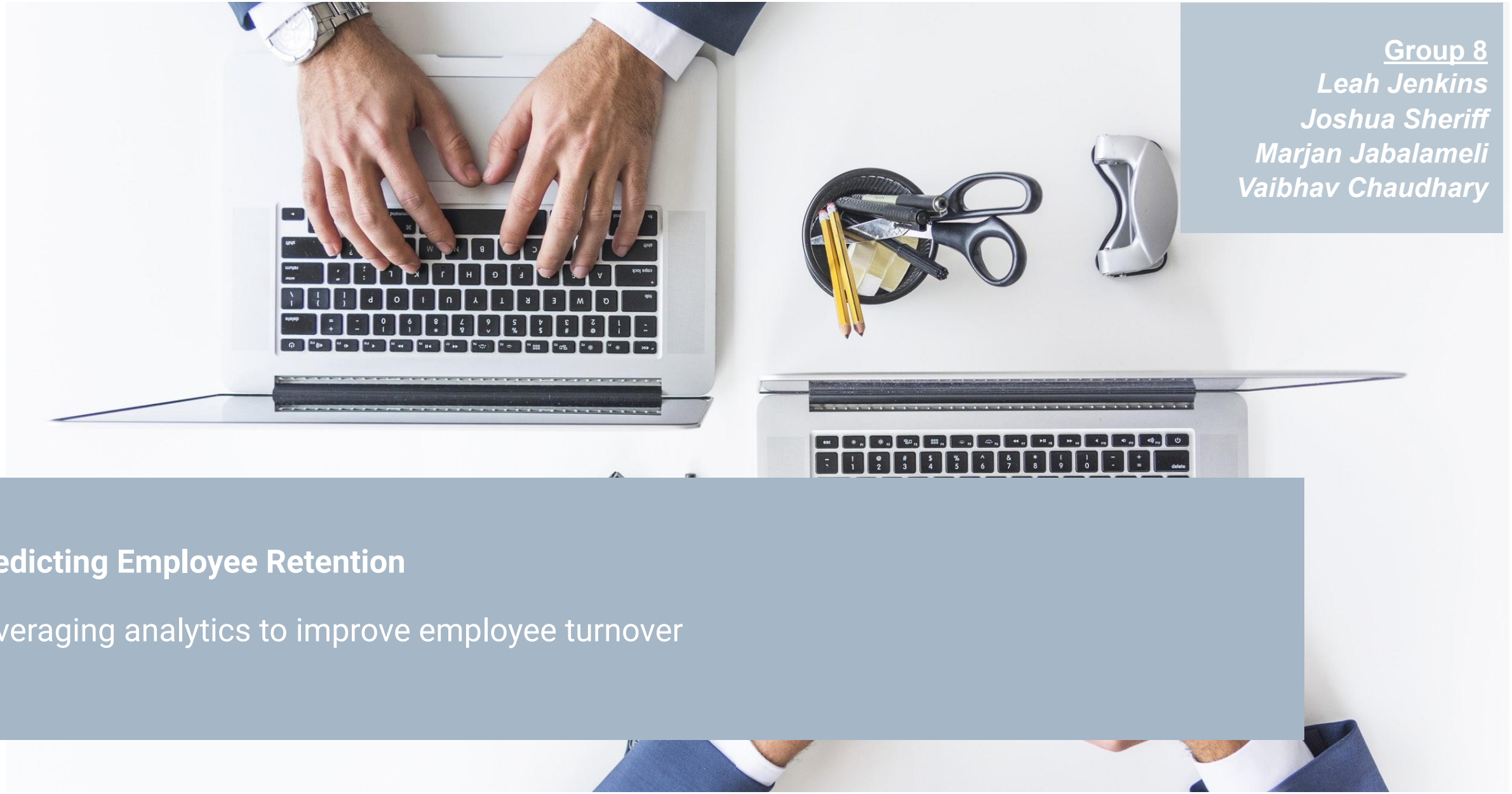


Group 8
Leah Jenkins
Joshua Sheriff
Marjan Jabalameli
Vaibhav Chaudhary

Predicting Employee Retention

Leveraging analytics to improve employee turnover

Agenda

- Business Problem & Objective
- Analytical Objectives
- Data Preparation
- Model Development
- Solution Deployment
- Performance Management
- Recommendations

17.8%

Industry average employee turnover rate per a 2016 Compensation Force Study

Source: Zenefits, ‘Does Your Company Have a Healthy Employee Retention Rate?’

Business Problem & Objective

Background

- Understanding what can drive employee turnover is of critical business importance
- Employee churn impacts:
  - Business' overall bottom line due to disruptions in project continuity
  - Can also lead to negativity amongst remaining staff
  - Result in lowered public perception and make it harder to attract and retain skilled talent in future

Business Problem & Objective

Stakeholder	Strategic Priorities	Key Questions	Success Criteria
CEO	Brand & Reputation Business Health	How do we reduce our HR Expenditure? How do we maintain a healthy culture?	Employee retention
President, Human Resources	Cost Management Profitability Operational Efficiency Risk Management	How to identify, monitor, and mitigate risks? How to achieve cost optimization? How to ensure fair staff compensation?	Clearly defined risk mitigation strategy Meet defined budget targets Policies and regulations followed
VP, Human Resources	Employee Acquisition Employee Retention Employee Experience Brand awareness Training Spend	How to acquire new Employees? How to reduce Employee attrition? How to reduce Training spend? How to optimize HR spend? How to improve Employee experience?	Acquisition rate Retention rate Training Expenditure Consultancy Commission rate.

Business Problem & Objective

Business Objective

- Objective of our research project is to investigate if can we predict which drivers will lead an employee to decide to leave the organization across the different departments

- Using a large sample of employee records (> 10,000) we set out to devise an algorithm that can predict whether a current employee is likely to leave or stay

Analytical Objectives

Data Model Goal

- Goal of data model is to determine how few variables will most accurately predict whether an employee will leave
- Devised 2 different algorithms to test this hypothesis and compare results

Model Success Criteria

- Success for our model is to get to a **prediction accuracy of 80% or higher** based on our algorithm
- Also sought for a precision of at least 65%

Analytical Objectives

Phase	Timeframe	Resources	Risks
Business Understanding	October 10 - October 14	Business experts Data scientists	Lack of support from senior leadership at company, delay in receiving data
Data Understanding	October 20th - October 31st	Data scientists Business experts	Technology issues, not enough data records to use for modeling
Modeling	October 29 - October 31st	Data scientists	Data problems, issues with modeling software
Evaluation	October 31 - November 2	Data scientists Business experts	Issues with software, issues with data model
Deployment	November 3	Data scientists Tech department	Inability to implement model,

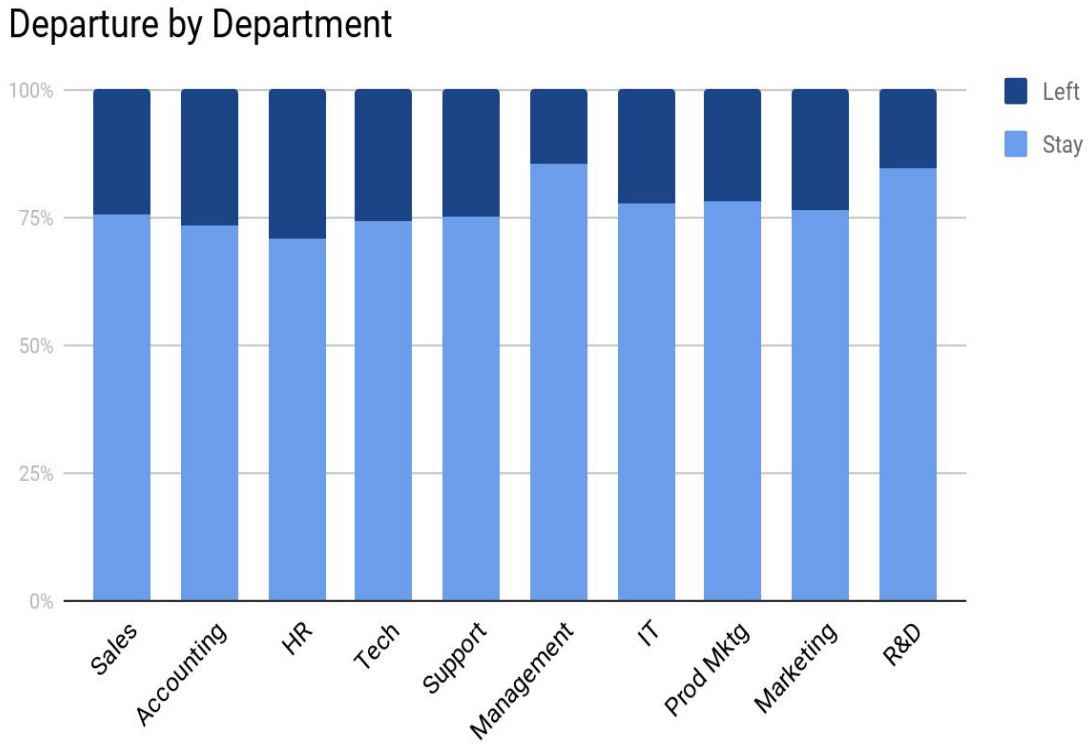
Data Preparation

- Acquired anonymized CSV file from HR with detailed records of more than 10,000 employees (10 features)
- Checked data for any issues/discrepancies to ensure uniformity in data types of variables that could create problems for processing
- Checked for any missing or incorrect values to ensure data was clean to proceed with

Assessing the Data

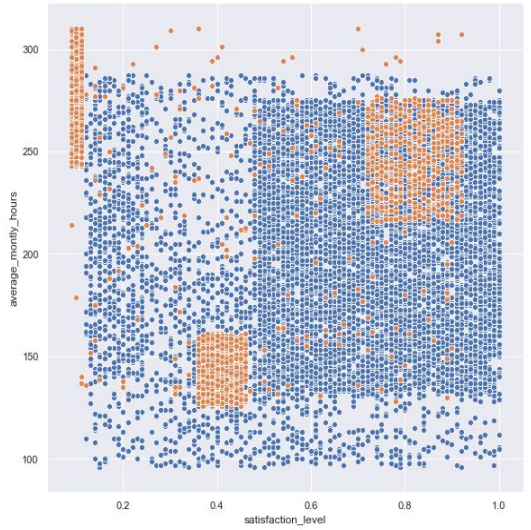
- Explored overall churn (24%) and rates by individual department and discovered **Tech (34%)**, **HR (29%)** and **Accounting (27%)** were highest across departments

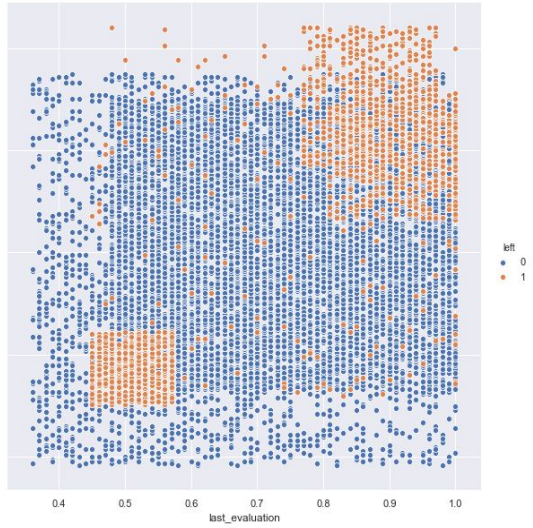
- Next we explored all of the 10 features to see if we could begin to see correlations between pairs of variables using Seaborn



Assessing the Data

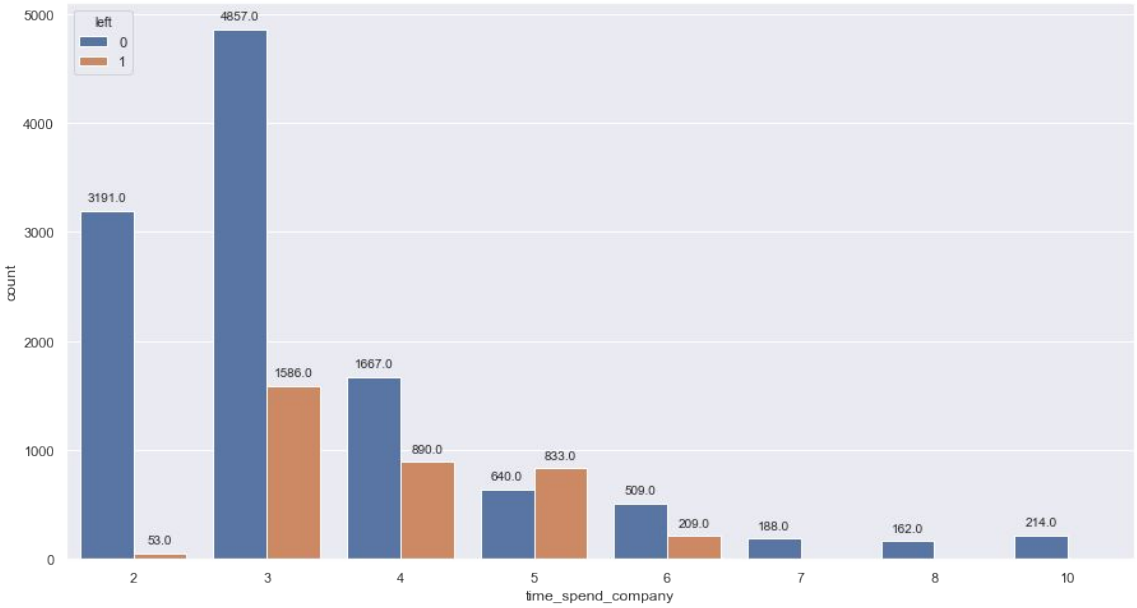
- Uncovered interesting clustering between average hours worked in the month and departure
- Saw similar cluster patterns when comparing job satisfaction level with departure - those who were rating satisfaction at very high level (over 80%) were often the ones who sought to leave





Assessing the Data

- Very high exit rate amongst employees who had spent 5 years in the organization - more employees at this stage left than even considered remaining with the company
- Thought this may be linked to not having received a promotion within 5 years
- The longer the tenure, the less likely an employee is to want to leave

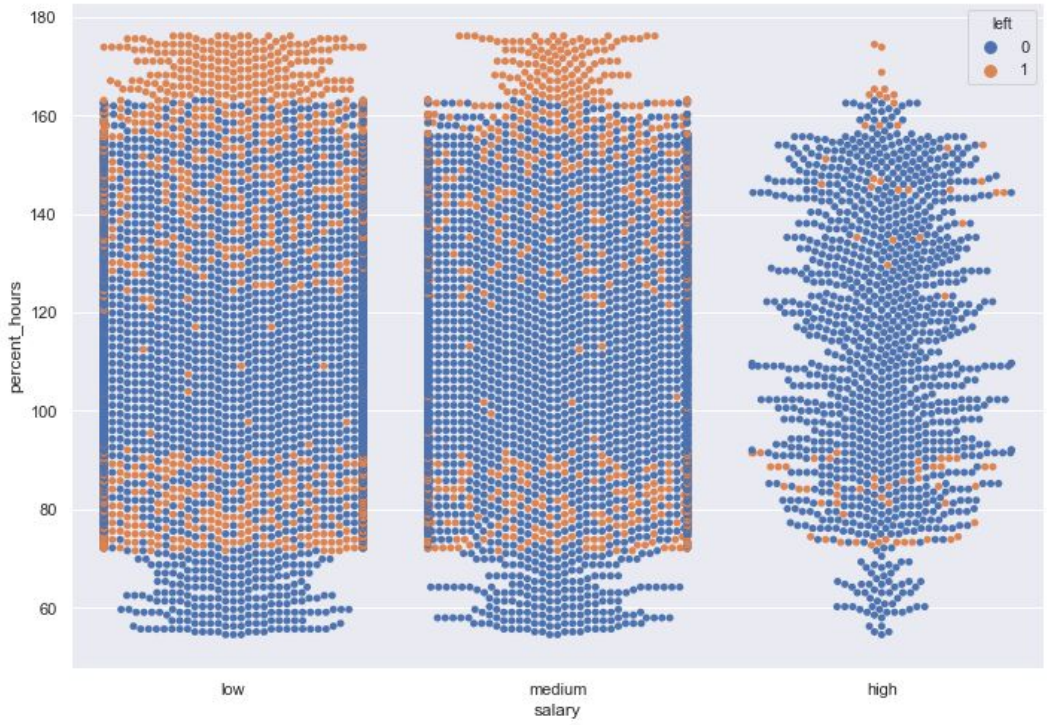


Assessing the Data

- Among remaining employees, we saw that they consistently were scoring above 0.5 on their evaluations, but departing employees appeared to be clustered around scoring an average evaluation (approx. 0.5) or were on the higher end (0.85 and up)
- Seems to suggest that those with higher evaluations felt they had outperformed and wanted to seek out opportunities elsewhere

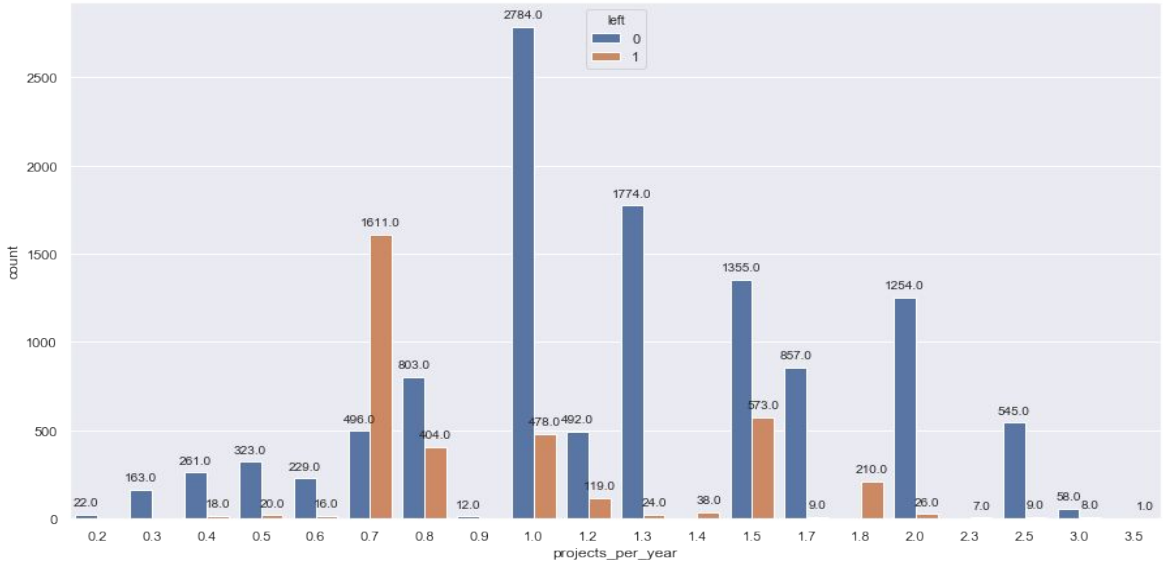
Assessing the Data

- Created a percent hours worked metric and used this to help visualize this across salary ranges
- Appears to be higher likelihood of departure amongst those employees who are working above normal hours
- Greater likelihood to leave amongst those working the least amount of work hours, especially those with low and middle salaries



Augmenting the Data

- Created our own derived variable to calculate a ratio based on the number of work projects completed based on the length of tenure in the organization
- This ratio shows that there were a large number of employees that left where they worked on average 0.7 projects per year



Integrate the Data

- For our research, we did not undertake any merging of data across tables and all aggregation of data was done solely in the data exploration phases of existing features
- We did not need to undergo any aggregations to our dataset for the purpose of our model

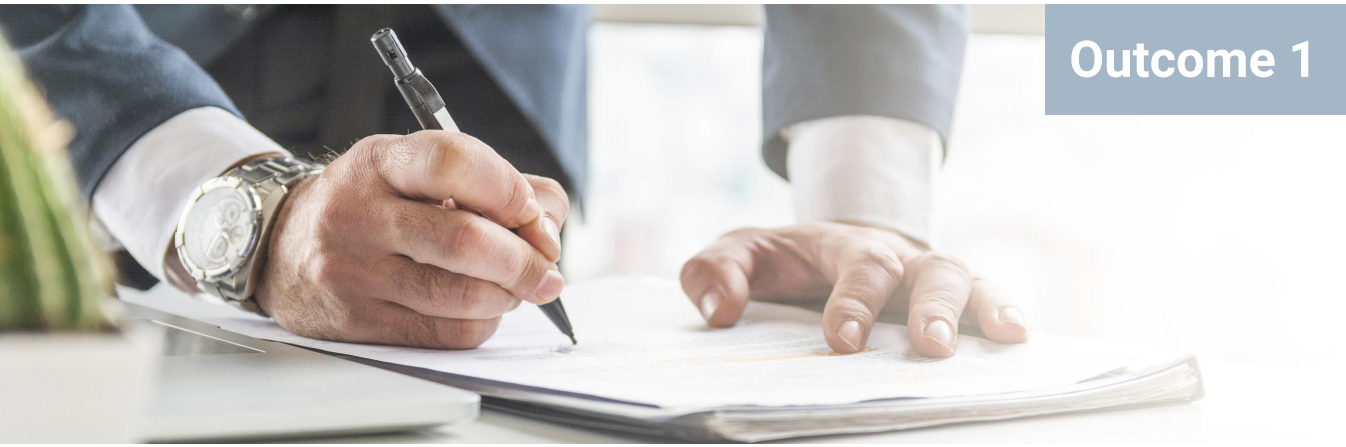
Model Development

- Utilized logistic regression algorithm, form of supervised learning
- 1 version utilized all features, other version used 4 key inputs:
  - Department
  - Satisfaction level
  - Average monthly hours
  - Last evaluation
- Both algorithms used Left (0,1) as output variable

Model Development

- We initially had to get dummies for the salary and department variables in order to be able to use them in our algorithms
- Next, we split out the dataset into our X features (independent variables) and our Y output variable (left or not). These were used to help train our algorithm
- After running the logistic regression, we attained a max prediction accuracy of 84%
- We then decided to explore using an SVM/SVC algorithm for our modelling

Comparison of SVC Versions



Outcome 1

Using SVC and all our input variables, we were able to attain an accuracy of 95% and a precision score of 88% (against those who left)

Outcome 2

Using SVC and only 4 select input variables, we were able to attain an accuracy of 86% and a precision score of 78% (against those who left)

Solutions Deployment

- Deployment plans will be decided once the executive team has had opportunity to review our findings & recommendations and secure internal buy-in

Performance Management

This phase will be decided based upon the solutions deployment

Performance Management

Recommendationsh

- Based on our use of both logistic regression and SVM/SVC models, we recommend in this case to use the SVC model as it had a higher accuracy and precision
- After seeing the difference in prediction accuracy, we have learned that there are other variables that have an impact towards predicting if someone will leave beyond just department, salary, last evaluation and average monthly hours.
- We recommend doing further iterations to further refine what the right combination of variables that would maintain an accuracy of 95% before implementing into the organization’s processes.

Thank You!

Any questions?