

Machine Learning and HR: Predicting Employee Attrition

Presented by Matt Schwartz

Contents

- 03 - Overview and Business Problem
- 04 - Data
- 05 - Methods
- 06 - Data Understanding
- 09 - Results
- 10 - Conclusions and Further Analysis

03

Overview and Business Problem

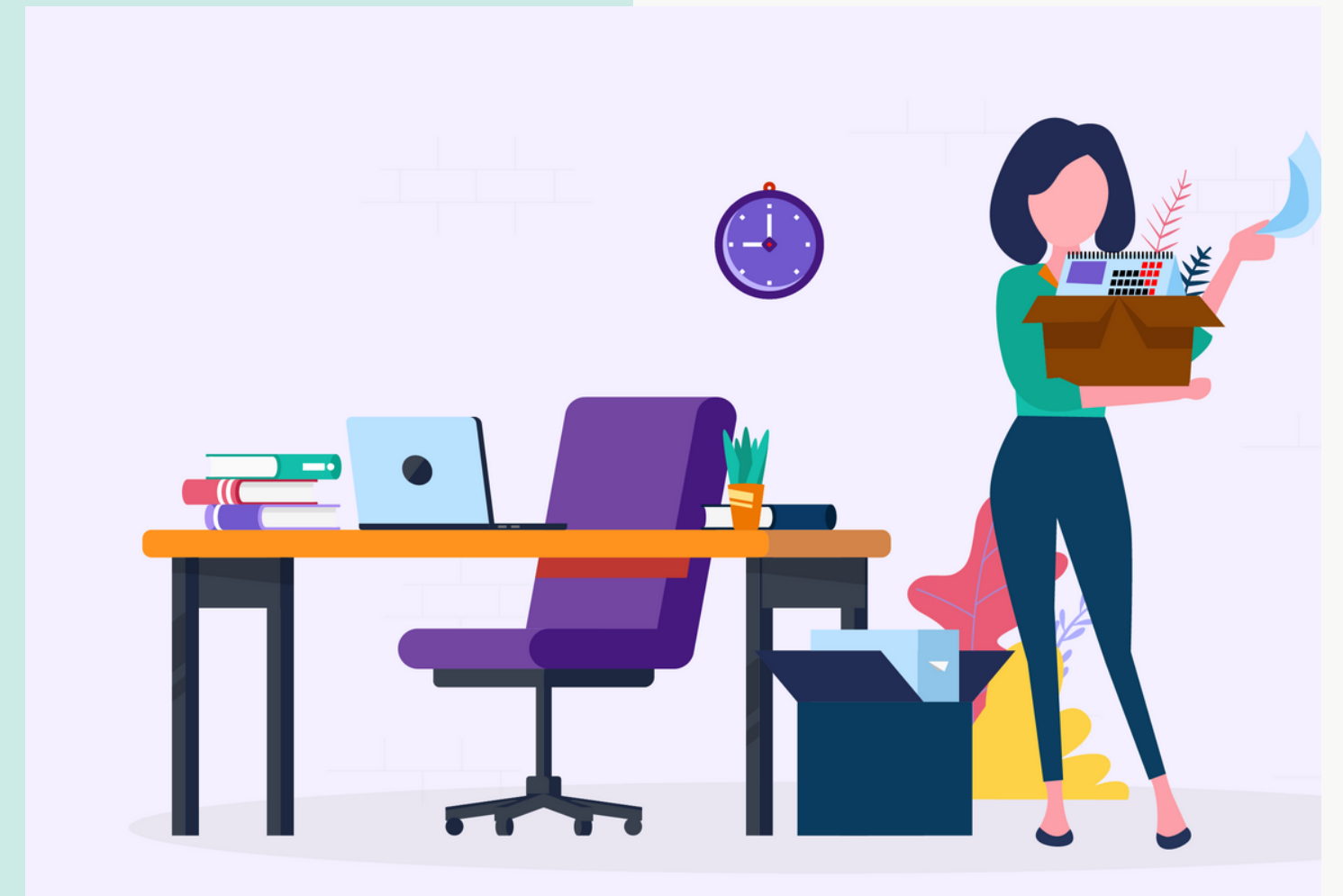


- Employee attrition is costly - can be up to 1.5x the salary of the employee who left
- How can we develop a model to predict which employees are likely to leave?
- What factors are most important in predicting employee attrition?

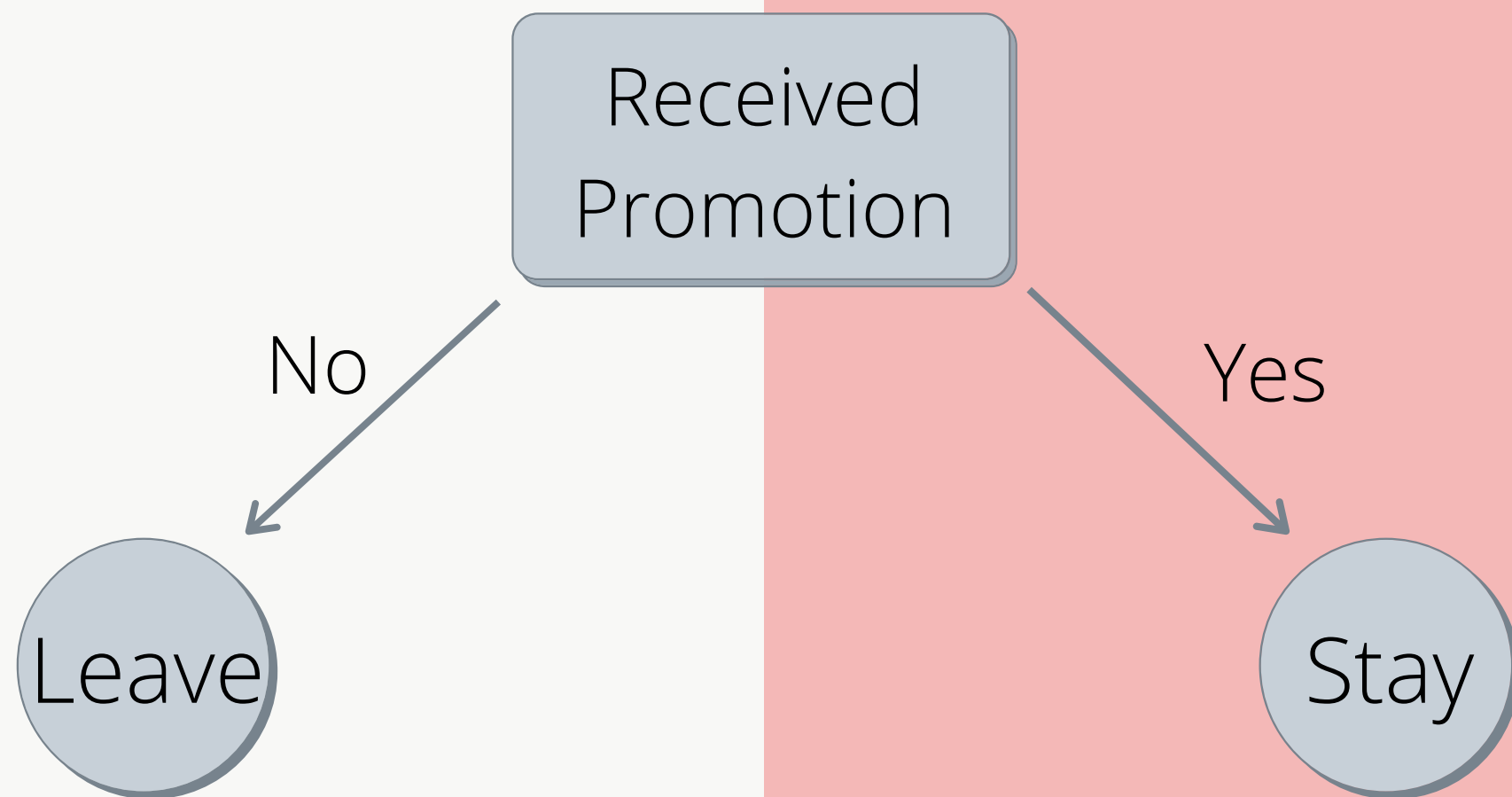
Data

04

- 1470 employees in a dataset created by IBM data scientists, 30 attributes per employee
- Some features include:
 - Age
 - Job satisfaction
 - Education
 - Total working years
 - Job role
 - Monthly income
 - Distance from home



Methods



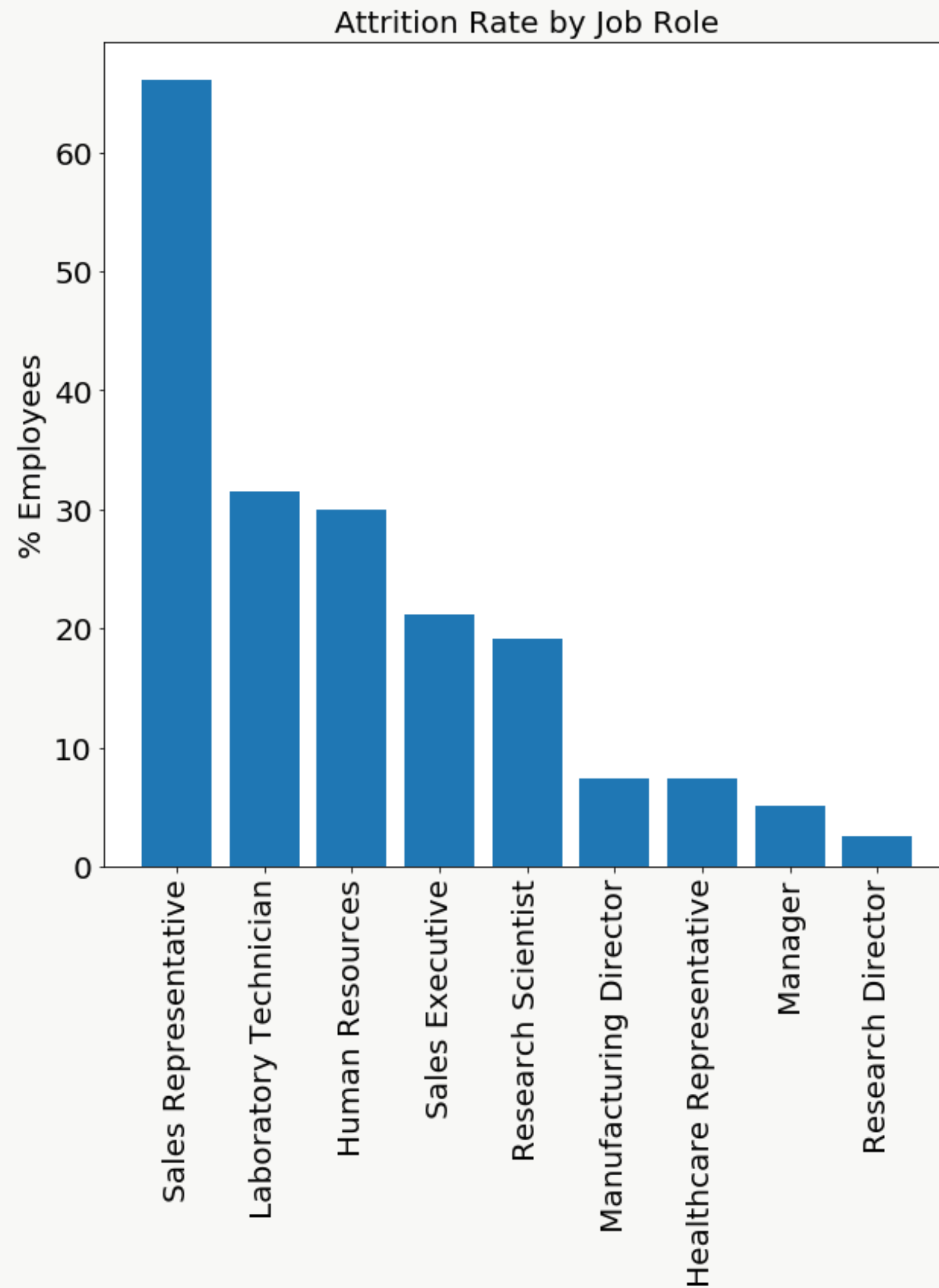
CLASSIFICATION MODELS

- Classification models are used when the target variable (the thing we're trying to predict) is binary
- Attrition is a binary target variable - each employee either left or didn't (1 or 0)

DEFINING A QUALITY MODEL

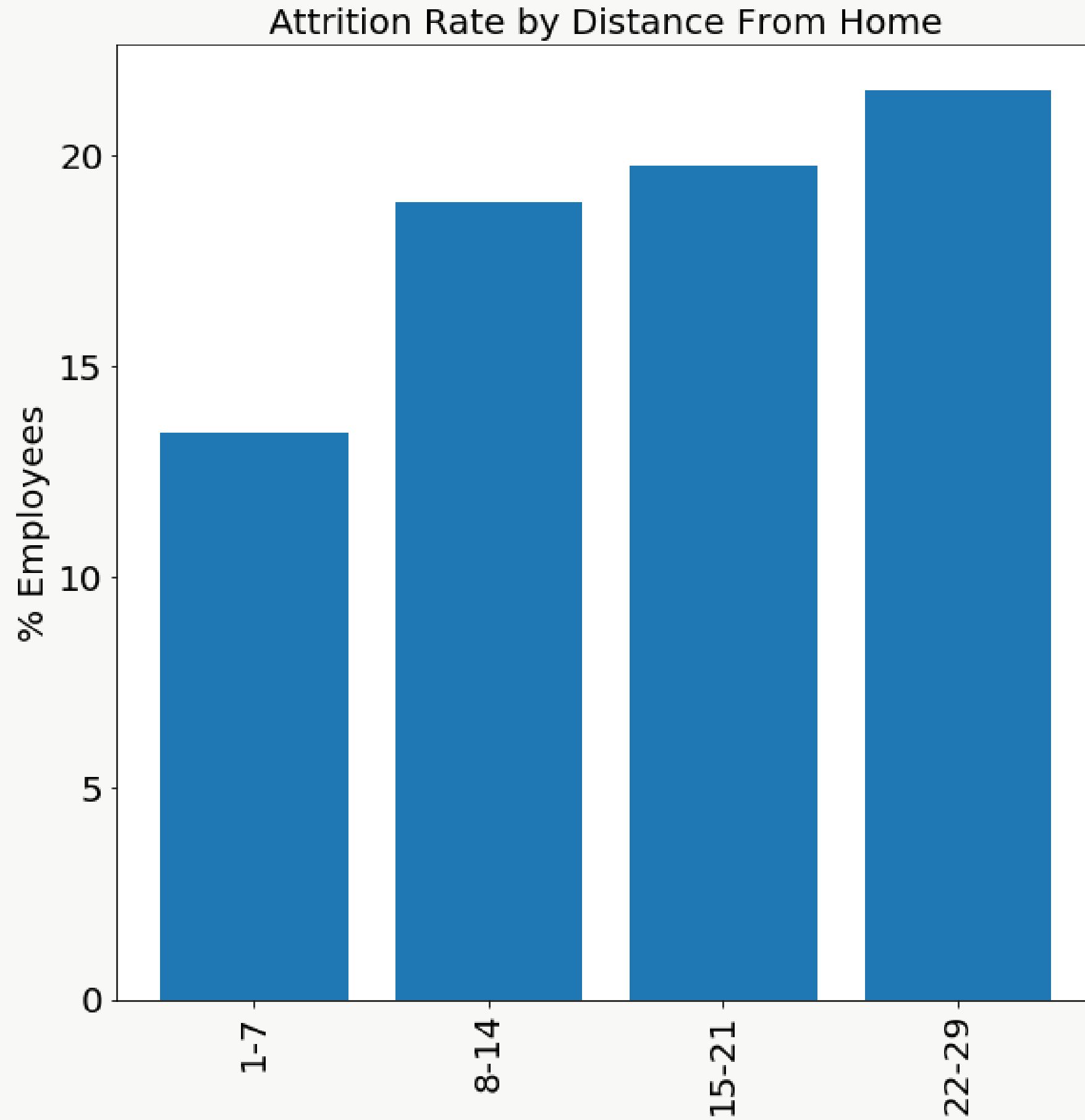
- Low false negative rate

06



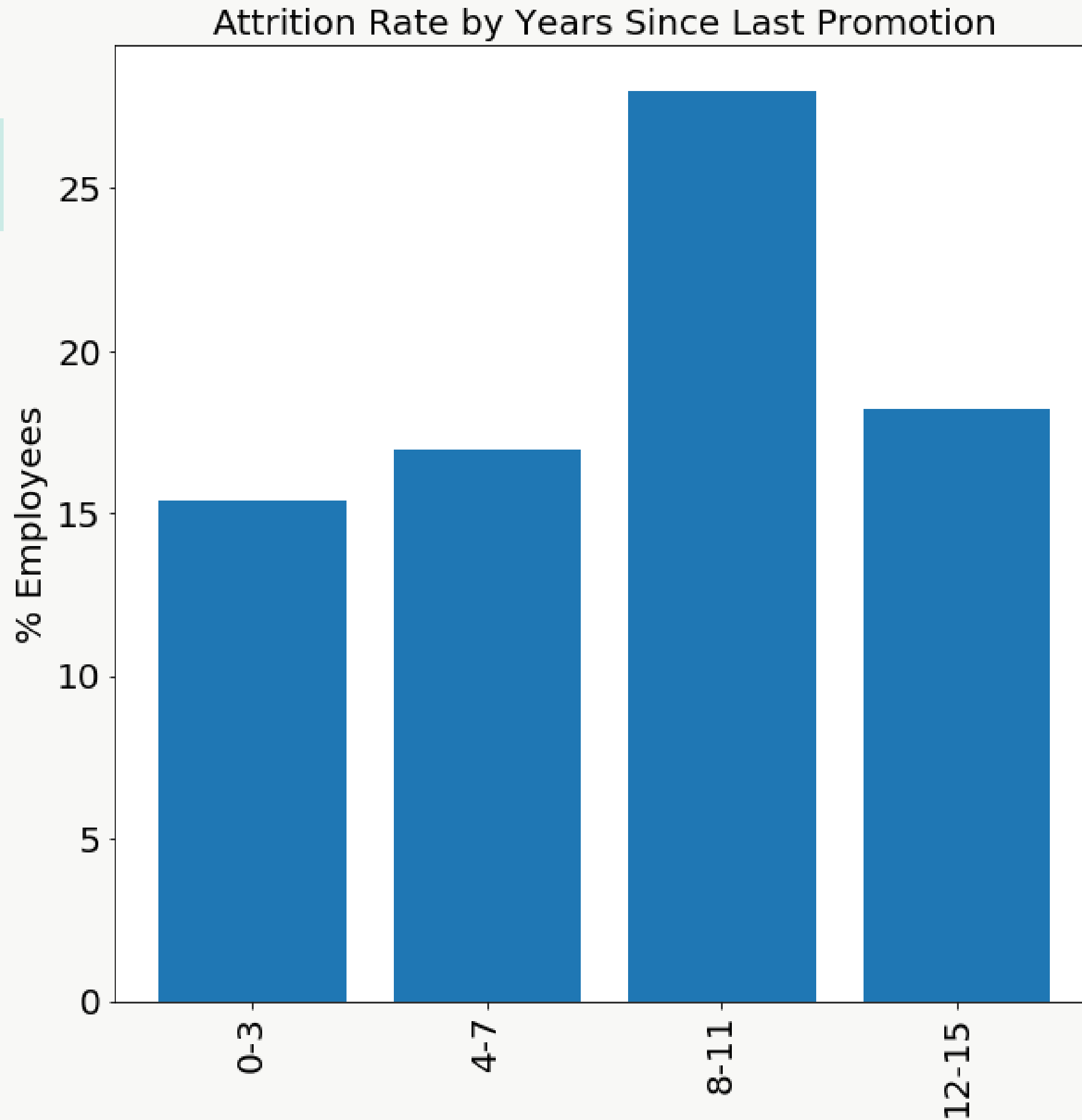
**Data
Understanding
- Attrition by
Job Role**

07



**Data
Understanding
- Attrition by
Distance From
Home**

08



**Data
Understanding
- Attrition by
Years Since
Last Promotion**

Results

09

- Model misclassified only 9 employees who actually left the company
- 5 most important features:
 - Years since last promotion
 - Frequent business travel
 - Working overtime
 - Distance from home
 - Job role = Sales Representative



Conclusions and Further Analysis

INVESTMENT

Identify and invest in employees you're at risk of losing

ADDRESS ISSUES

Top causes of attrition are easily addressable

USE ETHICALLY

Work to remove bias from hiring and firing decisions, don't misuse HR data and models used here

Presented By, Matt Schwartz

Flatiron Online Part-Time Data Science Cohort

Phase 3 Project - 3/28/2021

mtschant@gmail.com

**Thank You For
Your Time!**