Employee Retention Data Analysis Presentation

by Callum Hutchinson

Problem Specification and Analysis Roadmap

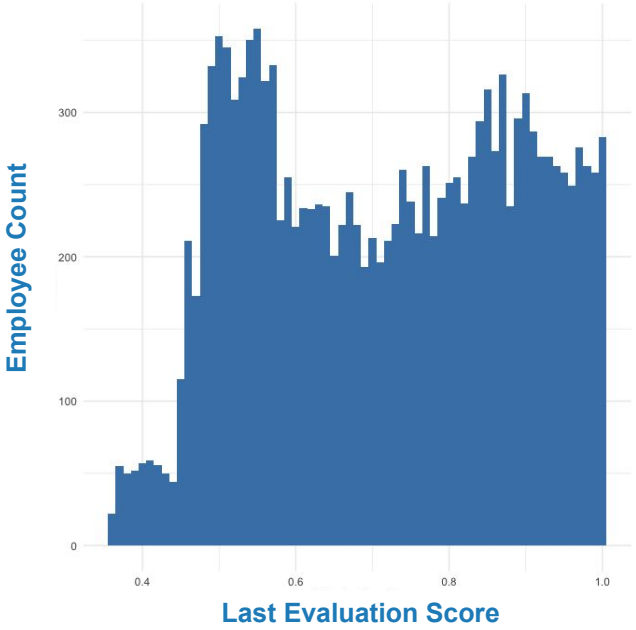
- Our client would like to understand how it can improve retention of its high performing employees and would like to be able to predict which employees are most likely to leave the company, so it can make operational changes to stop employees from leaving.

- Our analysis will therefore be broken into three parts:
  - **Part 1:** Exploratory Data Analysis
  - **Part 2:** Key Takeaways & Strategic Recommendations
  - **Part 3:** Predictive Modelling

(Analysis conducted using RStudio)

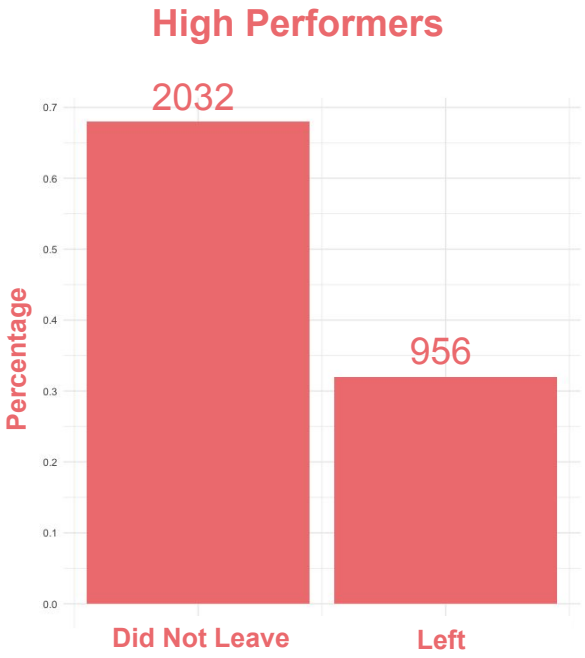
Part 1:
Exploratory Data Analysis

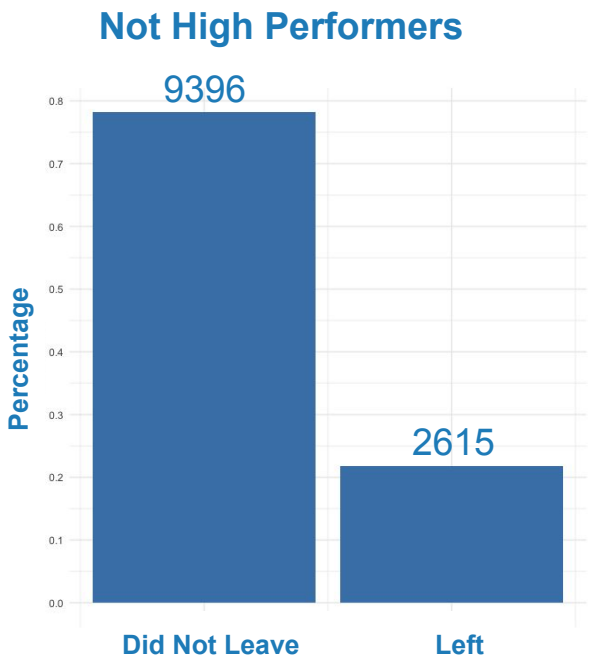
Who are the “high performing” employees?



- The minimum last evaluation score received was 0.36 and the maximum was 1.0
- From here on we will define “high performing” employees as those in the **top 20%** of the last evaluation score
- An employee with last evaluation score of 0.89 and above is a “high performer”
- **2,988** employees in the dataset are “high performers”

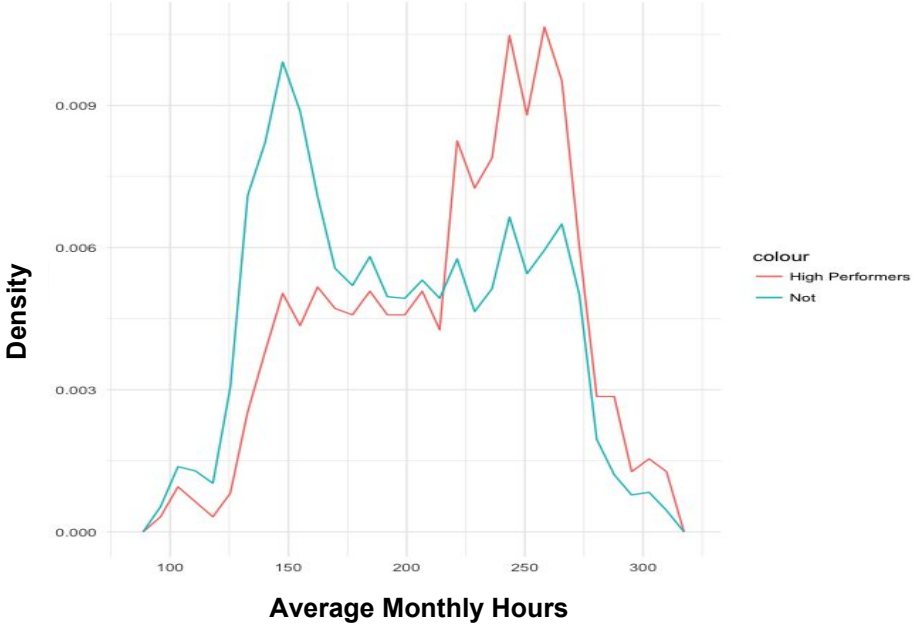
Are high performing employees more likely to leave?

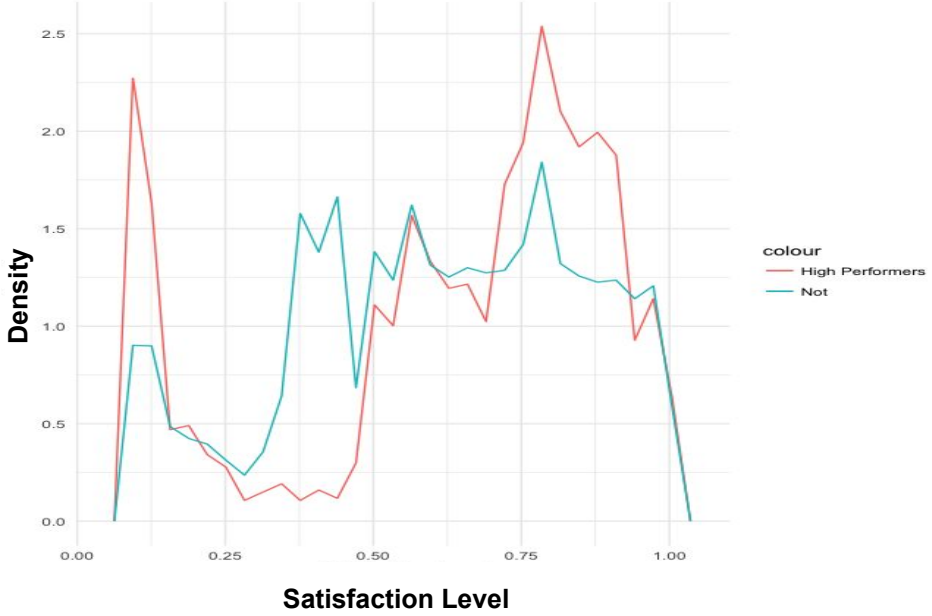




The probability of leaving given “high performing” status is **32%**, versus 22% for non “high performers”. Thus, high performing employees are more likely to leave than non high performing employees

How do high performers differ?

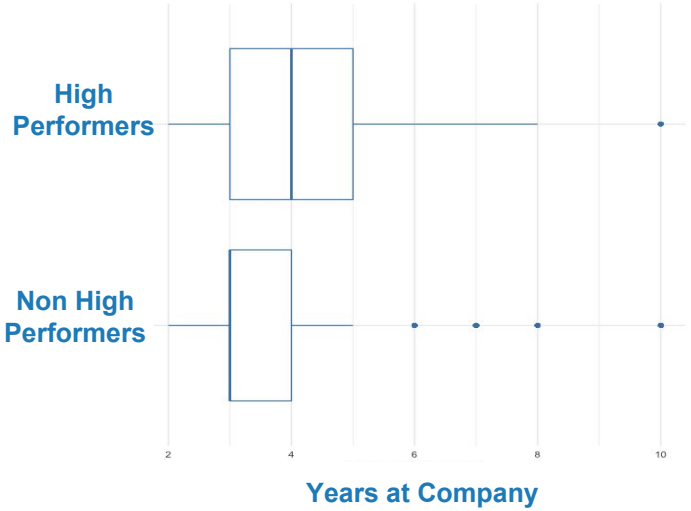




High performers typically work longer hours than their low performing peers

High performers are more polarised in their satisfaction ratings - they are more commonly very satisfied or very unsatisfied

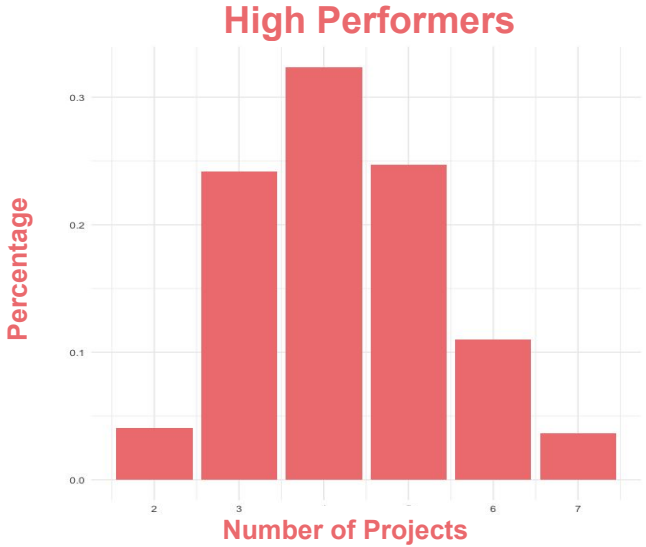
How do high performers differ?

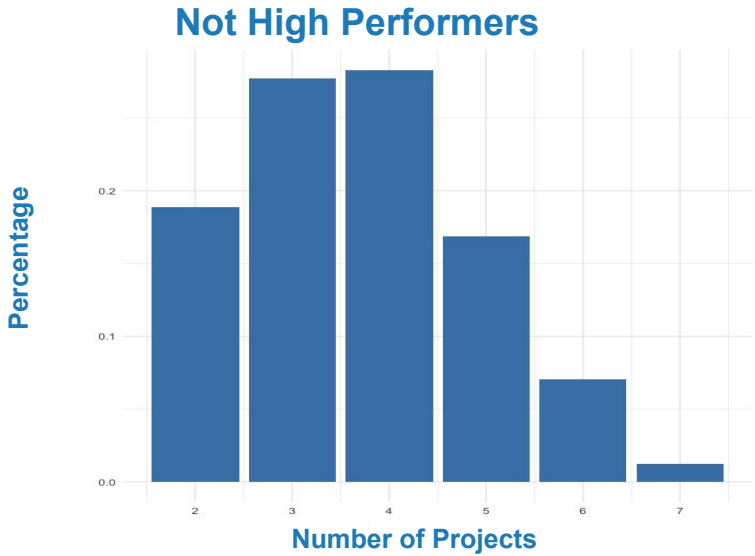


High performers have more often been working at the company longer...

...but they don't get paid more than low-performers

How do high performers differ?





High Performers Promoted in the Past 5 years:

1.9%

Non High Performers Promoted in the Past 5 years:

2.2%

High performers commonly take on more projects than their low performing peers… but are not promoted at a greater rate

Who is leaving the company?

A correlation analysis for high performing employees reveals several alarming relationships with left status:

- Satisfaction Level: -0.22****
- Number of Projects: 0.53****
- Average Monthly Hours: 0.52****
- Years at Company: 0.43****

High performing employees are more likely to leave the company if they have…

- Lower satisfaction levels
- Higher number of projects
- Higher average monthly hours
- More years spent at the company

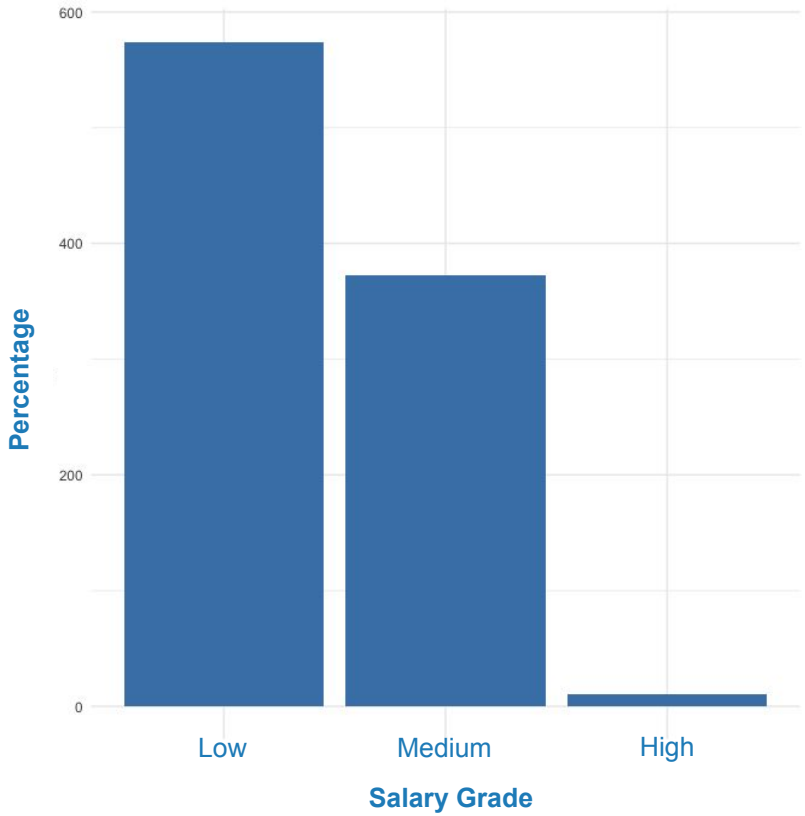
**** = Extremely significant

Who is leaving the company?

Two distinct clusters of high performing employees who have left the company emerge upon further analysis:

- Low satisfaction, highly worked employees
- High satisfaction, mid-to-highly worked employees

Who is leaving the company?



Of the high performing employees who have left the company, those in the low salary grade makeup 60%

Part 2:
Key Takeaways & Strategic Recommendations

Key Takeaways		Strategic Recommendations
1. High performers have often been working at the company longer, but do not command higher salaries than their lower performing peers	→	Adjust compensation structure or develop incentive programs to reward high performing and loyal employees
2. High performers commonly take on more projects and work longer hours than their low performing peers, but are not promoted at a greater rate	→	Evaluate and adjust promotion criteria to ensure that employees who take on a large responsibility are rewarded via career advancement
3. High performing, hard working employees are reporting extremely low satisfaction scores	→	Identify these employees and stage personal interventions to determine if career or work-life balance needs to be reevaluated

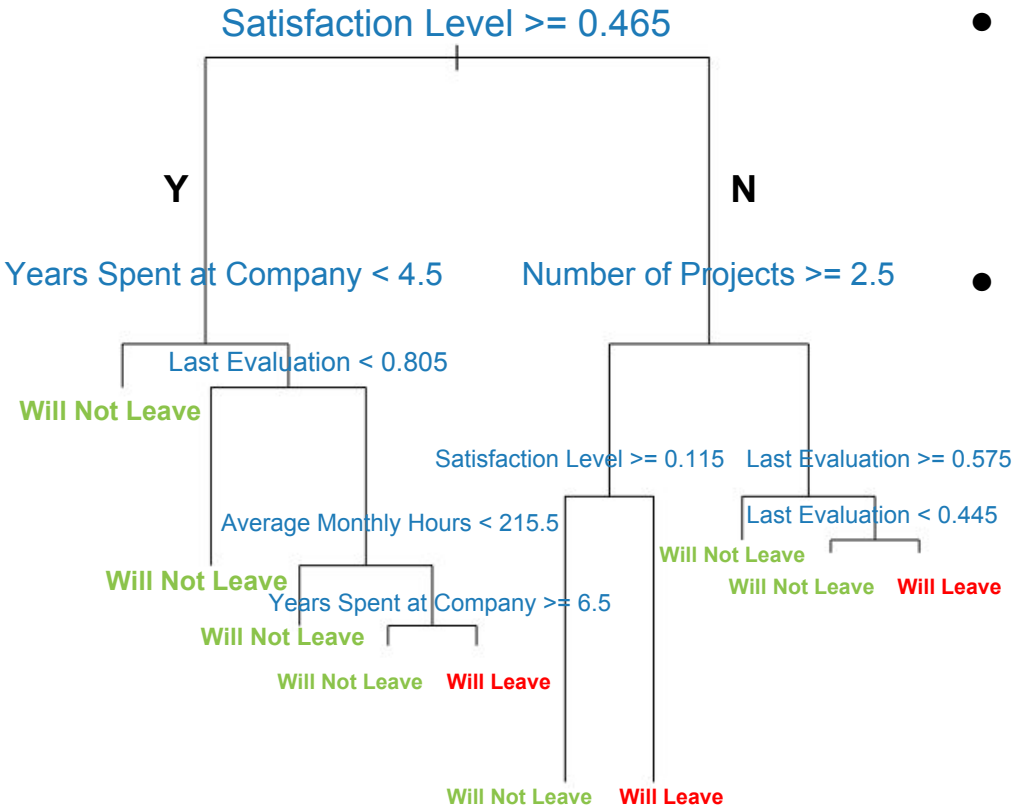
Part 3:
Predictive Modeling

Building a Predictive Model

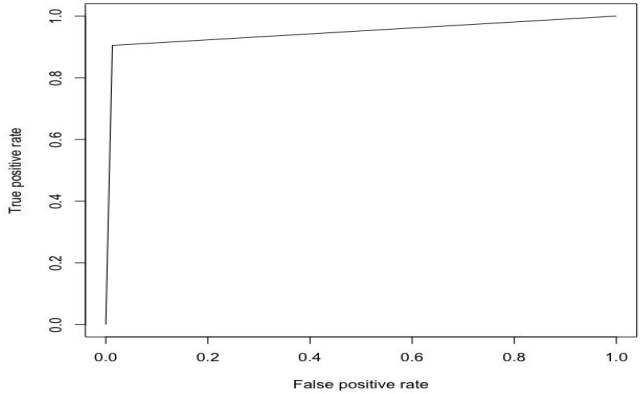
- Ultimately the client would like to predict which employees are most likely to leave the company. This capability is of particular strategic importance as it would allow the business to take preventative action for its most important employees based on an individual risk assessment.

- In order to achieve this classification task we will use a Classification Decision Tree due to its high accuracy, ability to model non-linear relationships and intuitive interpretation.
- After shuffling and splitting the data into 80% train, 20% test sets we can train the tree using Rpart.

Evaluating the Decision Tree



- The trained Decision Tree results in an intuitive, yes/no sequence of decision criteria, with the left branch representing the answer “Yes” and the right “No” to the previous question.
- When this model is tested against the training data set we obtain an impressive **96.7%** accuracy.



Type	Insights

The Decision Tree reveals three types of exiting employees:

1.	Satisfied, hard-working employees who have been at the company ~5 years.	→	It is possible these employees were “headhunted” by other companies, or were seeking a new challenge
2.	Very unsatisfied employees	→	Further, individual investigation is required to determine the source of unhappiness for these employees
3.	Unsatisfied employees with two projects and poor evaluations	→	These employees could be underworked and demoralised or just poor at their job