HR Analytics Case Study
For Predicting Employee Attrition

Submitted By:
Bhagyashree Barhate
Deepankar Kotnala
Gautami Ramesh Havele
Rohit Saini

Problem Statement and Business Objective

Problem Statement

Every year, around 15% of the employees of XYZ company leave the organization and hence are required to be replaced with the talent pool available in the job market. The management believes that this level of attrition(employees leaving, either on their own or because they got fired) is bad for the company, because of the following reasons -

- The former employees' projects get delayed, which makes it difficult to meet timelines, resulting in a reputation loss among consumers and partners
- A sizeable department has to be maintained, for the purposes of recruiting new talent
- More often than not, the new employees have to be trained for the job and/or given time to adapt themselves to the company

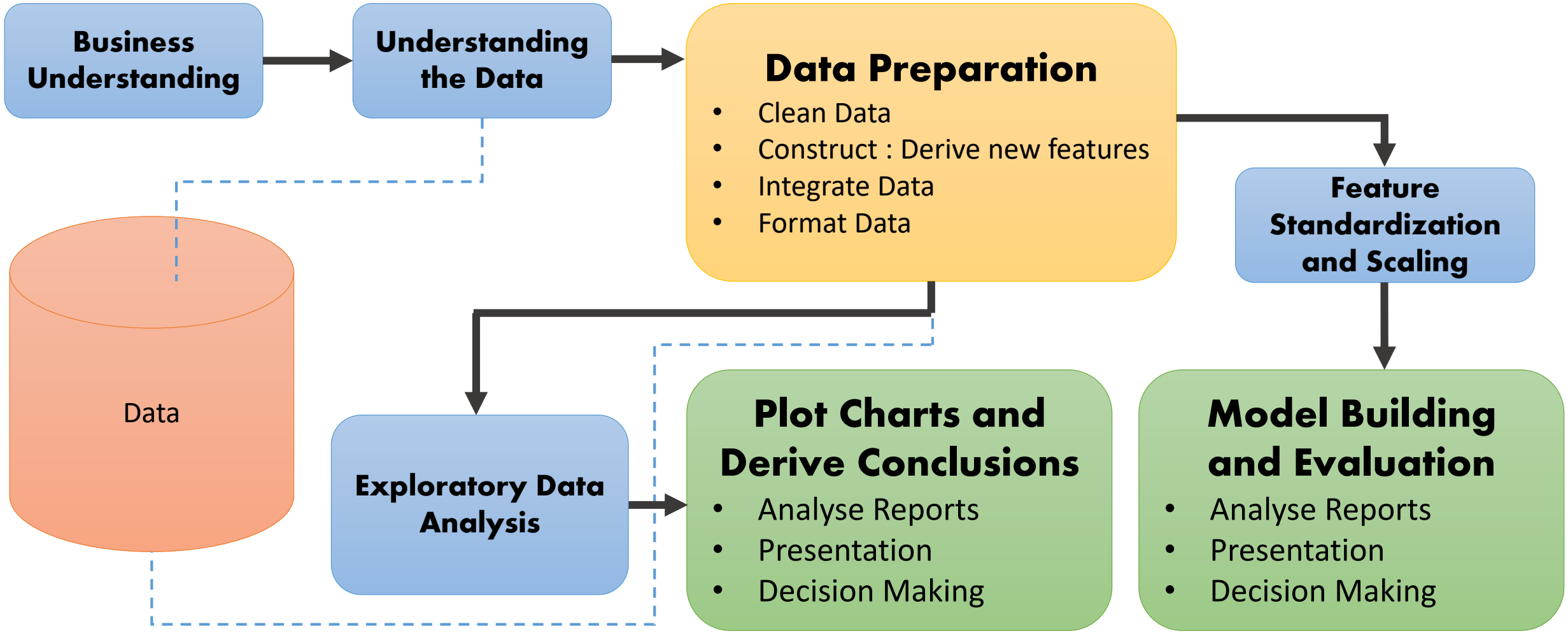
Business Objective

Build a model using Logistic Regression to find the probability of attrition.

Also, we need to understand what factors should be focused on by the management in order to curb attrition.

In other words, we need to find those variables which affect employee attrition and should be focused on right away by the management.

Problem Solving Methodology



Data Understanding

General Data

- DistanceFromHome
- Education
- EducationField
- EmployeeCount
- EmployeeID
- Gender
- JobLevel
- JobRole
- MaritalStatus
- MonthlyIncome
- NumCompaniesWorked
- PercentSalaryHike
- Age
- Attrition
- Over18
- BusinessTravel
- Department
- StandardHours
- StockOptionLevel
- TotalWorkingYears
- TrainingTimesLastYear
- YearsAtCompany
- YearsSinceLastPromotion
- YearsWithCurrManager

Employee Data

- EmployeeID
- JobSatisfaction
- WorkLifeBalance
- EnvironmentSatisfaction

Manager Survey Data

- Employee Id
- Job Involvement
- Performance Rating

InTime Data

- EmployeeID and InTime details of each employee

InTime Data

- EmployeeID and OutTime details of each employee

Data Understanding

- Converted In_time and out_time data from wide to long format in order to get the meaningful insights.

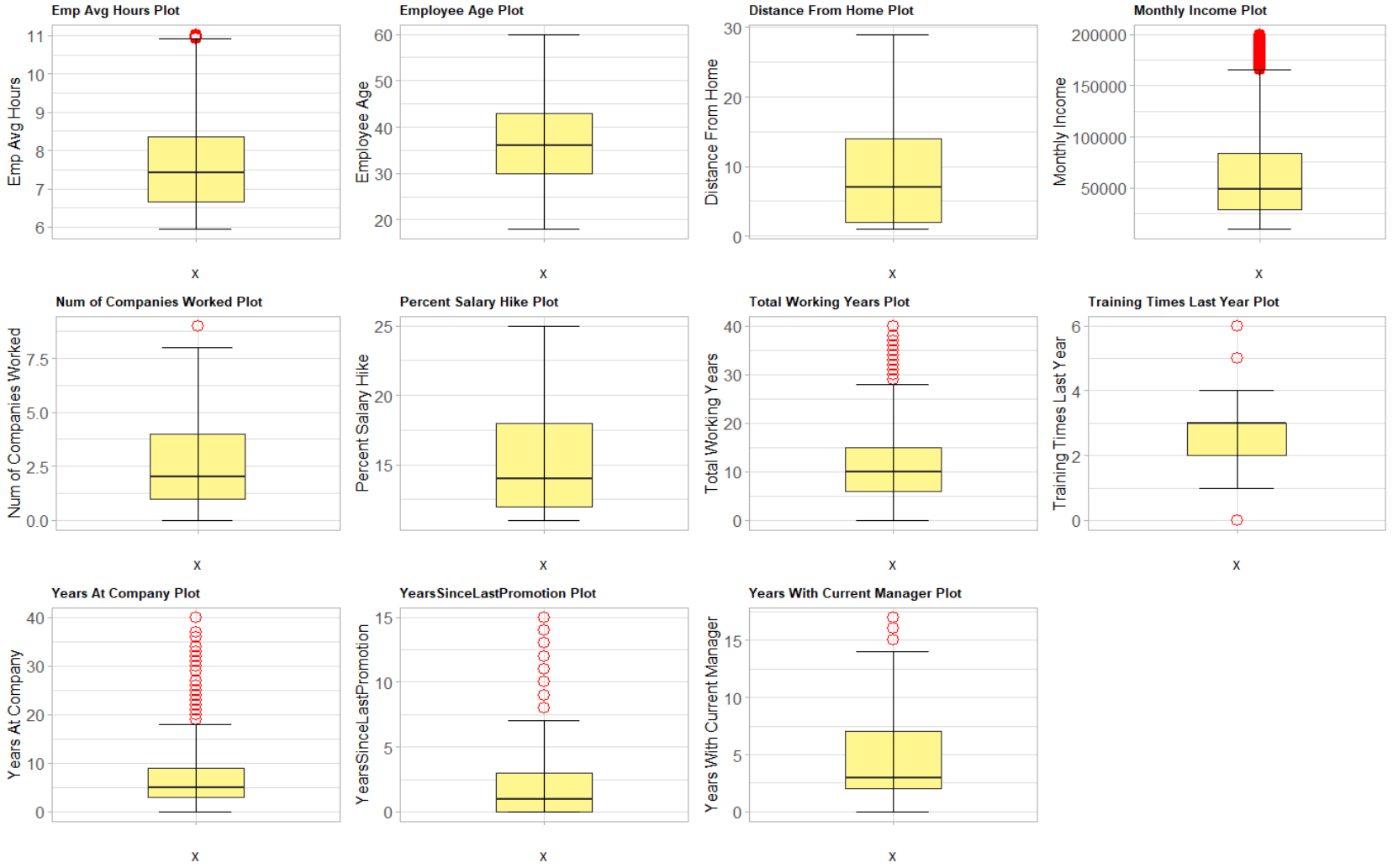
- Derived Variables :
  - No of leaves taken by employees
  - Average working hour per employee

- NA's and Outlier treatment:
  - Replaced the NAs in EnvironmentSatisfaction, JobSatisfaction, WorkLifeBalance with the median.

- Single constant value columns removed:
  - The columns employee count, Over18 and standard hours contains a constant value. Hence these were removed.
- Categorical Variables :
  - All categorical variables converted to factors and then created dummy variables for the model building

- Continuous Variables:
  - All continuous variables Scaled.

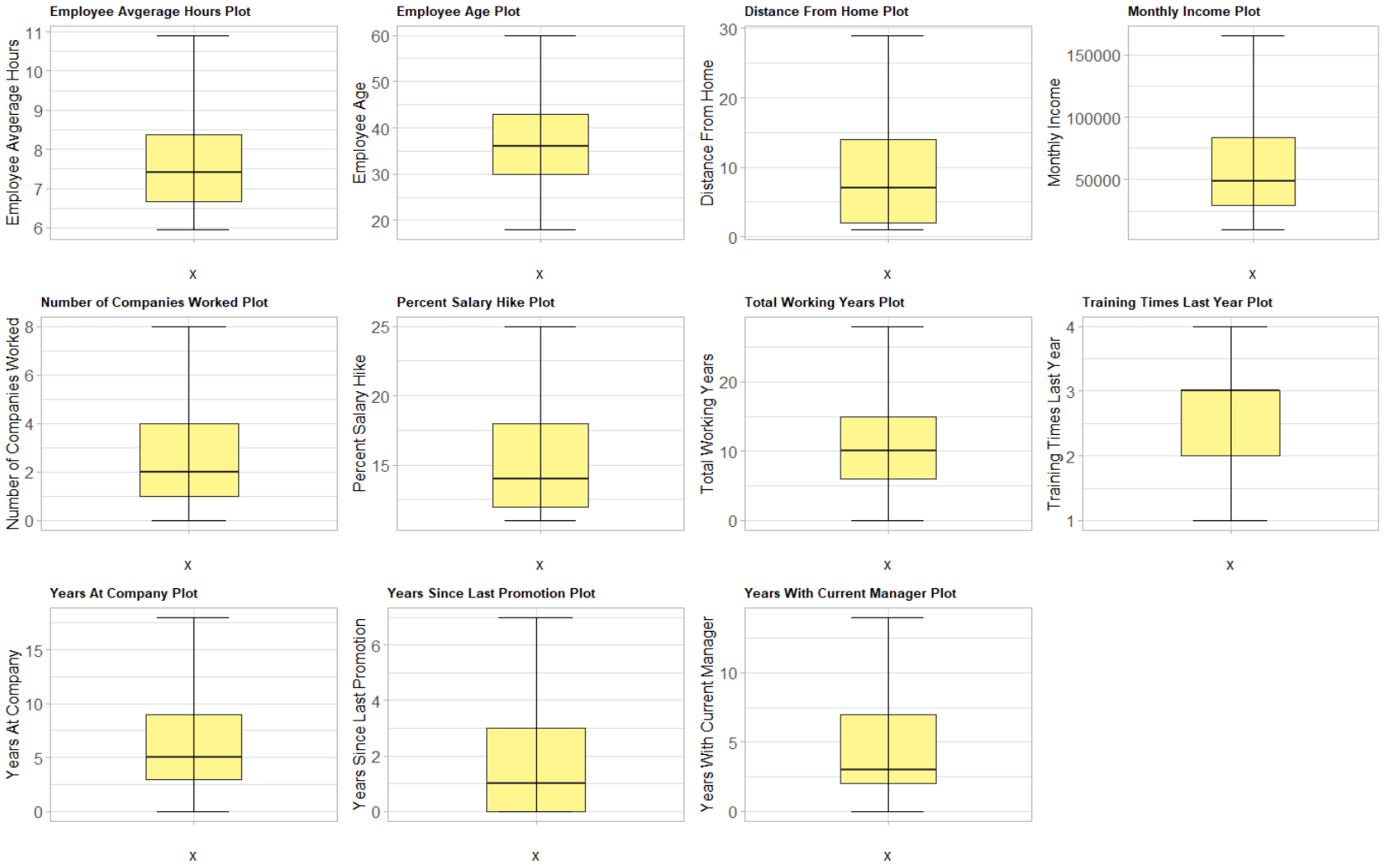
Analysis: Outliers



There were outliers present in:

- Employee Average Hours
- Monthly Income
- NumberOfCompaniesWorked
- TotalWorkingYears
- TraningTimesLastYear
- YearsAtCompany
- YearsSinceLastPromotion
- YearsWithCurrentManager

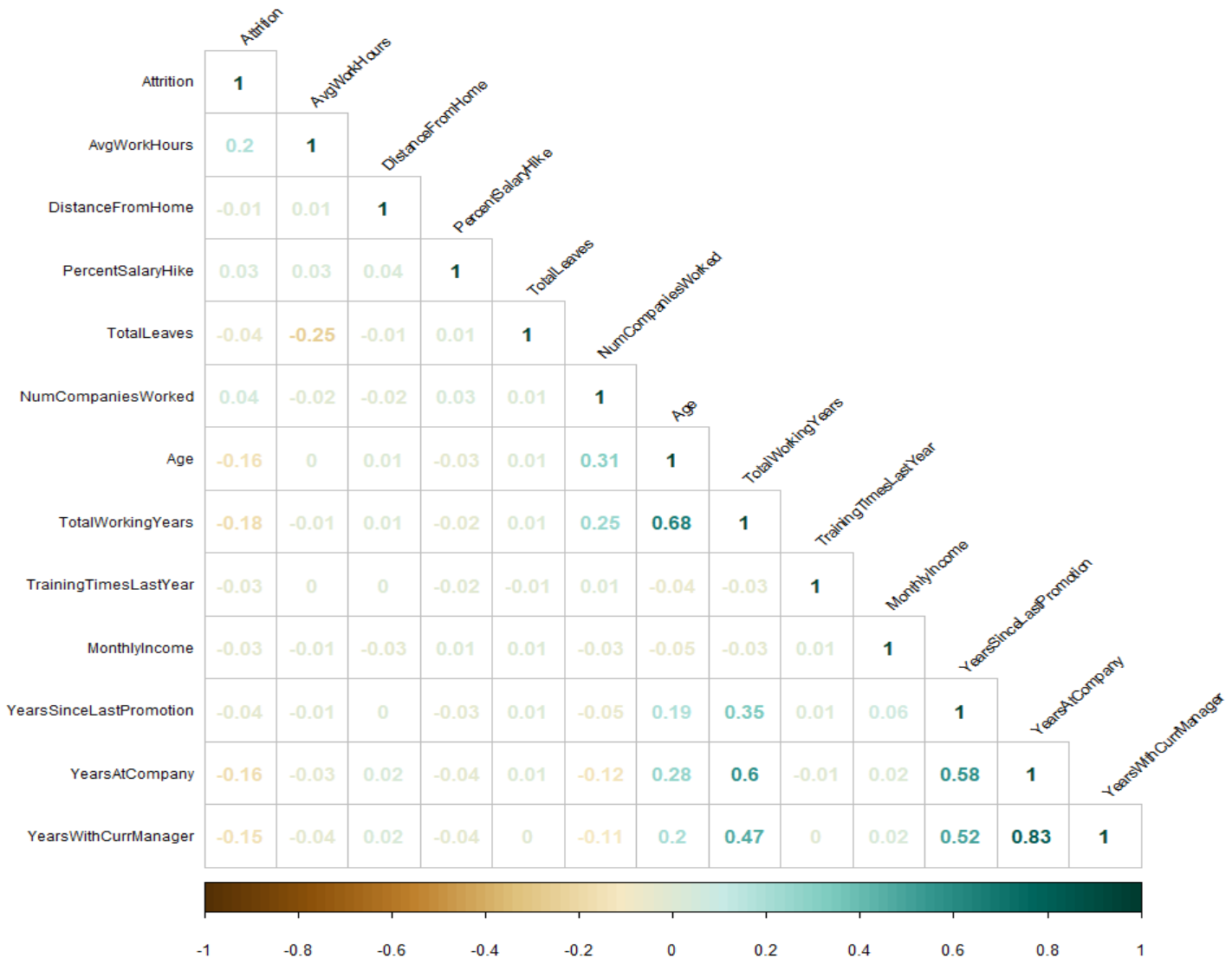
Analysis: Outliers Treated



We have treated all the outliers present in all these variables.

Outliers are handled by capping their values.

Analysis: Correlation Matrix



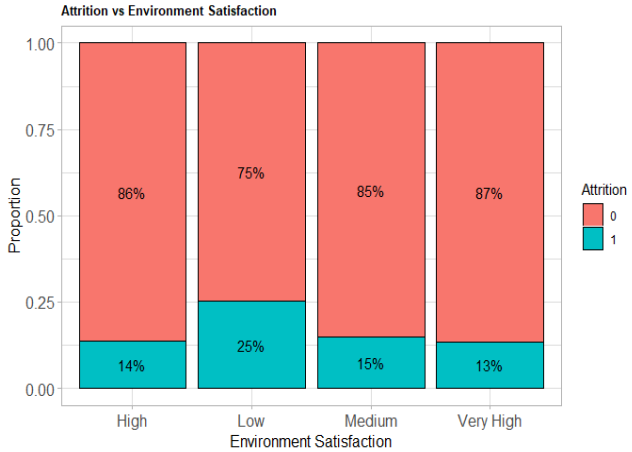
The correlation matrix tells us that:

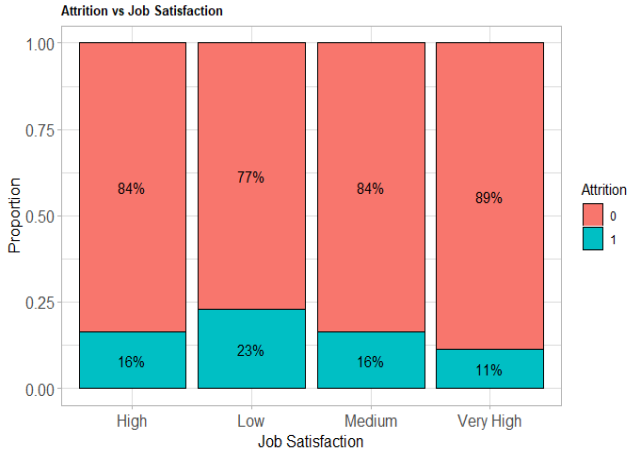
Attrition is negatively correlated to:

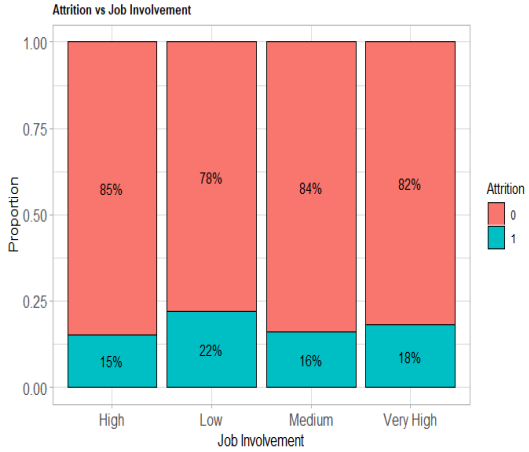
a. Age
b. Total working years
c. Years at company
d. Years with current manager
e. Total Leaves

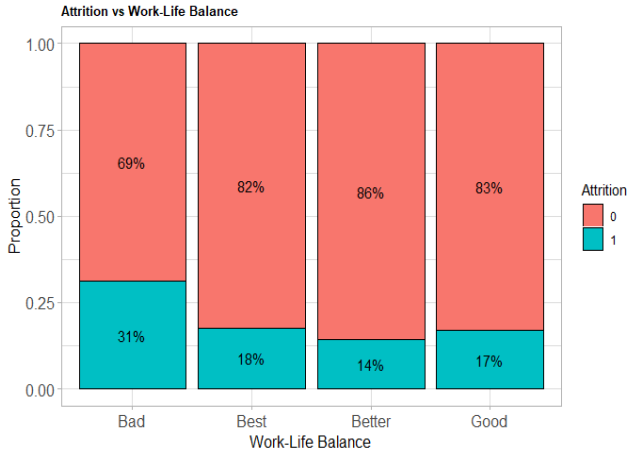
- The number of younger employees leaving the organization is high.
- The employees who work more with one manager are less likely to leave the organization.
- The employees who take less leaves(or who get less leaves approved) are the ones who leave the organization more often.

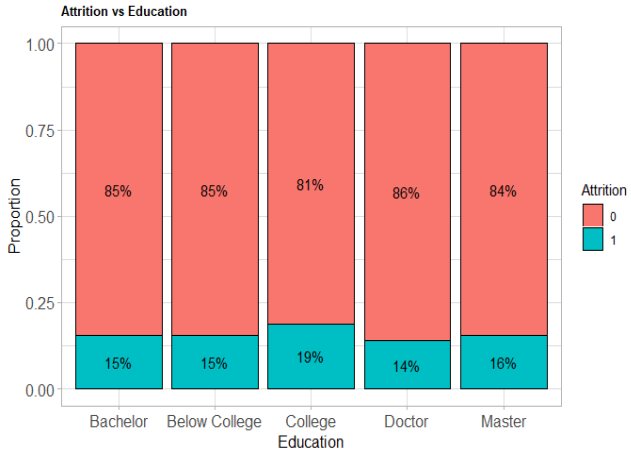
Analysis: Multivariate

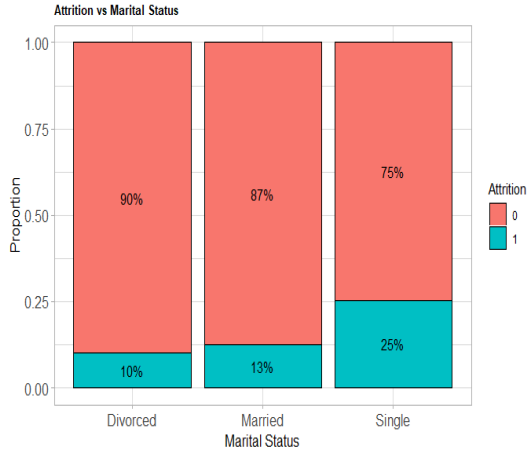






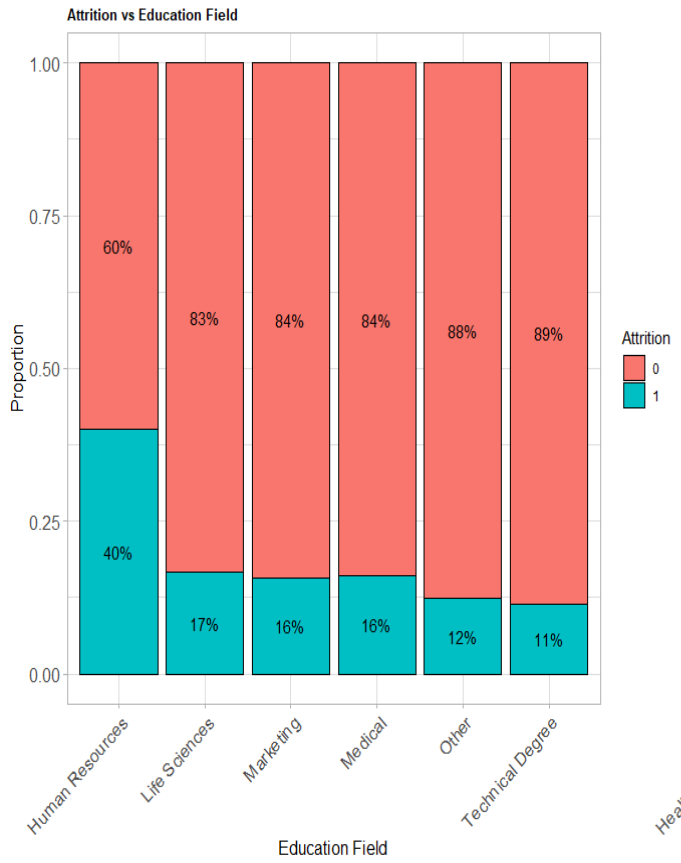


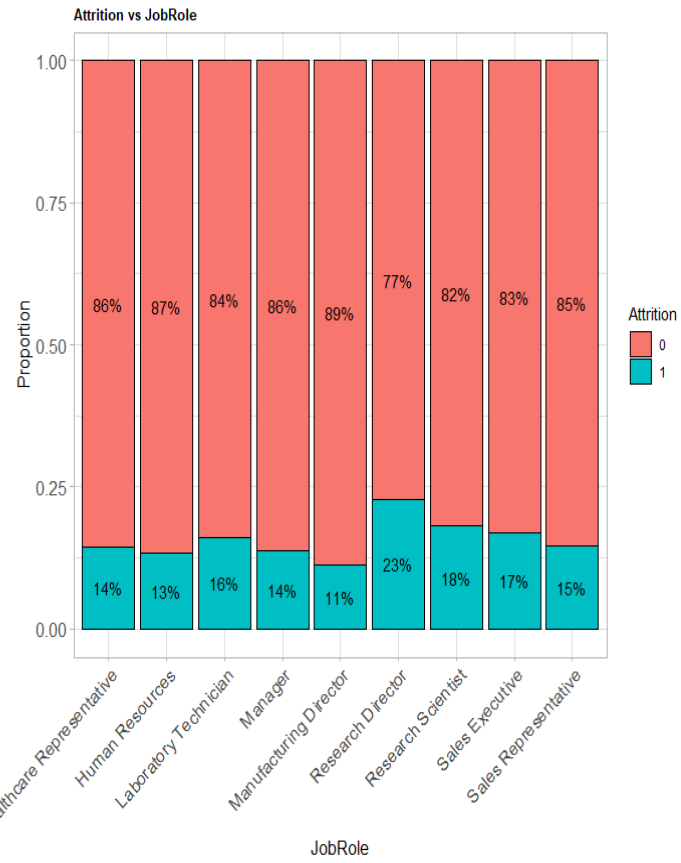


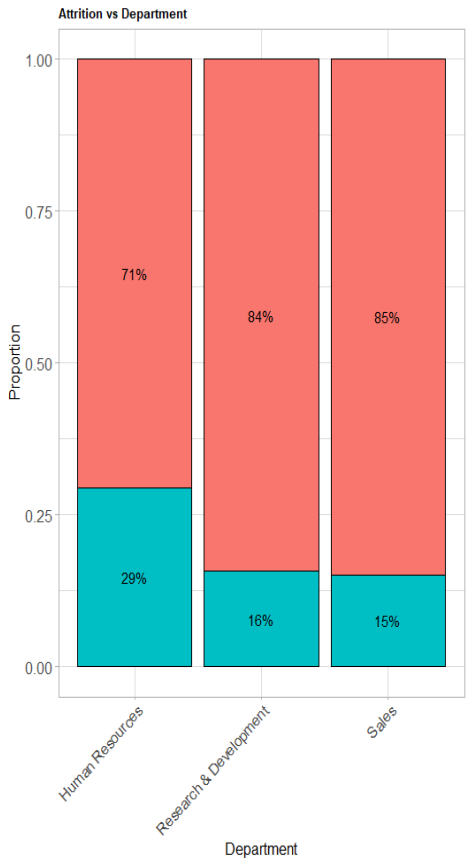


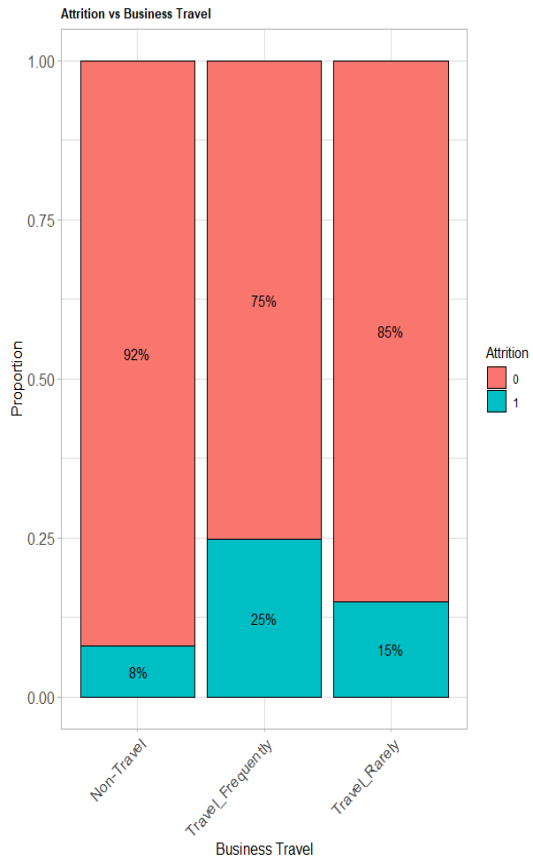
Employees with Low Environment Satisfaction, Bad Work-Life Balance, Low Job Satisfaction, Education – College, Low Job involvement, and Marital Status as Single are more likely to leave the organization.

Analysis: Multivariate



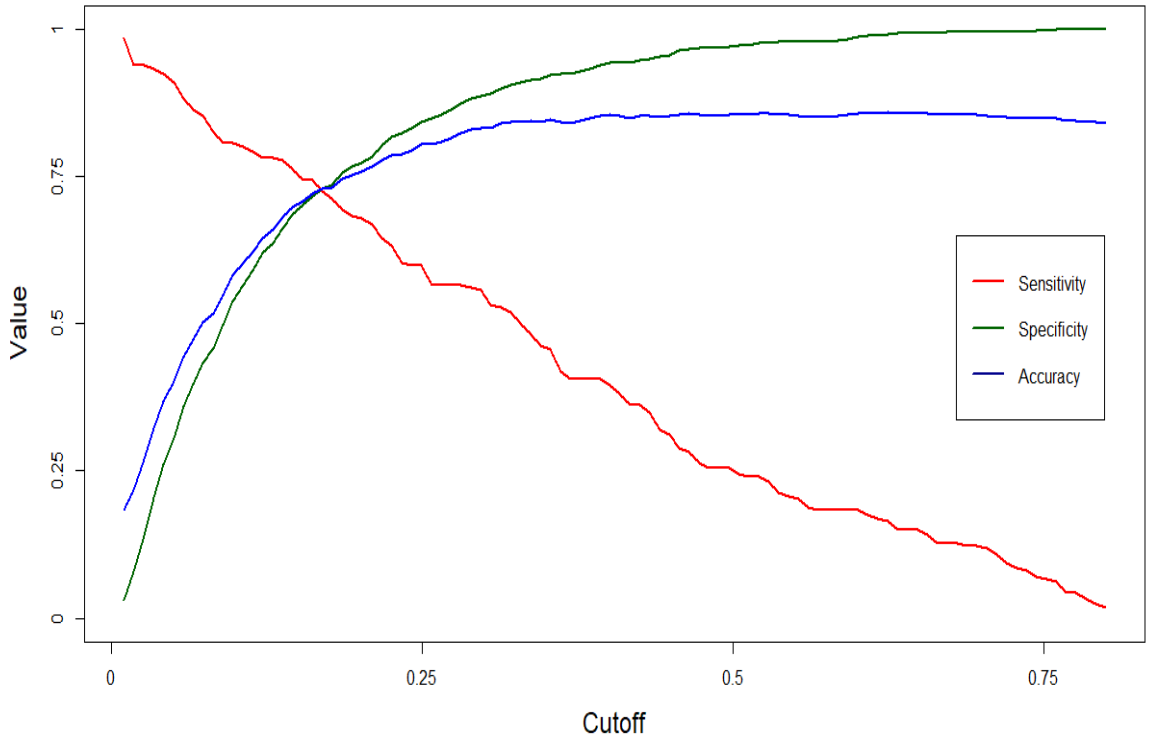


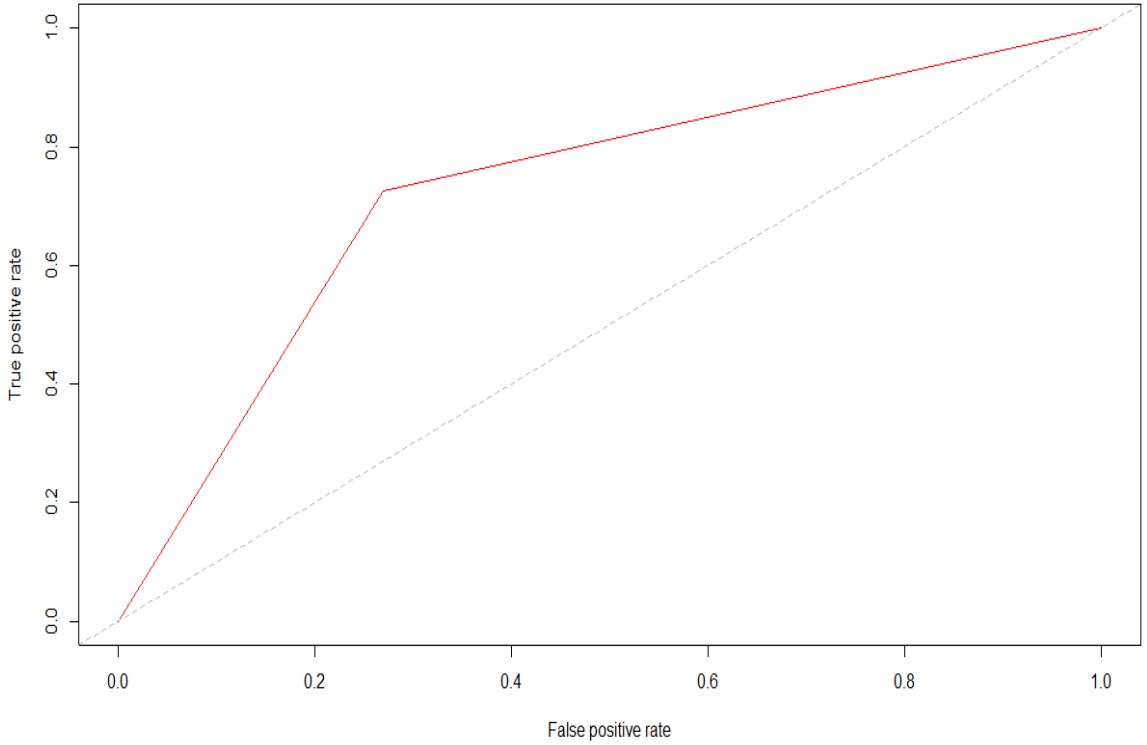




Employees with Education Field as Human Resources, Job Role as Research Director, Department as Human Resources, and who travel more frequently are more likely to leave the organization.

Model Prediction and Evaluation

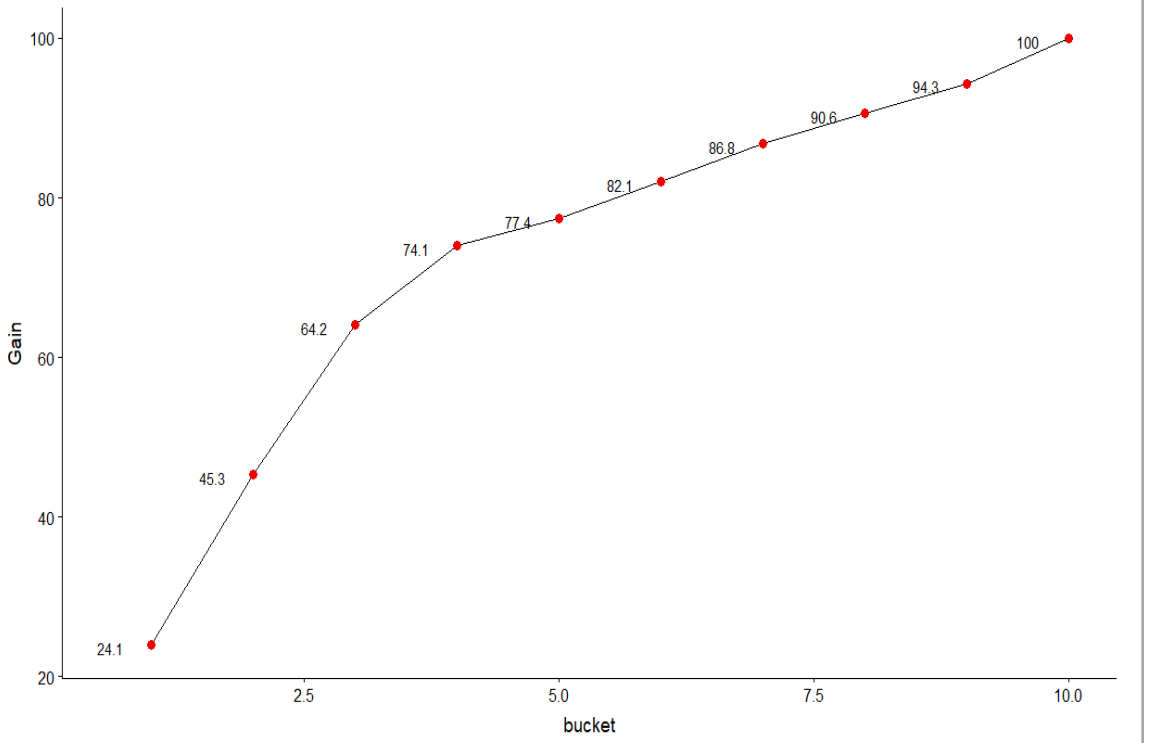


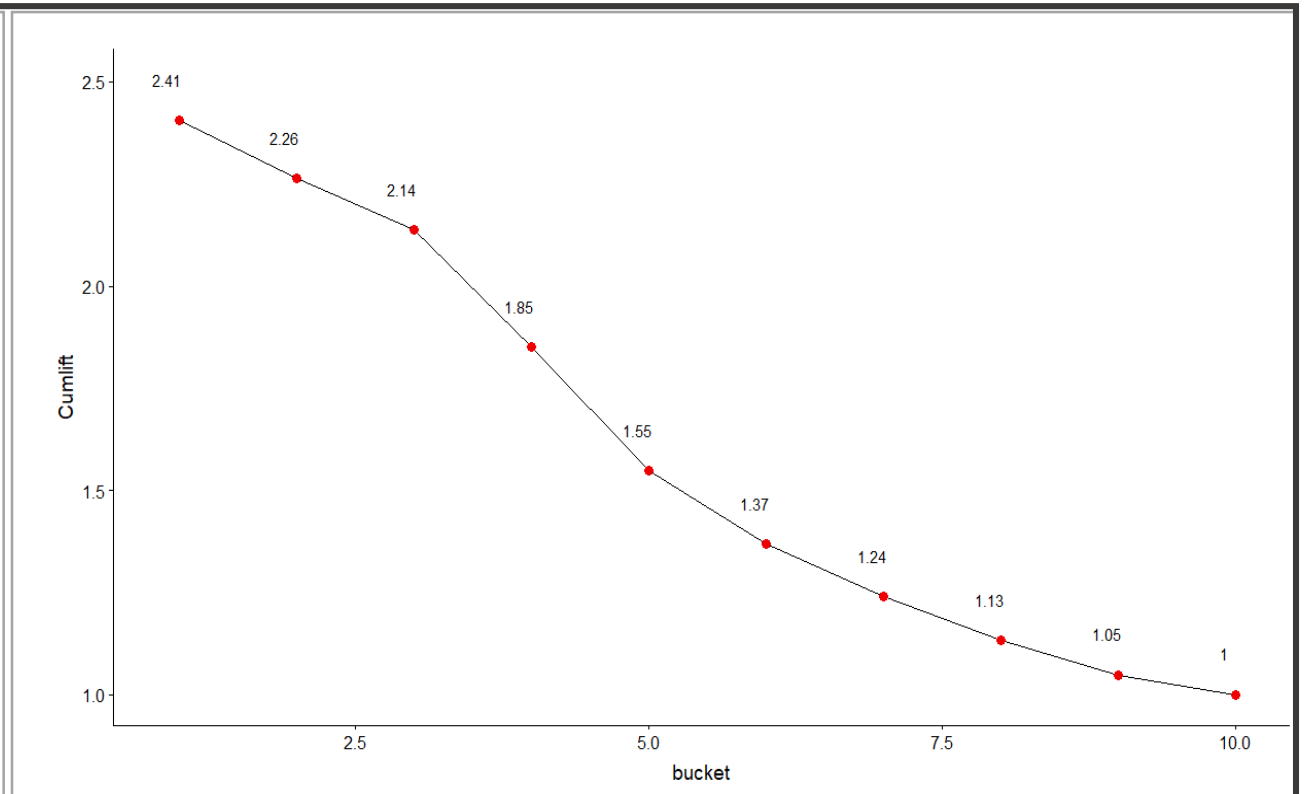


Accuracy : 73% -> How accurate the model is predicting the employee attrition.
Sensitivity : 72.64% -> proportion of true positive correctly predicted by the model as positive.
Specificity : 73.07% -> proportion of true negative correctly predicted by the model as negative.

Area Under Curve – AUC is used in classification analysis in order to determine which of the used models predicts the classes best.
AUC : 0.7285747

Model Evaluation





The model is predicting 74% at the 4th decile that means if we sort the employees according to probability , we will be contacting 74% of the employees who are likely to leave the company.(Among 40%)

The model's gain by the end of 3rd decile is 2.14 times that of a random model's gain at the end of 3 deciles. That means this model catches 2.14 times more attrition of employees than a random model would have caught.

Conclusion

Important factors Affecting Employee Attrition Rate:

As per the analysis, XYZ company should focus on the employees who exhibit following behaviour/attributes:

- Low Environment Satisfaction
- Bad Work-Life Balance
- Low Job Satisfaction
- Low Job involvement
- Marital Status as Single
- Education Field and Department as Human Resources
- Frequently Travel
- Less age (younger employees)
- Less number of companies worked
- Less work experience (fresher or lower level employees)

Accuracy	73%
Sensitivity	72.64%
Specificity	73.07%
AUC	0.73
KS Statistics	46%
Gain	74% @ 4th Decile
Lift	2.14 @ 3rd Decile

Thankyou