

# Attrition at XYZ

## (Prediction and Contributing Factors)

Group Name:

1. Kanupriya Kushwaha
2. Omkar Murthy Appala
3. Santhosh Dhanabal
4. Thogaruchiti Khasim Basha



# Abstract – Predicting Attrition and its contributing Factors

**XYZ** employs, at any given point of time, around 4000 employees. However, every year, around 15% of its employees leave the company and need to be replaced with the talent pool available in the job market.

## Objectives:

- Identify what factors must be focused on, in order to curb attrition.
- Highlight which of these variables is most important and needs to be addressed right away.
- Model the **probability of attrition** using a logistic regression.

Business  
Understanding

- Understanding the problem at hand (Identify factors to curb attrition at XYZ)

Data  
Understanding

- Understanding the data at hand (Employee data from the year 2015 - including general information, employee and management surveys; and employee in and out timings. Also a variable stating whether they left the firm.)

Data  
Preparation

- Clean and prepare the data so it can be effectively utilized for EDA and modeling

Exploratory  
Data Analysis

- Perform EDA to understand the distribution and relationship of variables

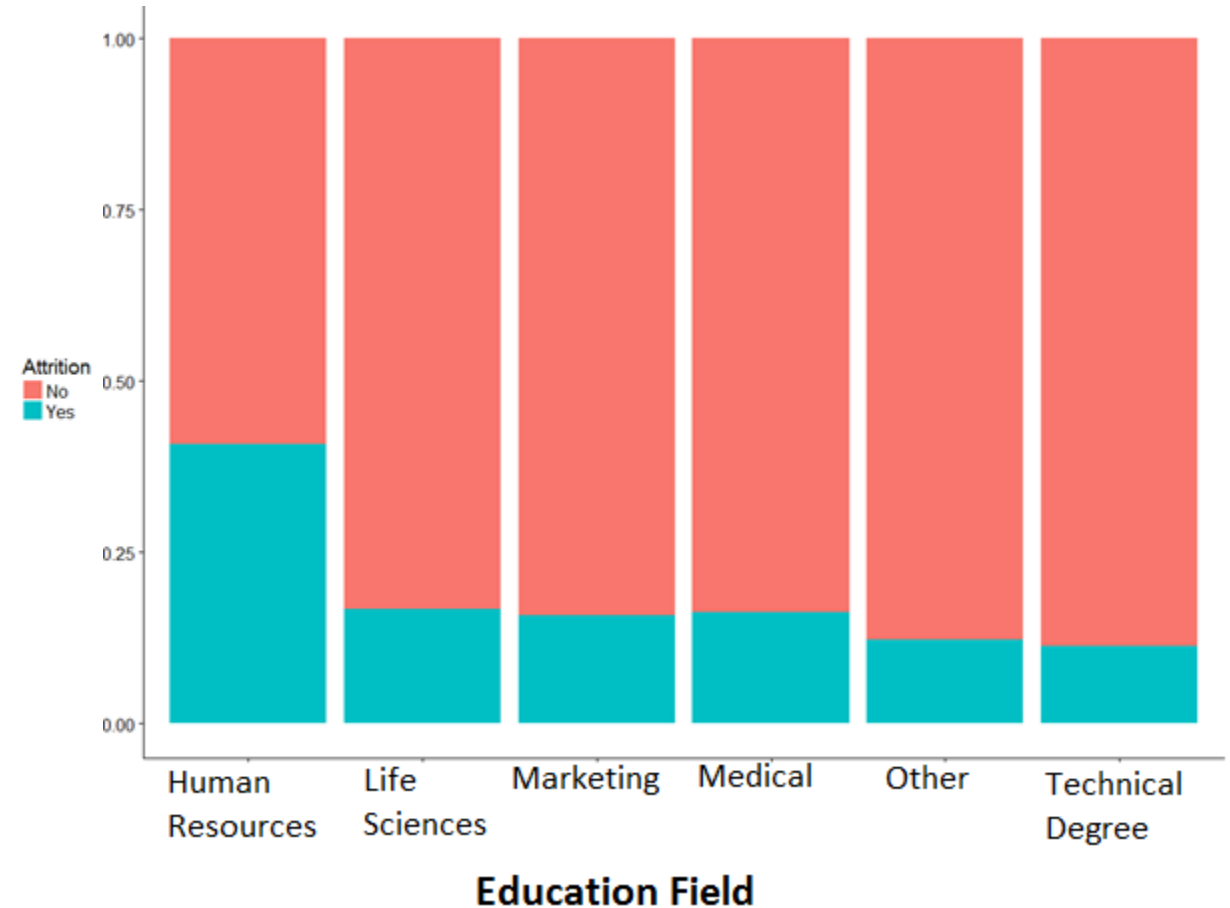
Model  
Building

- Build a logistic regression model for predicting the probability of attrition.

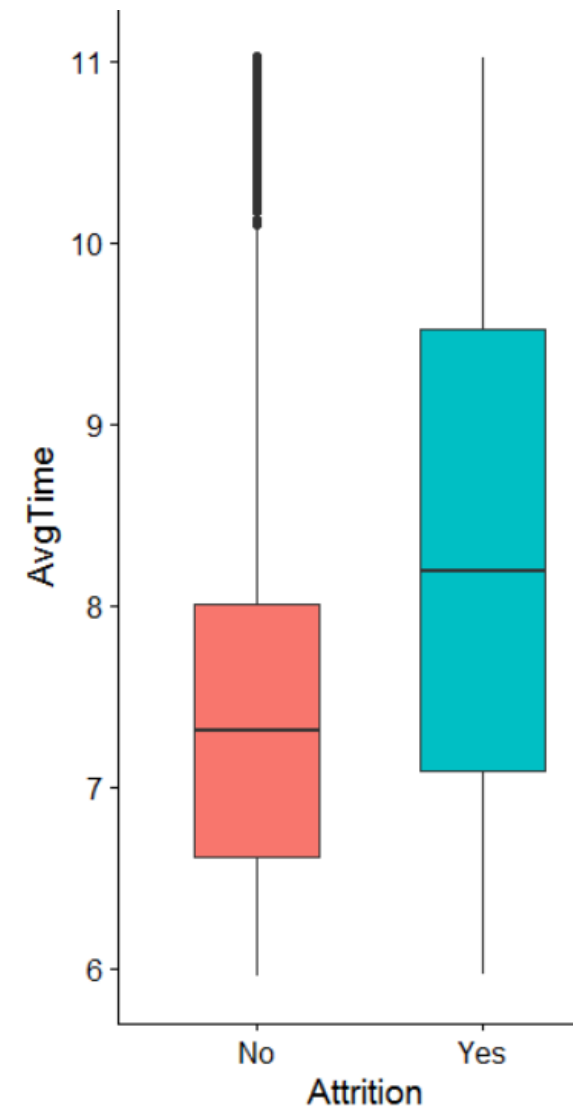
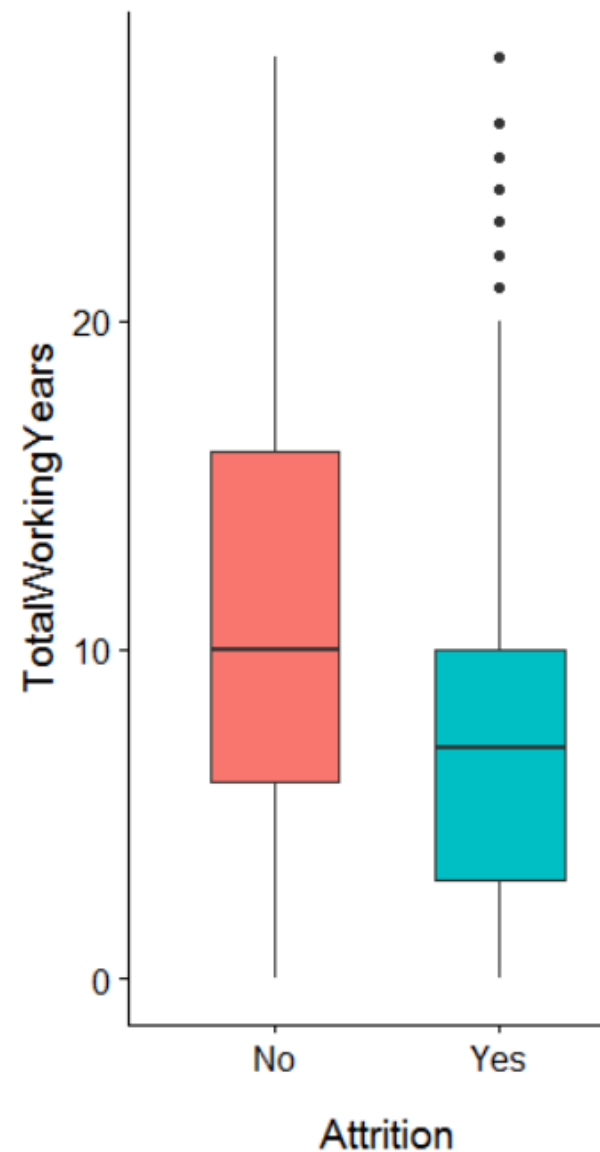
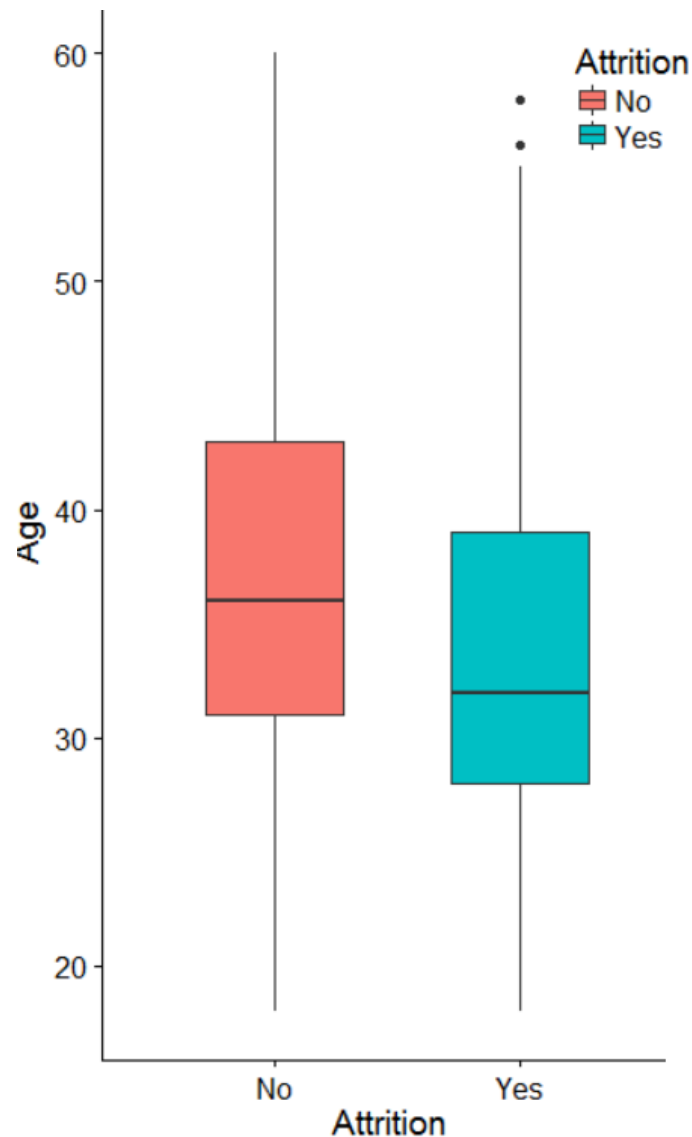
Model  
Evaluation

- Evaluate the final model using the correct principles and appropriate evaluation metrics.

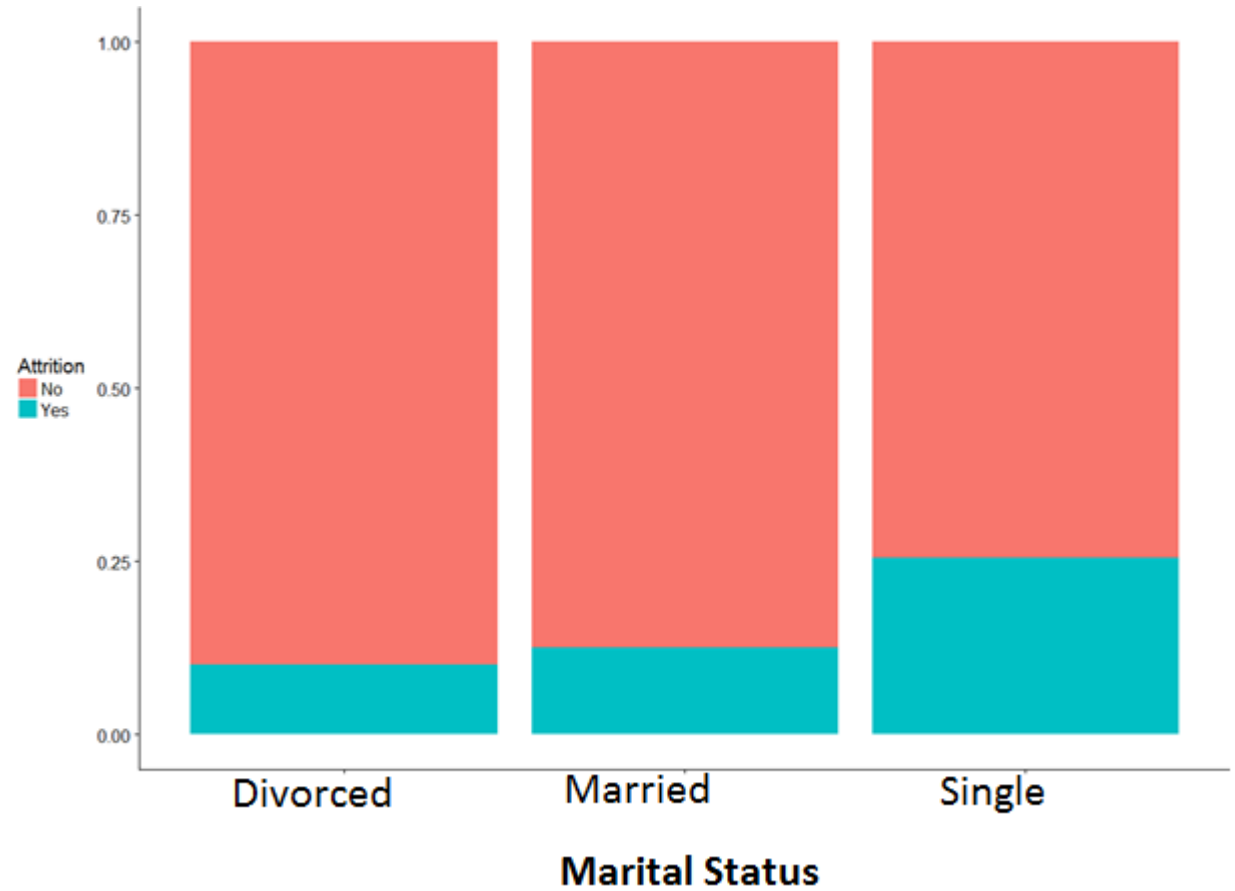
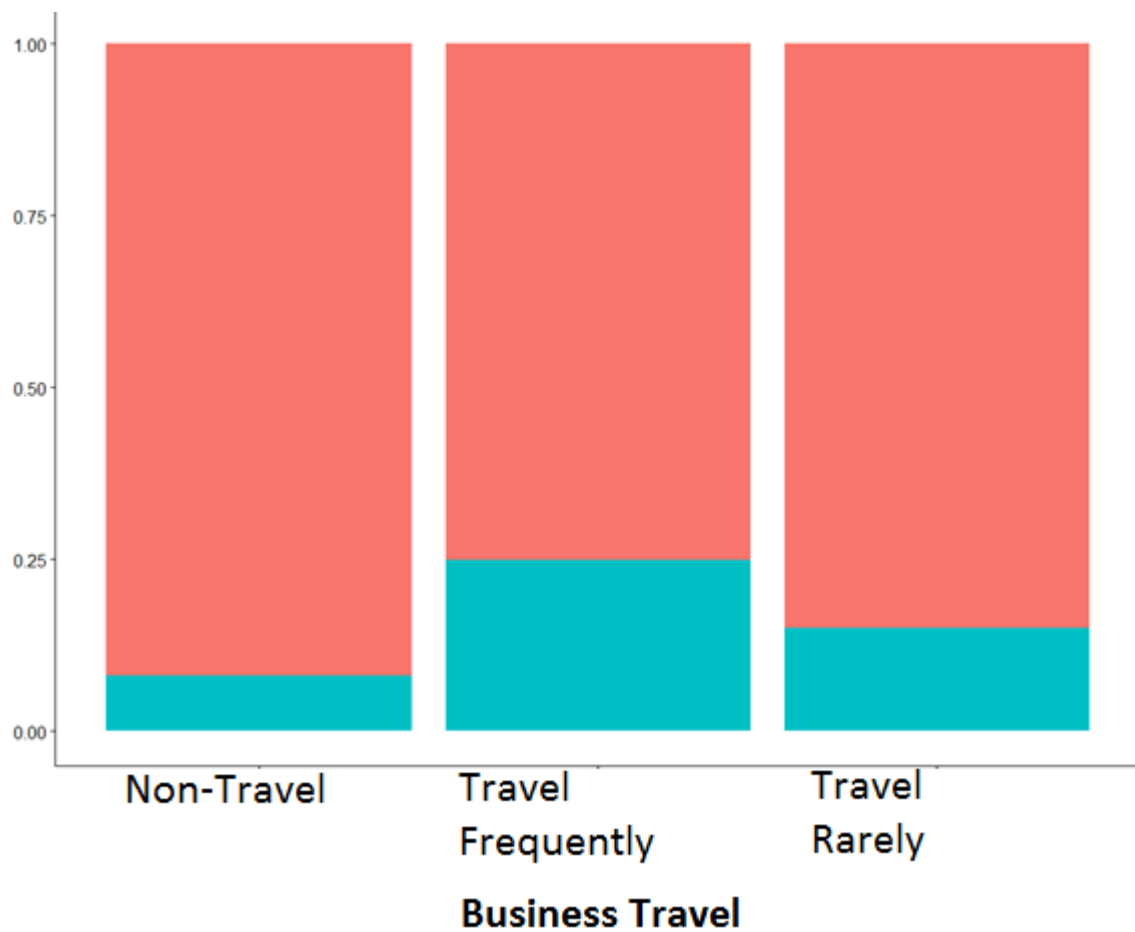
- Data loading, understanding and merging
- Data duplication identification and removal
- Removal of unnecessary columns that don't contribute to the analysis or modeling
- Missing value treatment
- Outlier treatment
- Introduction of new metrics
- Numeric data scaling
- Conversion of categorical data to numeric



- Employees with HR education or working in HR department are much likelier to leave.



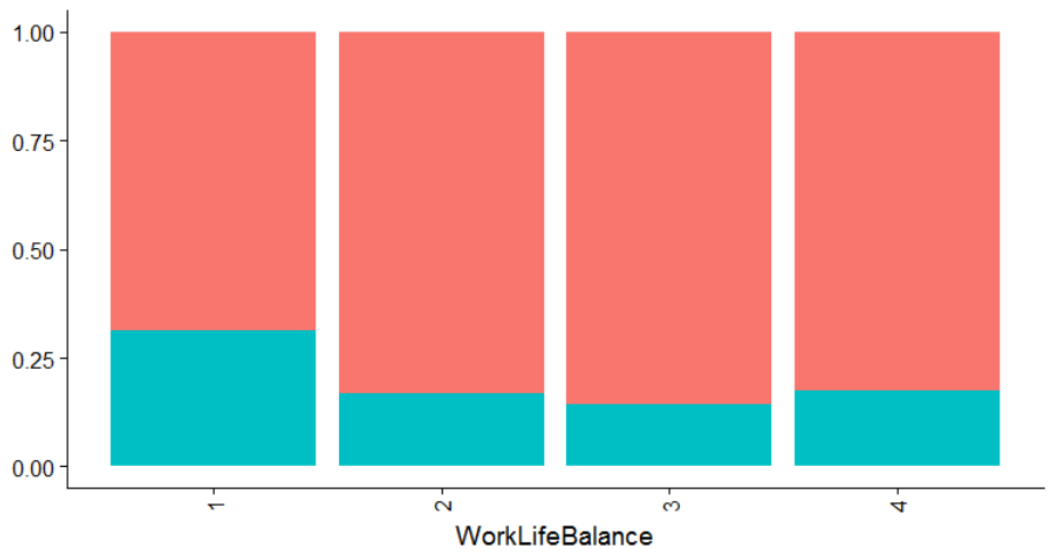
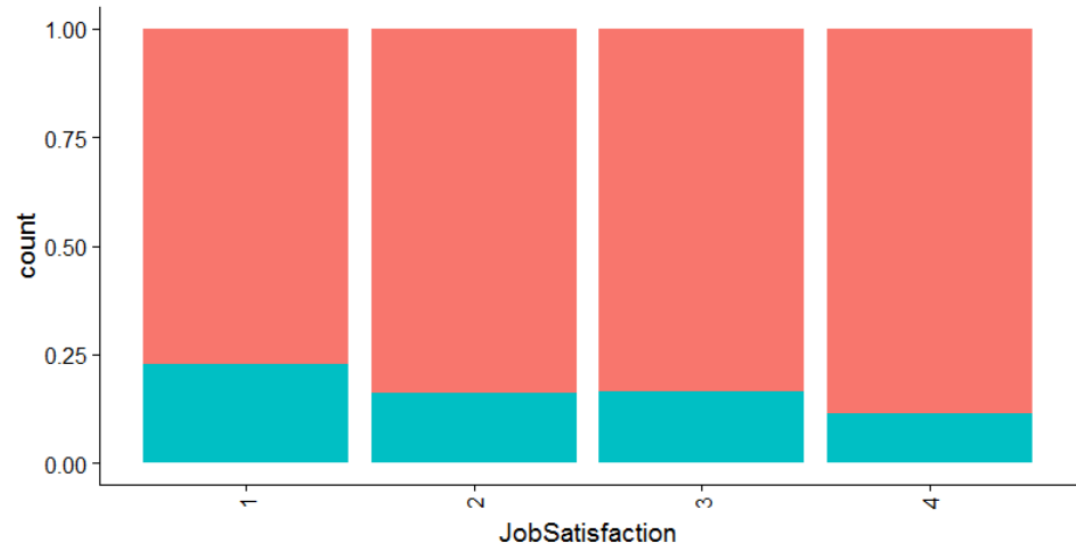
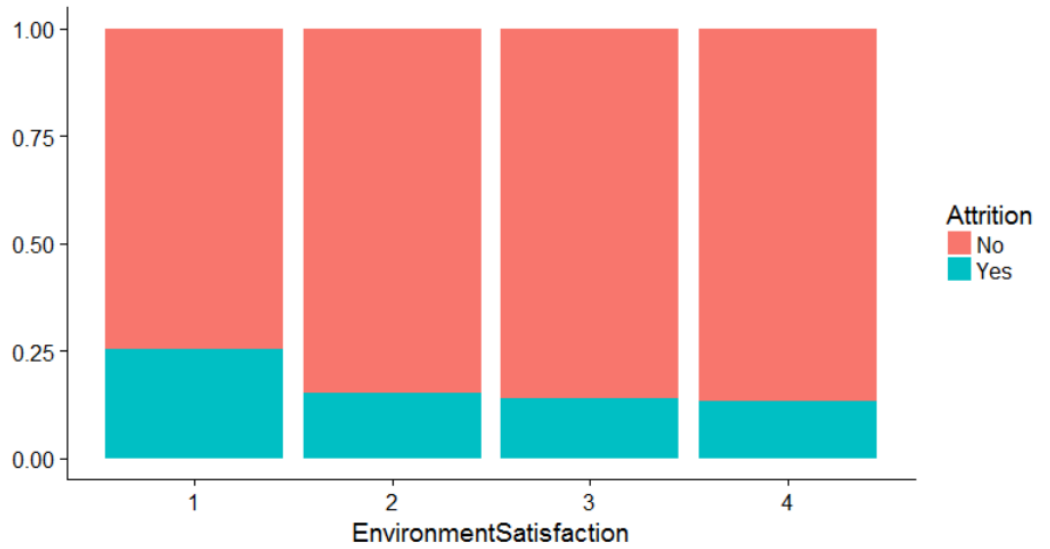
- Younger, less experienced employees and people who spend more time in office are likelier to leave.



- Employees who travel frequently are likelier to leave

- Single employees are likelier to leave

# Exploratory Data Analysis (Contd.)

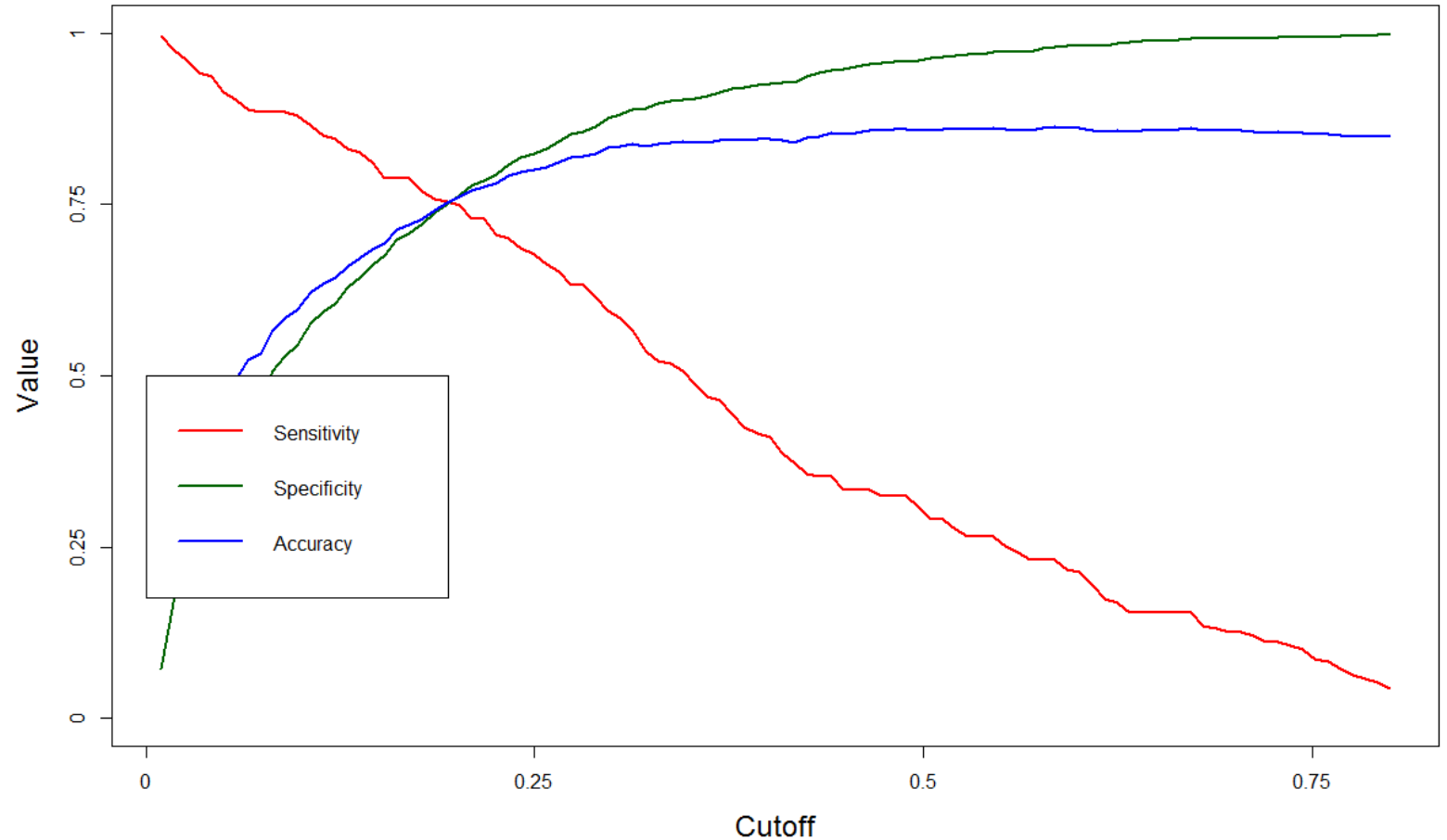


- Employees unsatisfied with their job, work environment or work-life balance are much likelier to leave.

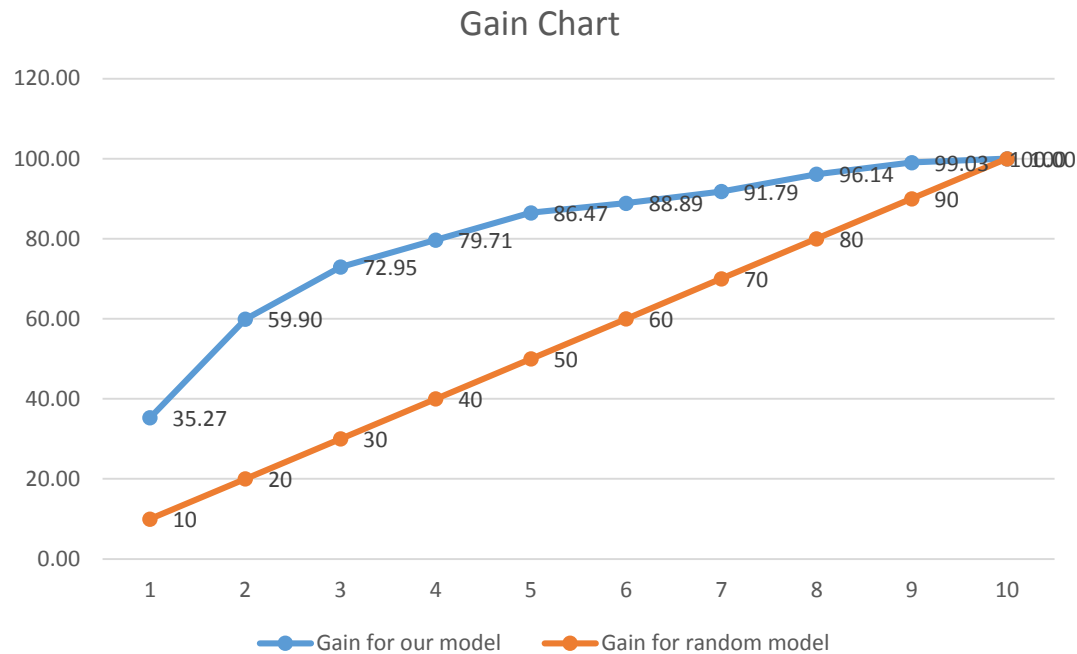


- A Logistic Regression Modeling was built to get the probability of an employee leaving the firm
- Various models were developed and iteratively fine-tuned; keeping in mind significance and multi-collinearity of variables.
- After a number of iteration, a model was identified which scored well on various model evaluation metrics.
- The following factors were identified as having impact on attrition:
  - Employees working in HR are likelier to leave
  - Employees who gave low ratings in surveys w.r.t work-life balance, job satisfaction or work environment satisfaction are likelier to leave
  - Employees who travel are likelier to leave
  - Young, single and less experienced employees are likelier to leave
  - People who spend more time in the office are likelier to leave
  - People who have not been promoted in many years are likelier to leave
  - If employees have been with the same manager for many years, they are less likely to leave
  - People who have already worked for a number of companies are likelier to leave
  - Employees who attended fewer trainings are likelier to leave
  - Employees in the role of Manufacturing Director are less likely to leave

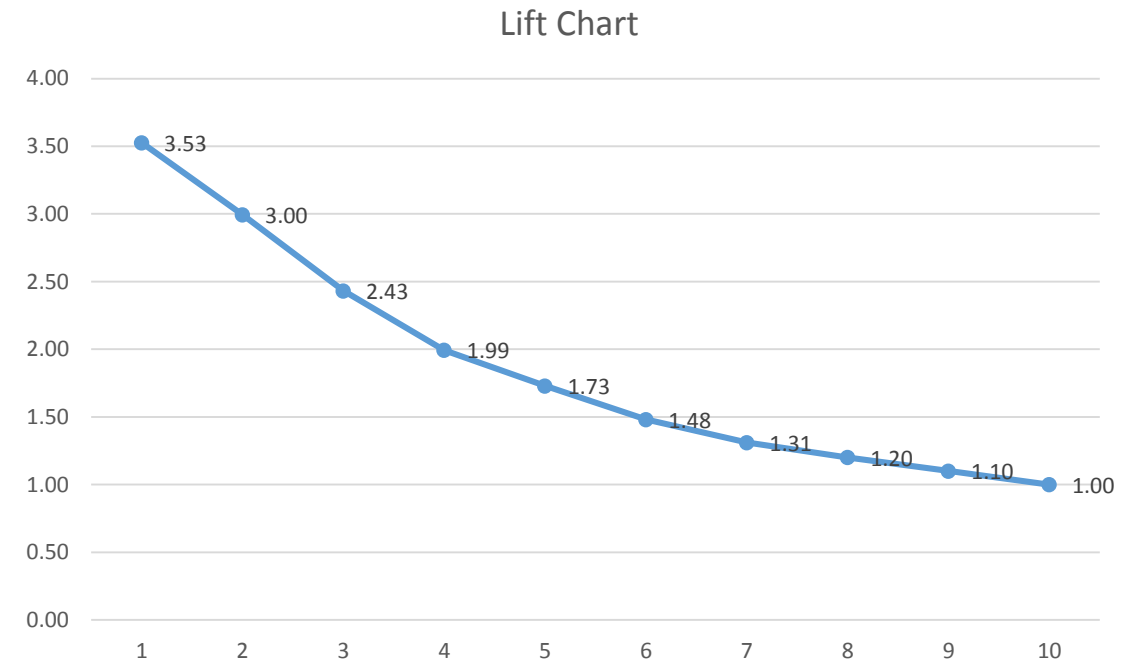
- The model was evaluated using various metrics.
- After creating an ROC chart, an optimal value of 19.35% was chosen to provide:
- Accuracy: 75.1% (The model can accurately predict 75% of the results)
- Sensitivity: 75.4% (The model can accurately predict 75.4% of employees who are going to leave)
- Specificity: 75.1% (The model can accurately predict 75% of employees who are going to stay with the organization)



## • Gain Chart:



## • Lift Chart



- The Gain and Lift charts show that the model is performing quite well when compared to a random model.
- KS-Statistic = **50.4%** at third decile (how well the model discriminates between Employees who are going to churn and those who are not)

- A logistic regression was performed on past employee data:
  - to calculate probability of an employee leaving the firm; and
  - to identify factors that have an impact on attrition.
- A model was built which predicts attrition at the firm with 75% Accuracy, 75.4% Sensitivity and 75% Specificity.
- If the firm targets top 40% of the employees identified by the model, it can accurately identify almost 80% of employees who are going to leave.
- Recommendations based on factors identified by the model are listed on the next slide.

## Recommendations:

- Focus on :
  - Employees in Human Resource Department
  - Employees who are giving low ratings in Employee Satisfaction Surveys
  - Employees who are traveling more often
  - Young, single and less experienced employees
  - Employees who have not been promoted in some time
  - Employees who have held numerous jobs before
- Aim to provide a better work-life balance and a more conducive work environment
- Discourage employees from staying back for extra hours
- Organize more trainings and incentivize employees to attend more trainings
- Encourage employees to stay with the same team/manager