HR ANALYTICS CASE STUDY

SUBMISSION

Group Name:
1.   Arpita Ghosh
2.   Biplab Ghosal
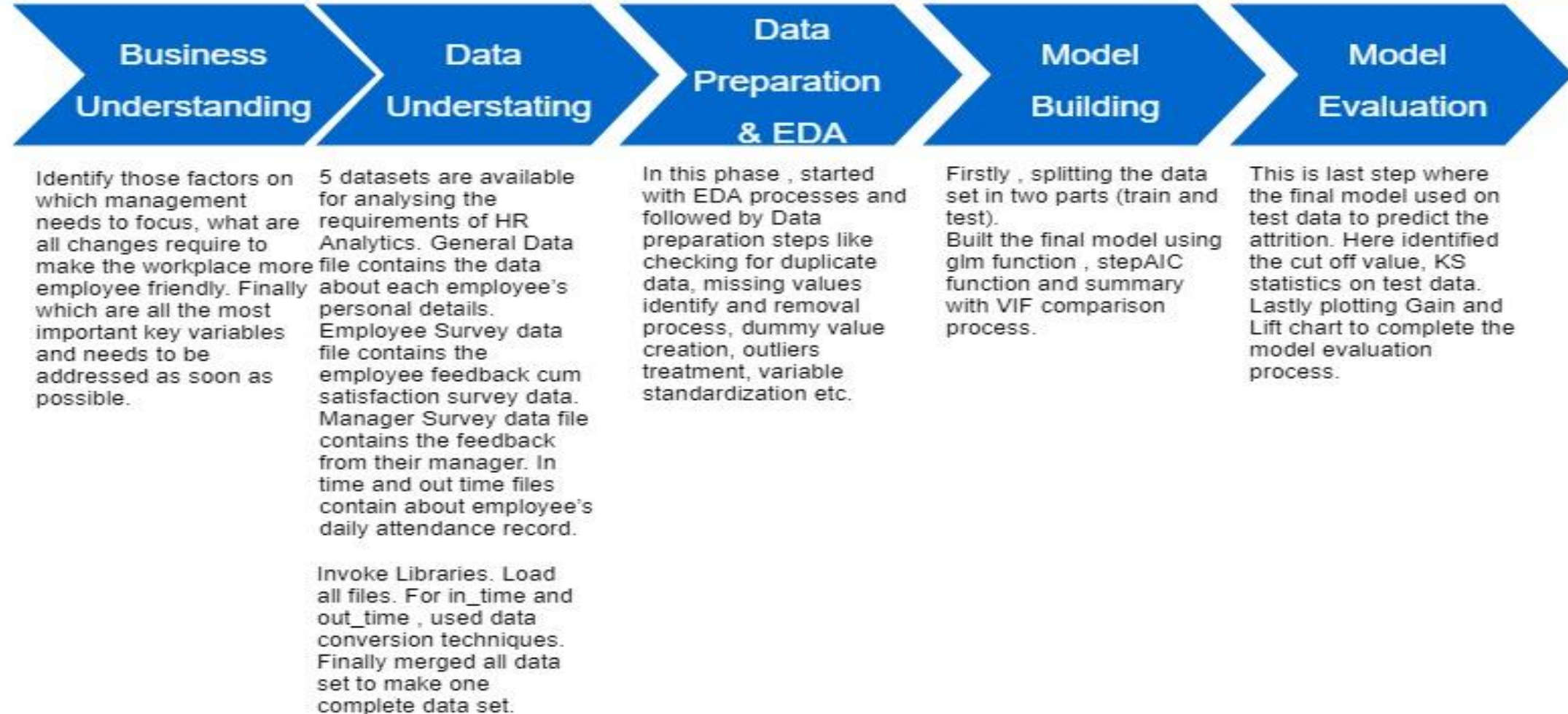3.   Jagannath Sen
4.   Pritam Pan

Abstract

A large company named **XYZ**, employs, around 4000 employees. However, every year, it is facing 15% of attrition rate. There are various reasons for this like project timelines , recruiting new talent to main sizable department and trained them for the job.
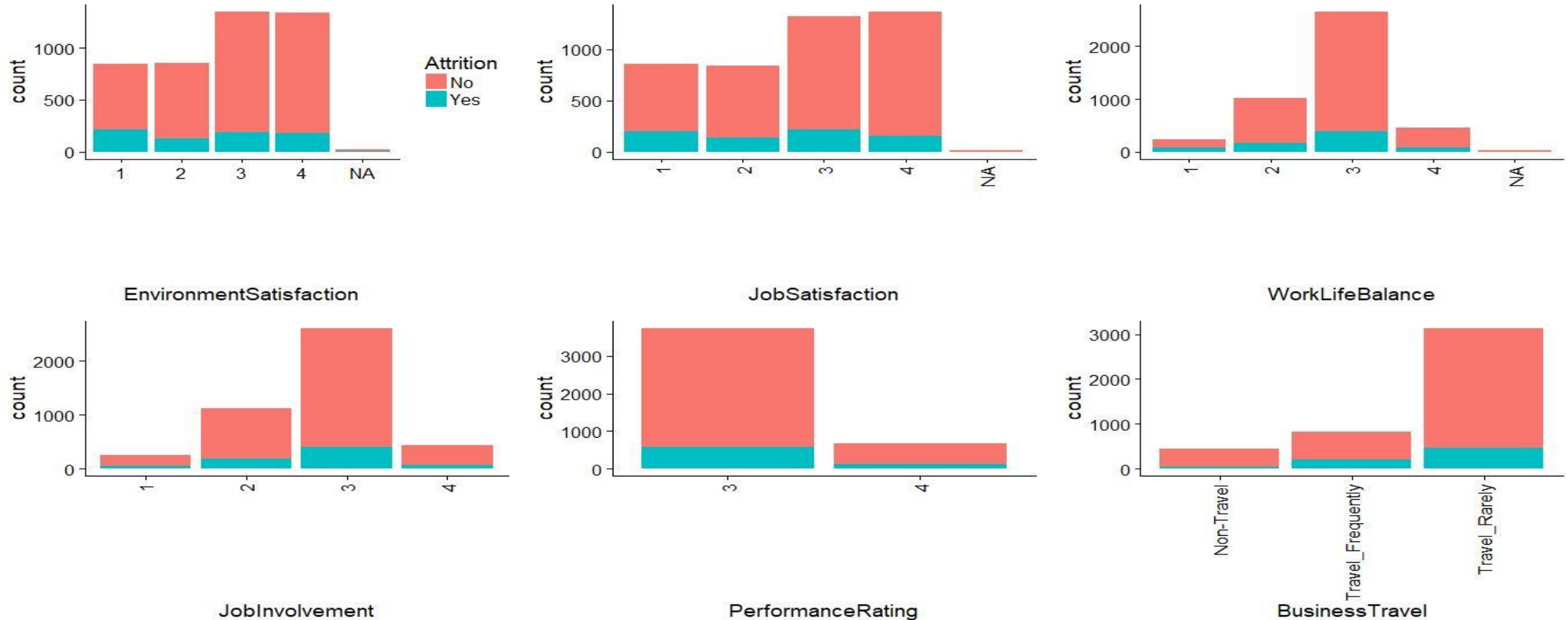
The primary objective is to identify those factors on which management needs to focus, what are all changes require to make the workplace more employee friendly. Finally which are all the most important key variables and needs to be addressed as soon as possible.

5 datasets are available for analysing the requirements of HR Analytics. General Data file contains the data about each employee's personal details. Employee Survey data file contains the employee feedback cum satisfaction survey data. Manager Survey data file contains the feedback from their manager. In time and out time files contain about employee's daily attendance record.

HR Analytics Analysis

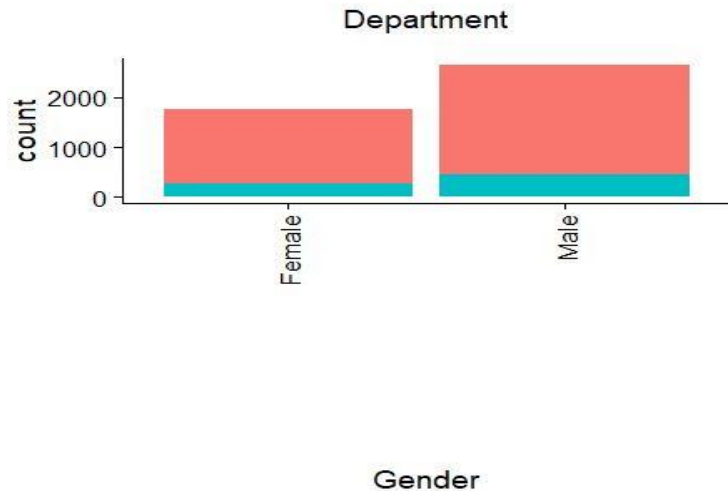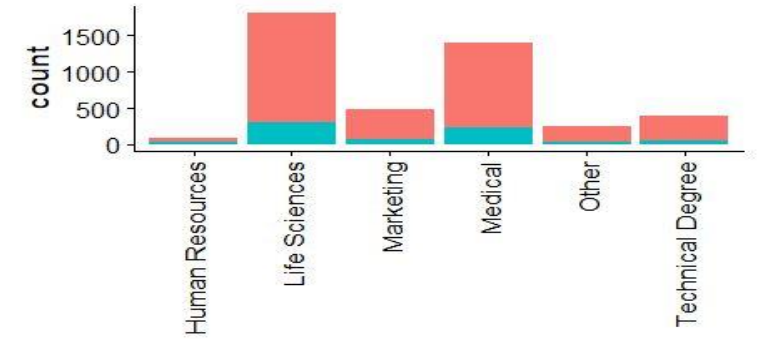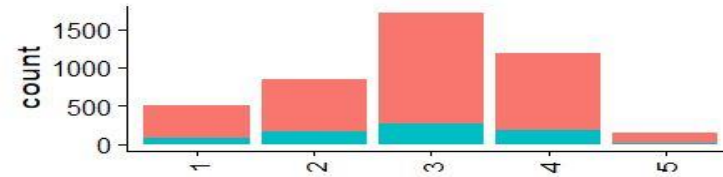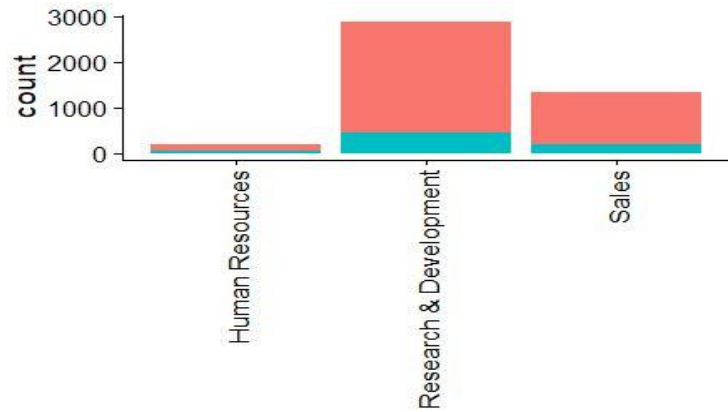| Business Understanding	Data Understating	Data Preparation & EDA	Model Building	Model Evaluation
Identify those factors on which management needs to focus, what are all changes require to make the workplace more employee friendly. Finally which are all the most important key variables and needs to be addressed as soon as possible.	5 datasets are available for analysing the requirements of HR Analytics. General Data file contains the data about each employee's personal details. Employee Survey data file contains the employee feedback cum satisfaction survey data. Manager Survey data file contains the feedback from their manager. In time and out time files contain about employee's daily attendance record. Invoke Libraries. Load all files. For in_time and out_time , used data conversion techniques. Finally merged all data set to make one complete data set.	In this phase , started with EDA processes and followed by Data preparation steps like checking for duplicate data, missing values identify and removal process, dummy value creation, outliers treatment, variable standardization etc.	Firstly , splitting the data set in two parts (train and test). Built the final model using glm function , stepAIC function and summary with VIF comparison process.	This is last step where the final model used on test data to predict the attrition. Here identified the cut off value, KS statistics on test data. Lastly plotting Gain and Lift chart to complete the model evaluation process.

From the above plots, we can visualize that Attrition might be depend on the following factors :
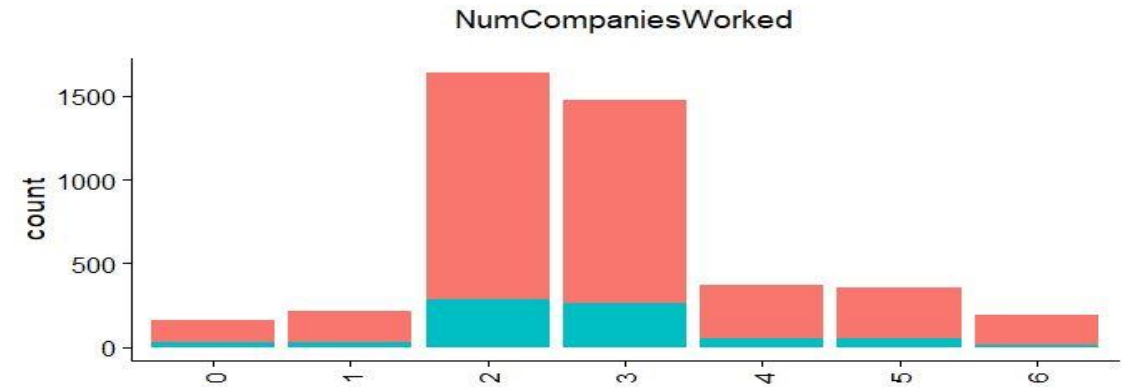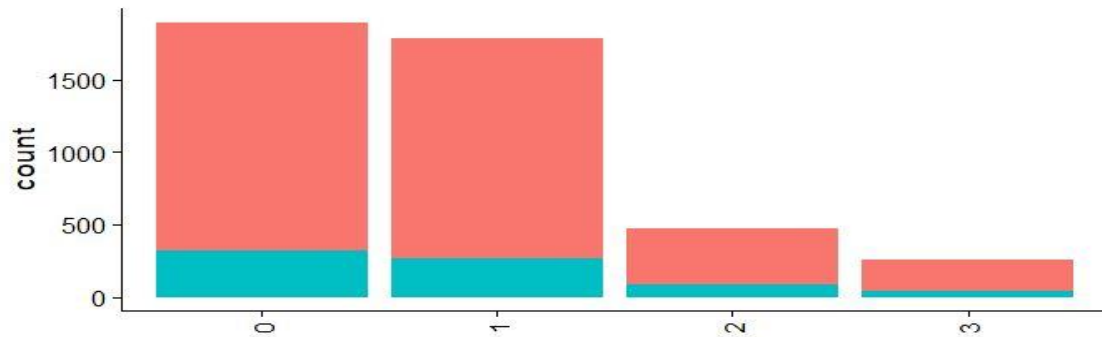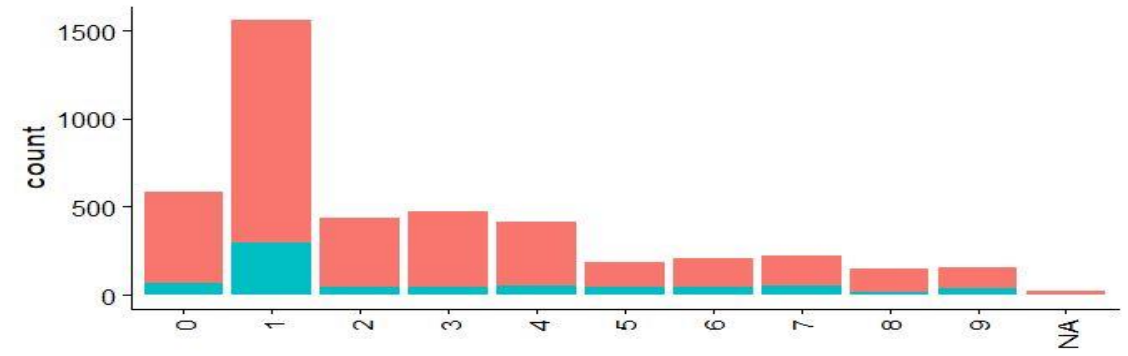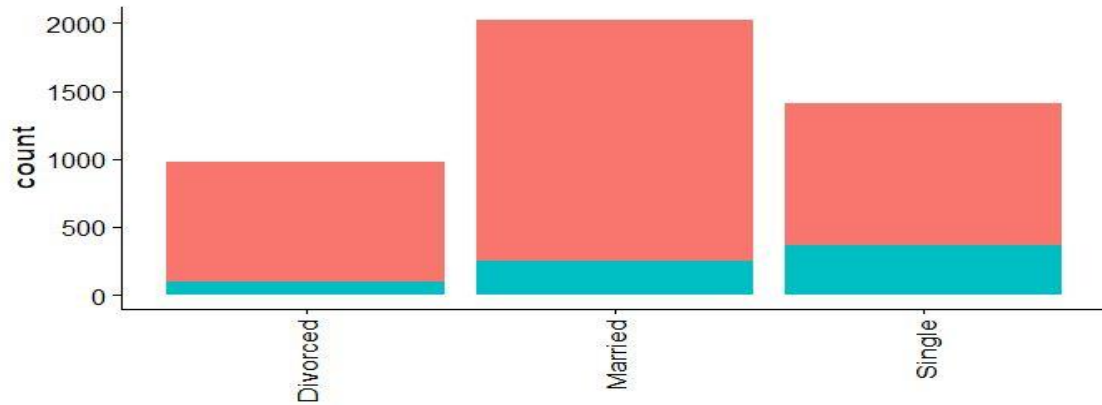Job Involvement, Performance Rating & Business Travel.
Environment Satisfaction, Job Satisfaction and Work Life Balance can also impact attrition.
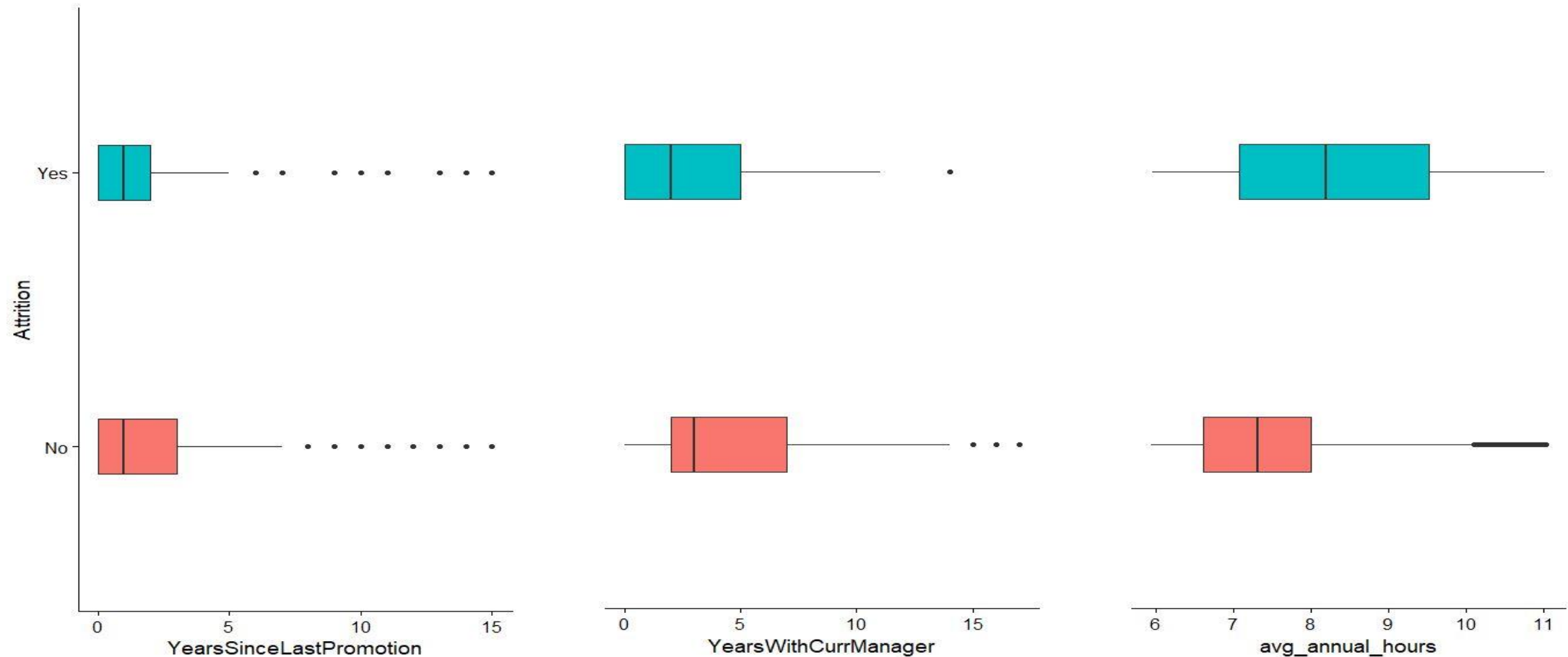
From the above plots, we can visualize that Attrition might be depend on the following factors :
Gender (Attrition rate in Male employee's are more compared to women), Job Level (Attrition rate is more in initial job levels) and Department (with R&D department has maximum attrition)

# EDA – How different factors affect attrition

From the above plots, we can visualize that Attrition might be depend on the following factors :
Marital Status (Most attrition rates among Single's), Number of Companies Worked (Attrition rate is more for employee's who have worked earlier in 1 more company).
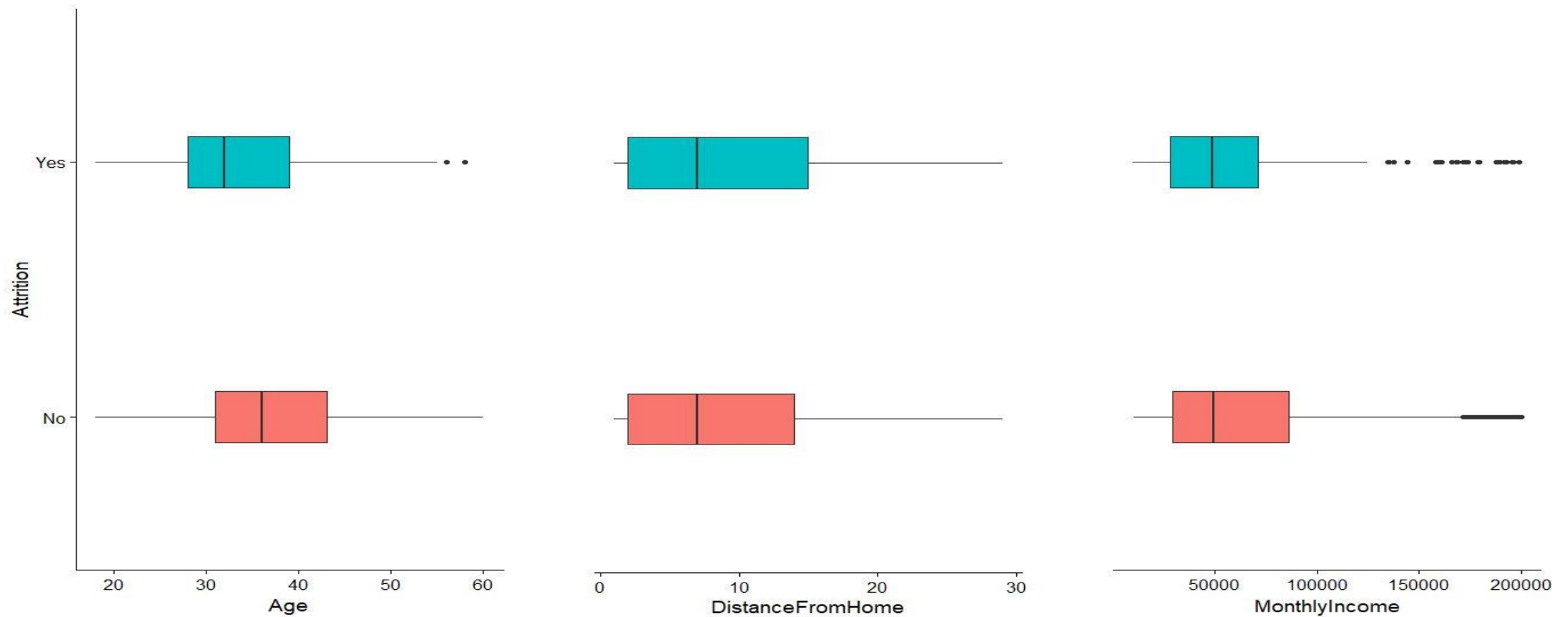
From the above plots, we can visualize that Attrition is more when average annual hours spent in office increases for a employee. Also less the number of years with current manager, more is the attrition.
We can also observe some outliers in the data.
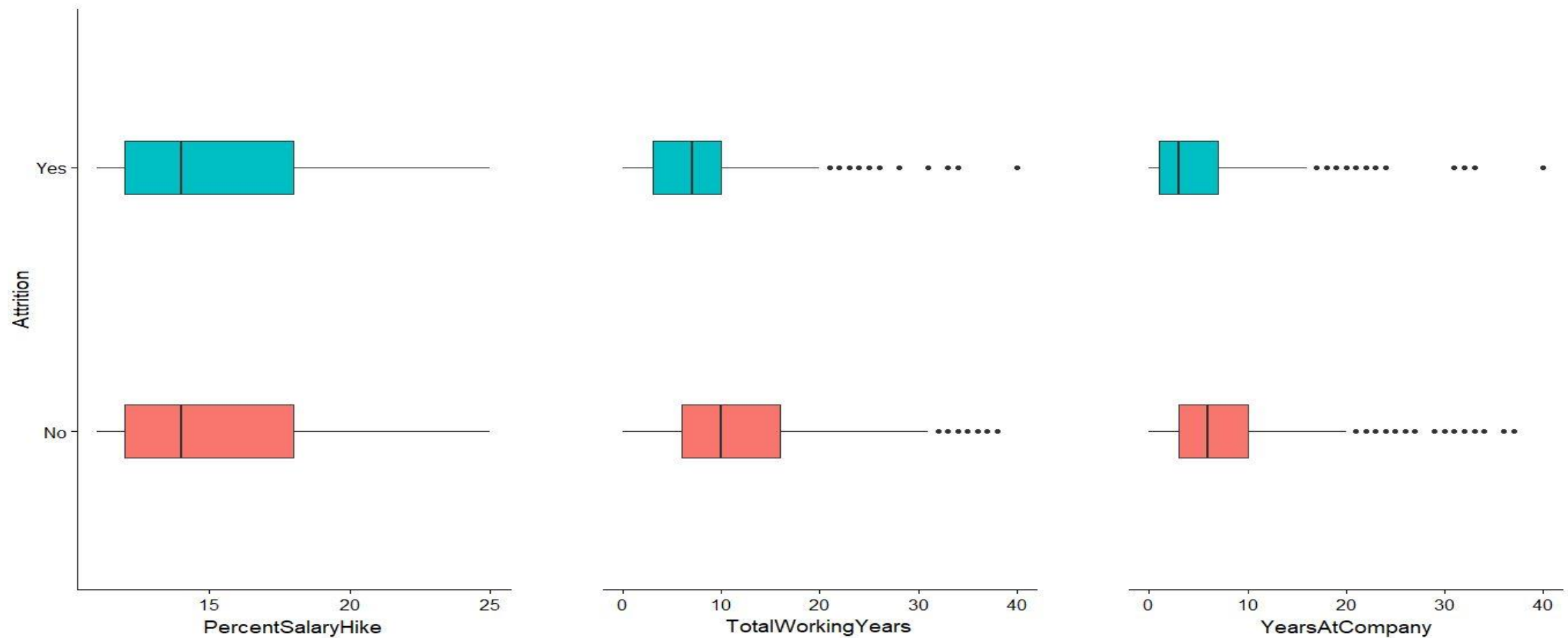
From the above plots, we can visualize that Attrition is more when total working years is less and also attrition is more when total number of years in current company is less.
We can also observe some outliers in the data.

From the above plots, we can visualize that Attrition is more when age of employee is less. It doesn't seem that attrition is highly dependent on distance from home or on monthly income.

Logistic Regression Model Creation

Based on the Logistic Regression Model created to predict whether an employee can probably leave the company or not, we found the following factors to be highly dependent on attrition rate.

- **Age** – Age of the employee

- **NumCompaniesWorked** – Number of companies the employee worked with

- **TotalWorkingYears** – Total working years of experience

- **TrainingTimesLastYear** – Number of trainings attended last year

- **YearsSinceLastPromotion** – Number of years since last promotion

- **YearsWithCurrManager** – Number of years with current manager

- **avg_annual_hours** – Annual average office hours

- **EnvironmentSatisfaction.xLow** – Low Environment satisfaction

- **JobSatisfaction.xLow** – Low Job Satisfaction

- **JobSatisfaction.xVery.High** – High Job Satisfaction

- **WorkLifeBalance.xBetter** – Better work life balance

- **JobRole.xManager** – Current job role of manager

- **JobRole.xManufacturing.Director** – Current job role of manufacturing director

- **MaritalStatus.xSingle** – Single employees

Logistic Regression Model Evaluation

We found the final model based on logistic regression model creation. In this section we will evaluate the performance of this model.

We will follow 3 different methods to evaluate the performance:

1. **Sensitivity, Specificity and Accuracy.**

At first, Attrition were assigned to all 1290 associates in the test data set. For this, a probability cutoff 0.5 was used. The model thus made, was very accurate (Accuracy = 85%), but it had a very low sensitivity (26%). Hence we created functions and iteratively found out best cut-off value of 0.1776. Thus the accuracy of the model decreased a little but sensitivity increased to 74%.
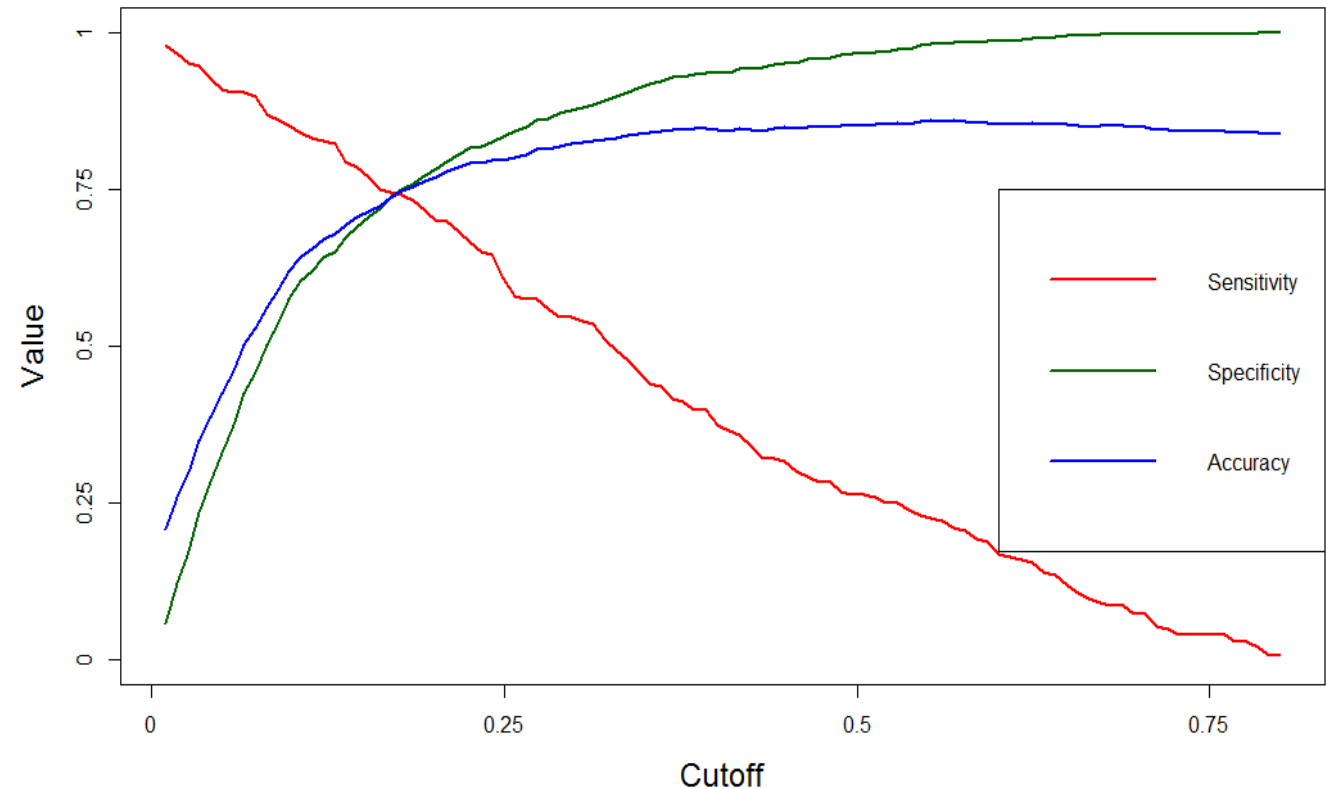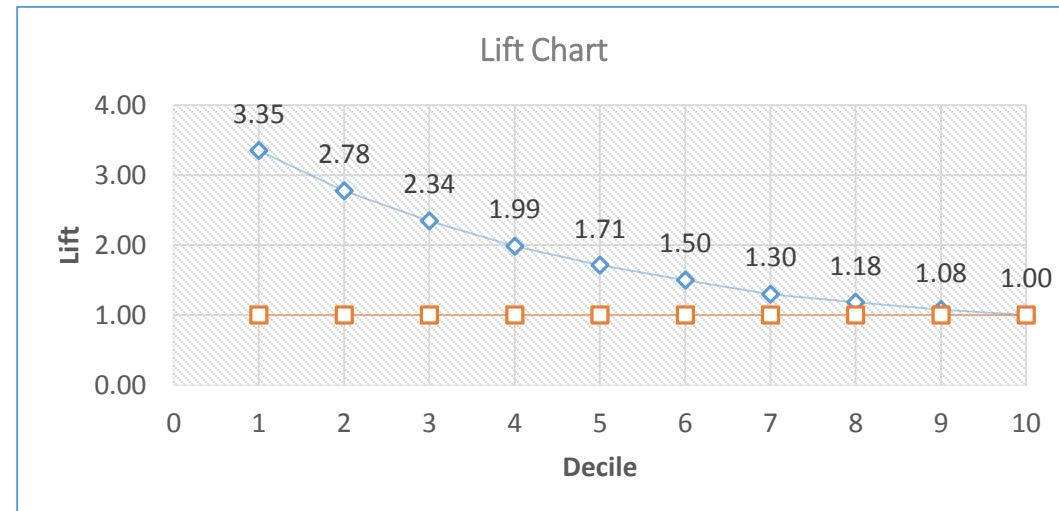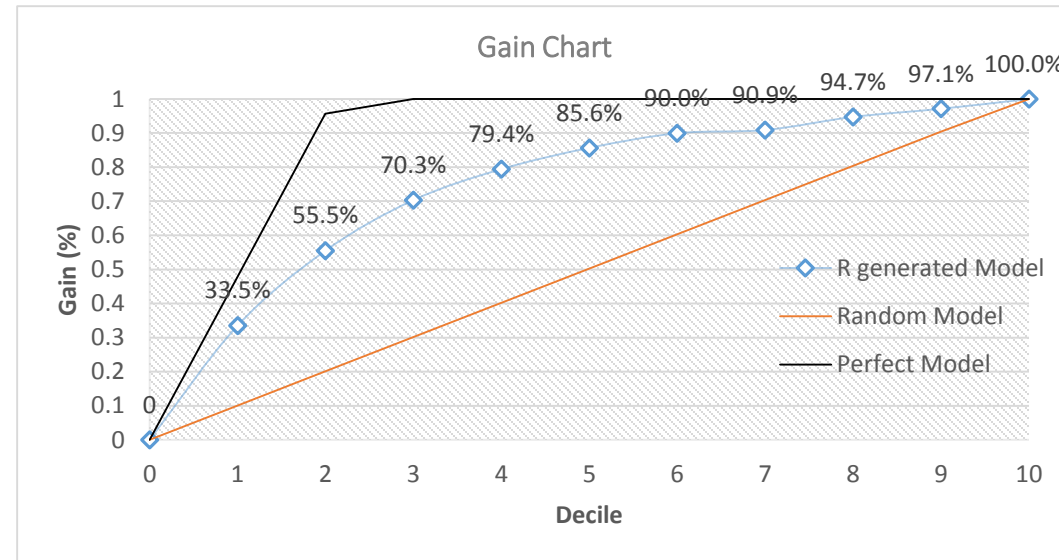
Logistic Regression Model Evaluation Cont.

2. Lift and Gain Chart

Second method is Lift and gain chart.

The gain for the final model is 79.4% by the 4th decile. It means if we sort based on probability, then 79.3% associates of top 40% associates of this sorted list, were likely to resign from the job

On the other hand, the lift is telling us the factor by which our final model is performed better than a random model. For instance, in our model's lift is equal to 1.99 by the 4rd decile, it means that our model's gain by the end of the 4rd decile is 1.99 times that of a random model's gain at the end of 4 deciles. In other words, our model catches 1.99 times more attrition than a random model would have caught.

Logistic Regression Model Evaluation Cont.

3. KS-statistic Value

In the last method we will evaluate KS –Statistic value.

Here we can find the value is more than 40% i.e. 48% for this particular model which is telling that the performance of the model is really good. A high KS statistic means that not only has all attrition at the top, it has all non-attrition at the bottom.

From all the above explained methods we can conclude that our final model has very good performance statistics i.e. our model can predict employee attrition accurately by 74%.

Decile	Observations	attrition	Cum- Attrition	% Cum-Attrition	Non- Attrition	Cum-Non-Attrition	%Cum-Non-Attrition	(%Cum-Attrition) - (%Cum-Non-Attrition
1	129	70	70	33.5%	59	59	5.5%	28.0%
2	129	46	116	55.5%	83	142	13.1%	42.4%
3	129	31	147	70.3%	98	240	22.2%	48.1%
4	129	19	166	79.4%	110	350	32.4%	47.0%
5	129	13	179	85.6%	116	466	43.1%	42.5%
6	129	9	188	90.0%	120	586	54.2%	35.7%
7	129	2	190	90.9%	127	713	66.0%	25.0%
8	129	8	198	94.7%	121	834	77.2%	17.6%
9	129	5	203	97.1%	124	958	88.6%	8.5%
10	129	6	209	100.0%	123	1081	100.0%	0.0%
Total	1290	209			1081			

Recommendations

After analyzing all the datasets using EDA and creating the model following are the recommendations to reduce attrition rate among employees of XYZ company.

- Job distribution by the team lead /manager should be evenly distributed among all the team members (average job involvement), which in turn will reduce average working hours for all the team members (considering everyone is average performer). This will help increasing job satisfaction.

- Introducing more awards, gifts and appreciation for young employees to keep them motivated.

- Single's more often leave the company, so by introducing good packages for intra company married couple may reduce attrition rate.

- Providing very good working environment to increase environmental satisfaction by introducing snacks, laptop and bean-bags or even providing gaming console for relaxation, resting room etc.

- In the initial job levels attrition level is higher, keeping the promotional cycle shorter in initial working years would decrease the attrition rate among employees.

- Frequent change of projects for a particular employee means less number of years with same manager, mostly in these cases performance evaluation and promotion get impacted. Keep team members for longer time in a particular project and motivate them with new responsibilities, recognition etc. is recommended.