HR Analytics

Attrition Analysis Prediction

Team Members:

1. Mohammed Vasanwala
2. Karthik Reddy Nareddy
3. Ravinder Reddy
4. Abhishek chouhan

Background – HR Analytics Case Study

Background

- A large company named XYZ, employs, at any given point of time, around 4000 employees. However, every year, around 15% of its employees leave the company and need to be replaced with the talent pool available in the job market. The management believes that this level of attrition (employees leaving, either on their own or because they got fired) is bad for the company which can delay project timelines, increases costs, & decreases overall productivity of the organisation

Business Objective

- Identify the important factors which drives employees to move out of organization
- Create an Machine learning model that can predict employee attrition probability
    - Algorithm Intended to use in designing ML model – Logistic Regression
- Recommend next steps to reduce employee attrition rate in the organisation

Data Understanding

Types of Variables

- Employee Time & Attendance Information
- Satisfaction variables captured in different dimensions from employee and manager (Survey Details)
- Past & Current Employee details (Experience, Role, Department, Performance etc..)
- Employee Educational qualifications

Employee Demographics	Current Employment Information	Time & Attendance Info	Salary & Performance	Survey (Employee & Manager)
• Age • Education • Education Field • Gender • Marital Status • Distance from Home • No. of Companies Worked • Years at Current Company	• Business Travel • Department • Job Level • Job Role • Trainings (Number) • Years with current manager	• Time In • Time Out • (by Day Wise)	• Monthly Income • Percent Salary Hike • Performance Rating • Stock level options • Years Since Last Promotion	• Job Satisfaction • Environment Satisfaction • Work life Balance

Data Preparation (1/3)

Variables – Converted from Continuous to Categorical (or) Large number of categories are binned into smaller set of categories based on Univariate & Bi-Variate Analysis

- Age, Number of Companies Worked, Years at Current Company, Total Working Years, Years since last promotion

Derived Variables from existing information

- Average working hours per day
- Overtime
- Salary Ratio (Monthly Income / Median Salary of respective job role)
- Number of Leaves
- Public Holidays

Missing Values Treatment

- Missing Values are 2.5% of overall dataset (4410 employees Information)
- Missing Values exist in Number of Companies Worked, Total Working Years, Environment Satisfaction, Job Satisfaction, Work life Balance
- Few of the missing values are imputed for Number of Companies Worked are imputed based on the existing details in Years at current company, Total work experience and rest are removed from dataset as the % of missing value rows are negligible proportionally
- Few of the missing values are imputed for Total Working Years are imputed based on the existing details in Years at current company, Number of companies worked and rest are removed from dataset as the % of missing value rows are negligible proportionally

Data Preparation (2/3) – Univariate & Bi-Variate Analysis

Based on the Uni-variate and Bi-variate analysis, following variables are binned to fewer categories to ensure the final model is stable and also, to enable for framing strategy for reducing employee attrition

Variable	Binned Categories
Age	18-25 Years 26-31 Years 32-40 Years >40 Years
Stock Options	No Stocks option 1 Stocks 2-3 Stocks
Years at Current Company	<2 Years 5-10 Years 11-20 Years 21-25 Years >25 Years
No. of Trainings attended (last year)	0-1 Trainings 2-3 Trainings 4-6 Trainings

Variable	Binned Categories
Marital Status	Single Others (Married & Divorced)
Work experience	1-2 Years 3-5 Years 6-10 Years 11-20 Years 21-30 Years 31-34 Years >=35 Years
No. of companies worked	<=1 Company 2-4 Company, >=5 Company
Years since last promotion	0-2 Years 3-7 Years >8 Years

Continuous Variables Treatment – Performed feature scaling for Distance from home, Monthly Salary, Salary ratio, Percent Salary hike

Data Preparation (2/3) – Key Assumptions & Approach

- As per above plots and data exploration, it is clear that data in "Total Working Years" and "Number of Companies Worked" has data descpriances and hence, following data transformation are performed with an key assumption that –

  "No. of companies worked is minimum - 1

  by considering current work experience in current organization
- The day no employee turns up in office and thus when no login time is available - it can be considered a holiday

Overall Modelling Approach is ➔

Understand The Metadata

Data Loading And Cleansing

Univariate & Bi-Variate Analysis

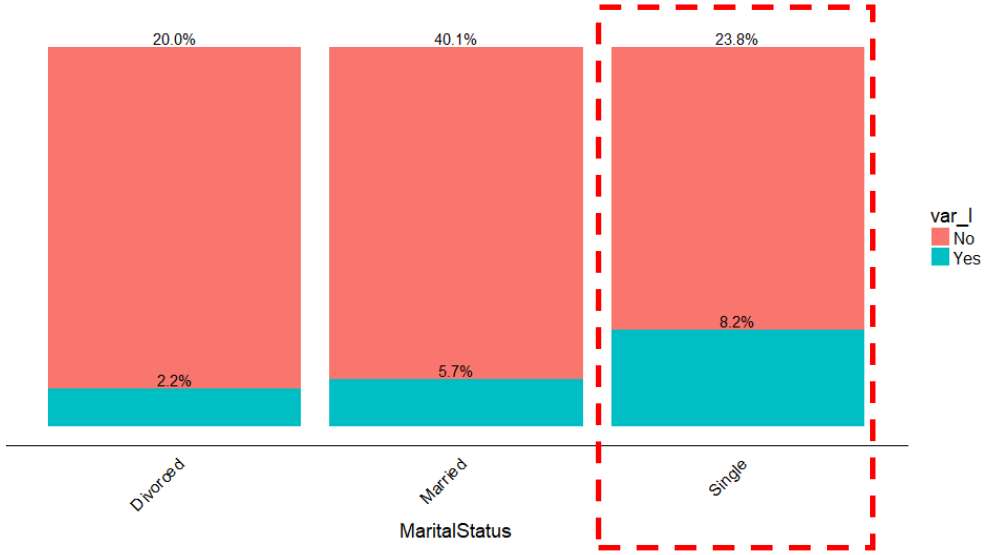
Model Building

Model Evaluation

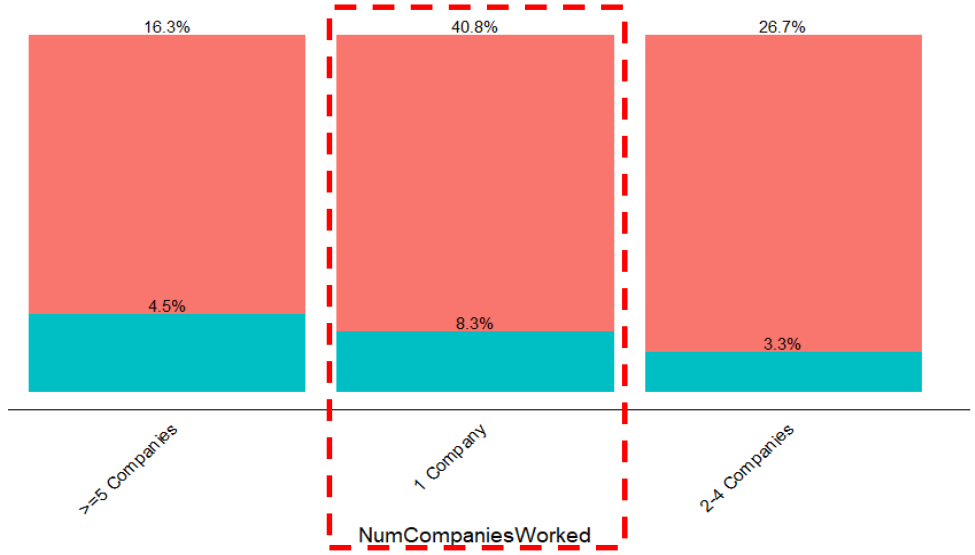
Attrition Rate – Analysis (1/3)

Overall Attrition Rate of XYZ organization – 16% among 4410 employees

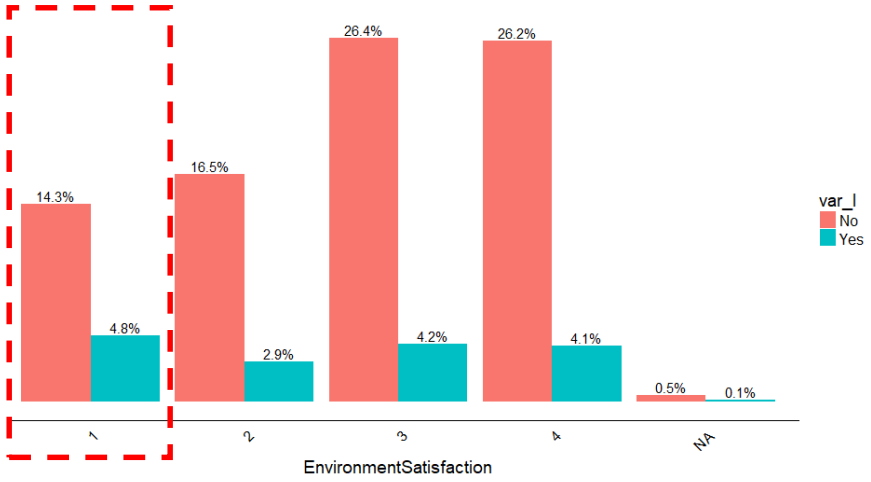
Legend:
Var_I → Attrition Rate
This is "% of Count"



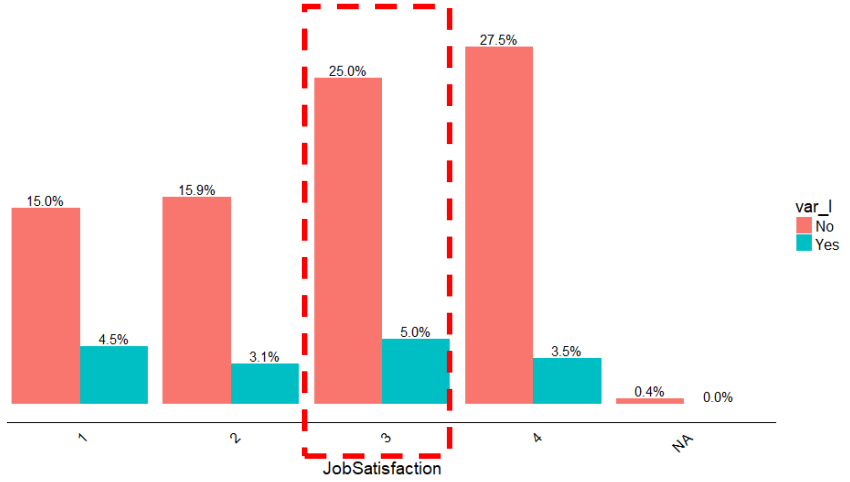
Unmarried Employees has higher Attrition Ratio



Freshers tend to leave move out of organization more

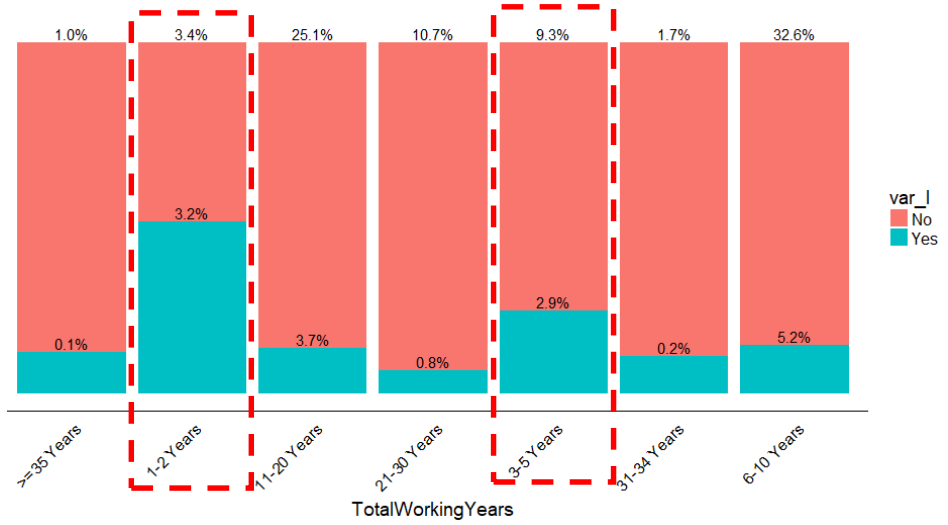


Indicates lower Environment Satisfaction has higher attrition

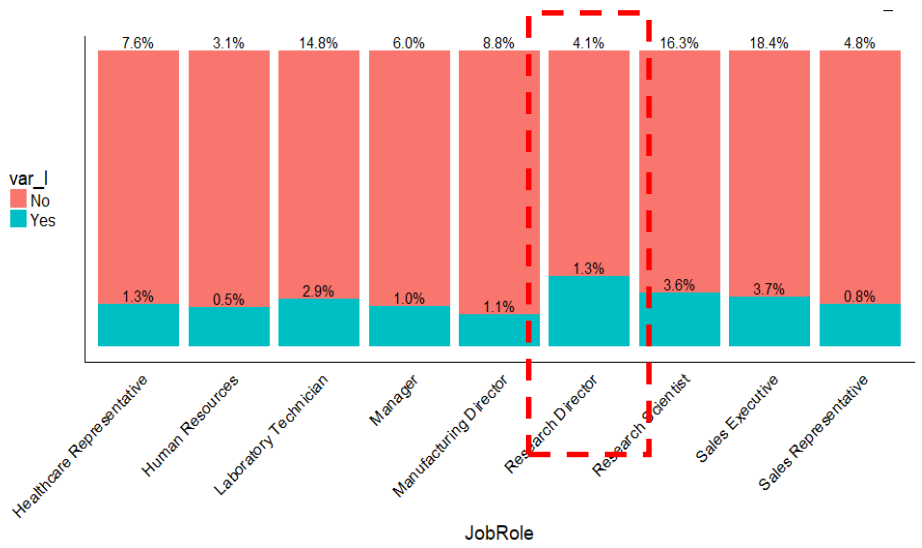


Relatively, 30% of Employees rated high but 17% tend to get churn

Attrition Rate – Analysis (2/2)



Higher attrition rates when Work Ex is 1-2 or 3-5 Years

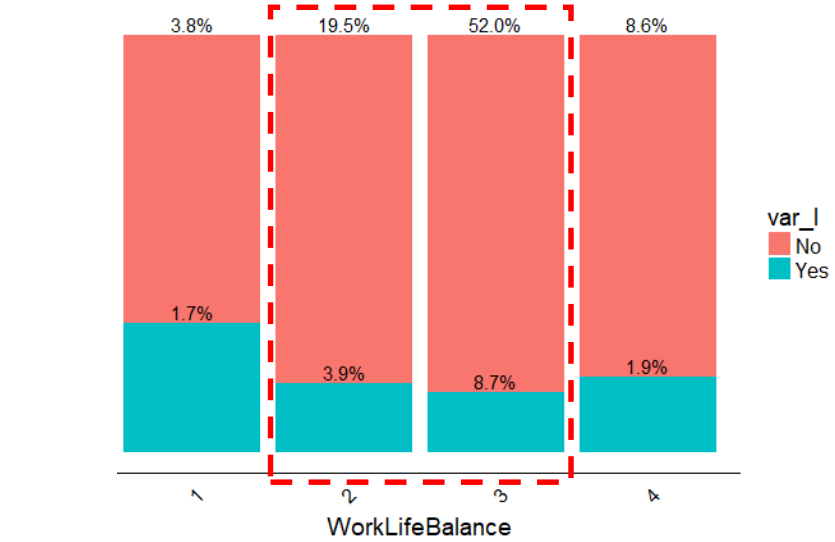


Research Directors has higher attrition rate relative to other Job roles

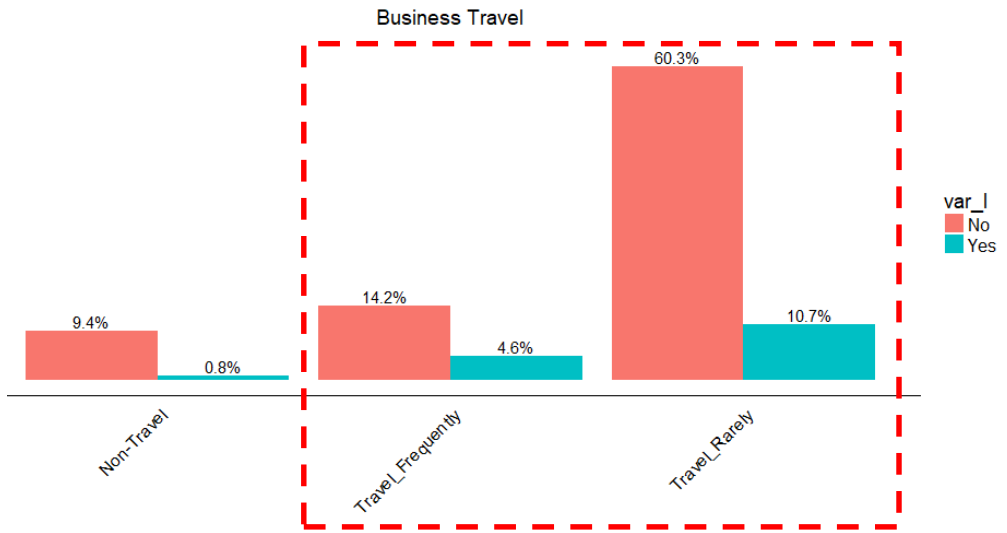
Legend:

Var_l → Attrition Rate

Y Axis is "% of Count"

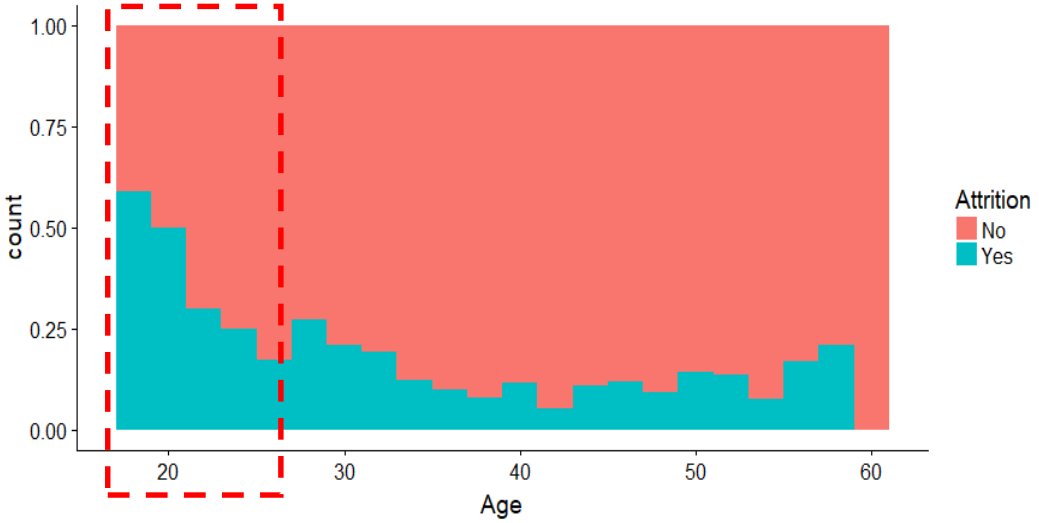


Work Life Balance – Good / Better seems to have low churn

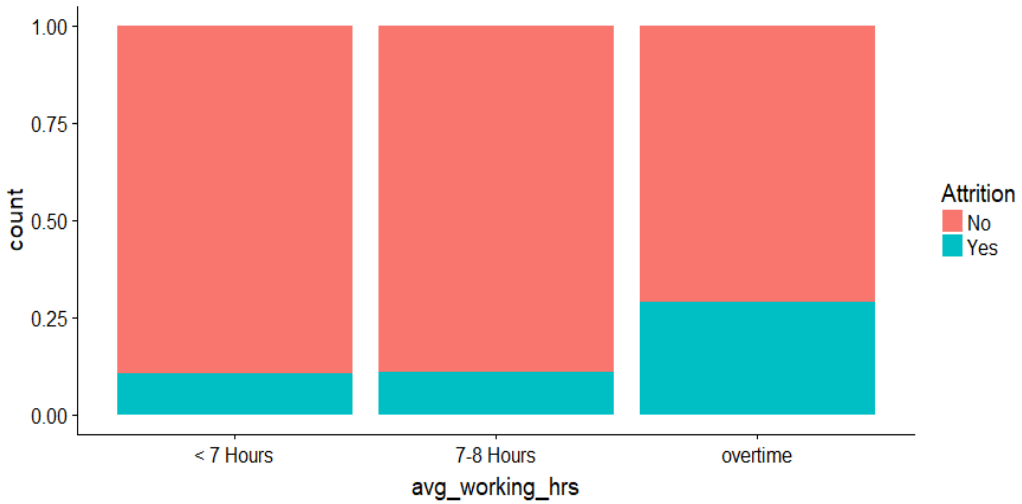


Travelling has an impact on Attrition rates

Attrition Rate – Analysis (3/3)

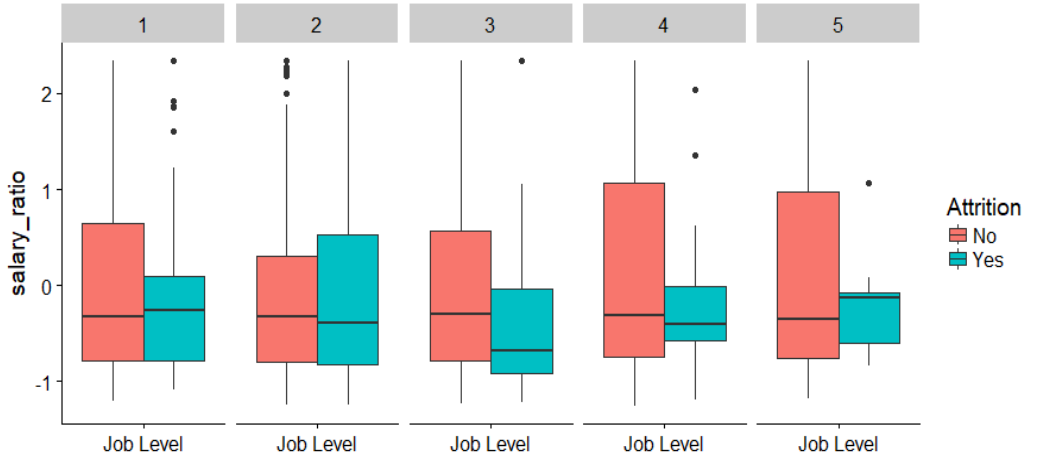


Age < 25 Years has higher Attrition Ratio

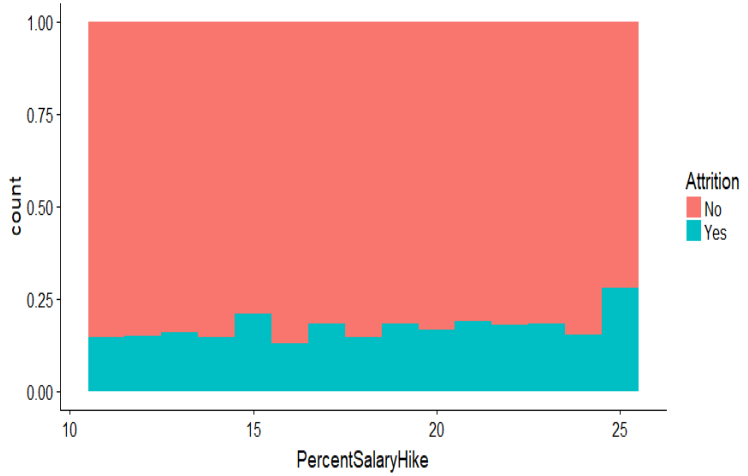


Overtime seems to have an impact over attrition rates

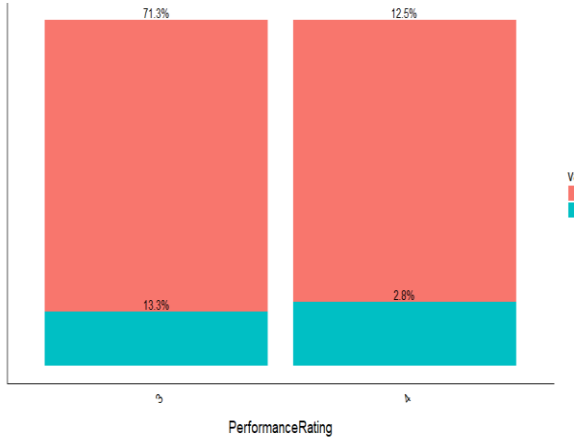
Legend:
Var_I → Attrition Rate
Y Axis is "% of Count"



>50% of churned employees in JobLevel-3 has lower salary than median salary of JobRole-3. Need to check whether it has real impact on churn through modelling



%hike don't have any impact on Attrition rate



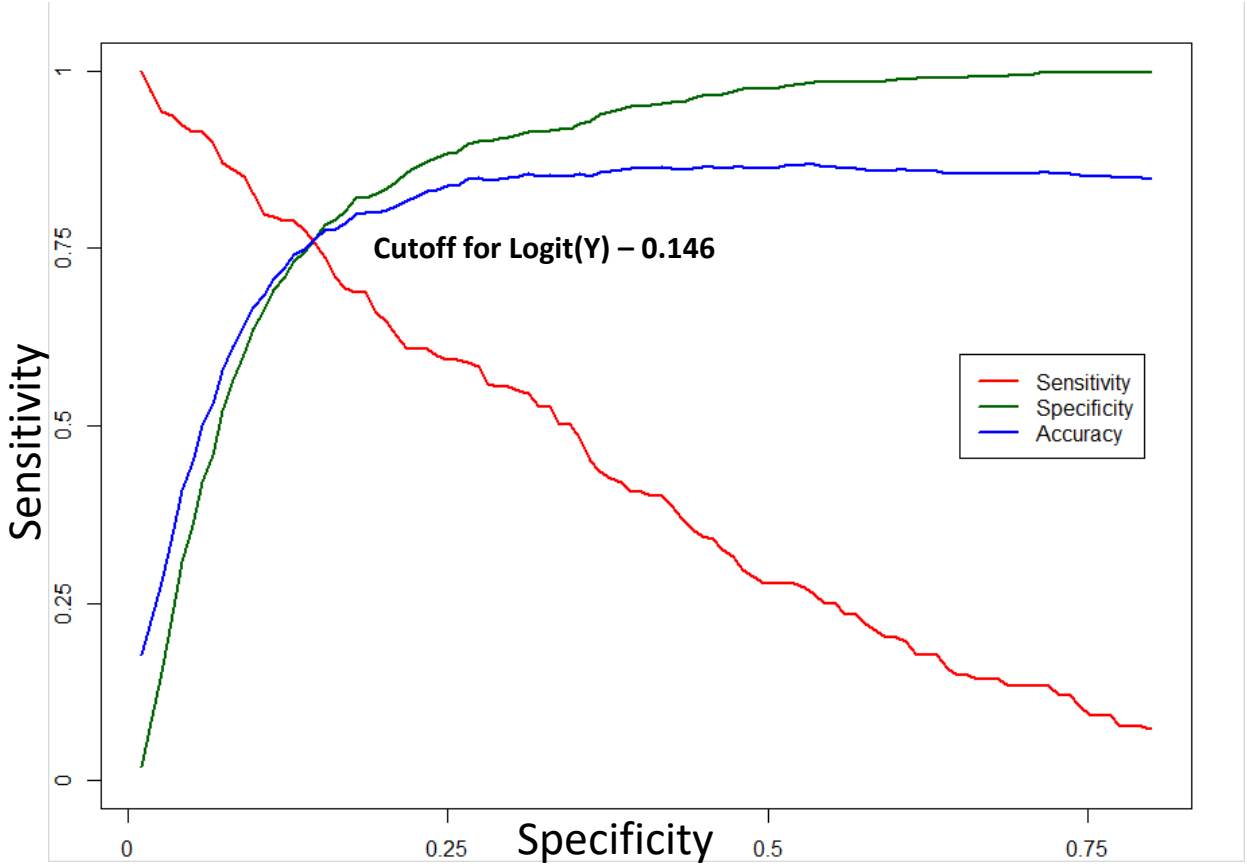
No Impact of Performance Rating

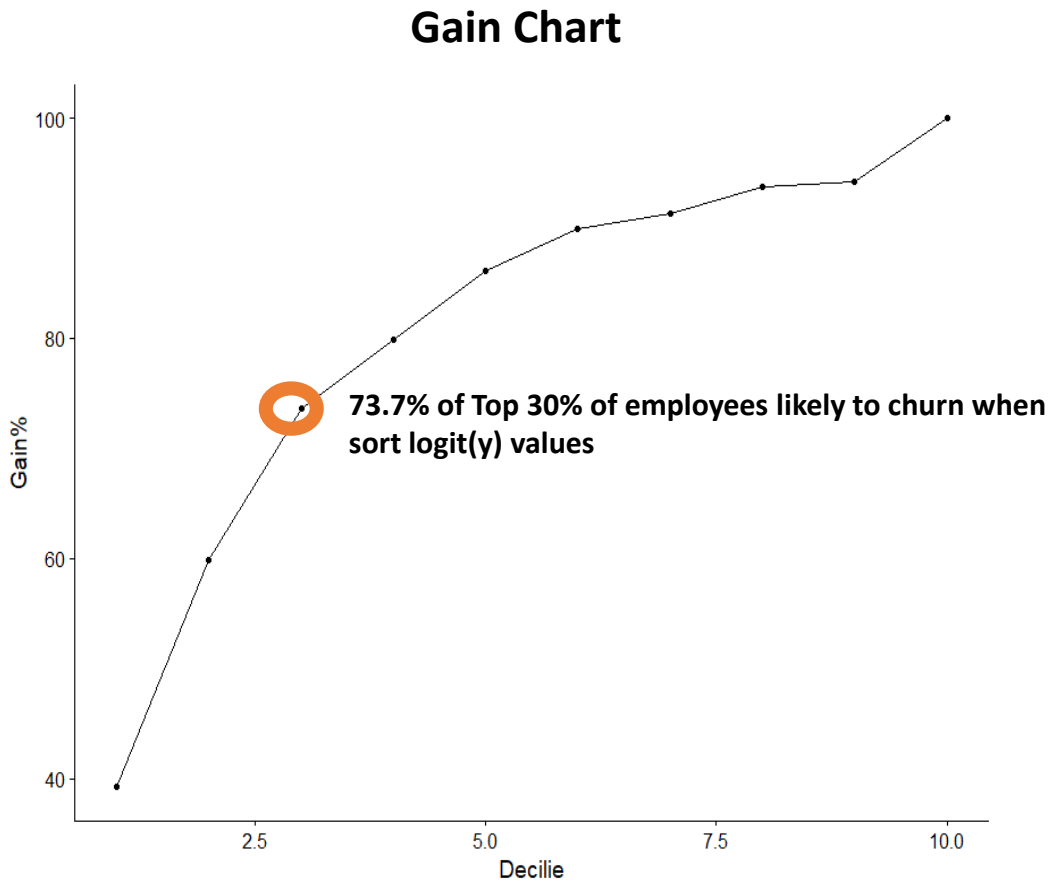
Final Prediction Model for Attrition Rate Model (1/2)

After Performing Logistic Regression on the data, logit(employee churn) is summation of values of below column – A (Intercept*Feature)

Intercept * Feature	Possible Feature Values	Interpretation
-0.9769	-	Model with no predictor variables is the estimated log odds of being employee churned
0.8498*(Marital Status)	Unmarried – 1 Married or Divorce - 0	Employee churn probability increases when unmarried
2.2134*(Work Experience 1-2 Years)+ 0.6663*(Work Experience 3-5 Years)	Work Exp is 1-2 Years – 1, Else 0 Work Exp is 3-5 Years – 1, Else, 0	Total Work Experience of 1 – 2 Years or 3-5 Years seems to get churned more
-1.2364*(No. of Companies Worked – 1 + -0.9321*(No. of Companies Worked 2-4)	No. of Companies Worked is 1 – 1, Else 0 No. of Companies Worked is 2 to 4 – 1, Else 0	No. of Companies Worked is < 4 then, probability of attrition rate decreases
1.5055*(Overtime)	If worked overtime then 1, else 0	If employees are working overtime then probability of attrition increases
-0.4323*(Work Life Balance is 3)	Work life balance is 3 then 1, else 0	If employees has good worklife balance then lesser the probability of churn
-0.7160*(Job Satisfaction is 2) + -0.6470*(Job Satisfaction is 3) + -1.1530*(Job Satisfaction is 4)	Job Satisfaction is 2 then 1, else 0 Job Satisfaction is 3 then 1, else 0 Job Satisfaction is 4 then 1, else 0	If employees has good Job satisfaction then lesser the probability of churn
-0.6981*Environment Satisfaction is 2 + -0.9217*Environment Satisfaction is 3 + -1.1413*Environment Satisfaction is 4	Environment Satisfaction is 2 then 1, else 0 Environment Satisfaction is 3 then 1, else 0 Environment Satisfaction is 4 then 1, else 0	If Job environment is good then employees to stay with the organization
0.7905*Job Role is Research Director	If research director 1, then else 0	If Job role is research director, then probability of attrition rate is more
0.8714* Frequently Travels on Business Purpose	If Travels frequently then 1, else 0	If any role that has travelling frequently then chances of churn is more
0.9025 * Age is 18-25 years + 0.7293 * Age is 26-31 years	Age is between 18-25 Years 1, else 0 Age is between 26-31 Years 1, else 0	If Age is between 18-31 Years, then probability of employees switching jobs

Final Prediction Model for Attrition Rate Model (2/3)





Model Validation

Accuracy – 76.3%

Sensitivity – 75.6%

Specificity – 76.5%

KS-Statistic – 52% (which is good)

Conclusion

Based on the model, we could understand that,

Negative Impact Variables

- Employees with Age of 18-31 Years tend to churn more
- Employees with Total Experience of <5 Years with No. of companies worked is 4 (which indicates freshers) churns more
- Total Experience & No of Companies Variables indicate that freshers or experience < 1 years tend to move out of organization
- Employees who are unmarried tend to get churn more
- Employees working overtime tend to churn more
- Employees who travels frequently tend to get churned more

Positive Impact Variables

- If Work environment, Job satisfaction & Work life balance is high then employees tend to stay with organization

Conclusion – Model clearly indicating organization need to focus on freshers (or) lesser work experience employees and prepare :

- HR Strategies to Improve Work environment, Job Satisfaction & Work life balance of these employees
- Re-look at hiring processes to ensure potential candidates are selected who are aligned to organization values & environment
- Ensure HR policies are designed such that “Overtime” should not be considered as the prefer option (One can provide Incentives but it impacts work life balance and hence, it might benefit few employees only so, penalize managers whose team works overtimes). Also, Research indicates overtime impacts productivity of employees too