

HR ANALYTICS CASE STUDY

Team Members:

1. Deepshikha Raghuwanshi
2. Husain Madraswala
3. Piyush Sharma
4. Richa Jasuja

Overview

Problem Statement

Company XYZ faces a high attrition rate, around 15% on an average every year & for the year under consideration, the attrition rate recorded was approx. 16% leading to the following issues:

- Project delays and missing deadlines, costing the company its reputation.
- New employee trainings or time taken to acclimatise to the company.
- Maintain a sizeable Talent Acquisition team for recruiting new talent

Business Objective

- To understand the underlying factors that lead to this high attrition.
- Recommend steps for talent retention.

Process Methodology

Data Cleaning & Preparation

- Fixing all different data formats, extracting tangible information & collating all the data in one common pool, based on unique key, for further Analysis
- Treating the collated data for outliers, missing values etc.
- Removing redundant columns – Standard hours, Age above 18 years, Employee Count

Data Analysis

- Scaling the continuous variables to avoid misinterpretation.
- Converting characters to factors.
- Creating dummy variables for more than 2 levels of factors.
- Plotting variables for a visual understanding of their nature and their interaction with other variables.
- Calculating correlations between variables for insights and model building

Logit Model building & testing

- Analyzing the nature of the predicted variable, in this case "Attrition".
- Separating training and Testing Data sets
- Build the Logistic Regression Model
- Run the model on the "Test Data"
- Confusions matrix creation to derive model accuracy
- Computation of KS statistics, Gain & Lift charts

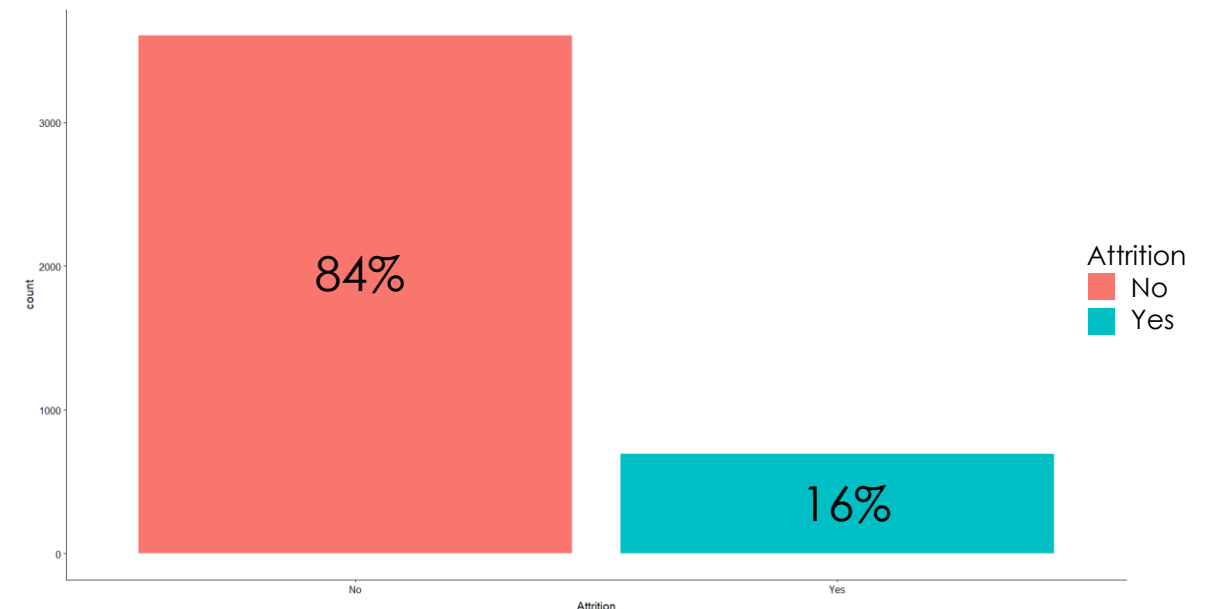
Business Insights

Derive Business Insights to find contributing factors and recommend correction measures.

Insights derived from Initial screening

- Time sheets show only 1 year worth of data hence we can assume that all independent data sets are for 1 Fiscal year only.
- There are a total of 12 Holidays in the given FY.
- Rest all NA's in the time sheets are considered as leaves taken by employees. This insight is derived by matching time sheets.
- Daily working hours and Average yearly working hours for each employee is derived from this data set.
- The derived data is merged into the main data set.
- Redundant variables are identified and removed.
- Total NA values from other variables in the merged data set are identified as shown in table.
- Since % of NA values are significantly low and cannot be imputed due based on business understanding, we have removed those observations.
- For the Current data set the yearly Attrition is 16 % out of total 4300 employee records.

VARIABLES	NUMBER OF NA'S	% OF NA'S
NumCompaniesWorked	19	0.4 %
TotalWorkingYears	9	0.2 %
EnvironmentSatisfaction	25	0.5 %
JobSatisfaction	20	0.4 %
WorkLifeBalance	38	0.8 %

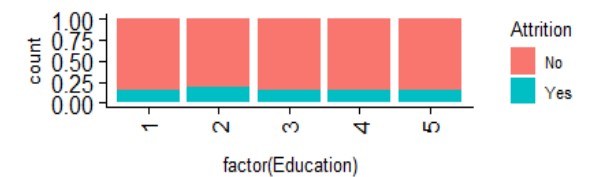
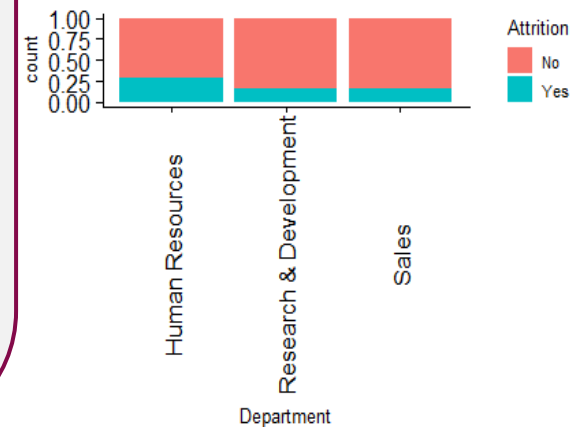
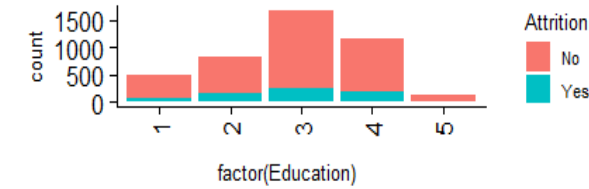
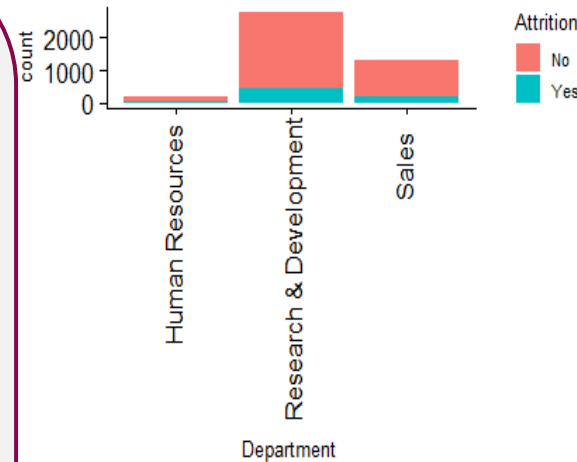


Visualizing Data Independency & Interdependency

Impact of Department, Job level & Education

Based on these visualizations it can be observed that,

- R&D Department has maximum employee strength followed by sales and HR but in terms of ratio of Attrition by department, HR has highest attrition.
- Employees who are at Junior levels 1 & 2 have higher attrition rates, 12 % out of total 16 % attrition, but the ratio of attrition is almost constant, which implies that in addition to job level 1 & 2, company should also keep a close watch on middle level management employees (level-3)
- Employees who are Graduates have the highest attrition followed by Masters & college level's, 6 %, 4.3 % & 3.6 % respectively out of total 16% attrition, but the ratio of attrition is almost constant so % attrition is the only factor for consideration.

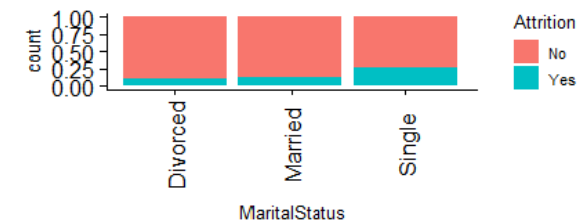
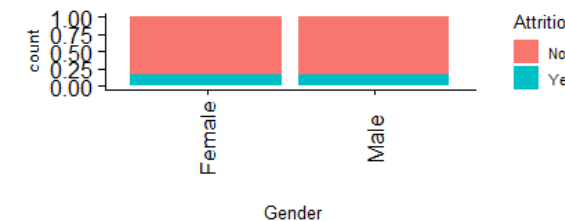
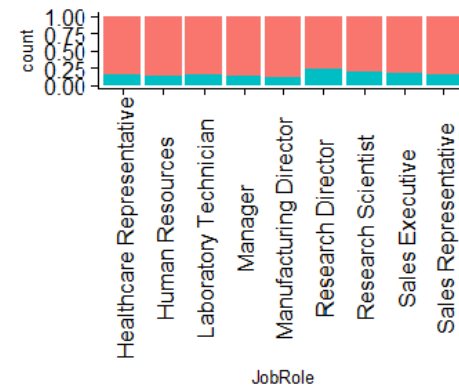
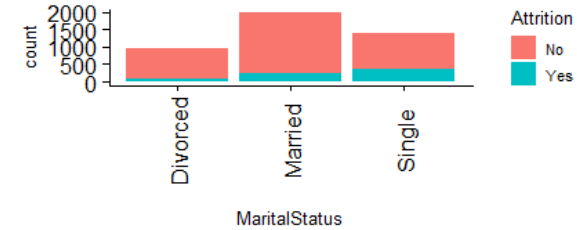
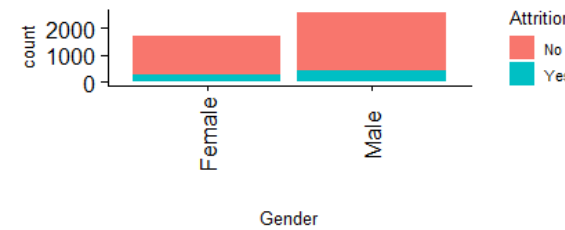
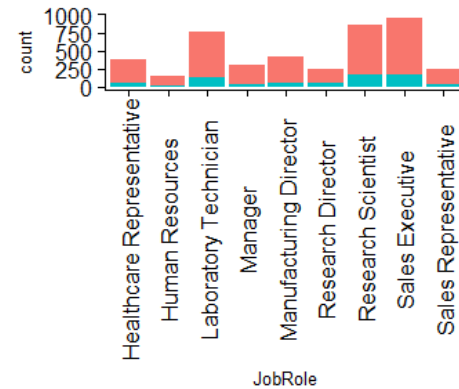


Visualizing Data Independency & Interdependency

Impact of Job Role, Gender & Marital Status

Based on these visualizations it can be observed that,

- Out of all Job roles, the top three with maximum manpower are Sales Executive, Research scientists & Laboratory Technicians. And coincidentally these roles also have the maximum attritions, 3.8 % , 3.7 % & 2.8 % respectively. Which is a clear watch out for the company. Nothing very evident from ratio of Attritions.
- Gender Seems to have no considerable impact on Attrition level.
- Employees having single marital status have highest attrition 8 % out total 16 % attrition. Which is 50% of total attrition. This ratio is also evident in the chart showing ratio of Attrition as well which is a clear call out for the company.

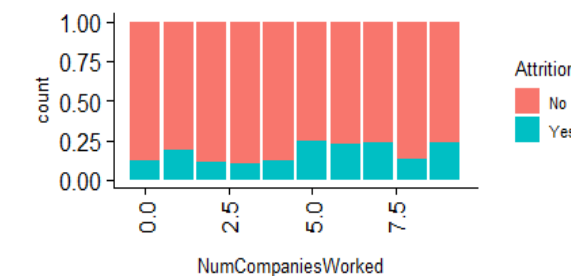
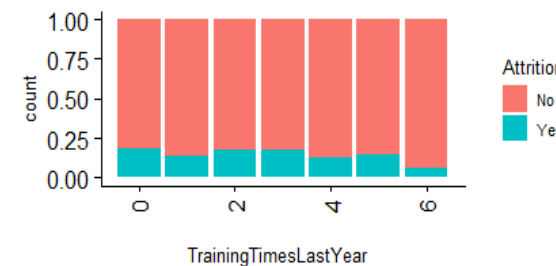
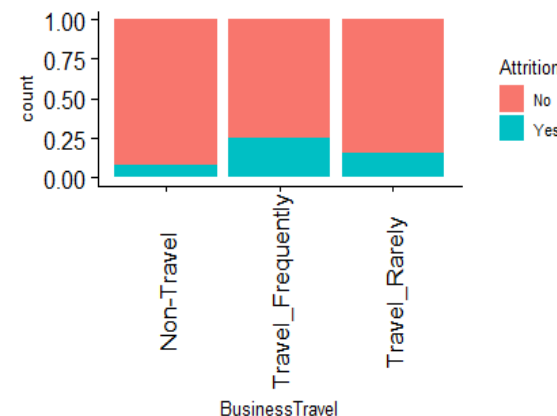
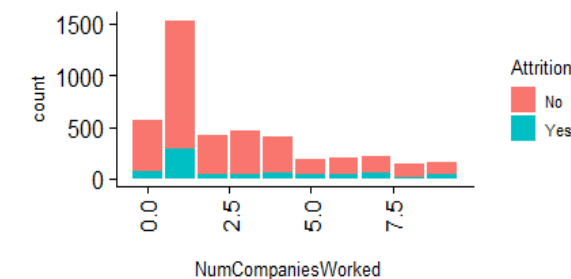
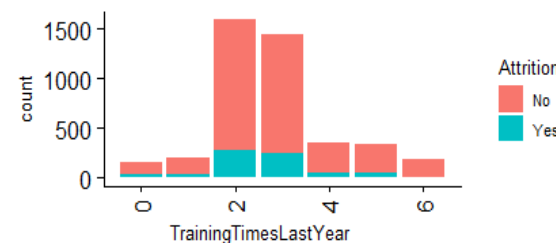
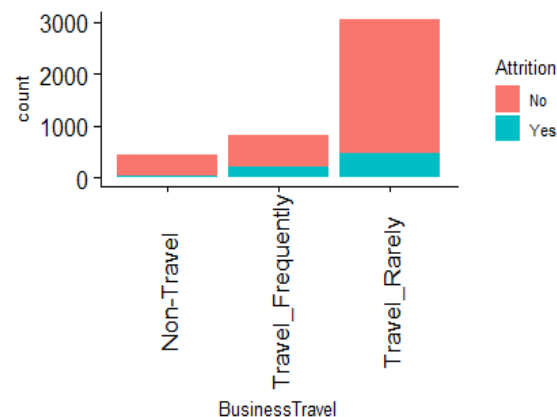


Visualizing Data Independency & Interdependency

Impact of Travel Frequency, Training & No. Companies worked for

Based on these visualizations it can be observed that,

- Although Employees travelling rarely seem to have higher attrition visually, but in terms of ratio of attrition. Frequent travelers have 10 % higher attrition compared to less frequent travelers.
- Training times seems to have inconsistency in terms of % of Attrition and its ratio, hence no considerable insights can be derived.
- In terms of % of attrition, employees who have worked for only 1 company prior XYZ, have higher rate of attrition. This also points towards the facts that the current employee strength in the company comprises of maximum employees in this category. Although it should be noted that there is a higher ratio of attrition in employees from 5 to 7 & 9 and above categories which is also a watch out.

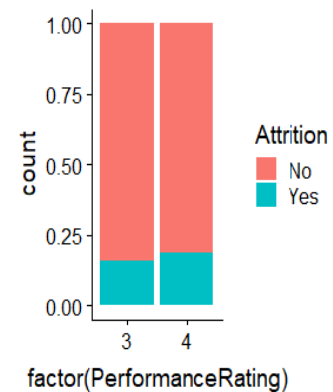
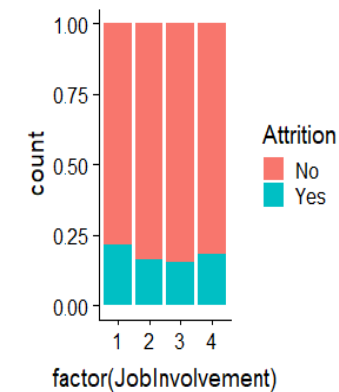
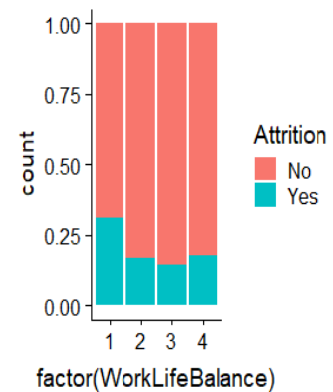
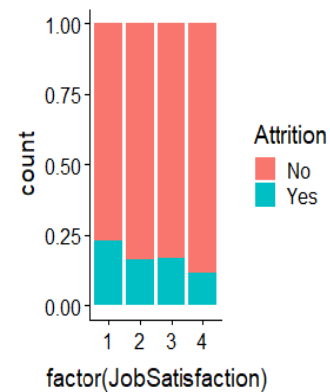
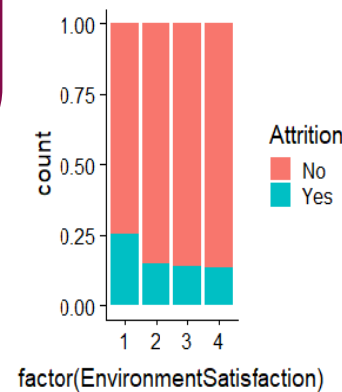
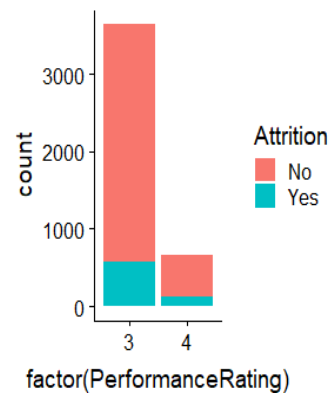
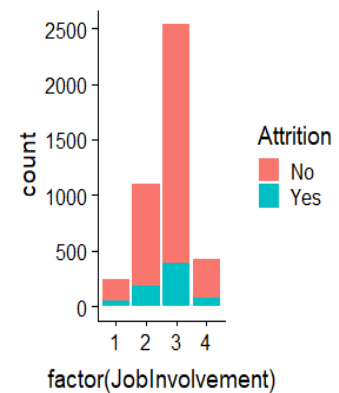
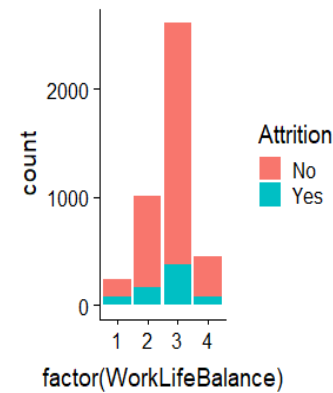
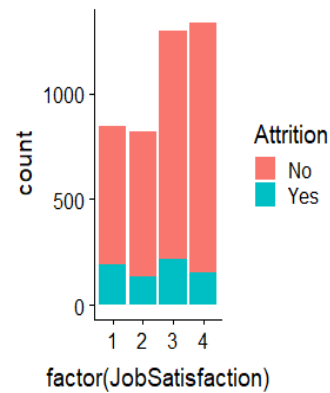
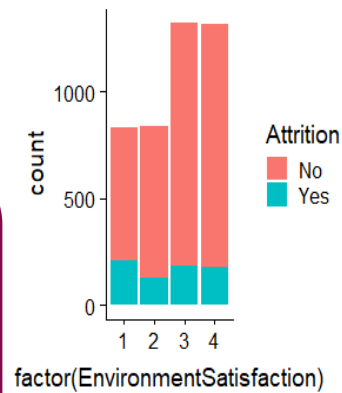


Visualizing Data Independency & Interdependency

Impact of Work Environment & Job satisfaction, work life balance and other factors

Based on these visualizations it can be observed that,

- Work environment & Job satisfaction seems to have a considerable impact on Attrition,
- Similarly Bad work life balance seems to be a contributing factor.
- Lower involvement also leads to higher attrition.



Visualizing Data Independency & Interdependency

Impact of Average work hours, Monthly Income & Work-Home distance

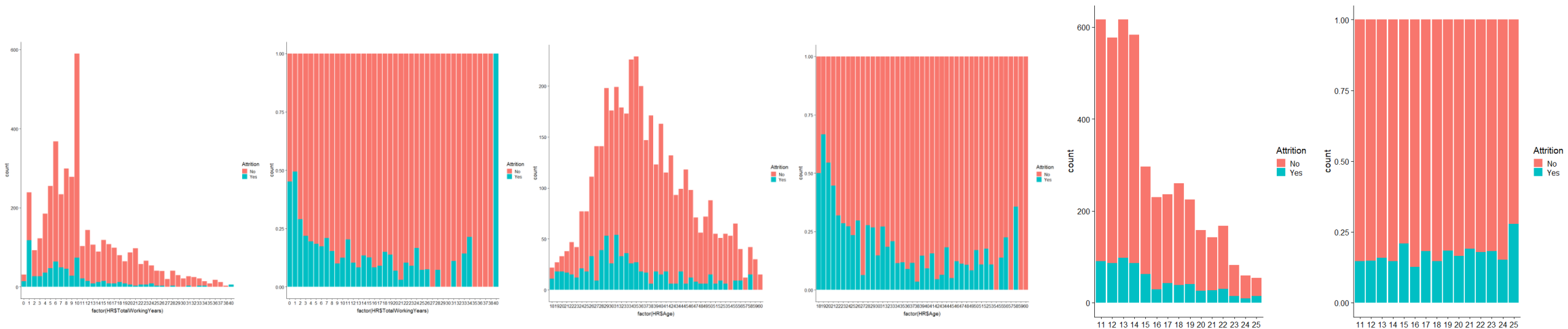


Based on these visualizations it can be observed that,

- Most people work in the range of 6 to 8 hours while those working overtime are in the range of 8 to 11. These employees have a higher Attrition rate compared to others
- People working in a salary range of above 150K seem to have the highest attrition. Possible indicator of Higher performers moving to better opportunities. Clear watch for the company to keep a close watch on retaining top performers.
- Distance of work place from home does not seem to have any significant impact on the Attrition rate based on visualizations.

Visualizing Data Independency & Interdependency

Impact of Work experience, Age & Percentage of Salary Hike.

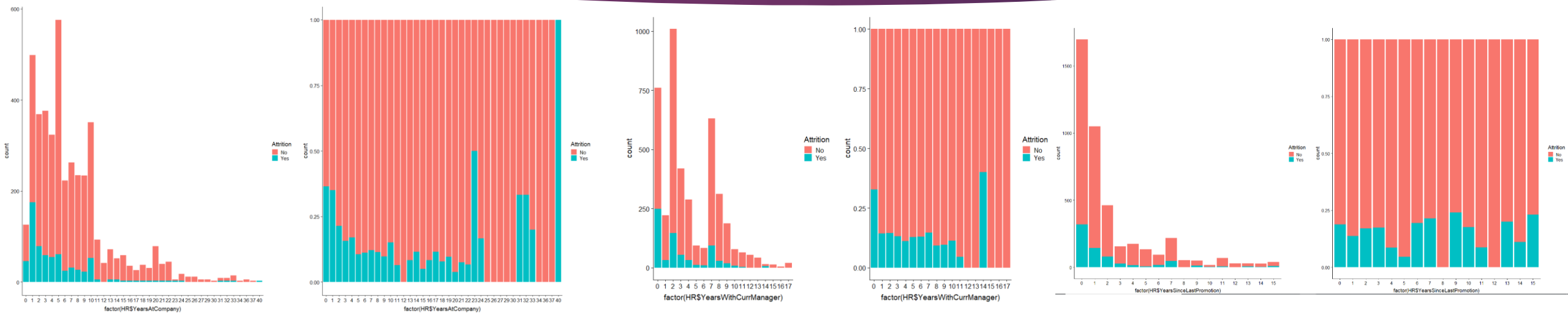


Based on these visualizations it can be observed that,

- There is spike of people leaving after 1, 6 & 10 year tenure. Clearly the ones leaving after 1 year are looking to move up faster by going to better opportunities but there is a possibility that the ones leaving after 5 years or 10 years could be moving with the expectation of a promotion or better salary hike in their next job.
- Interestingly employees below 33 years of Age have a comparatively higher Attrition rate, post 33 yrs. The rate of attrition seems low which indicates employees looking for stability. Moving towards 60 years, we see higher attrition at 58 yrs. indicating that people might be opting for an early voluntary retirement.
- For percentage of salary hikes, most people fall in category of 11 to 14% salary hike. Then count gradually decrease. Although Attrition seems to be proportional across all the salary hike%, with maximum being at 25% hike, indicating that these employees are exceptional performer and can better negotiate their role/Salary in the market and leave.

Visualizing Data Independency & Interdependency

Impact of Years at company, years with current manager & years since last promotion .



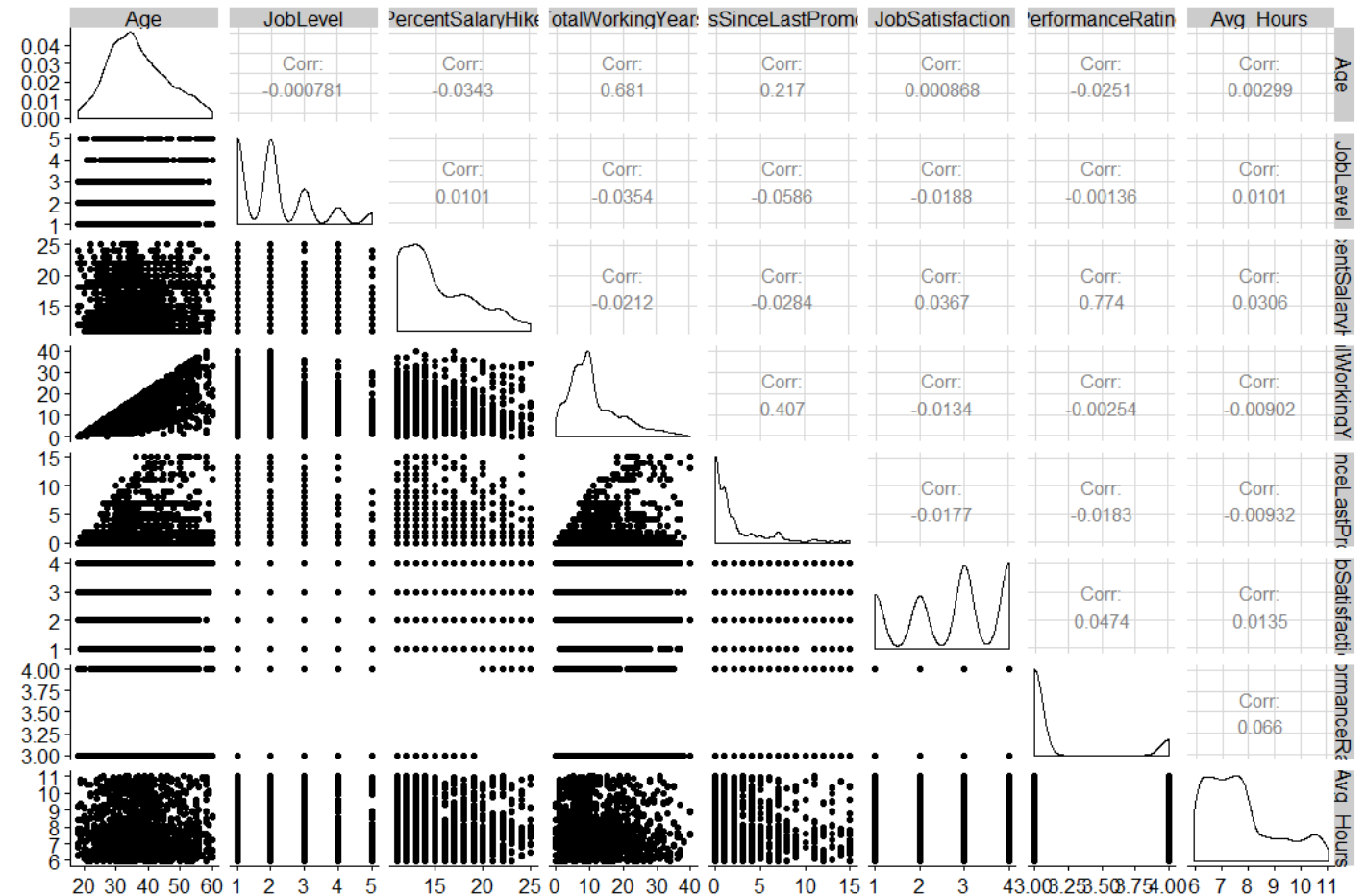
Based on these visualizations it can be observed that,

- Maximum employees are with the tenure of 0 to 10 years and this is evidently the prime range for high attritions.
- It seems evident that the longer an employee spends time with the Current manager the lower his chance of attrition.
- The general tendency seems to be that people immediately leave after receiving a promotion or after 6 to 7 years since last promotion indicating that they might be expecting a promotion but not receiving hence choosing to move out for better opportunities.

Correlation Matrix between Continuous Variables

Based on this matrix it can be observed that

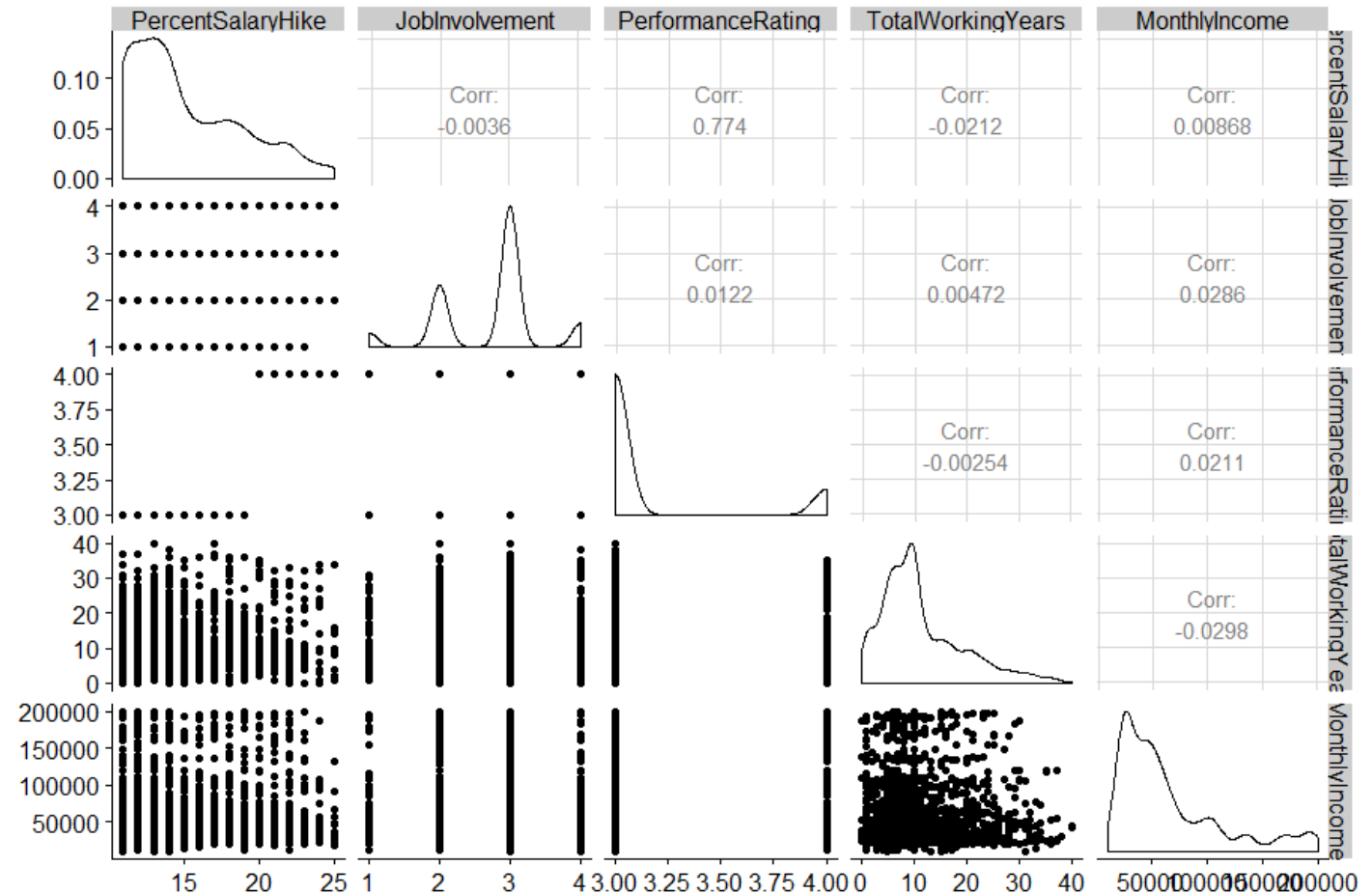
- Total working years and age show high correlation of .681, which is obvious - as the person age, work experience will accumulate



Correlation Matrix between Continuous Variables

Based on this matrix it can be observed that

- There is a High correlation between Performance Rating and % Salary Hike - .774



Model Building

Pre-Modelling Stage

- In this stage, we prepared the input variables which were to be given to the model
- We normalized the continuous variables. We scaled them using scale function in R
- We found the nature of the target variable to decide the model to be used
- Here our target variable is Attrition which is categorical in nature and hence we used logistic regression model
- Attrition rate found in our dataset is 16% only
- 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

Modelling Stage

- 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

Model Building

We checked the model against Train dataset to see how well it has predicted the known values. Here we see that there is no clear demarcation between the Yes and Nos. This is because our dataset had only 16% attrition rate. We will now run the model against the test dataset to see what probability of attrition we can correctly predict.

FINAL MODEL

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.66766	0.10194	-16.36	< 2e-16
Age	-0.33293	0.08025	-4.148	3.35E-05
BusinessTravelTravel_Frequently	0.74657	0.13295	5.615	1.96E-08
EducationFieldHuman Resources	1.41416	0.32634	4.333	1.47E-05
JobRoleManufacturing Director	-0.8752	0.21727	-4.028	5.62E-05
MaritalStatusDivorced	-1.15654	0.16332	-7.081	1.43E-12
MaritalStatusMarried	-0.80914	0.12413	-6.518	7.10E-11
NumCompaniesWorked	0.34492	0.05917	5.83	5.56E-09
TotalWorkingYears	-0.56402	0.10712	-5.265	1.40E-07
TrainingTimesLastYear	-0.2036	0.05793	-3.515	0.00044
YearsSinceLastPromotion	0.65971	0.07718	8.547	< 2e-16
YearsWithCurrManager	-0.52919	0.08592	-6.159	7.31E-10
Environmentsatisfaction	-0.33681	0.0559	-6.026	1.68E-09
Jobsatisfaction	-0.37215	0.05643	-6.595	4.26E-11
WorkLifeBalance	-0.23595	0.05585	-4.225	2.39E-05

Model Building

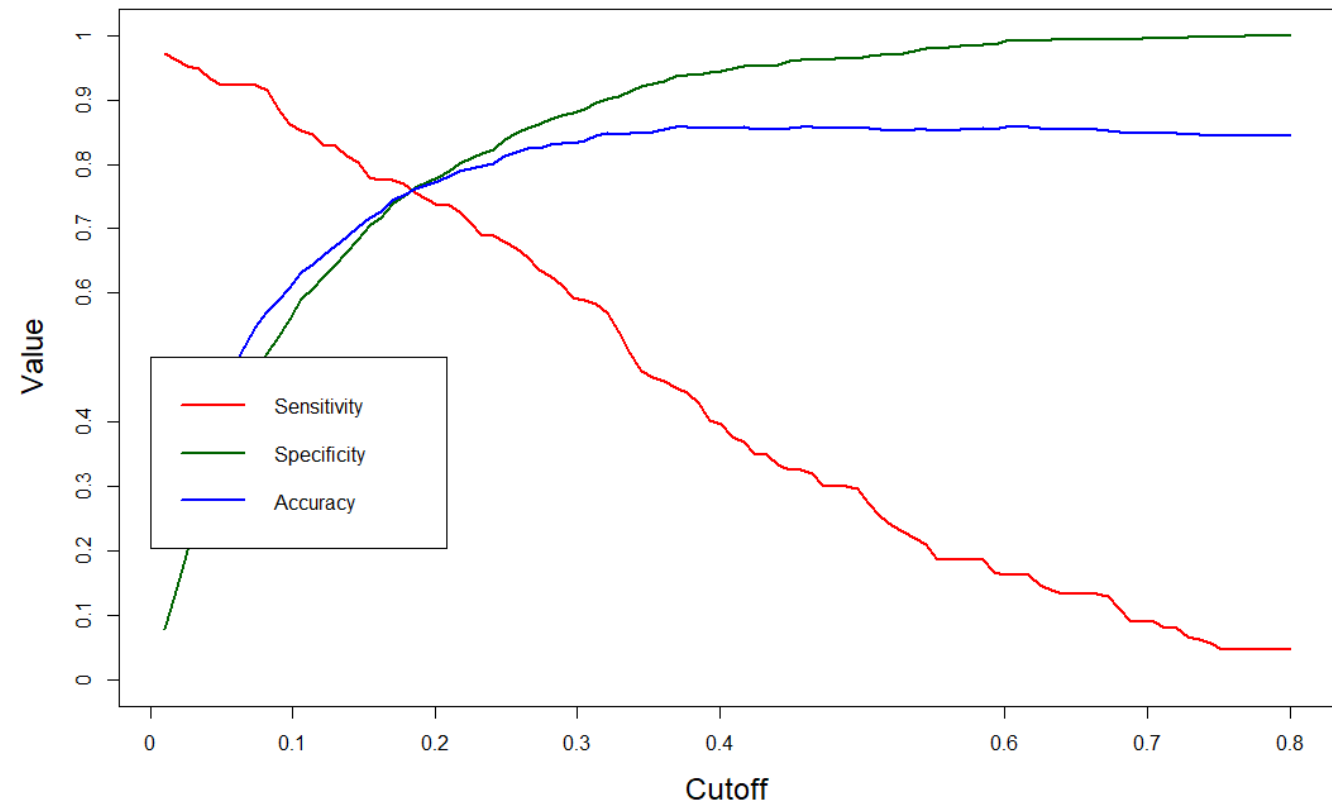
Post-Modelling Stage

- We ran the model against the test dataset
- We performed the following model validations –

1. Finding Accuracy, Specificity and Sensitivity through Confusion Matrix

- In order to find a suitable probability cut-off, we checked the Accuracy, sensitivity and specificity for 1% to 80% probability values
- The optimum cut-off probability is the one where the value of specificity and sensitivity are close to each other. Here we have taken a safe range of 0.01. Cut-off probability was found to be ~0.185

Specificity - 0.76
Sensitivity - 0.75
Accuracy - 0.76



Model Building

Calculating KS Statistics

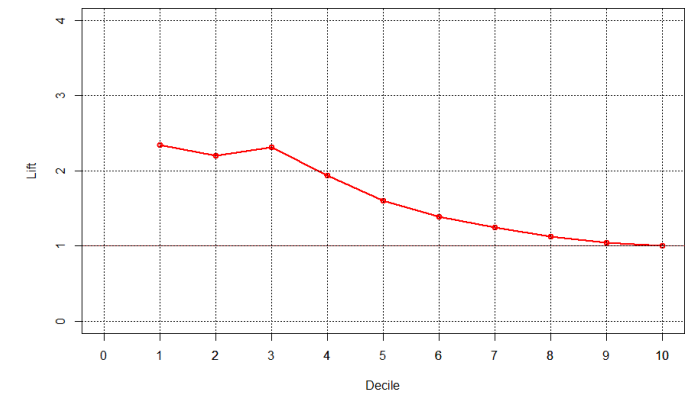
- KS statistics measures the degree of separation between positive and negative distribution
- The optimal value of KS statistics for a good model should lie between 40-60 and should be within first 3 deciles
- KS-Statistics for our model is 0.519

Gain and Lift Charts

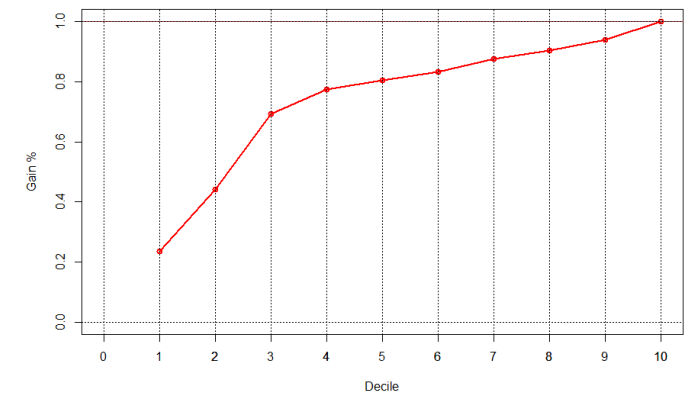
- It helps to measure the effectiveness of the model by calculating the percentage of events captured in each decile

Decile	Total	Total Resp.	Cumulative Resp.	Gain	Cumulative Lift
1	129	49	49	23.4	2.34
2	129	43	92	44.0	2.20
3	129	53	145	69.4	2.31
4	129	17	162	77.5	1.94
5	129	6	168	80.4	1.61
6	129	6	174	83.3	1.39
7	129	9	183	87.6	1.25
8	129	6	189	90.4	1.13
9	129	7	196	93.8	1.04
10	129	13	209	100	1

Lift Chart



Gain Chart





The underlying factors for attrition

Factors Considered	
Age	Job Role – Sales Exec
Business Travel – None	Marital Status – Divorced
Business Travel – Frequent	Marital Status – Married
R&D Department	No. of companies worked in
Education	Stock Option Level
Education Field – HR	Total Working years
Education Field – Life Sciences	Training Time Last year
Education Field – Medical	Years Since Last Promotion
Job level	Years With Current Manager
Job Role – HR	Environment Satisfaction
Job Role – Manager	Job Satisfaction
Job Role – Manufacturing Director	Work Life Balance
Job Role – Research Director	Performance Rating
Job Role – Research Scientist	Average Working Hours



The underlying factors for attrition

FACTORS important to an employee

Age

Marital Status – Married

Marital Status – Divorced

Education field - HR

Business Travel - Frequent

Avg. Work Hrs

Job Satisfaction

Environment Satisfaction

Total Working Years

Years Since Last Promotion

Training

Work Life Balance

Years with Current Manager

Factors responsible for JOB SATISFACTION & PROFESSIONAL GROWTH

Job Role – Manufacturing Director

Companies Worked with

Business Insights

Important Variables indicative of Increasing Attrition

Factors	Conclusions	Recommendations
Number of companies worked	If an employee has switched companies 5 to 7 times, they are more likely to leave the current company	Company should focus more on the individual's aspirations and check if they are inline with the companies aspirations before hiring such employees. They should also ask for a reason for switch from these employees to prevent attrition from the current company
Marital Status	Those who are single are more likely to leave the company	-
Age	Employees below the Age of 33 years are more likely to leave the company	Company should provide suitable opportunities to these employees and reward the critical ones as part of their retention policies.
Travelling Frequency	Those who travel frequently / less frequently are likely to quit the company. This may be because they do not like to travel and they are forced to or they have found better opportunities in other places	Company should keep a continuous check on employee's aspirations here and see if they are happy with travelling. They should also keep a tab on what other opportunities in the market and keep pace with the changing workforce dynamics
Average time spent at office each day	Those who spend more time at office each day are more likely to quit. The reason could be work pressure and frustration	Company should ensure that work in the team is uniformly divided and one employee should not be burdened with too much responsibilities . Knowledge sharing should be increased within the team so that there is no pressure on one employee in absence of another.

Business Insights

Important Variables indicative of Lowering Attrition

Factors	Conclusions	Recommendations
Work Life Balance	Employees having a good , better and best work life balance are likely to stay	Company should ensure that the employee enjoys personal and family time apart from work. They can roll out some holiday offers to the employees once in a year or two. They should be a bit lenient on leaves given to the employees
Job Satisfaction	Those who are having a medium, high and very high job satisfaction are likely to stay with company	Company should have a mechanism in place where the employees can share the feedback of the work they are doing. Based on this they should work on the negative ones and promote the positive ones
Environment Satisfaction	Those with medium, high and very high satisfaction are likely to continue long with the company.	Company should ensure conducive atmosphere to work where the employee are motivated. Employees should be given rewards in terms of money or coupons from time to time. Company should also ensure that the workspace is clean and safe for the employees to enjoy working.
Department	Those working on research and development and sales are more likely to stay	Company should ensure that whenever there are vacancies in these department, they should check on movement of employees internally as opposed to hiring someone outside the company
Job Role	Those who are Manufacturing directors are more likely to stay	-
Total working years	More is the working year, lesser are the chances to quit the company	Company should look to utilize the experience of such professionals
Years with current Manager	The more the employee spends time on working with the current manager, more likely are they to stay	Company should ensure a healthy relation between the employees and the manager. The manager should easily be approachable and should be considerate on the employees goals. This will help the company retain the employees



Thank You