HR Analytics Case Study : Employee Attrition

Submitted by :

1. Ajay Joshy
2. Asher Peter Babu
3. Jithin Thomas
4. Prakash Babu

Date: 20/08/2017

Introduction

Problem Statement

- A company XYZ with around 4000 employees has an annual attrition rate of 15%. The company wishes to understand the factors contributing to the attrition.

Goal

- To understand the factors contributing to employee attrition and to make changes to the work place to minimize it.

Analysis approach

- Based on the employee datasets provided a logistic regression model is generated for predicting the probability that an employee moves out of the organization.
- An optimum set of variables are filtered after each iteration, based on their VIF and significance rating as given by the model summary, to build the next iteration of the logistic regression models.
- The final set of variables used to build the final model are the most influential factors causing employee attrition

Why Reduce attrition?

- The cost of turnover. Expenses involving talent acquisition and training the employee.
- The loss of company knowledge, trade secrets
- Interruption in the customer service, delay in projects
- Loss of reputation and goodwill of the company due to missing deadlines
- Effort and resources spent on regaining the efficiency

Data Understanding and Presentation

Understanding the data

Three categories of data is provided
1. **Demographic Information -** Includes the personal attributes of the employee such as age, marital status, educational qualifications, experience etc

2. **Survey results of the employee** – Information sourced from the employees through surveys like environment satisfaction, work life balance

3**. Employee Ratings** – Performance rating, job involvement rating, salary hike percentage etc

These information is merged with EmployeeID as key.
The merged data frame has 4410 observations with 30 variables each

…Cont

Preparing the data

- **Missing values imputation**
  NAs are present in EnvironmentSatisfaction,JobSatisfaction,WorkLifeBalance,NumCompaniesWorked, TotalWorkingYears column, The missing values are imputed with the respective mode values

- **Change categorical variables to more meaningful format**

employee$EnvironmentSatisfaction : 1-low, 2-Medium, 3-High, 4-Very high
employee$JobSatisfaction : 1-low, 2-Medium, 3-High, 4-Very high
employee$WorkLifeBalance : 1-Bad, 2-Good, 3-Better, 4-Best
employee$JobInvolvement : 1-Low, 2-Medium, 3-High, 4-Very High
employee$PerformanceRating : 1-Low, 2-Good, 3-High, 4-Very High
employee$Education : 1-Below College, 2-College, 3-Bachelor, 4-Master. >4 - Doctor
employee$JobLevel : 1-JLevel1, 2-JLevel2, 3-JLevel3, 4-JLevel4, 5-JLevel5
employee$StockOptionLevel : 0-SOLevel0, 1-SOLevel1, 2-SOLevel2, 3-SOLevel3
employee$Attrition : "Yes"-1, "No"-0

...Cont

Preparing the data

- **Outlier treatment**

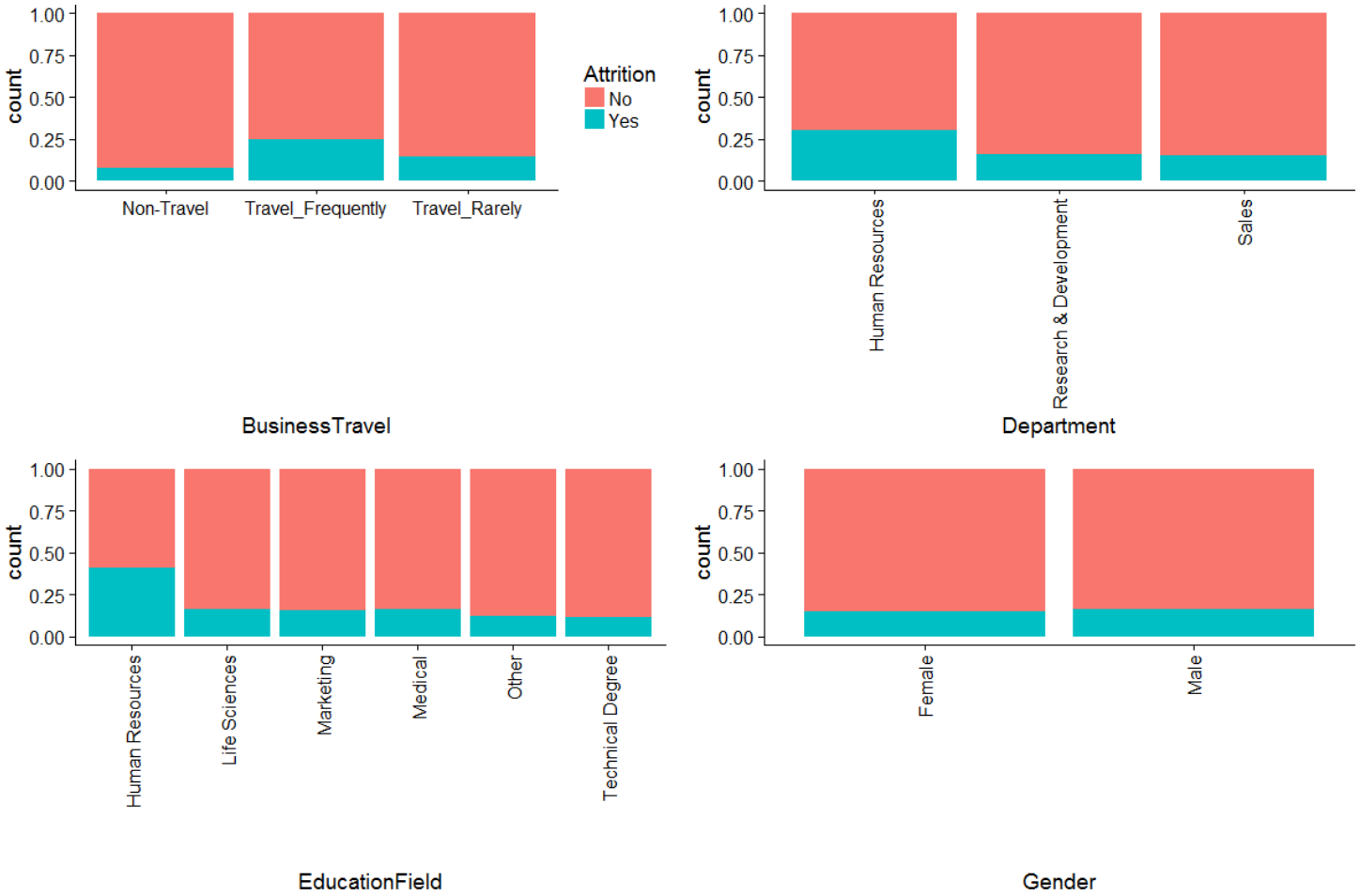
  Box plots stats are used to check for outlier in numeric variables
  Outliers are capped at 2% and 95% if there is a sudden jump in the values

  Following variables are treated for outliers :

"DistanceFromHome","MonthlyIncome","NumCompaniesWorked","PercentSalaryHike","StandardHours","TotalWorkingYears","TrainingTimesLastYear","YearsAtCompany","YearsSinceLastPromotion","YearsWithCurrManager","timedf"

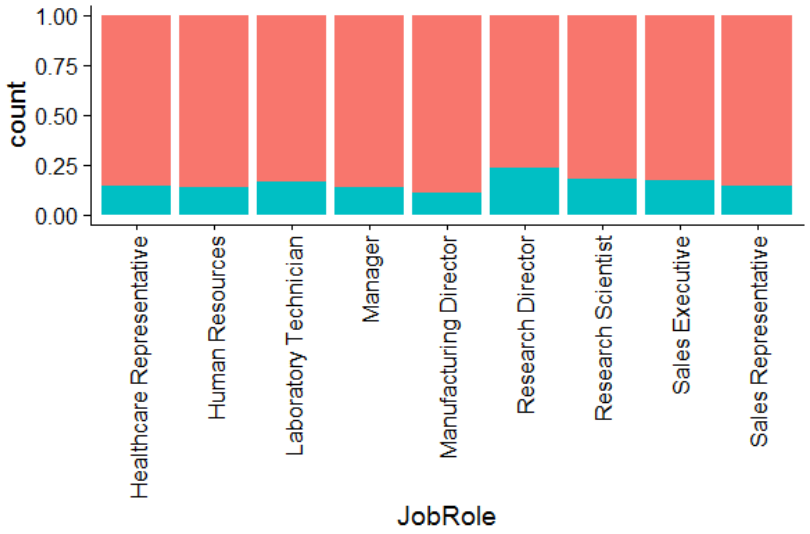
- **Feature standardisation**
  The numerical variables are normalized to prevent the different scales of values from skewing the prediction model
- **Dummy variables**
  All the categorical variables of the employee data frame are factorised and dummy variables created for them. These dummy variables are merged with the final employee frame.
  70% of the dataset is used for training the model and 30% for testing
- **Derived variables**
  timedf – difference between in_time and out_time
  avg_working_hrs – daily average working hours
  total_leaves_taken – the total number of leaves taken by the employee

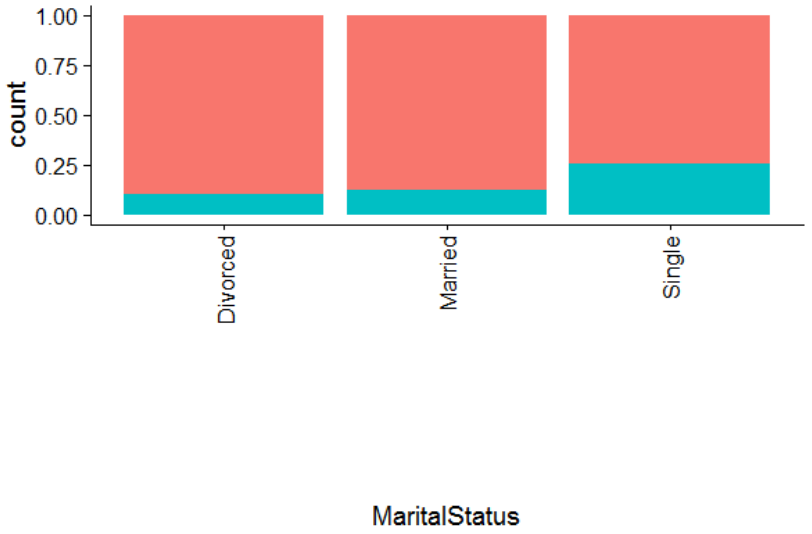
EDA-categorical

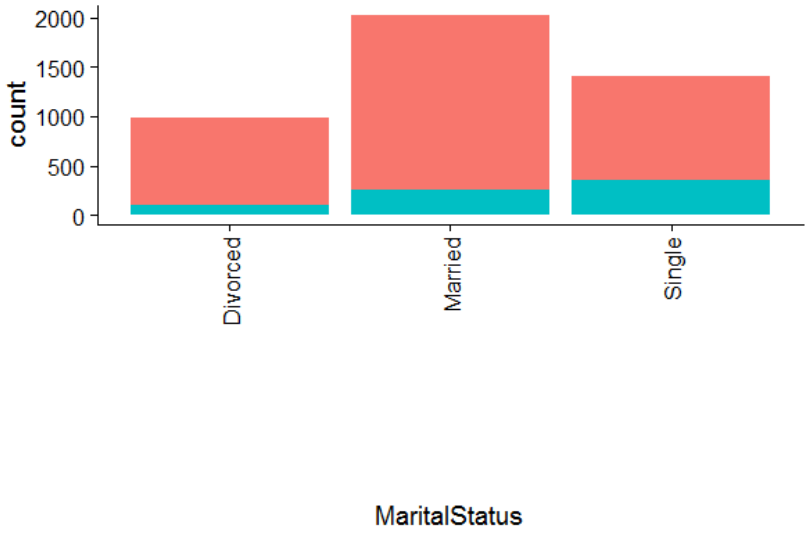


- Employees who travel frequent are more prone to leave the organization
- HR department sees the highest proportion of attrition
- The employees who have studied human resources shows a higher trend of leaving the company
- The gender wise comparison shows that there is no real disparity in the percentage of women/men leaving the organization

EDA-categorical

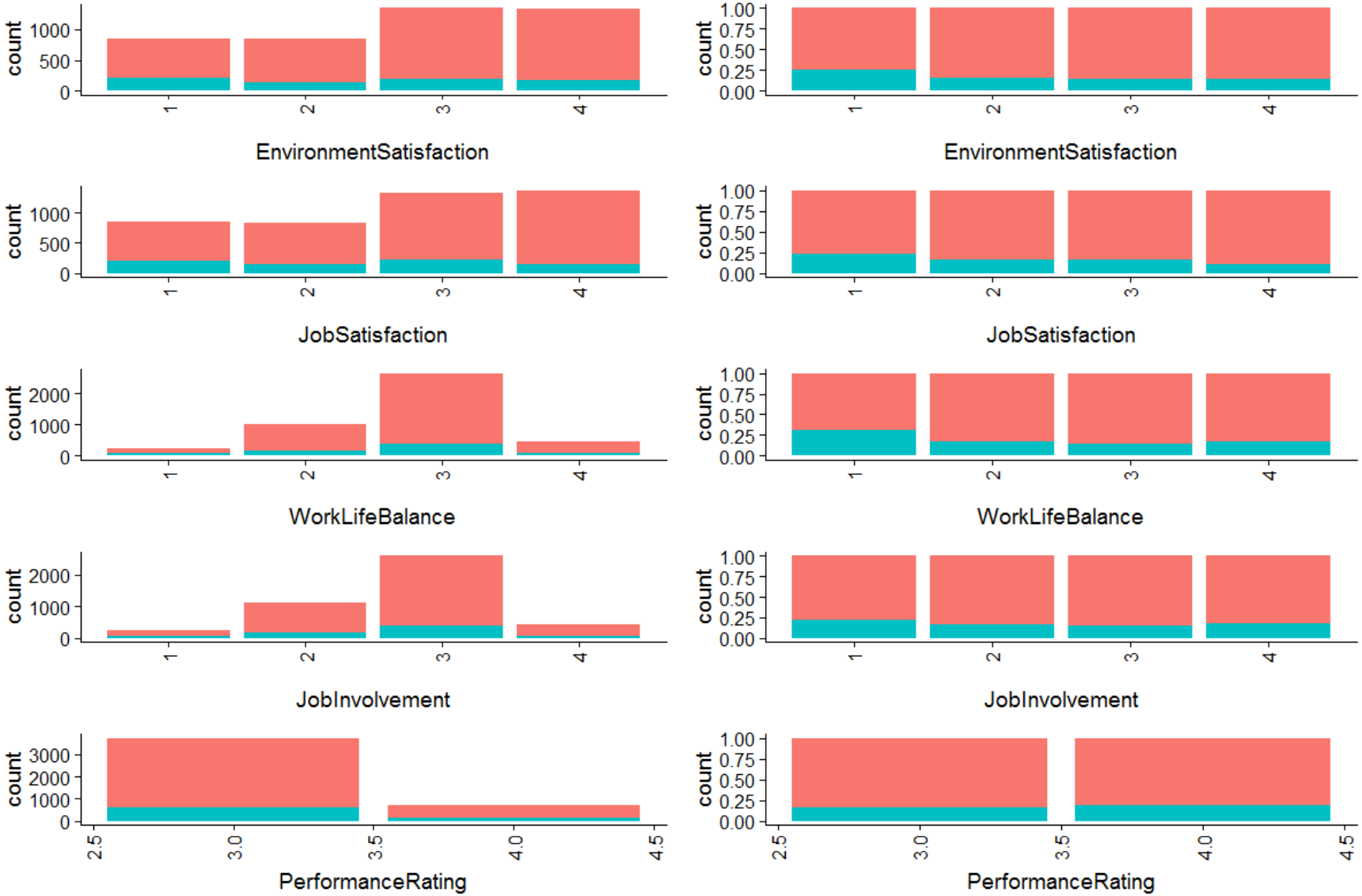






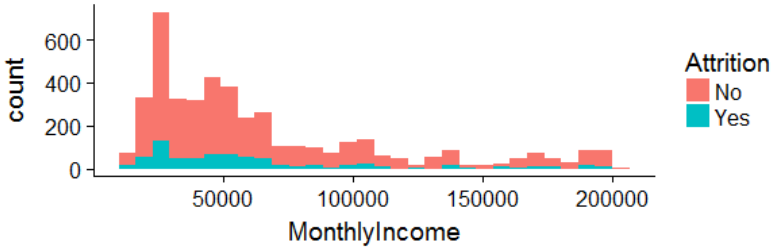
- Attrition rate among single employees are higher compared to other categories
- Manufacturing director and manager are the job roles with least attrition rates
- Attrition rate is significant in research jobs like Research director and Research scientist

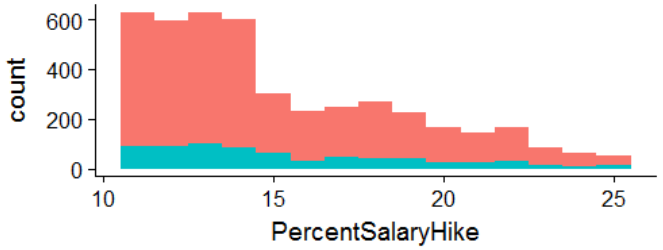
EDA-categorical

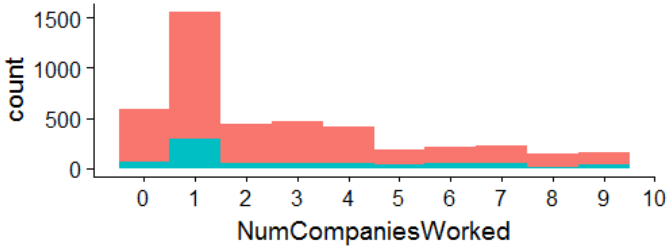


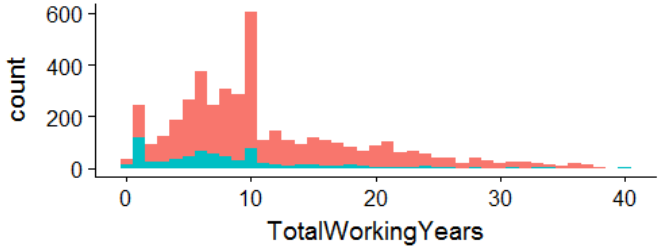
- Employees who have given 1 rating seems to leave more
- Attrition rate is highest amongst employees who have rated job satisfaction as 1
- Over 25% of employees who have given least rating for work life balance leaves
- Job involvement seems to have very similar trends across all categories
- Majority of the employees leaving the organization has a performance rating < 3.5

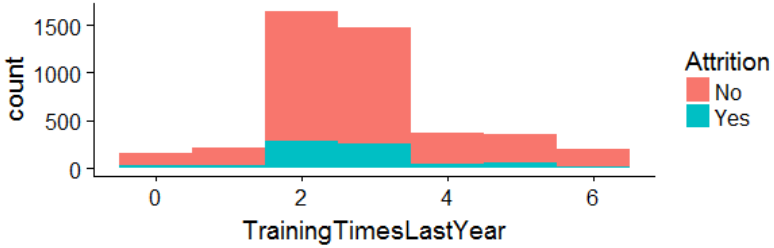
EDA -numerical

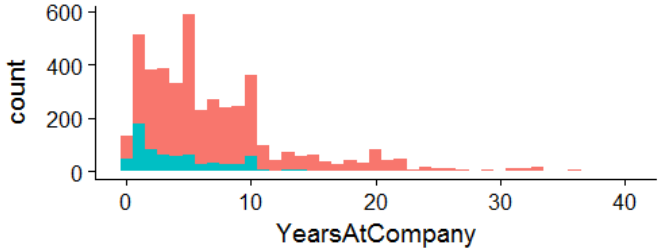


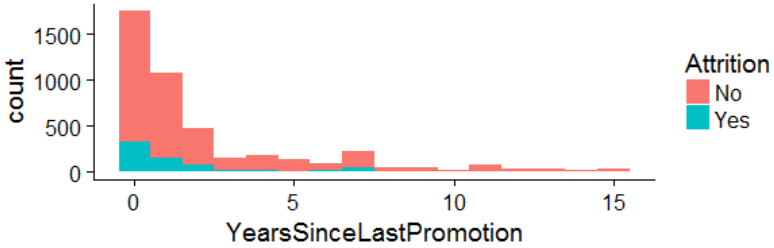


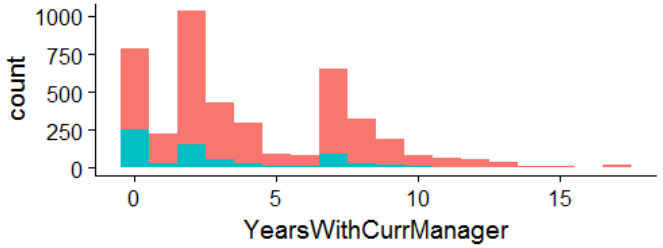












- **Monthly Income** – bulk of employees in the company have a payscale < 70000, max attrition count is observed in this range though the proportion is same throughout
- **NumCompaniesWorked** - Employees who have worked previously with 2-4 companies have the highest probability to stay
- **Training** – The more the number of training is given the higher the chances to stay back
- **PercentSalaryHike** – The trend is more or less steady but employees with the highest hike seems to show a greater rate of attrition
- **TotalWorkingYears** - Greater the working years the more likely he is to stay
- **YearsAtCompany** - Follows similar trend as TotalWorkingYears
- **YearsWithCurrManager** – The highest percentage of employees leave before the initial year with the current manager

Logistic Regression

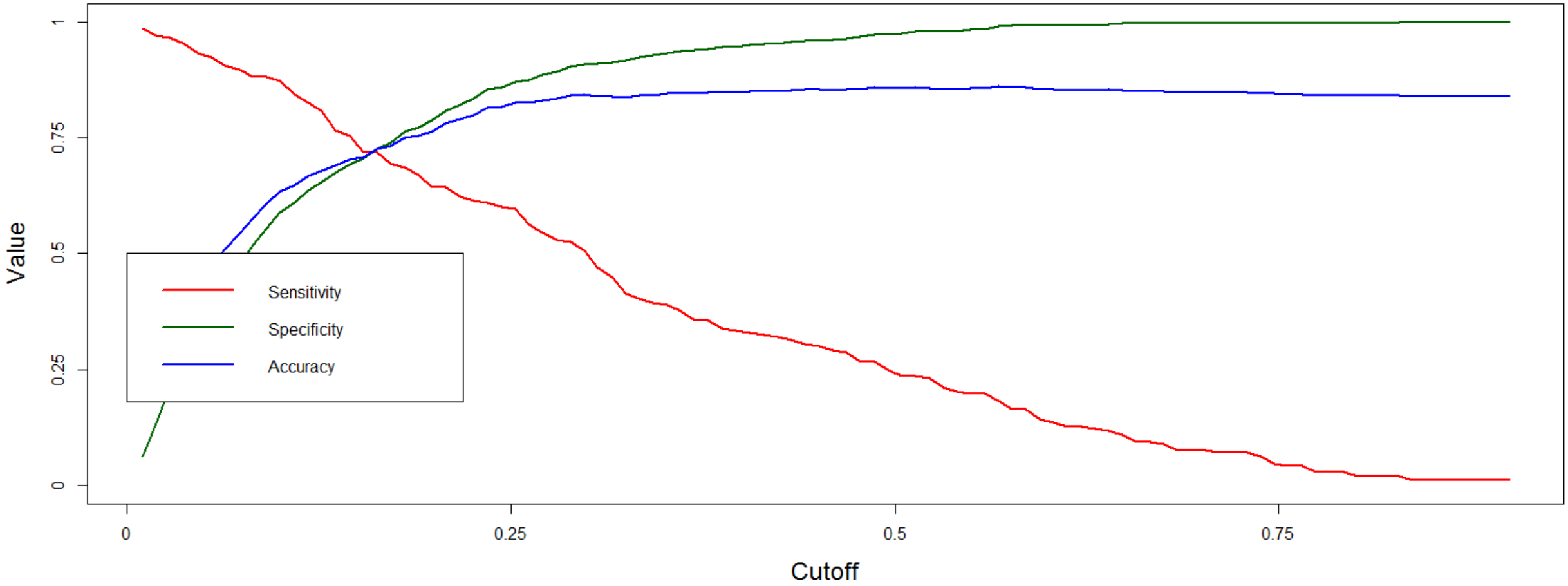
- Logistic regression is performed by fitting a Generalized Linear Model with family as binomial
- Multi collinear variables were removed using the step AIC function and by considering vif values > 2 of lesser significant variables.
- 17 models were trained until all insignificant and highly collinear variables are removed. Following are the variables in the final model

1. TrainingTimesLastYear : Number of trainings received in the previous year. Coeff : -0.20422
2. BusinessTravel.xTravel_Rarely : Indicates whether business travel is rare or not. Coeff : 1.11153
3. JobSatisfaction.xLow : Indicates whether job satisfaction is low or not. Coeff : 0.51695
4. Age : Age of the employee. Coeff : -0.33233
5. JobSatisfaction.xVery.High : Indicates whether job satisfaction is high or not. Coeff : -0.63670

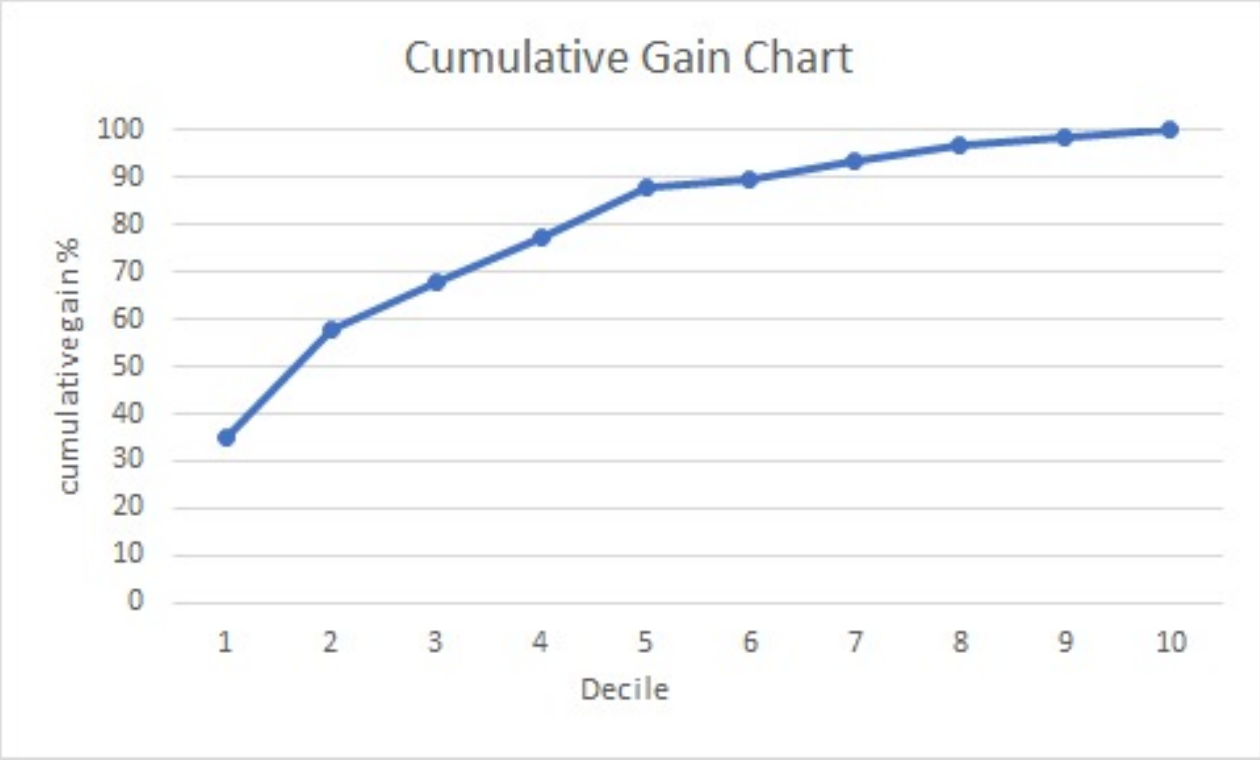
…Cont

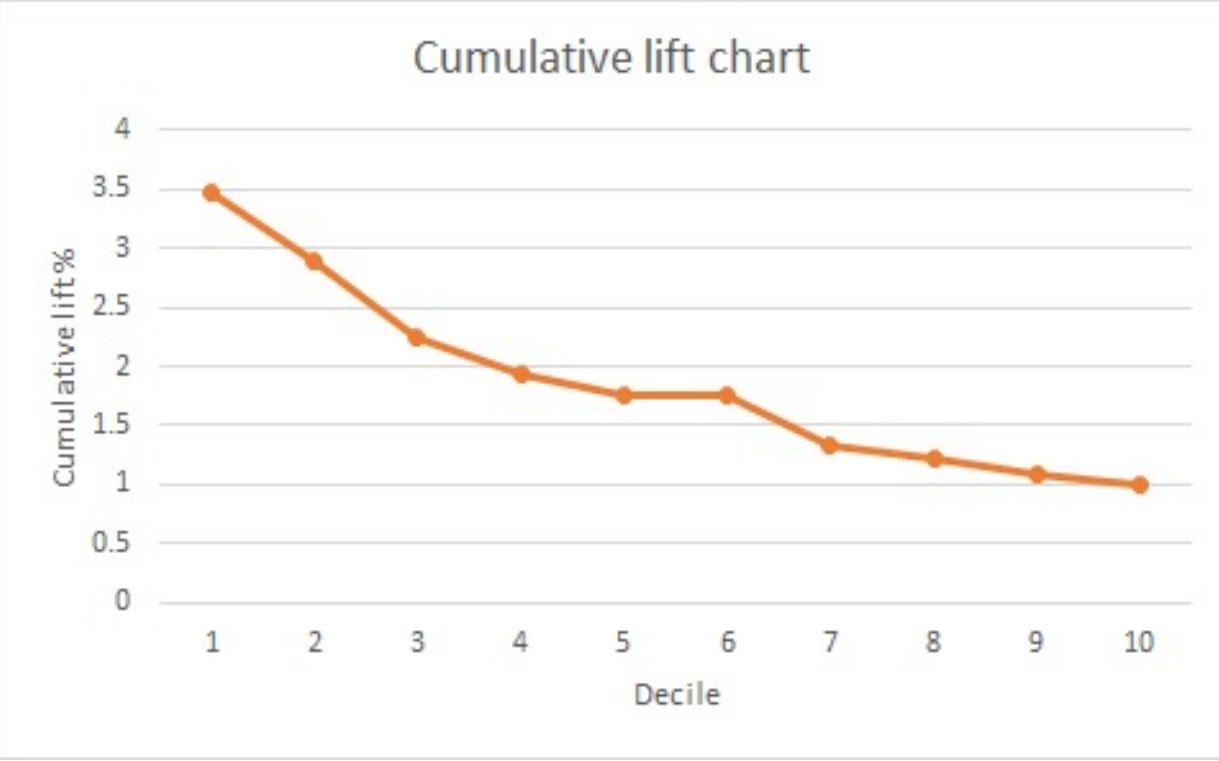
6. Department.xResearch...Development : Indicates whether employee's department is R & D. Coeff : -1.15065
7. Department.xSales : Indicates whether employee's department is sales. Coeff : -1.21833
8. YearsWithCurrManager : Number of years under the current manager. Coeff : -0.46373
9. TotalWorkingYears : Total years of work experience. Coeff : -0.50607
10. WorkLifeBalance.xBetter : Indicates whether or not the work life balance is better. Coeff : -0.37680
11. YearsSinceLastPromotion : Number of years since last promoted. Coeff : 0.47935
12. EnvironmentSatisfaction.xLow : Indicates if satisfied with the work environment or not. Coeff : 1.06251
13. NumCompaniesWorked : Number of companies previously worked with. Coeff : 0.38750
14. BusinessTravel.xTravel_Frequently : Indicates whether business travel is frequent or not. Coeff : 1.83336
15. MaritalStatus.xSingle : Whether or not the employee is single. Coeff : 1.01687
16. avg_working_hrs : Average working hours per day. Coeff : 0.52644

Model cut off value = 0.1628283



Gain and lift Charts





- Top 3 deciles will cover nearly 70 % of the employees who will leave the company
- Top 2 deciles - 57.74648
- Top 3 deciles - 67.60563

- Top 2 deciles have a lift of 2.887324
- So the top 2 deciles will contain 2.88 times the number of employees who would leave as compared to the top 2 deciles of a dataset without a model

Model Summary

- Null deviance: 2728.0 on 3086 degrees of freedom
- Residual deviance: 2123.5 on 3069 degrees of freedom
- AIC: 2159.5
- Cut off probability for attrition to happen is set at 16.28%.
- Confusion matrix for the testing data set with 1323 observations

Attrition	Actual No	Actual Yes
Predict No	807	60
Predict Yes	303	153

- Confusion Matrix Statistics

Accuracy	Sensitivity	Specificity
72.56%	71.83%	72.70%

- Using this model around 70% of employees who has a high chance of leaving the company can be identified
- KS Test measures to check whether model is able to separate events and non-events. In probability of attrition(employee attrition) model, it checks whether the attrition model is able to distinguish between employees going to stay and leave the organization.
  In our case, KS score is 44.45

Note:
The coefficient estimate corresponding to an independent variable in the model summary is the value by which "log of odds of attrition" being equal to 1 will change. If the value is positive then the odds of attrition increase otherwise the odds reduce

Decreases Attr		Increases Attr	
Department.xSales	-1.21833	BusinessTravel.xTravel_Frequently	1.83336
Department.xResearch...Development	-1.15065	BusinessTravel.xTravel_Rarely	1.11153
JobSatisfaction.xVery.High	-0.6367	EnvironmentSatisfaction.xLow	1.06251
TotalWorkingYears	-0.5060	MaritalStatus.xSingle	1.01687
YearsWithCurrManager	-0.4637	avg_working_hrs	0.52644
WorkLifeBalance.xBetter	-0.37680	JobSatisfaction.xLow	0.51695
Age	-0.33233	YearsSinceLastPromotion	0.47935
TrainingTimesLastYear	-0.20422	NumCompaniesWorked	0.38750

Conclusion

Based on the insights from the EDA and the regression model following are the suggestions to reduce the attrition rate in the company:

Dont’s

1. Business travel needs to be moderated as necessary. As too much travel, as well as not providing enough travel opportunities contributes to attrition.
2. Work environment should be good as it is a deciding factor for leaving the company. The shared assumptions, values, and beliefs which governs how employees behave in an organization must be set as per employee demographic and company values.
3. Higher average daily working hours may cause employee fatigue leading the employee to contemplate quitting the job, hence paid vacations, mandatory holidays, less overtime working must be encouraged to keep the employee in a productive state.
4. Job satisfaction is a factor that is integral to the employees self esteem and happiness. This may be dependent of many factors like nature of the job, job recognition, compensation etc. The company must help the employee to fulfil this need if he is to stay longer. Low job satisfaction has a positive effect on attrition while high job satisfaction has a negative effect, hence this factor must be given due consideration
5. The more years an employee goes without promotion the more likely he is to quit. So he is deserving the company must consider promoting him. Also the company needs to look at restructuring the current career path to ensure employees are sufficiently motivated
6. The company must be careful of hiring people who have worked in > 3-4 companies as they show a tendency to leave often

Do’s

1. Employees in the Sales and R&D departments tend to stay longer
2. If the employee works under the same manager for a longer period, they are more likely to stay. So avoid frequent manager switches unless requiredAlso few of the managers under whom employees quit within a year or so, needs to be analysed and provided with necessary managerial skills.
3. Good work life balance must be encouraged. Frequent late night meetings, working on weekends, late nights must be discouraged. Proper roadmap and planning can help reduce these issue. There are many development methodologies like Agile that are followed to address these concerns
4. The older the employee the more likely he is to stay. So while hiring younger employees a certain degree of attrition is to be expectedFactors needed to retain younger employees need to be further analysed providing necessary challenges, growth and motivation to continue to stay with the organization
5. Giving adequate training every year has a positive effect on employee’s to stay back. Encourage trainings and skill development sessions.
6. The longer the employee stays with a company the more likely they are to stay longer. This is a positive effect that helps cut attrition rate further if the company implements other measures to reduce attrition