HR CASE STUDY

SUBMISSION

Members name:

Ritu NA

Sairam Sikhakolli

Santosh Kumar Raya

Vasanth Rayudu TH

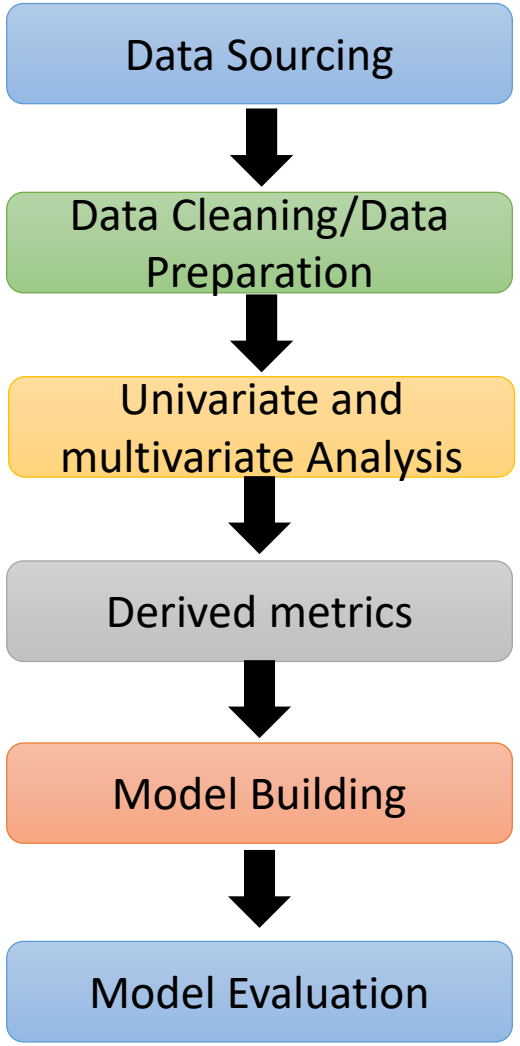
Abstract

- HR Analytics is carried out on a XYZ company to understand the driving factors for attrition. As 15% employees leave the company annually either voluntarily or begin fired away. This is very bad for the company as this results in negative company reputation and project deadlines getting delayed.
- As a result of the analysis, will be able to present the driving factors behind employees leaving the company.

Problems:

- Delay in projects making it difficult in meeting the timelines resulting in a reputation loss among clients.
- New employees have to be trained for the job and/or given time to acclimatize themselves to the company.
- A sizeable department has to be maintained , for the sole purpose of recruiting new resources.

Problem solving methodology | Logictic Regression Model



- Data is collected from the XYZ company.

- Checked for duplicate values, removed columns with whole NA and 0 values.
- Converted date columns to standard date format.
- Removed outliers on continuous variables

- Performed univariate analysis for all the categorical and continuous variables

- Derived the total hours, average hours and total no of leaves taken by the employees from the input data

- Treat the variables accordingly.
- Created dummy variables for the categorical values
- Scaling is performed on the continuous variables
- Glm is used to perform logictic regression

- The created model is tested and evaluated based on various metrics like Accuracy, Specificity, Sensitivity, ROC curve, KS Statistic, Gain and Lift charts….

Data Understanding

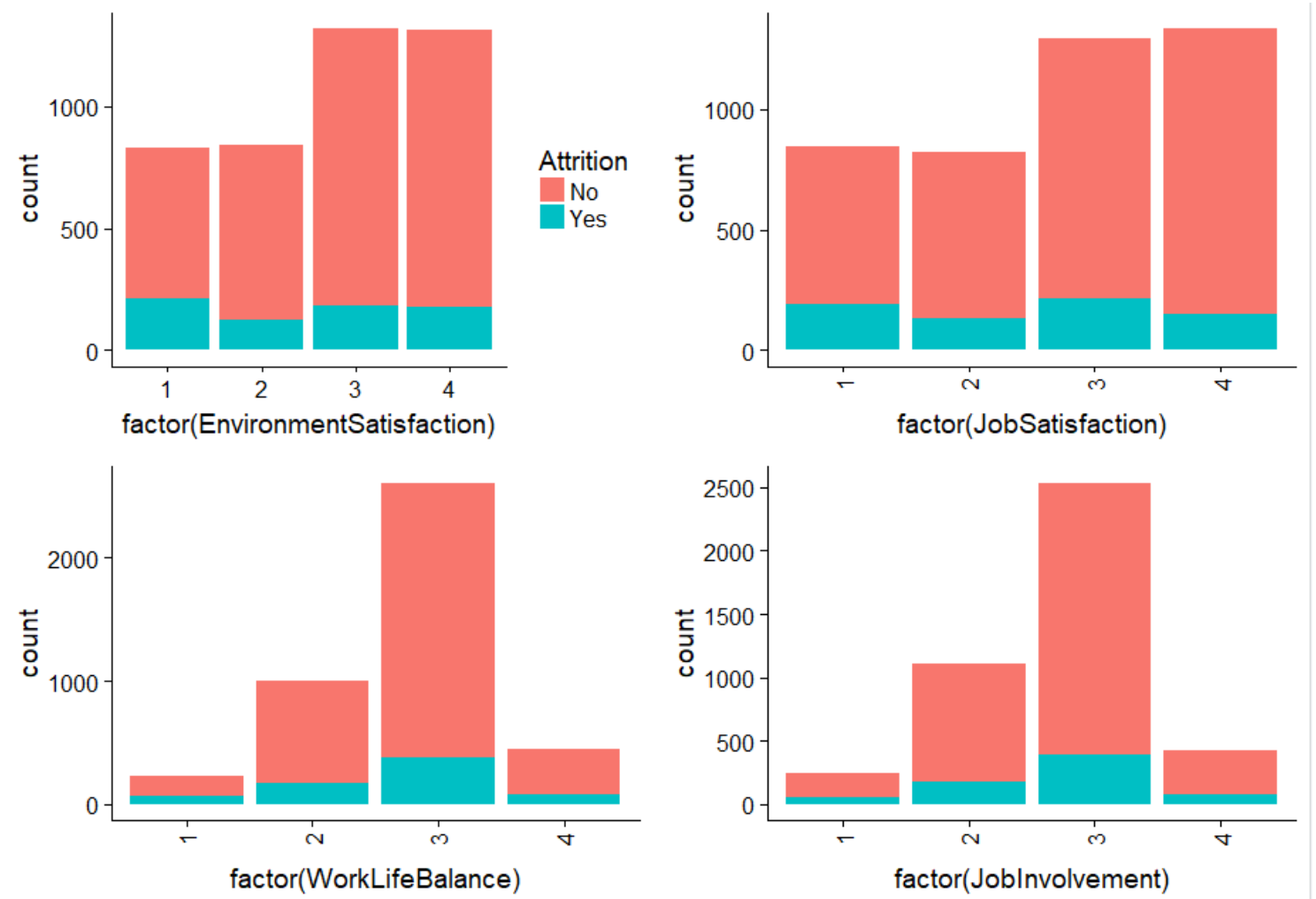
Data is provided under 5 categories

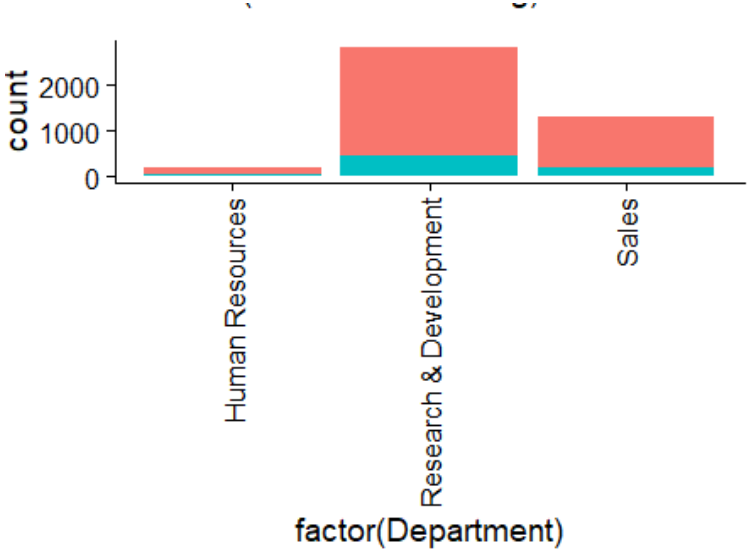
- Employee Info data - This is the main data having all the information of the employees in the professional front.
- Employee Survey data - This describes the employees behaviour towards the company/self via various factors like work satisfaction, Environment satisfaction, Work life balance etc,…
- Manager Survey data - This tells us the employees performance reviewed by the manager himself.
- In time - This data set has the check in time of all the employees spread across all the days. NA here can represent time off taken by the employee or a public holiday.
- Out time - This has the logout time of all the employees spread across same duration to in time.

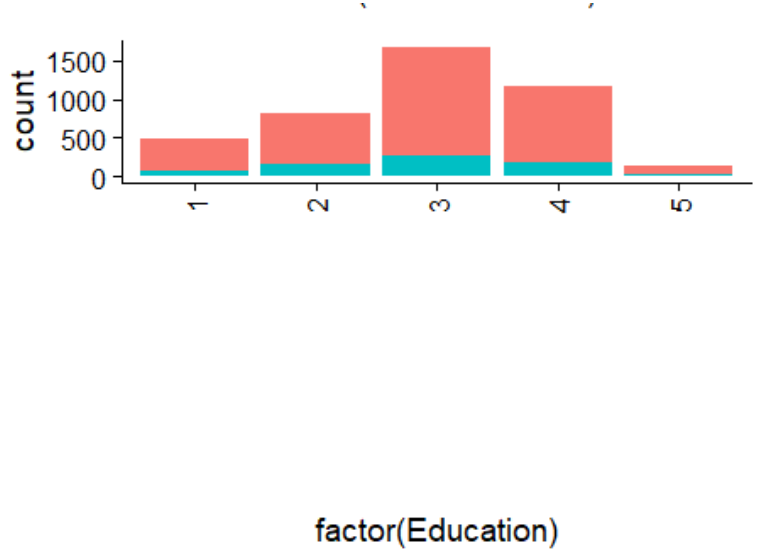
- Attrition here is the target variable and based on the data we have, we need to predict a model to identify significant factors for attrition and use the same to avoid this in the coming future.
- Attrition rate was found out to be 16.16%, this means around 111 employees left the company among 4410

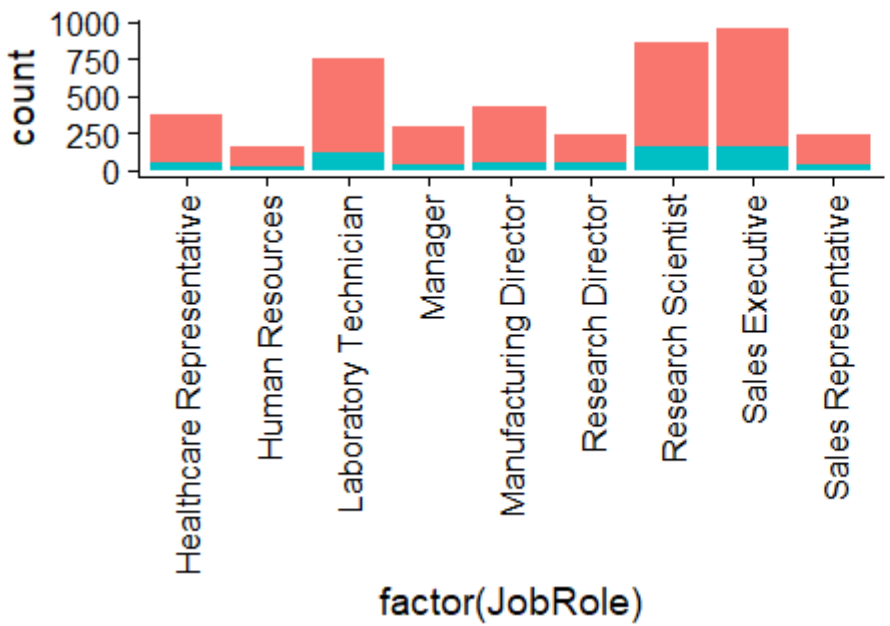
Analysis on Categorical

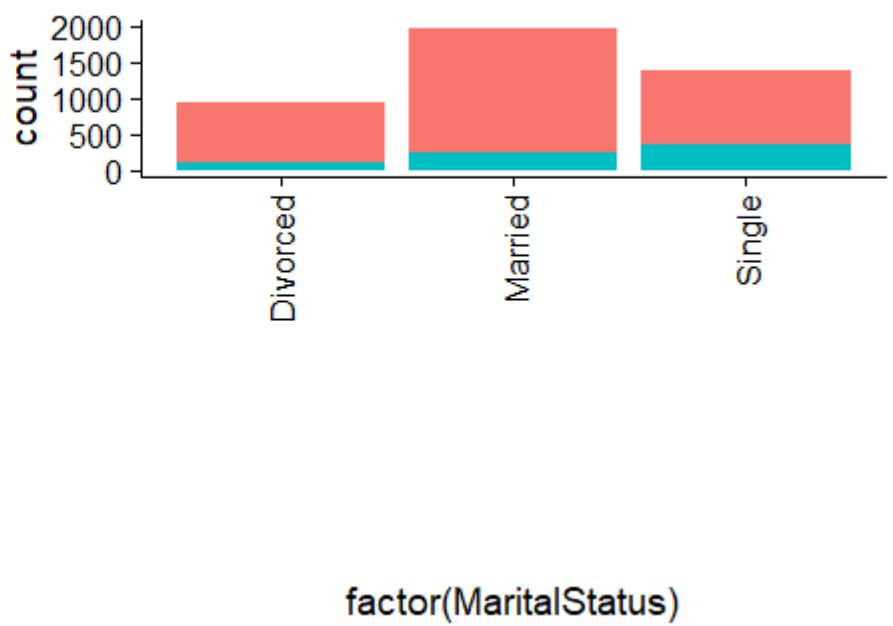
- We plotted graphs on various categorical variables based on attrition
- We can see the less the Environment Satisfaction, Job Satisfaction, WorkLifeBalance more is the attrition rate.
- Also, people who are single are more likely to leave the company compared to the married and divorce.




count
2000
1000
0
Human Resources
Research & Development
Sales
factor(Department)

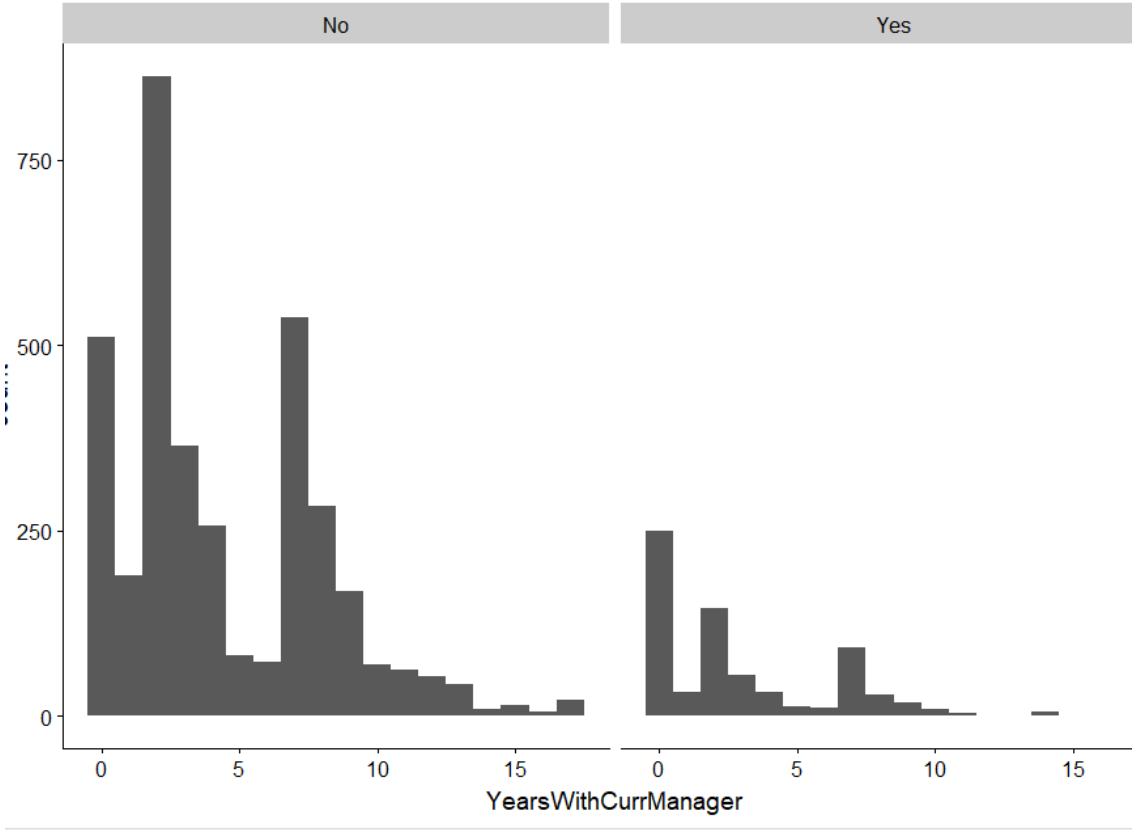

count
1500
1000
500
0
1
2
3
4
5
factor(Education)

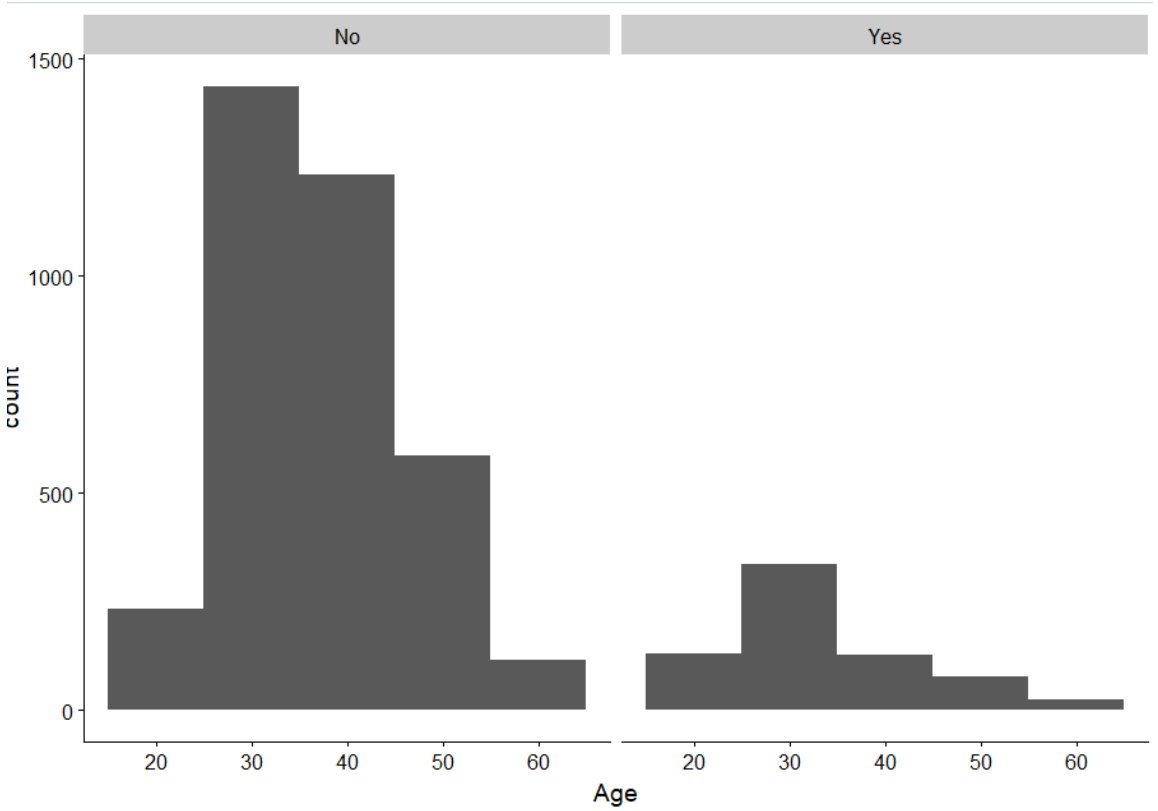

count
1000
750
500
250
0
Healthcare Representative
Human Resources
Laboratory Technician
Manager
Manufacturing Director
Research Director
Research Scientist
Sales Executive
Sales Representative
factor(JobRole)


count
2000
1500
1000
500
0
Divorced
Married
Single
factor(MaritalStatus)

Continuous Variables

- The primary continuous variables effecting the attrition are yearwithcurrmanager and Age
- The more years the employee with same manager, the less the employee is intended to leave





Model Building

- After data cleaning and preparation comes the data modelling stage. We decide the model based on the nature of our target variable.
- Here the target variable is Attrition which is categorical in nature. So we are using logistic regression technique in predicting the significant variables.
- We scaled our continuous variables and created dummy variables for our categorical variables.
- Now the data is ready for modelling. We divided the data into train and test in the ration of 70:30
- We used Generalized linear model I,e. glm function for logistic regression model to predict the attrition.
- We also used VIF to check for collinearity and eliminate the variables causing them
- As they result in model unstability.
- Later, after the vif value is less than 2, we checked for p value and eliminated variables causing insignificancy.

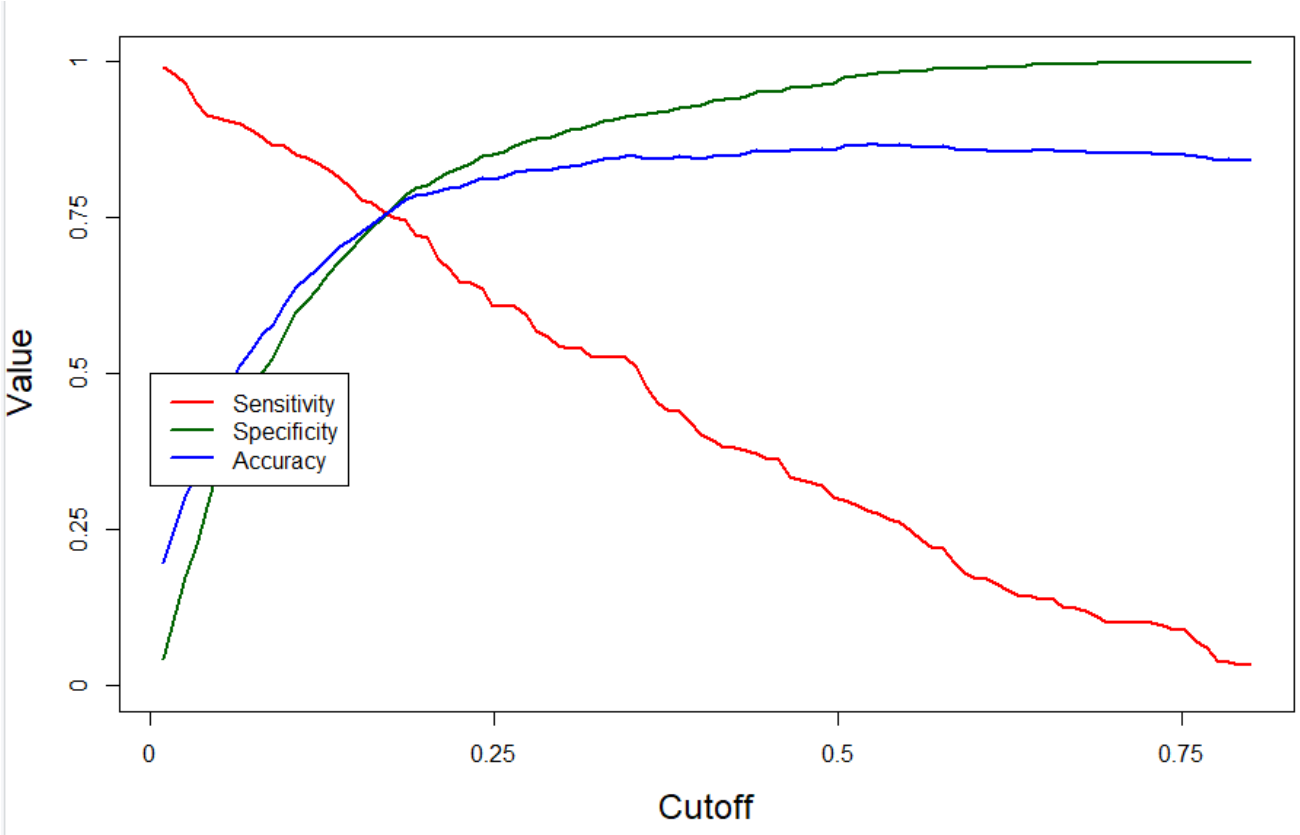
FINAL MODEL

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.47131	0.12528	-19.726	< 2e-16	***
Age	-0.55995	0.06535	-8.568	< 2e-16	***
NumCompaniesWorked	0.28225	0.05927	4.762	1.92e-06	***
YearsSinceLastPromotion	0.53930	0.07234	7.456	8.95e-14	***
YearsWithCurrManager	-0.66657	0.08256	-8.074	6.83e-16	***
extend_hrs	1.56799	0.11517	13.614	< 2e-16	***
BusinessTravel.xTravel_Frequently	0.69379	0.13411	5.173	2.30e-07	***
EducationField.xHuman.Resources	1.50369	0.31246	4.812	1.49e-06	***
JobRole.xManufacturing.Director	-0.80367	0.21877	-3.674	0.000239	***
MaritalStatus.xDivorced	-1.07717	0.16128	-6.679	2.41e-11	***
MaritalStatus.xMarried	-0.81969	0.12486	-6.565	5.21e-11	***
EnvironmentSatisfaction.x1	0.85396	0.13212	6.463	1.02e-10	***
JobSatisfaction.x1	0.68763	0.13221	5.201	1.98e-07	***
WorkLifeBalance.x1	1.24124	0.20329	6.106	1.02e-09	***

Model Evaluation

- Our final model has 13 highly significant variables now. This was achieved by eliminating the insignificant variables and checking VIF values.
- The model is now predicted on the test data set.
- Now the evaluation can be done using several factors. We will be considering factors like Accuracy, Specificity, Sensitivity through Confusion Matrix.
- We found out that the optimal probability cut off is 0.1696
- Using the cut off value we figured out the accuracy, specificity and sensitivity values.

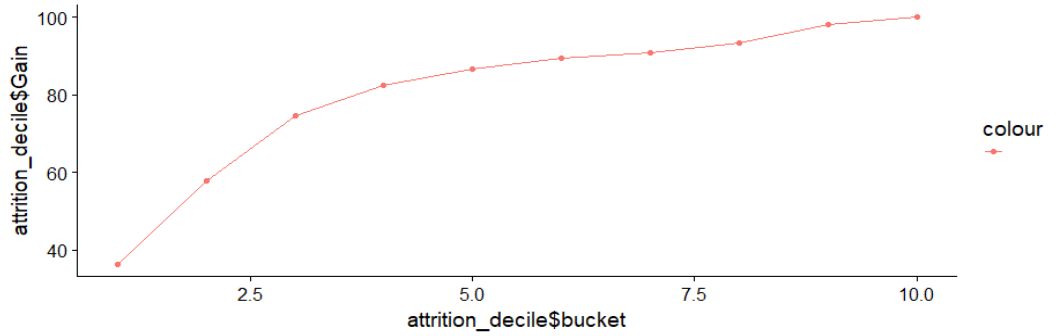


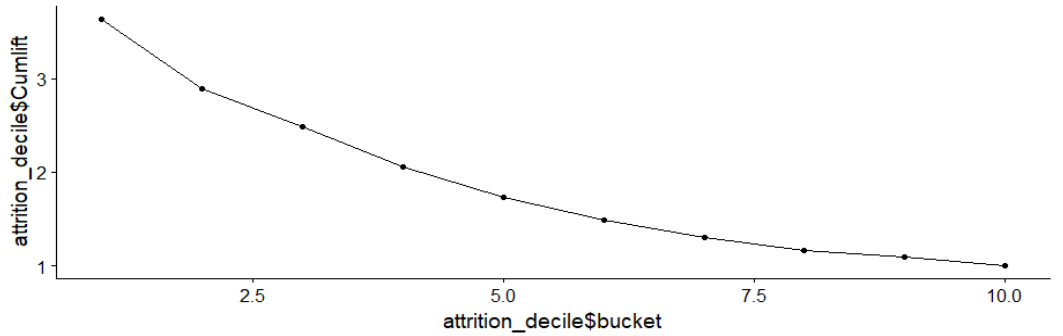
Accuracy : 0.7527 Sensitivity : 0.7608 Specificity : 0.7512

KS Statistic and Gain & Lift charts

- The KS Statistic for our model is 0.512
- Which shows the model is very good in distinguishing the employees who leave the company.
- The first four deciles is covering more than 80% of population with 75% accuracy
- The model has an increasing Gain and decreasing Lift

	bucket	total	totalresp	Cumresp	Gain	Cumlift
	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	129	76.	76.	36.4	3.64
2	2	129	45.	121.	57.9	2.89
3	3	129	35.	156.	74.6	2.49
4	4	129	16.	172.	82.3	2.06
5	5	129	9.	181.	86.6	1.73
6	6	129	6.	187.	89.5	1.49
7	7	129	3.	190.	90.9	1.30
8	8	129	5.	195.	93.3	1.17
9	9	129	10.	205.	98.1	1.09
10	10	129	4.	209.	100.	1.00





Conclusion/Recommendations

FACTORS	REASON	RECOMMENDATIONS
NumCompaniesWorked	The more the no of companies the employee have worked, the tendency of leaving the company is more	Through background check has to performed and ignore people who change companies frequently
YearsSinceLastPromotion	The less the recognition of employee the more is the employee attrition	Company has to take care of proper recognition of talents and motivating them
YearsWithCurrManager	The less the employee is with the same manager the more is the attrition	Company has to take care the reporting structure and make sure it is stable
extend_hrs	The more the employee extends the work hours the more is the probability of leavin the company	If the employee extends, proper allowances has to be given and extension has to be avoided except few circumstances
Age	More the age of an employee the less is the leaving the company	Even for the youngsters employee has to motivate them and provide better oppurtunities
BusinessTravel.xTravel_Frequently	The more the employee travels frequency the more is the chances of leaving the company	Company has to understand if the employee is good enough to travel and keep this rotational among otherresources
EducationField.xHuman.Resources	People from HR backgroung have more chances of leaving the company	Good recognition and provide them with better career opputunities
JobRole.xManufacturing.Director	The greater the position of the employee in the company the less is the chances of employee moving away	People with high grade tend to stay longer
MaritalStatus.xDivorced	Divorced people shows a negative correlation indicating they tend to stay in the company	People who are single are more likely to leave the company
MaritalStatus.xMarried	Married people shows a negative correlation indicating they tend to stay in the company	People who are single are more likely to leave the company
EnvironmentSatisfaction.x1	People with low level of satisfaction tend to leave the company	Company should be employee friendly and needs to equip a good HR team to resolve this
JobSatisfaction.x1	People with low level of satisfaction tend to leave the company	Company should be employee friendly and needs to equip a good HR team to resolve this
WorkLifeBalance.x1	People with unstable work life balance tend to leave the company	Company should be employee friendly and needs to equip a good HR team to resolve this

Thank you