

HR Analytics Case Study

Yogesh Shinde

Abstract

Background

- A large company named XYZ, with 4000+ employee strength has around 15% attrition rate. 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 -
 1. The former employees' projects get delayed, which makes it difficult to meet timelines, resulting in a reputation loss among consumers and partners
 2. A sizeable department has to be maintained, for the purposes of recruiting new talent
 3. More often than not, the new employees have to be trained for the job and/or given time to acclimatize themselves to the company

Business Objective

- Analyze the sample data provided to understand key variables that affect attrition and identify key factors contributing towards higher attrition. The results thus obtained will help management to understand what changes they should make to their workplace, in order to get most of their employees to stay

Data Understanding & Cleaning

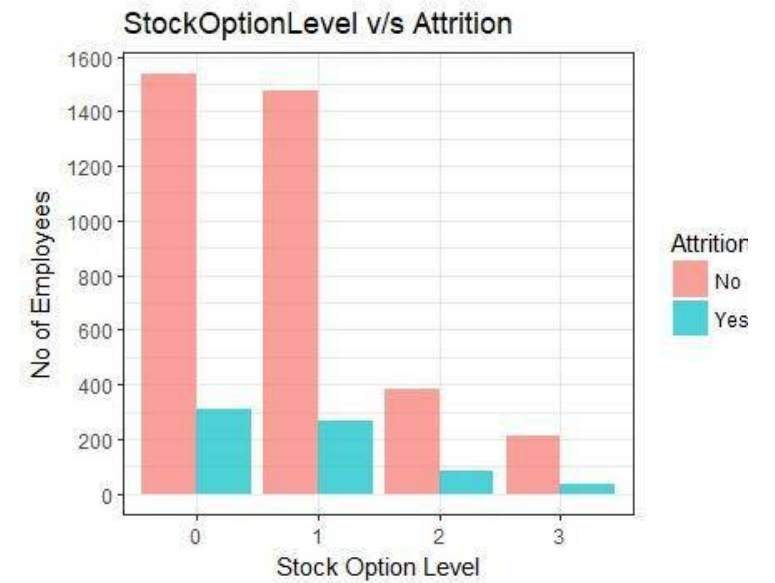
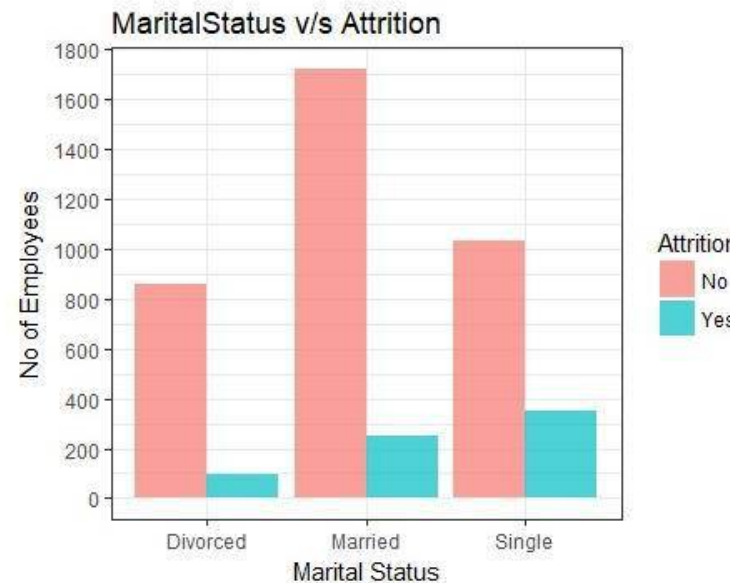
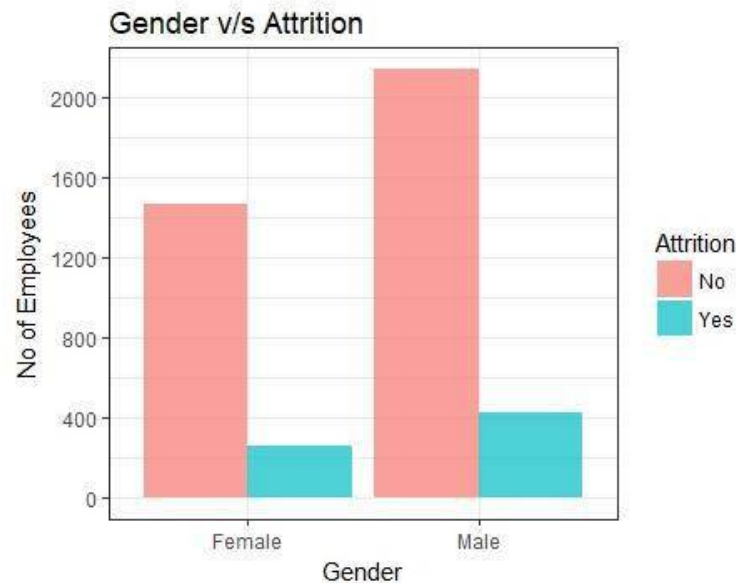
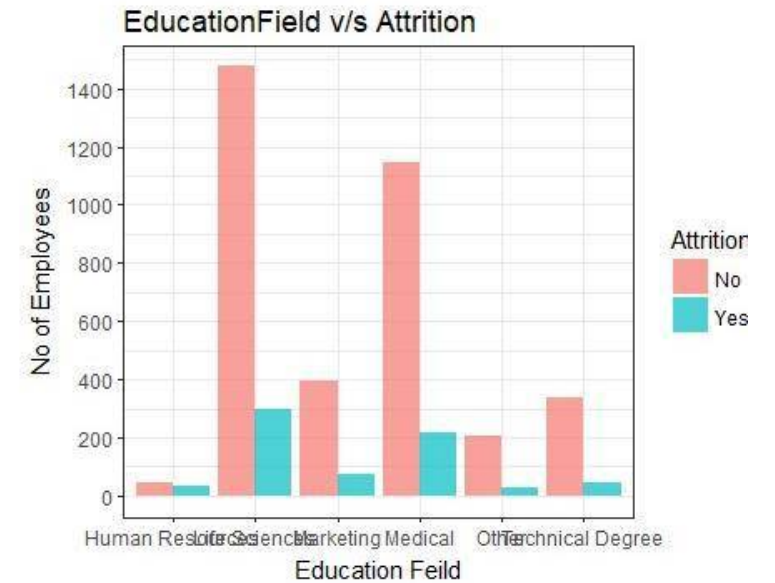
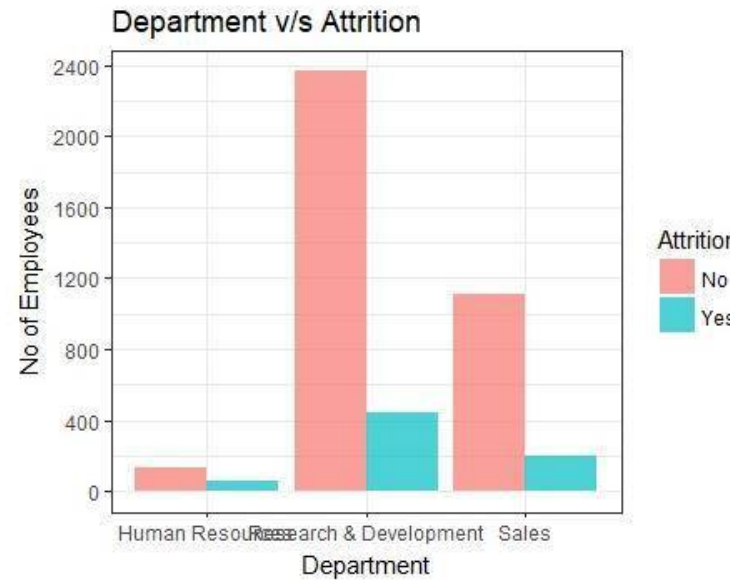
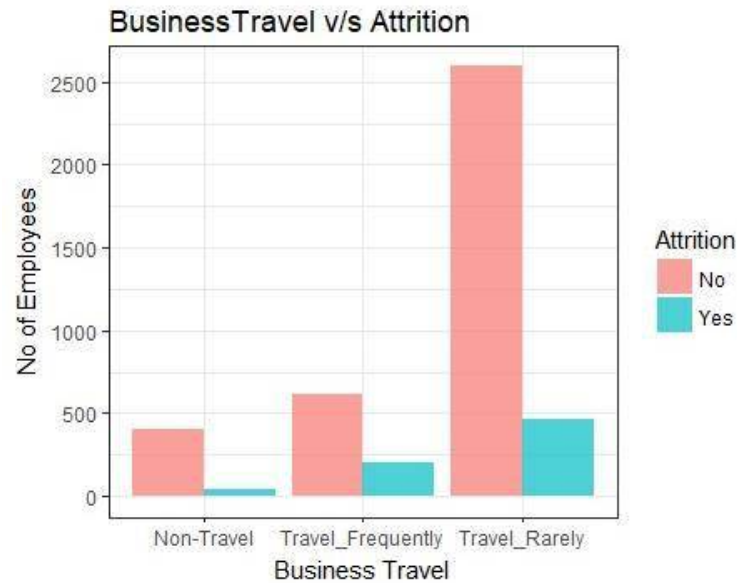
The Meta Data consists of 5 Data files

- ❖ Employee survey data – Consists of behavioral data of employee for e.g. Environmental and job satisfaction and work life balance
- ❖ General data – This datasets consists of general data of the employees like age, gender, department, education, attrition etc. Nearly 16% of employees in dataset (711 out 4410 employees) left the company
- ❖ Manager survey data – Consists feedback of job involvement, performance rating.
- ❖ In time – Contains In time of all the employees for a period of 1 year from 1st January 2015 to 31st December 2015
- ❖ Out time - Contains Out time of all the employees for a period of 1 year from 1st January 2015 to 31st December 2015

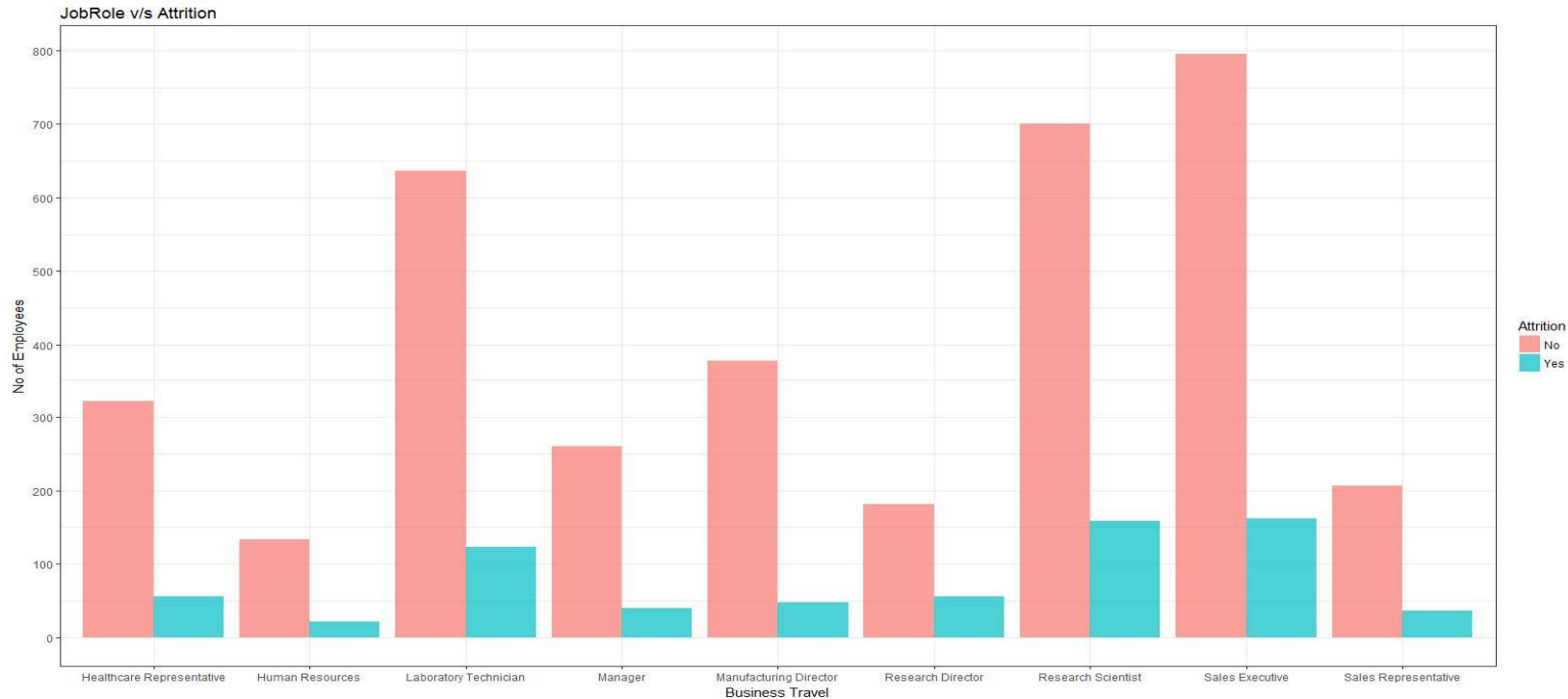
Data Cleaning

- ❖ Removed all the columns having only NA values in in & out data frames
- ❖ Created derived variables
- ❖ Replaced / remove NA values with appropriate substitute values

EDA – Univariate & Bivariate Analysis (Categorical Variables)

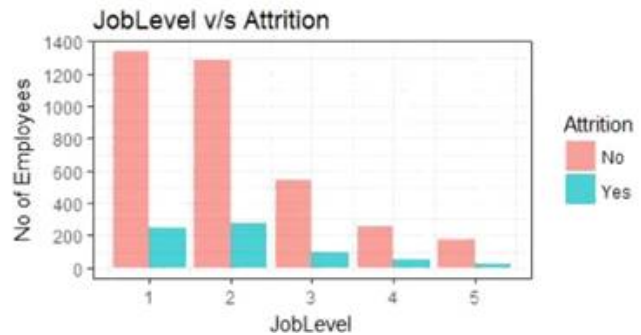
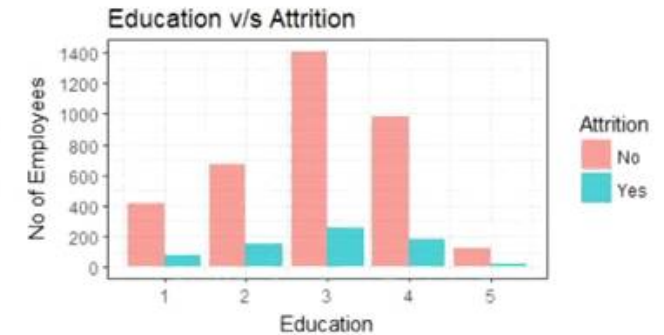
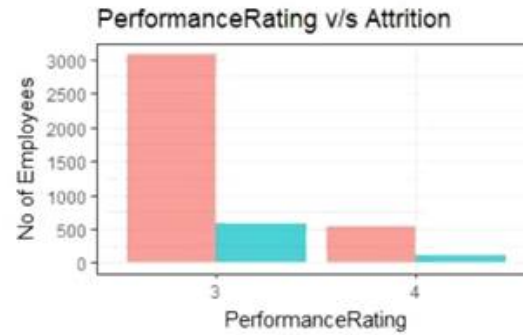
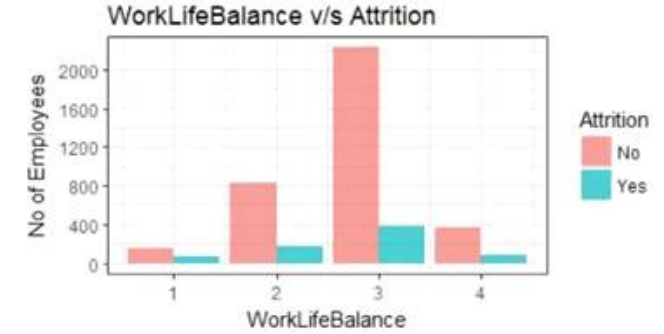
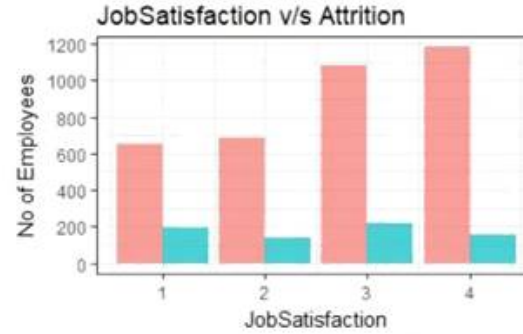
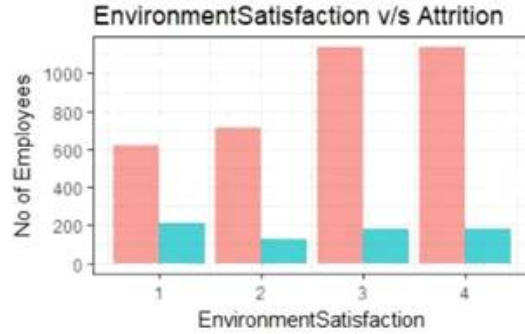


EDA – Univariate & Bivariate Analysis (Categorical Variables)



- ❖ Frequent business travel have higher attrition
- ❖ Research and development department has more attrition, HR department has maximum percent of attrition
- ❖ Education field "Life Science has highest attrition, HR is prone to highest percent of attrition
- ❖ Male attrition is higher than female
- ❖ Employees with marital status as Single have higher attrition
- ❖ Attrition of Sales Executive is highest in job roles, Research and development has highest percent of attrition

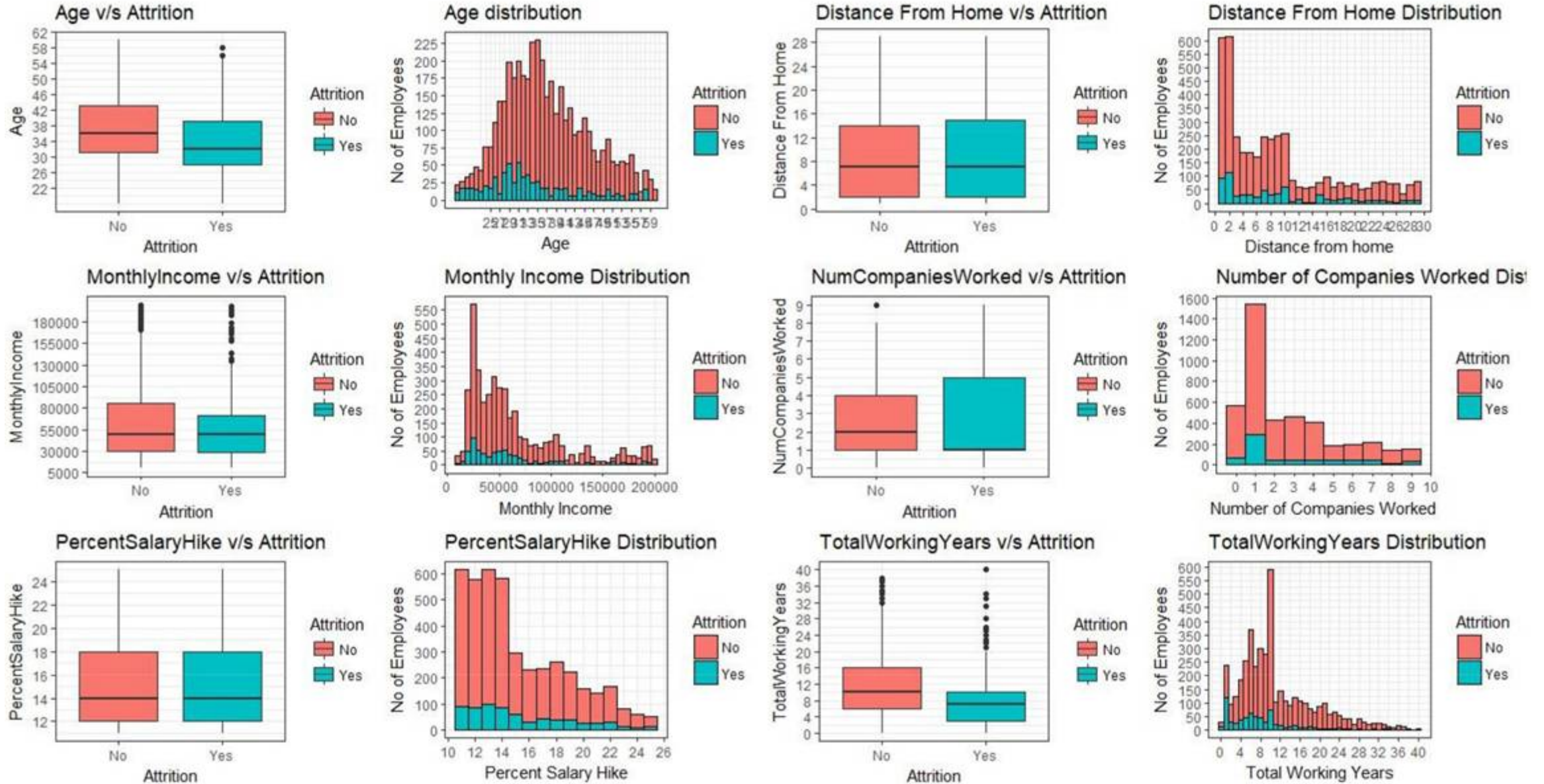
EDA – Univariate & Bivariate Analysis (Numeric Variables)



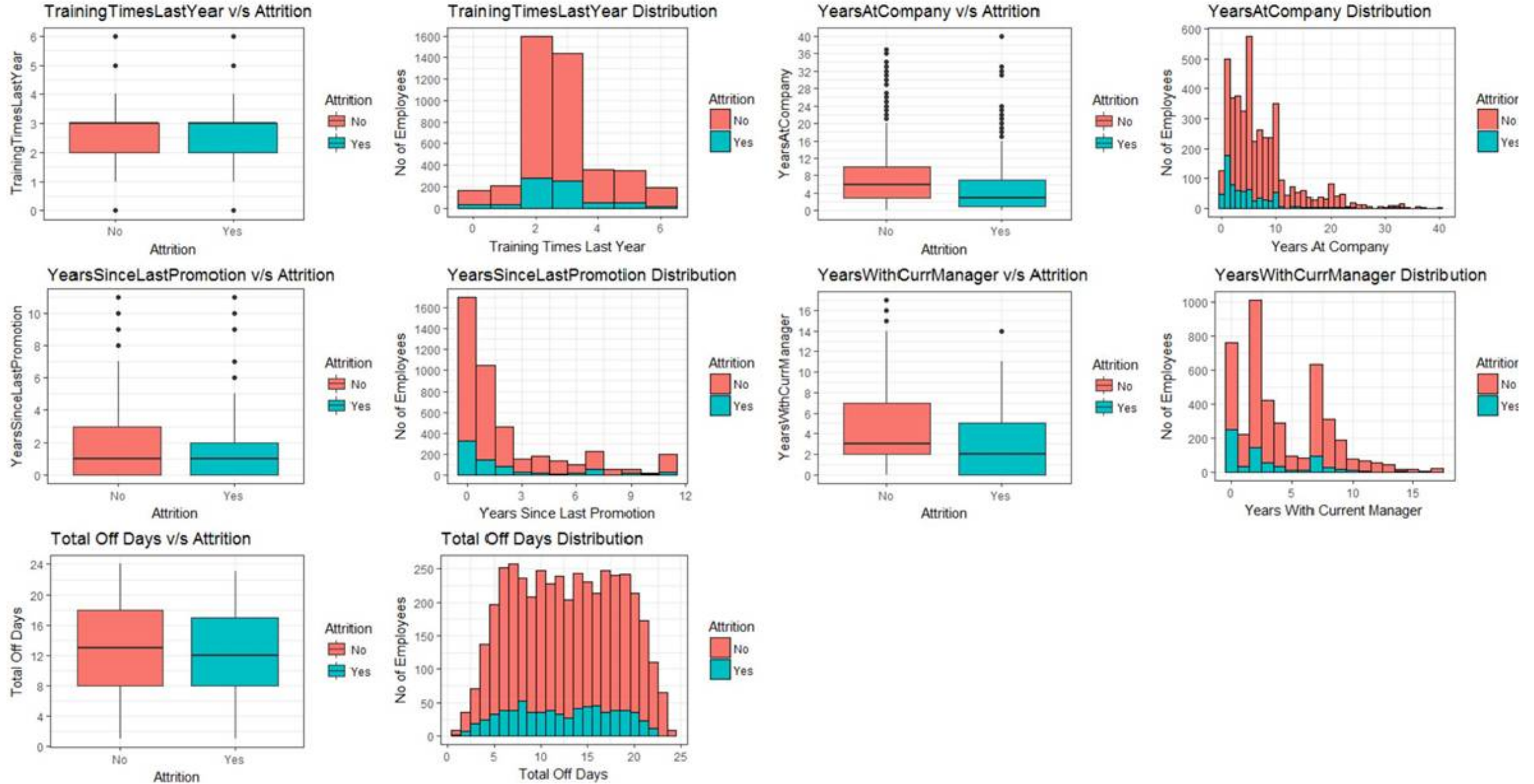
EDA – Univariate & Bivariate Analysis (Numeric Variables)

- ❖ Low work environment satisfaction leads to higher attrition
- ❖ Job Satisfaction has high impact on attrition, percent attrition decreases as job satisfaction increases
- ❖ Similarly work-life balance has high influence on attrition, lower the work-life balance higher is the attrition
- ❖ Performance rating, Education, Job level doesn't have significant influence on attrition

EDA – Univariate & Bivariate Analysis (Continuous Variables)



EDA – Univariate & Bivariate Analysis (Continuous Variables)



EDA – Univariate & Bivariate Analysis (Continuous Variables)

- ❖ Employees with age<32 have higher rate of attrition
- ❖ Majority of employees stay closer and attrition rate among them is also higher
- ❖ Lower the Monthly Income higher is the attrition rate
- ❖ Very high attrition rate with freshers (total experience =1)
- ❖ High attrition rate with employees who have spent 2 years with company
- ❖ Employees with 2 & 3 trainings have highest rate of attrition
- ❖ Years at company have high outliers and needs to be treated
- ❖ Attrition rate is higher among the employees who are promoted recently

EDA – Data Preparation

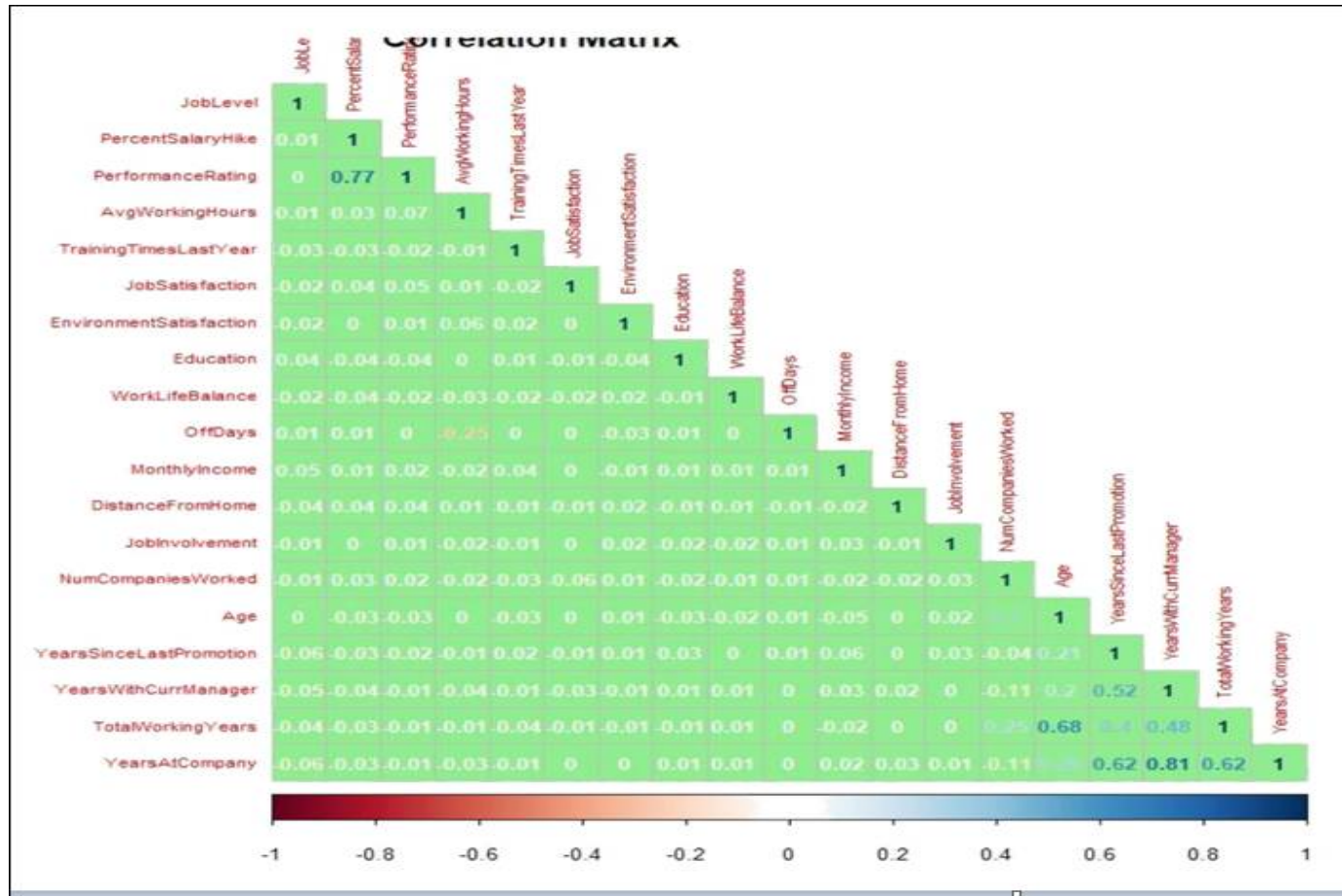
Outliers Treatment

- ❖ “Monthly Income” had 6.5% outliers on high income side which were addressed
- ❖ “Total Working Years” outliers with high values were addressed without impacting attrition as most of the employee records are on lower side
- ❖ Around 2.6% of outliers for “Years at Company” were addressed
- ❖ There were 2.9% of outliers for “Years Since Last Promotion” outliers all on higher side years that were addressed

Preparing data for applying Model

- ❖ In this stage, prepared the input variables which were to be considered as an input to the model
- ❖ Normalized the continuous variables, also scaled them using scale function in R
- ❖ During pre-modelling, our target variable is Attrition which is categorical in nature and hence we used logistic regression model
- ❖ We created dummies for categorical variable as part of variable reduction technique
- ❖ We split the data into train and test data in ratio - 7:3

Correlation between Variables



- Years at company and Years with current manager are highly correlated (0.81)
- Years at company and Years since last promotion and Total working years are highly correlated
- Total working years and age are highly correlated

Model Building

- ❖ We used generalized linear model to build our model to predict the attrition
- ❖ We used step AIC to get the starting model. Here we are predicting how Attrition behaves with respect to other variables. Since step AIC suggests the optimal model based on Akaike Information Criteria (model having lowest AIC value is preferred), there are some insignificant variables included in the model
- ❖ We used VIF (Variable Inflation Factor) to check the correlation between the variables
- ❖ Finally we used p-values to remove insignificant variables until all the left variables had very less p-value

Model Building: Final Model

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9318	-0.5665	-0.3463	-0.1641	3.7436

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.65714	0.09981	-26.621	< 2e-16	***
EnvironmentSatisfaction	-0.38457	0.05564	-6.912	4.79e-12	***
JobSatisfaction	-0.39961	0.05669	-7.050	1.79e-12	***
WorkLifeBalance	-0.25130	0.05475	-4.590	4.43e-06	***
Age	-0.27289	0.07887	-3.460	0.000541	***
NumCompaniesWorked	0.37553	0.05772	6.506	7.74e-11	***
TotalWorkingYears	-0.66086	0.10504	-6.292	3.14e-10	***
TrainingTimesLastYear	-0.21490	0.05795	-3.708	0.000209	***
YearsSinceLastPromotion	0.64065	0.07863	8.147	3.72e-16	***
YearsWithCurrManager	-0.56046	0.09281	-6.039	1.55e-09	***
AvgWorkingHours	0.58916	0.05430	10.850	< 2e-16	***
BusinessTravel.xTravel_Frequently	0.78866	0.13043	6.046	1.48e-09	***
JobRole.xManufacturing.Director	-0.94578	0.22317	-4.238	2.26e-05	***
MaritalStatus.xSingle	1.04290	0.11574	9.011	< 2e-16	***

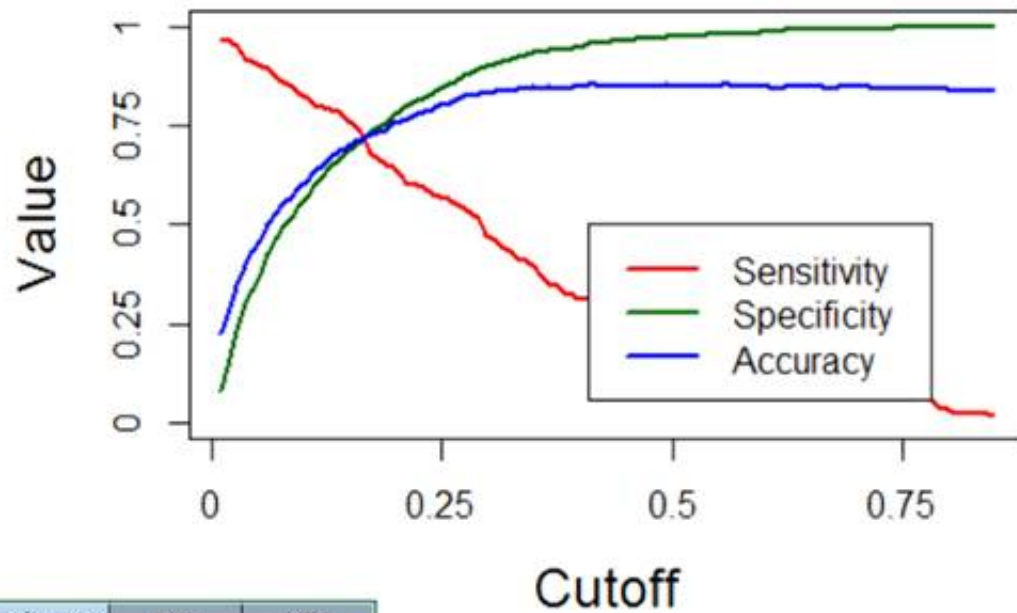
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

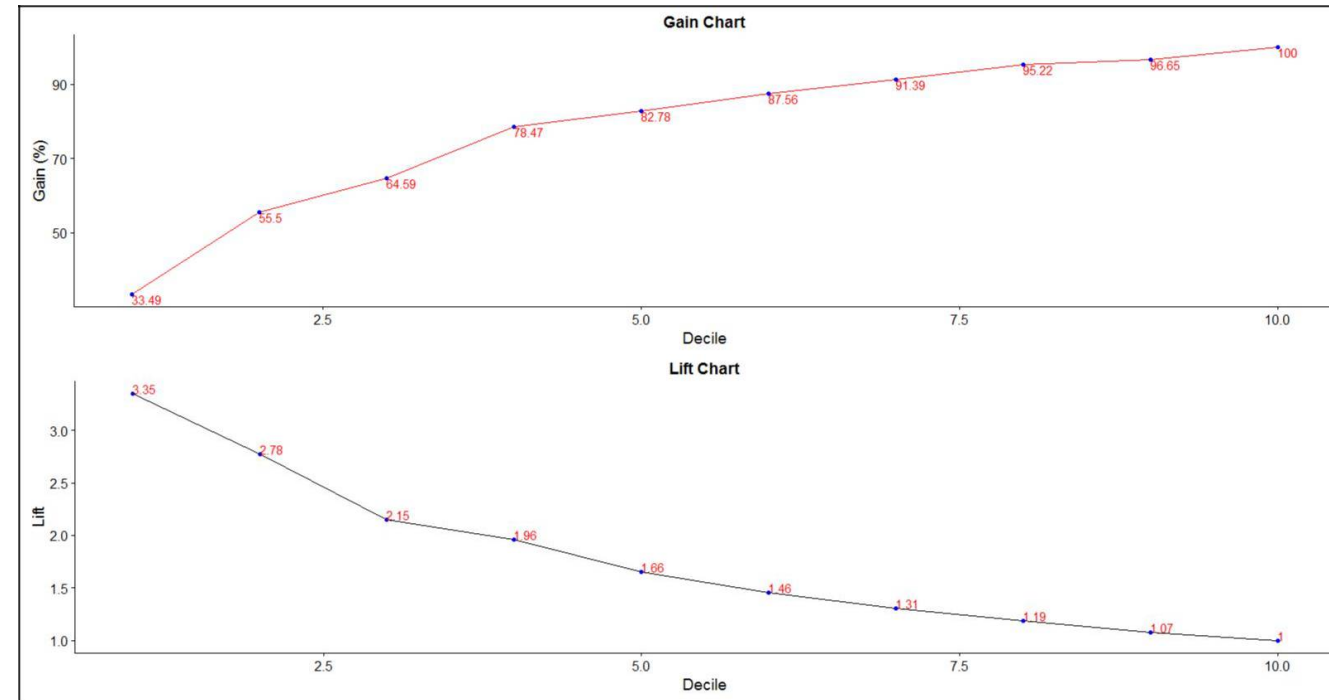
Null deviance: 2670.1 on 3015 degrees of freedom
Residual deviance: 2076.7 on 3002 degrees of freedom
AIC: 2104.7

❖ Final Model was identified after 14 iterations

Model Evaluation



Prediction	NO	YES
NO	1059	168
YES	24	41



- ❖ Executed the final model against the test dataset
- ❖ Performed the following model validations
 - ❑ Finding Accuracy, Specificity and Sensitivity through Confusion Matrix
 - ❑ To find a suitable probability cut-off, we checked the Accuracy, Sensitivity & Specificity for 1% to 85% probability
 - ❑ The optimum cut-off probability is the one where the value of specificity and sensitivity are close to each other.
 - Cut-off :: 0.162723
 - Sensitivity :: 0.7320574
 - Accuracy :: 0.7167183
 - Specificity :: 0.7137581

Impact on Attrition - Variables from Final Model

Variable	Coefficient	Impact on Attrition
EnvironmentSatisfaction	-0.38457	Negative coefficient : Higher the Environment Satisfaction --> Lesser is attrition probability
JobSatisfaction	-0.39961	Negative coefficient : Higher is the JobSatisfaction --> Lesser is the attrition probability
WorkLifeBalance	-0.2513	Negative coefficient : Higher is Work-Life Balance --> Lesser is the attrition probability
Age	-0.27289	Negative coefficient: Higher the Age --> Lesser attrition probability
NumCompaniesWorked	0.37553	Positive coefficient : Employee switching jobs frequently --> Higher attrition probability
TotalWorkingYears	-0.66086	Negative coefficient: Higher the Total Working Years Experience --> Lesser the attrition probability
TrainingTimesLastYear	-0.2149	Negative coefficient: More Number of Trainings --> Lesser is the attrition probability
YearsSinceLastPromotion	0.64065	Positive coefficient : Later the Promotion --> Higher attrition probability
YearsWithCurrManager	-0.56046	Negative coefficient: Higher the Years with Same Manager --> Lesser attrition probability
AvgWorkingHours	0.58916	Positive coefficient: More Working Hours per Day --> Higher attrition probability
BusinessTravel.xTravel_Frequently	0.78866	Positive coefficient : More Frequent Business Travel --> Higher attrition probability
JobRole.xManufacturing.Director	-0.94578	Negative coefficient : Job Role as Manufacturing Director --> Lesser attrition probability
MaritalStatus.xSingle	1.0429	PositiveCoefficient : Single Employees --> Higher attrition probability

Conclusions & Recommendations

- ❖ Environment Satisfaction, Job Satisfaction and Work-Life Balance have high influence on attrition. So initiatives taken to improve these parameters will lead to lesser chances of employees leaving the company.
- ❖ On an average the more an employee works overtime, there are more chances that employee leaving the company. So bring a work culture to avoid over working.
- ❖ If an employee works with the same manager for a longer period of time, then there are lesser chances that employee will leave the company. So avoid frequent changes in reporting managers
- ❖ Male employees and single employees are more prone to leaving the company. So build healthy ratio of male-female and single-married employees
- ❖ Frequent Business Travel have negative impact on employees which then leave company. So avoid frequent travels by enabling employees with various communication tools like Skype, remote infrastructure management etc.
- ❖ Last but not least appreciation always keep employees engaged with company. It is evident from the correlation between Promotion and Attrition. So build a culture to appreciate and acknowledge the performance.

THANK YOU !