

HR Analytics CASE STUDY

SUBMISSION

Group Members:

1. Amit Shekhar
2. Ankush K R
3. Aparna Sastry
4. Suman Kundu



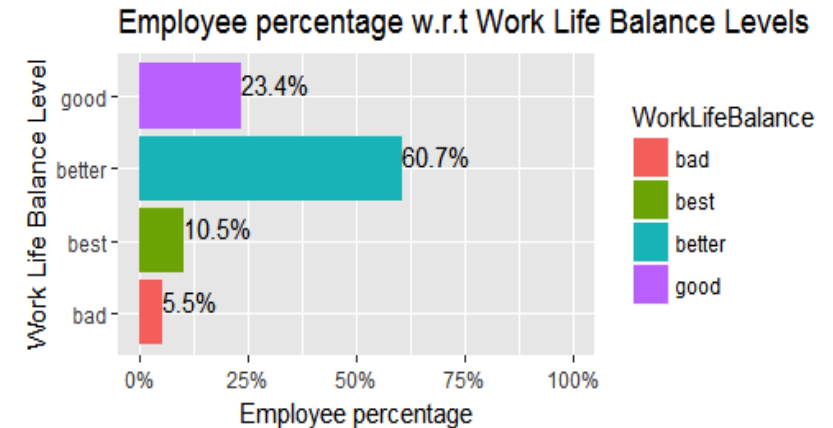
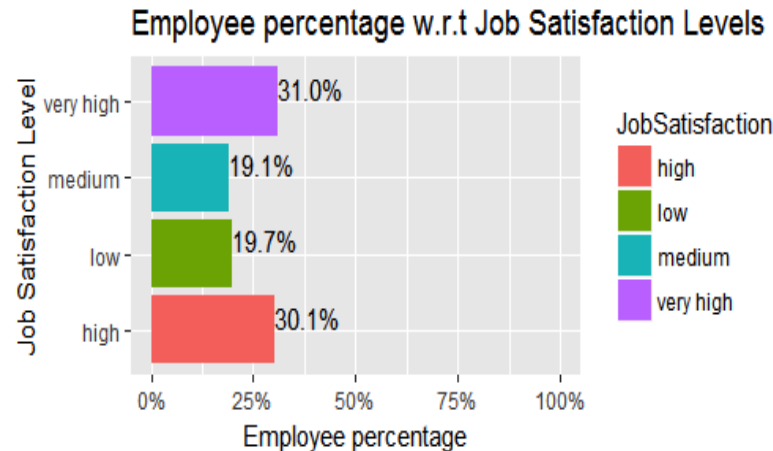
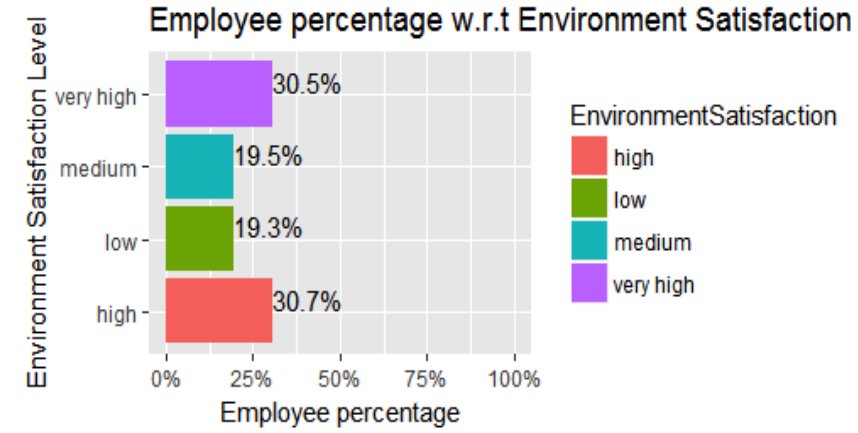
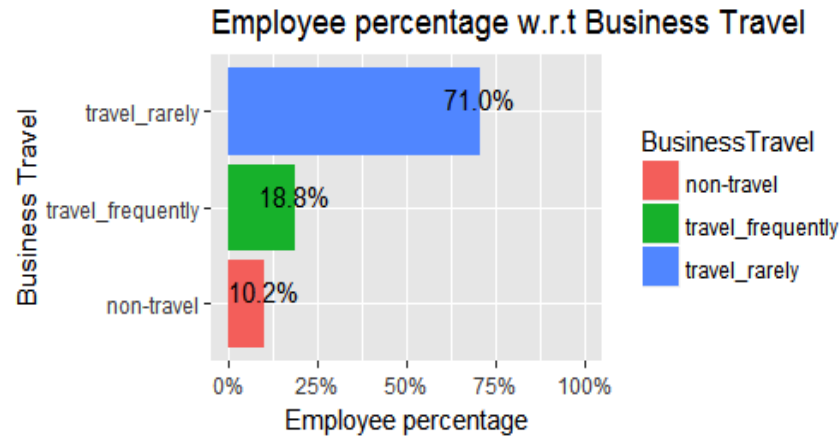
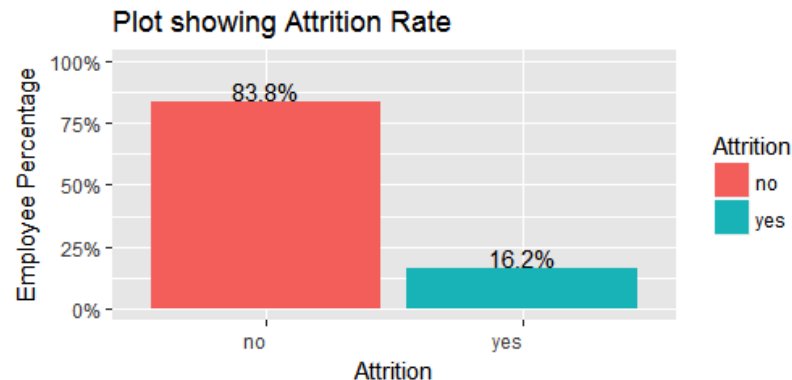
Problem Statement: We are required to model the **probability of attrition** using a logistic regression. The results thus obtained will be used by the management to understand what changes they should make to their workplace, in order to get most of their employees to stay.

Approach: Logistic Regression technique approach is used for modelling the probability of attrition(based on the following steps)

- **Data Understanding** - The given data set is already provided along with data dictionary for understanding the data fields.
- **Data Preparation**
- **EDA – Univariate and Bivariate analysis**
- **Model Building**
 - Missing Value Imputation
 - Outlier Treatment
 - Standardising scales of continuous variable
 - Dummy variable creation for categorical variables
 - Model Building
- **Model Evaluation**
 - Accuracy, Sensitivity, and Specificity
 - KS Static
 - Lift and Gain chart

- Data Understanding : There are 5 data frames - employee_survey_data, manager_survey_data, general_data, in_time, out_time
 - **employee_survey_data** gives details of the Work Environment Satisfaction Level, Job Satisfaction Level and Work life balance level for each employee of the organisation.
 - **manager_survey_data** gives details of the Job Involvement Level & Performance rating for last year as given by the manager of each employee of the organisation.
 - **general_data** gives details of a number of attributes for each employee in the organisation. There are 10 Continuous Variables & 14 Categorical Variables in the general_data dataset.
 - **in_time and out_time** datasets : **in_time** gives the date & time when employee enters the office for each employee. **out_time** dataset gives the date & time when employee leaves the office for each employee.
 - Data spreads for the complete year of 2015.
 - There are 12 columns which are company marked holidays and have NA throughout all rows in both in_time & out_time. The other NA values imply that the employee has taken a leave of absence from work.
 - The average number of working hours per month for each employee are calculated using the in_time & the out_time datasets.
- Data Preparation
 - Data Preparation for in_time & out_time and calculation of average monthly working hours for each employee has been done. Transformed the dates into a format more suited.
 - Data Preparation for employee_survey_data, manager_survey_data & general_data are done by removing the NA values wherever required, creating the dummy variables, standardizing the values, etc.

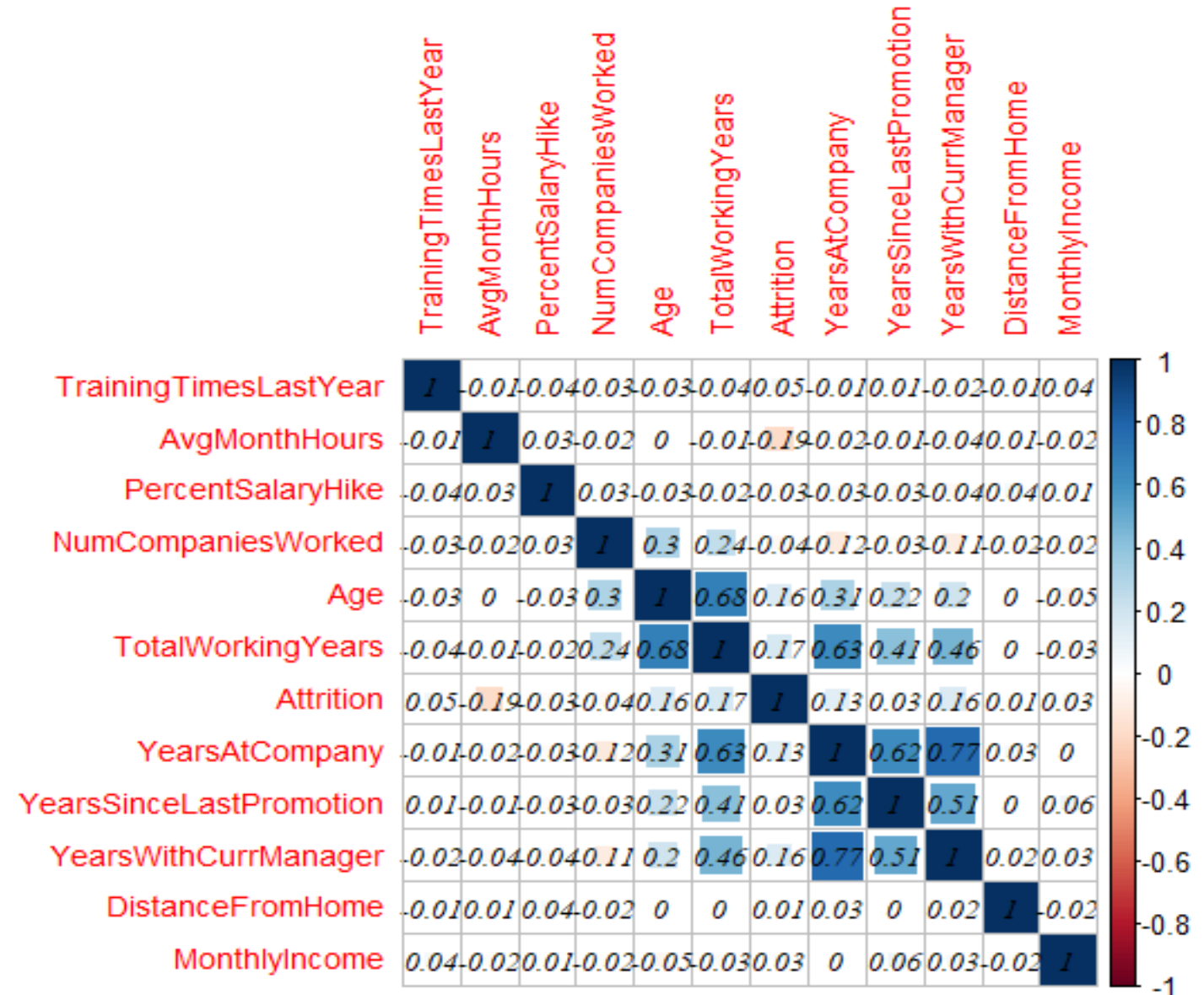
- **Plot showing Attrition Rate** : The percentage of attrition in the last year is 16.16% **Plot showing Employee percentage w.r.t Business Travel**: Majority(70.95%) of the employees travelled rarely for business purposes in the last year. 18.81% of employees have travelled frequently in the last year.
- **Plot showing Employee percentage w.r.t Environment Satisfaction levels**: Majority (30.67%) have high Work Environment Satisfaction. 19.30% have low Work Environment Satisfaction.
- **Plot showing Employee %age w.r.t. Job satisfaction levels**: Majority(31.02%) have very high level of Job Satisfaction. 19.70% have low level of Job Satisfaction.
- **Plot showing Employee percentage w.r.t Work Life Balance**: # Majority(60.67%) have better Work life balance, followed by 23.37%. 10.47% have best and Very few(5.49%) have bad Work life balance.



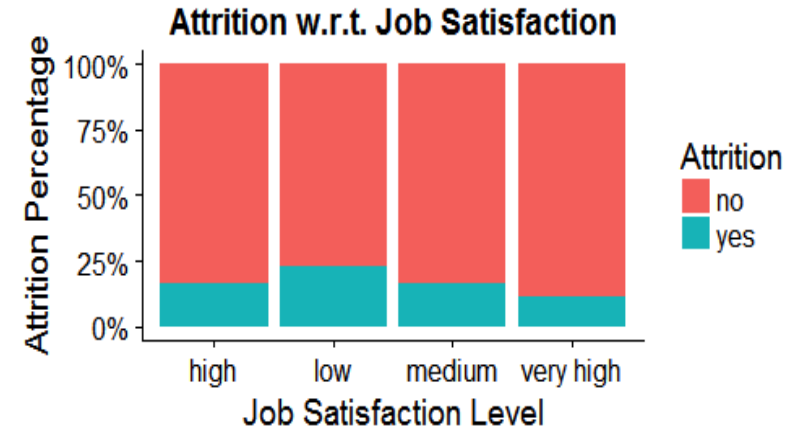
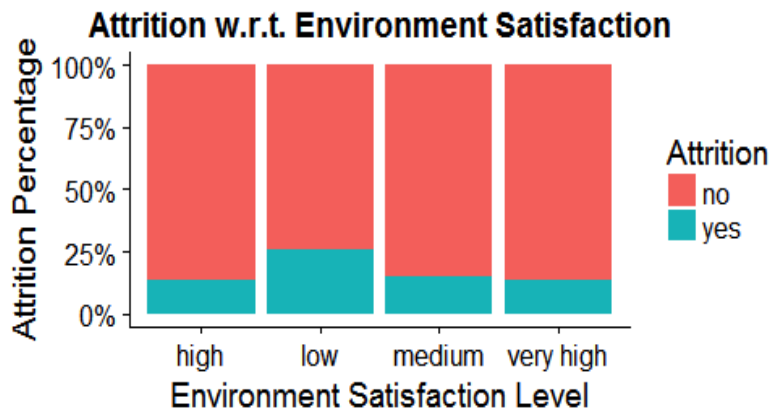
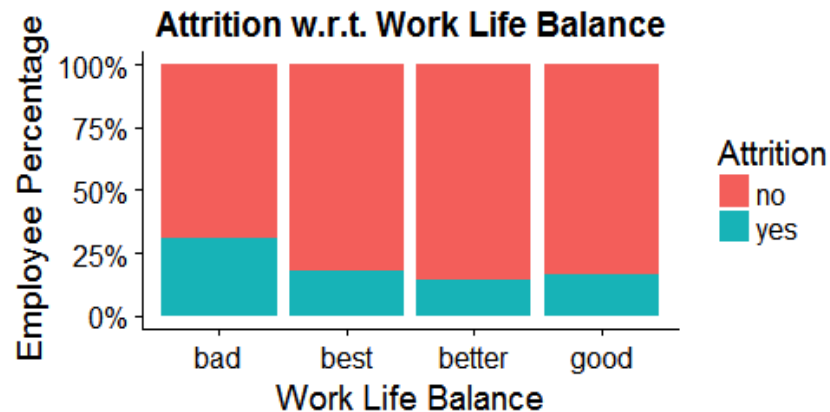
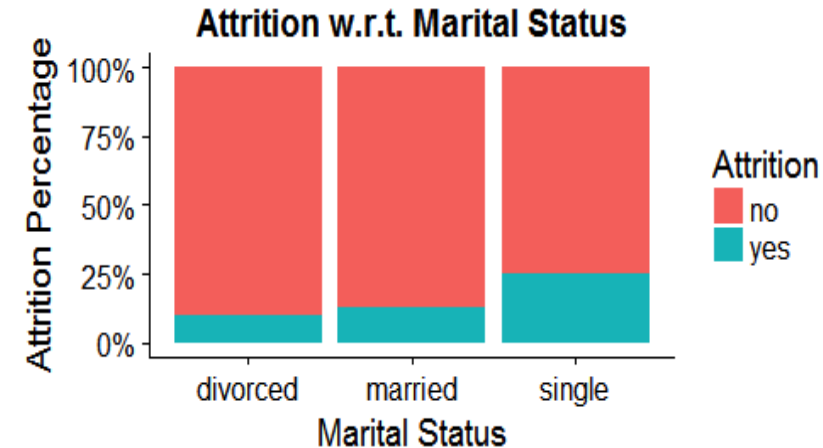
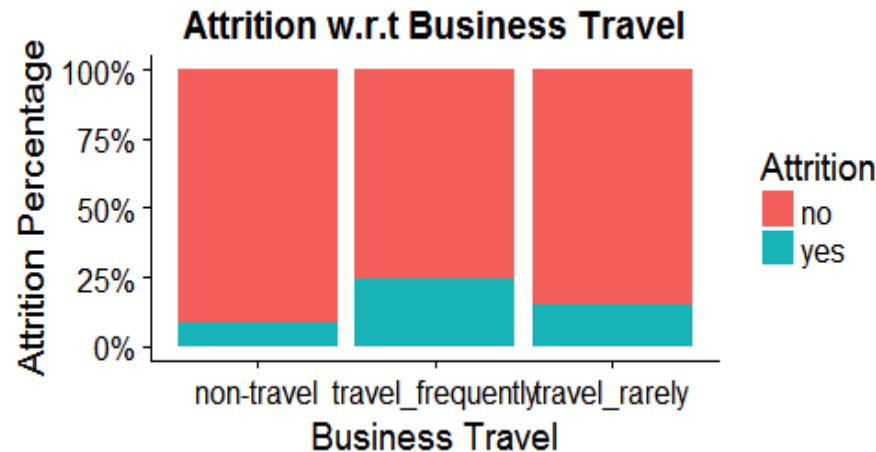
Correlation Plot

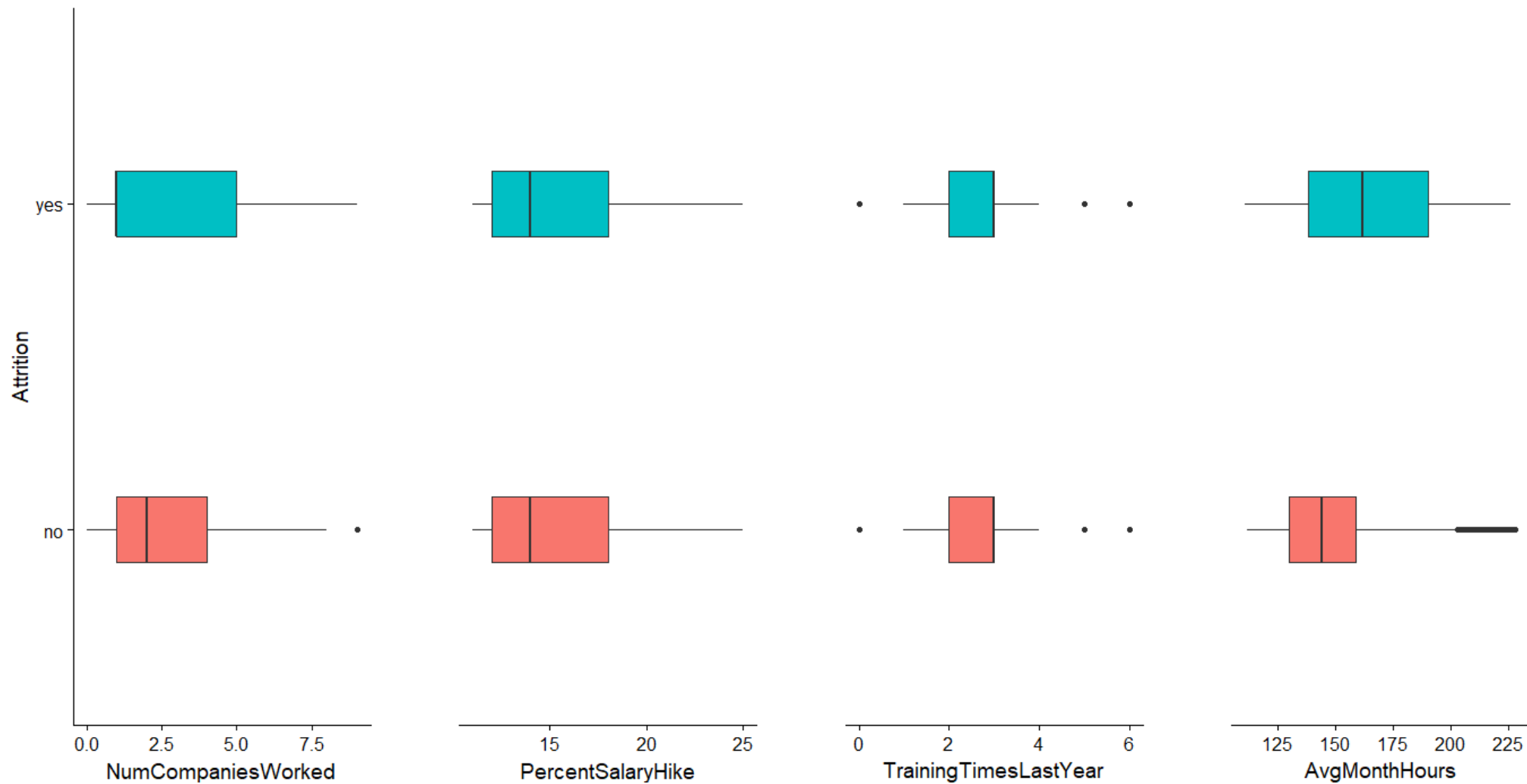
There is some impact of the 'Attrition' variable on other continuous variables.

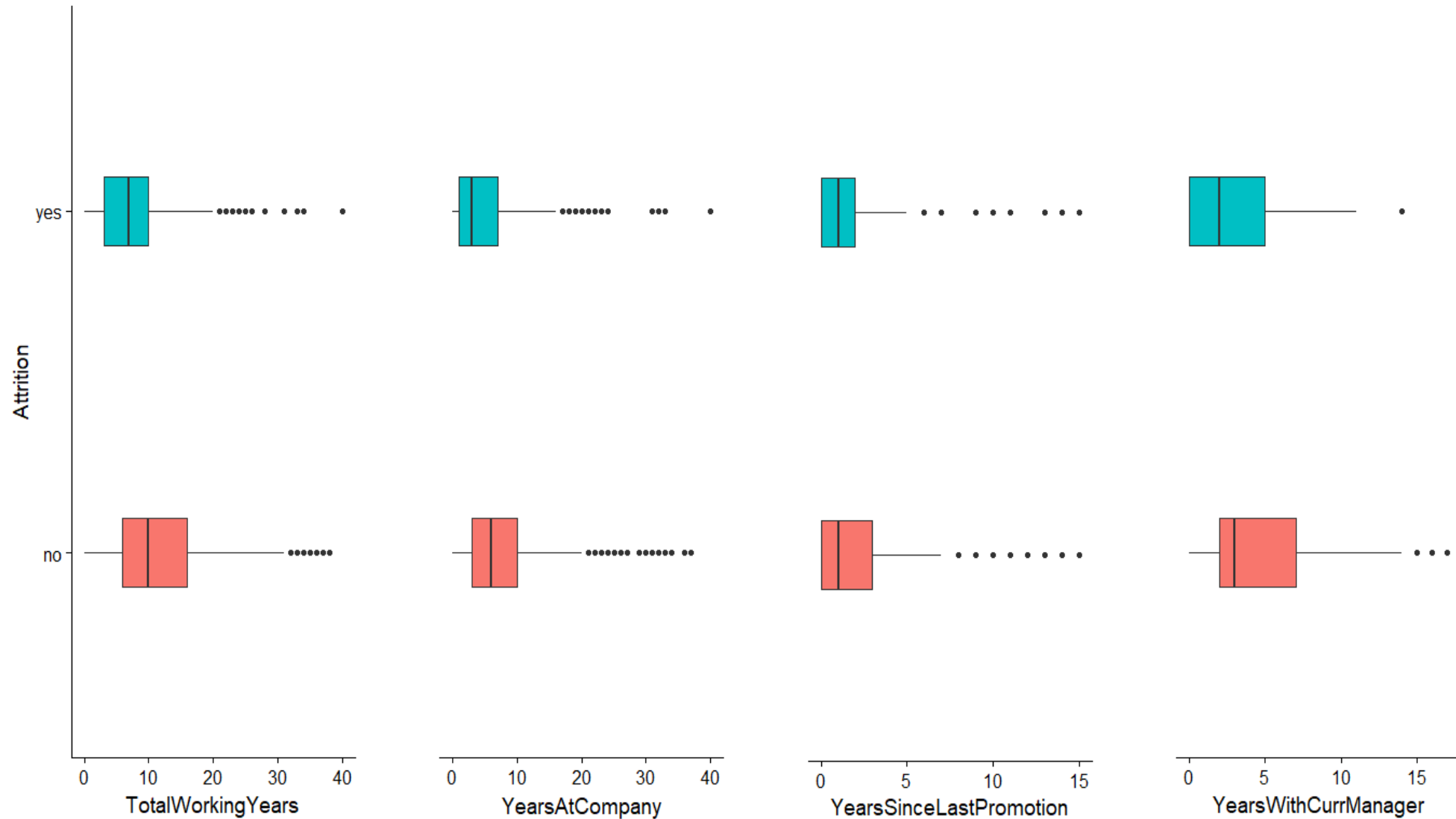
- The variables 'YearsAtCompany', 'YearsWithCurrManager' & 'YearsSinceLastPromotion' are correlated.
- The variable 'TotalWorkingYears' is also correlated with the following variables:
 - 'YearsAtCompany', 'YearsWithCurrManager', 'YearsSinceLastPromotion' & 'Age'.
- The variable 'AverageMonthHours' & 'Attrition' are negatively correlated, although not significantly. Attrition rate for people who work more per more is high.



- **Proportion of Attrition w.r.t Business Travel:** Rate of Attrition is higher in the employees who travel frequently.
- **Proportion of Attrition w.r.t Marital Status:** Rate of attrition is higher for the employees who are single followed by employees who are married.
- **Proportion of Attrition w.r.t Environment Satisfaction Level:** Rate of attrition is higher for the employees with Low Environment Satisfaction Level.
- **Proportion of Attrition w.r.t Job Satisfaction Level:** Proportion of attrition is higher for employees with Low Job Satisfaction Level.
- **Proportion of Attrition w.r.t Work Life Balance:** Rate of attrition is higher for the employees with Bad Work Life Balance Level.





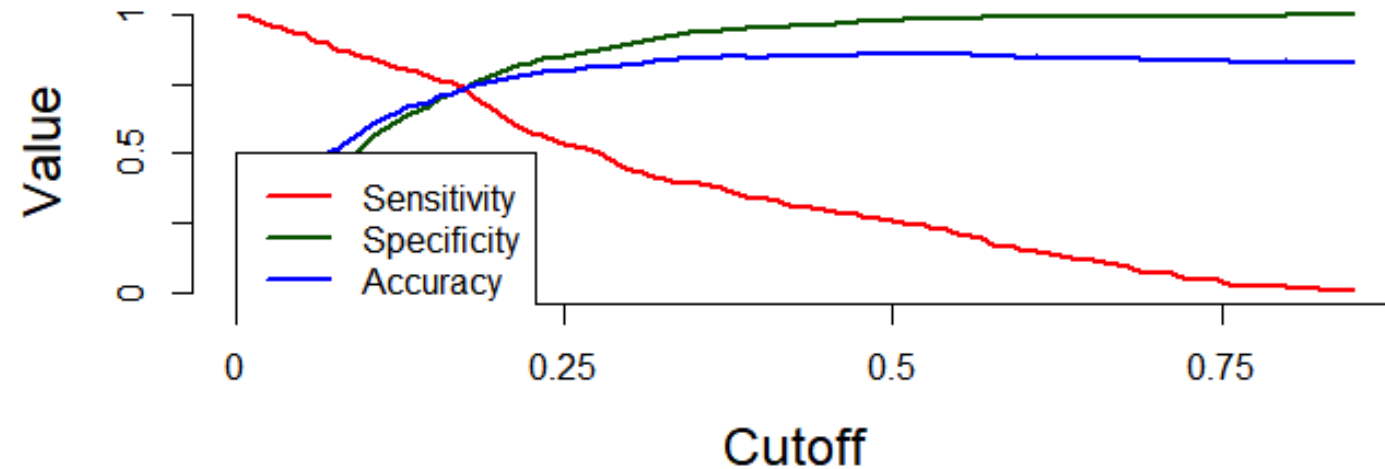
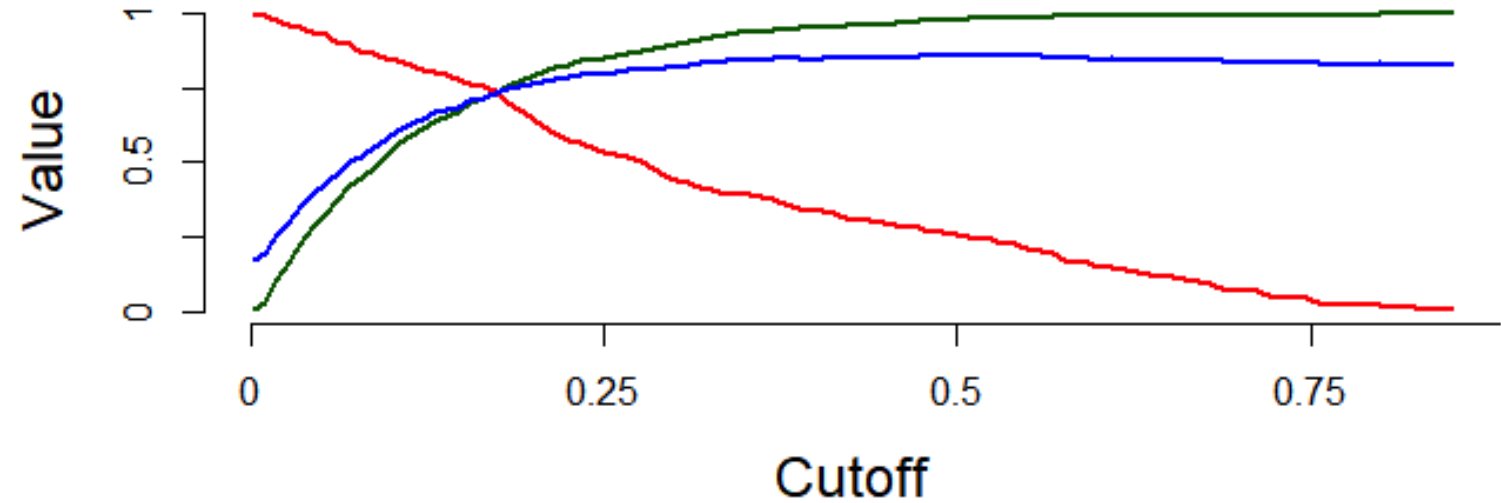




Model Building

- The significant variables which have an impact on attrition rates are :
 - NumCompaniesWorked
 - TotalWorkingYears
 - TrainingTimesLastYear
 - YearsSinceLastPromotion
 - YearsWithCurrManager
 - AvgMonthHours
 - BusinessTravel - frequently
 - MaritalStatus - single
 - Environment Satisfaction - low
 - JobSatisfaction - low
 - WorkLifeBalance - better

- Based on the final model on test data and choosing a cut-off value of 0.1725 for final model, we have arrived at –
 - Accuracy - 0.7364341 (73.6%)
 - Sensitivity - 0.7409091 (74.0%)
 - Specificity - 0.735514 (73.5%)
- KS Statistic - 47.64% (0.4764231)



Some of the recommendations based on the results for changes that Management should make in order to get most of their employees to stay are mentioned below:

- The attrition tends to increase with increase in number of companies worked for by the employee
- The number of times an employee has been trained last year has an effect of the attrition rate especially if the employee is single.
- The majority of attrition happens for employees who have lesser work experience.
- The attrition rate tends to decrease if employees are promoted frequently.
- Employees who have higher average of monthly working hours tend to leave the company
- Employees who travel frequently have a higher attrition rate
- The company should look at improving the work environment by engaging employees in various activities
- Employees' personal growth has to be addressed for improving their job satisfaction levels
- Work Life balance is also a factor which affects the attrition rate.