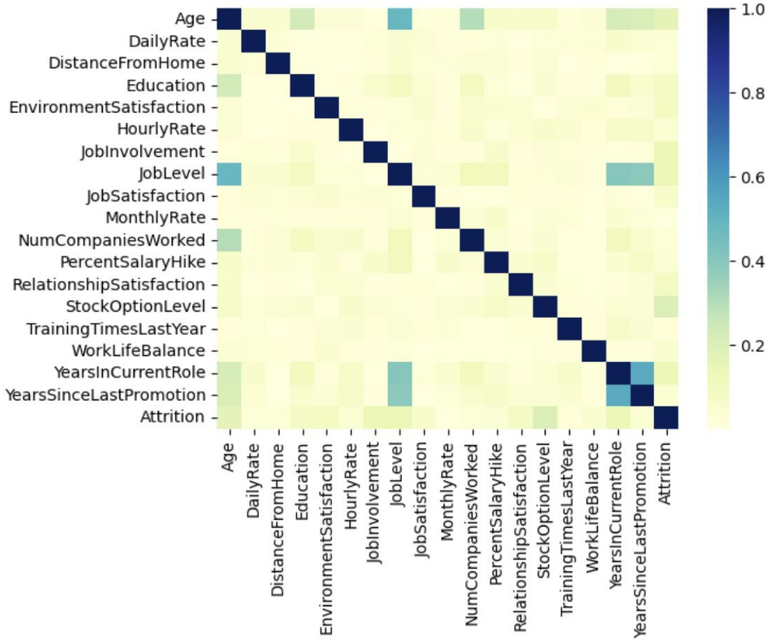
Correlation Matrix



Drop:

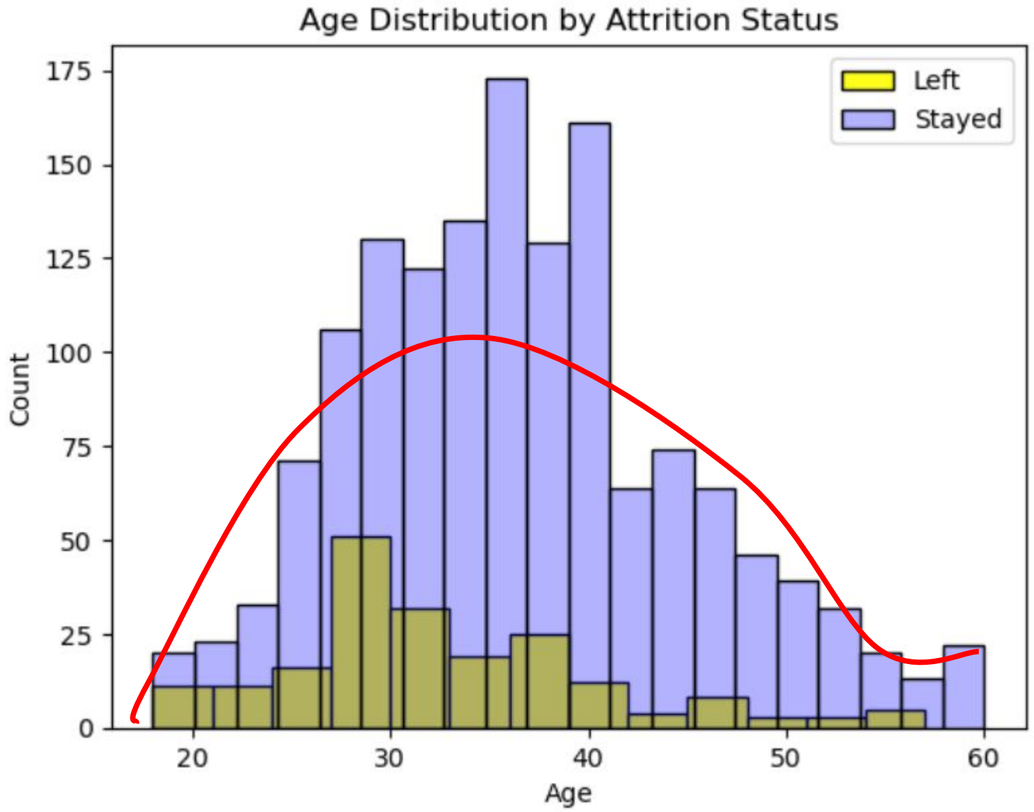
- MonthlyIncome
- TotalWorkingYears
- PerformanceRating
- YearsWithCurrManager
- YearsAtCompany

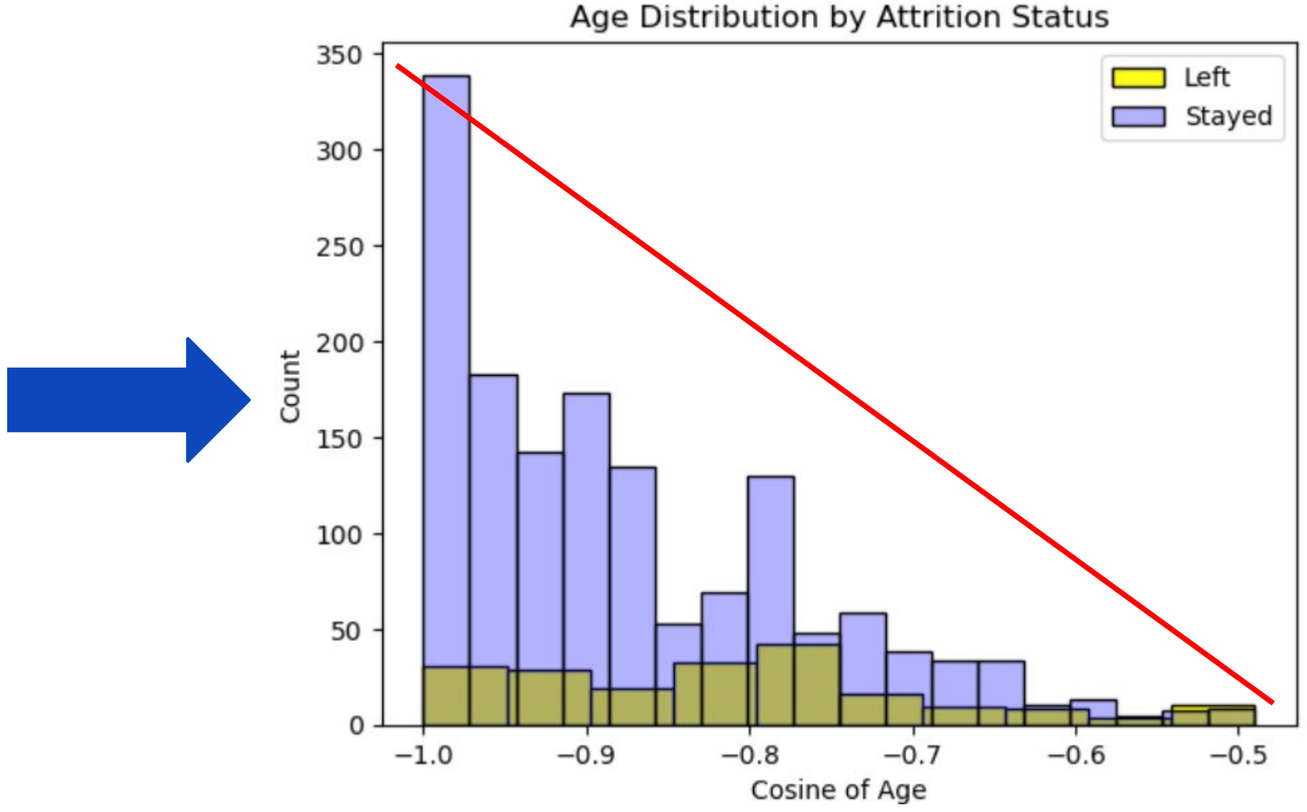
Final Features



Feature Creation

Converting the Age into a Linear Feature





Feature Creation

ONE-HOT-ENCODING (OHE) AND STANDARD SCALING

- Data may be represented as **words**, **letters**, or **symbols**
- Scaling is a common preprocessing step as most Machine Learning algorithms only process numerical data and require **standardized** Numerical Variables.
- We did **OHE** on **Categorical Variables** and **Standard Scaling** on **Numerical Variables**

Models Trained

Logistic Regression

Baseline Model

Gaussian Naive Bayes

Simple - Efficient

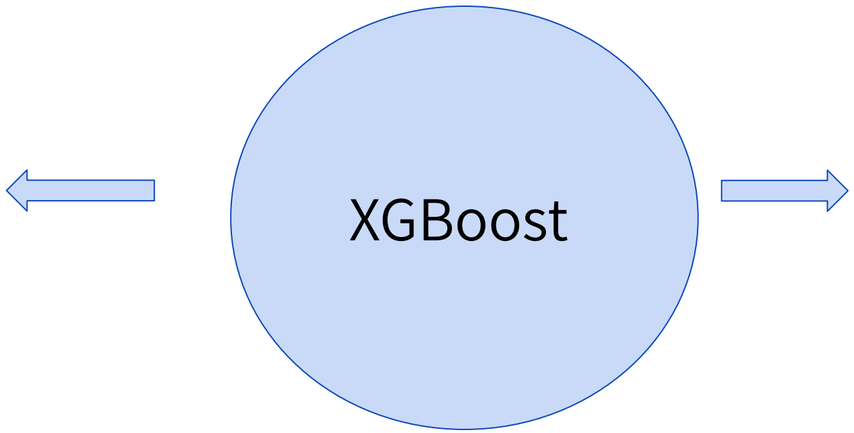
Random Forest

**Flexible - Powerful -
For High Number of Features**

Why did We Choose the Model?

Informative by offering the possibility to extract information on **variable importance**

Highly **versatile** and can be applied to Binary Classification tasks



Exceptional speed due to its ability to perform **parallel computation**

How Many Employees will Leave Next Year?

ID	Predictions
1677	1
1678	0
1679	0
1680	0
1681	1

THRESHOLD = **0.12**

Predictions	Amount
0	808
1	311

311 is the Number of workers leaving the company

Final Conclusions

Profile of **predicted leavers**:

- Average Age: **32 years**
- Gender Ratio: **1.9** Male vs. Female
- Job Satisfaction: **2.6** out of a range of 1 to 4
- Distribution of Employees by Departments
  - Sales : **27 %**
  - Human Resources : **4 %**
  - Research & Development : **70 %**

Proposed Actions:

1. Implement **employee benefits** especially **targeting young professional in R&D**, such as company phones for private usage, a company gym, or free meals
2. Invest in employee development by creating a **career plan** for young professionals
3. Foster a **positive work environment** by defining strong company values

Thank you !