INVESTMENT CASE STUDY

SUBMISSION

Prateek Pathak

Abstract

Problem Statement:

- XYZ company employs 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 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

Objectives of the Analysis:

- Model the probability of attrition using a logistic regression.
- Identify changes they should make to their workplace, in order to get most of their employees to stay.

Solution Approach:

- R is used for Data preparation, from raw data sources.

Data Cleaning Approach

- Remove all the columns which have 'NA' or 0 values.
- Remove all the columns which have only 1 value throughout.
- Remove all the columns which are redundant for analysis.
- Calculate Daily Average hours spent in office from In and Out time sheets.
- Derive new columns for further analysis.
    1. Employee Satisfaction: Sum of all employee survey data for each employee. As a continuous variable gives more scope for analysis
    2. Manager Survey: Sum of all manager survey data.
    3. Overtime (Yes/No) : If employee is spending more than the standard time of 8 hrs in office.

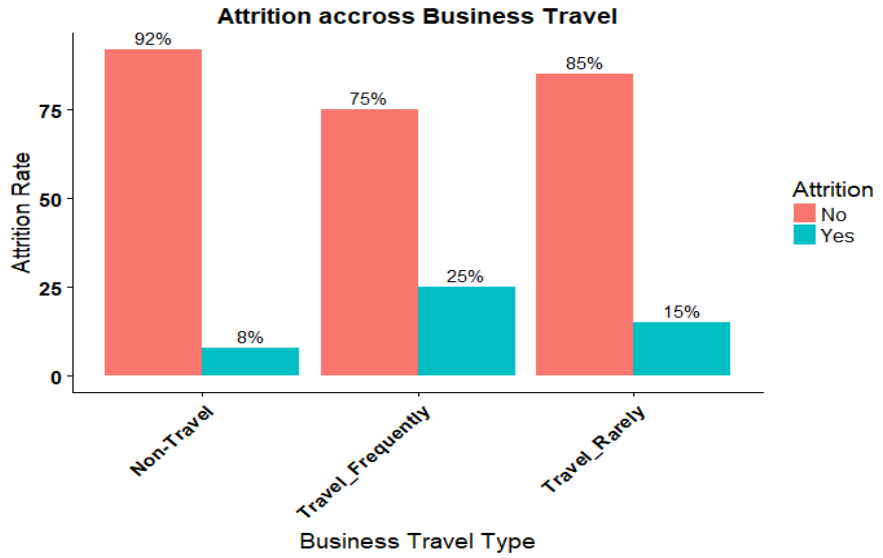
Data Description

1. Data contains **4410** unique employee loan data.
2. Two Timesheets are provided for In and Out time of employee.
3. Employee survey data is provided.
4. Manager Survey data is provided.
5. The employee data is quite unbalanced for modeling with only **16%** employees leaving.

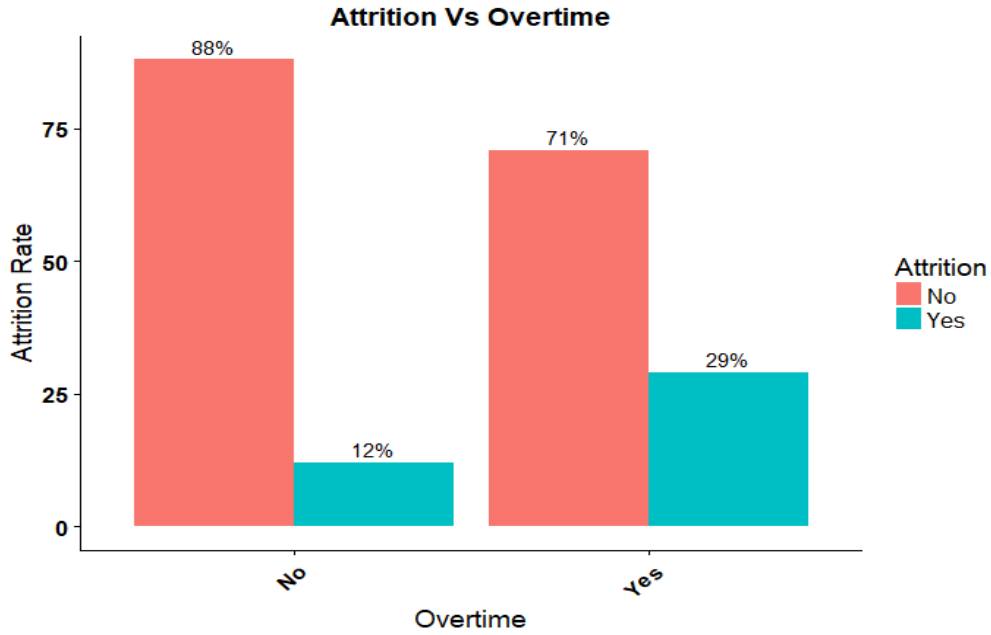
NOTE : After complete univariate and bivariate analysis following slides showcase the highlights. Rest of the plots and analysis summary can be looked up in R file.

Univariate Analysis – Exhibit 1

- Business Travel: Employee who have to travel frequently have a high rate of attrition.

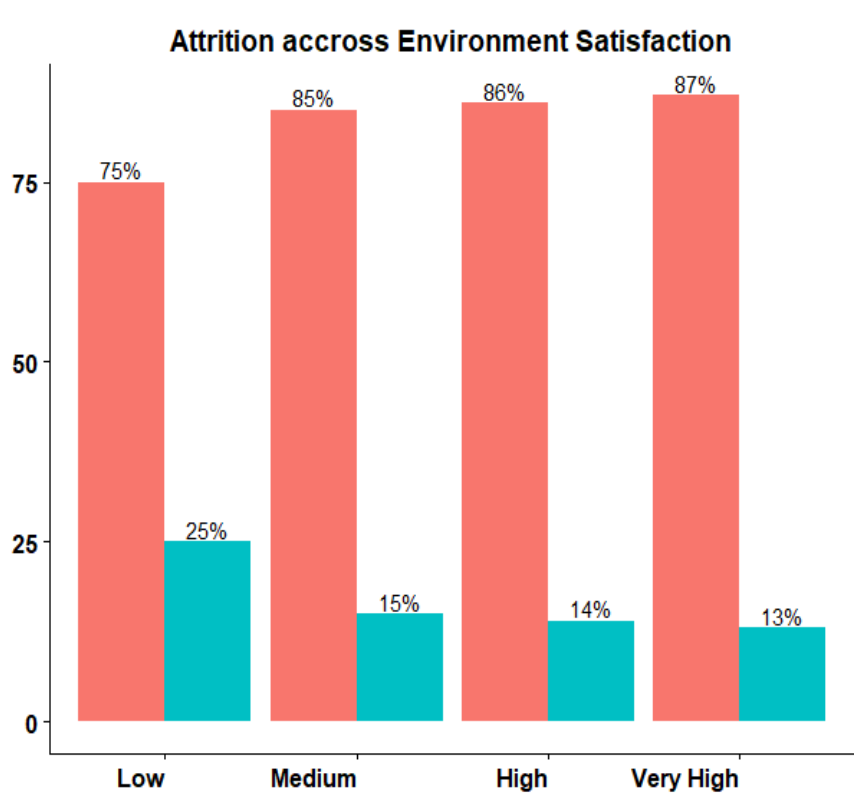


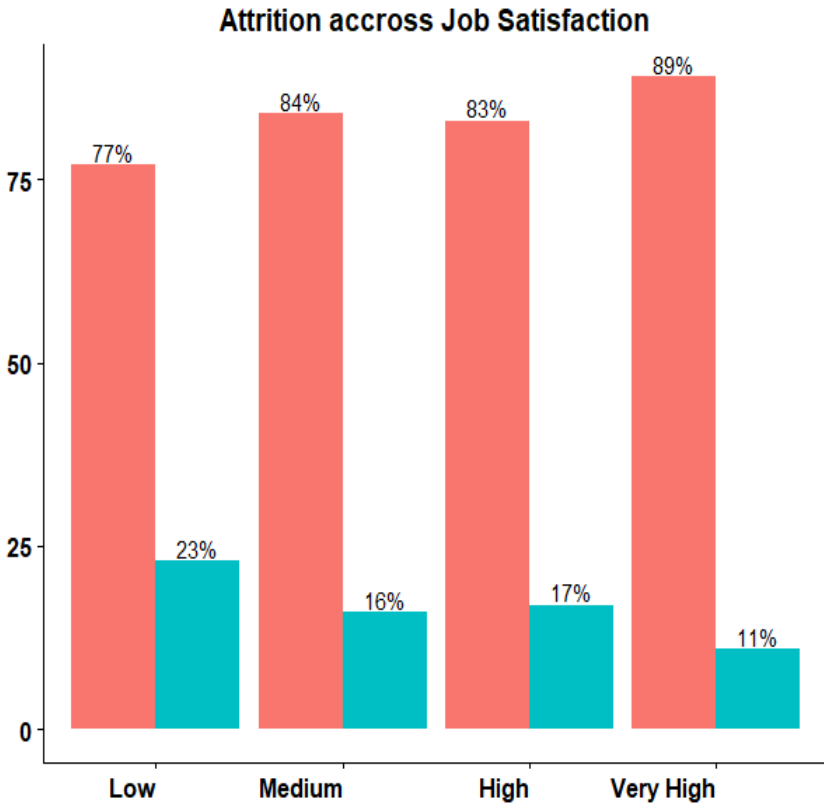
- Employees spending more than standard hours in office i.e overtime have a higher tendency to leave.

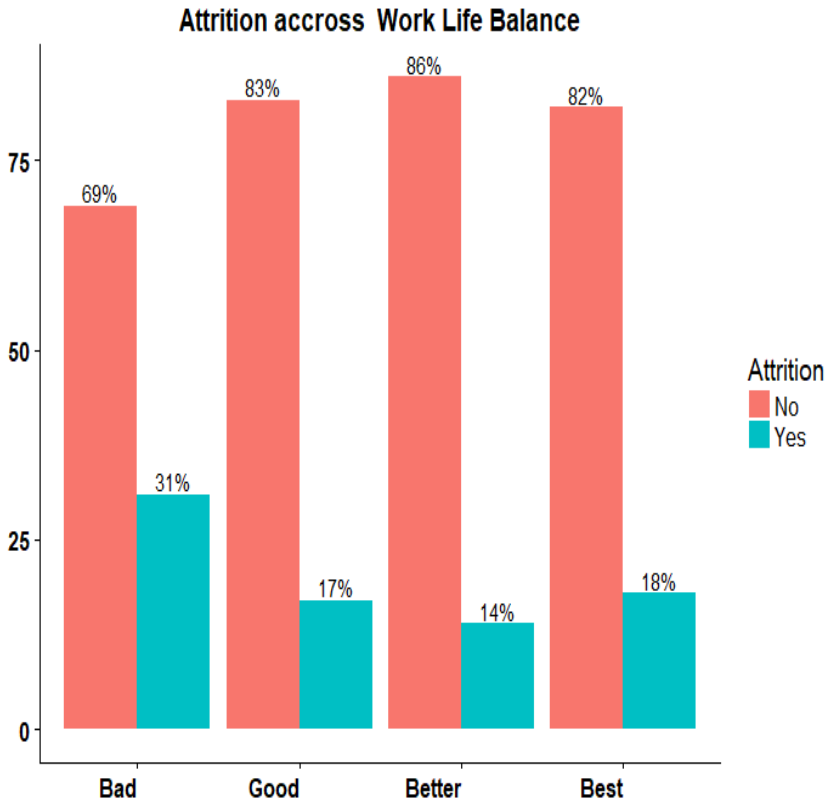


Univariate Analysis – Exhibit 2

Across all categories of Employee survey, employees responding poorly to survey have a higher attrition rate.

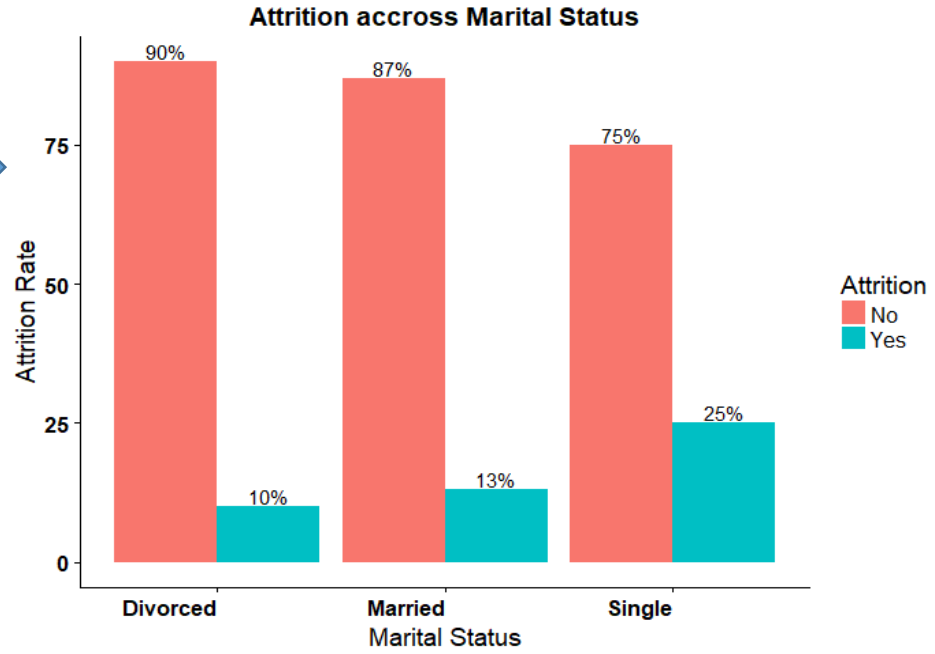


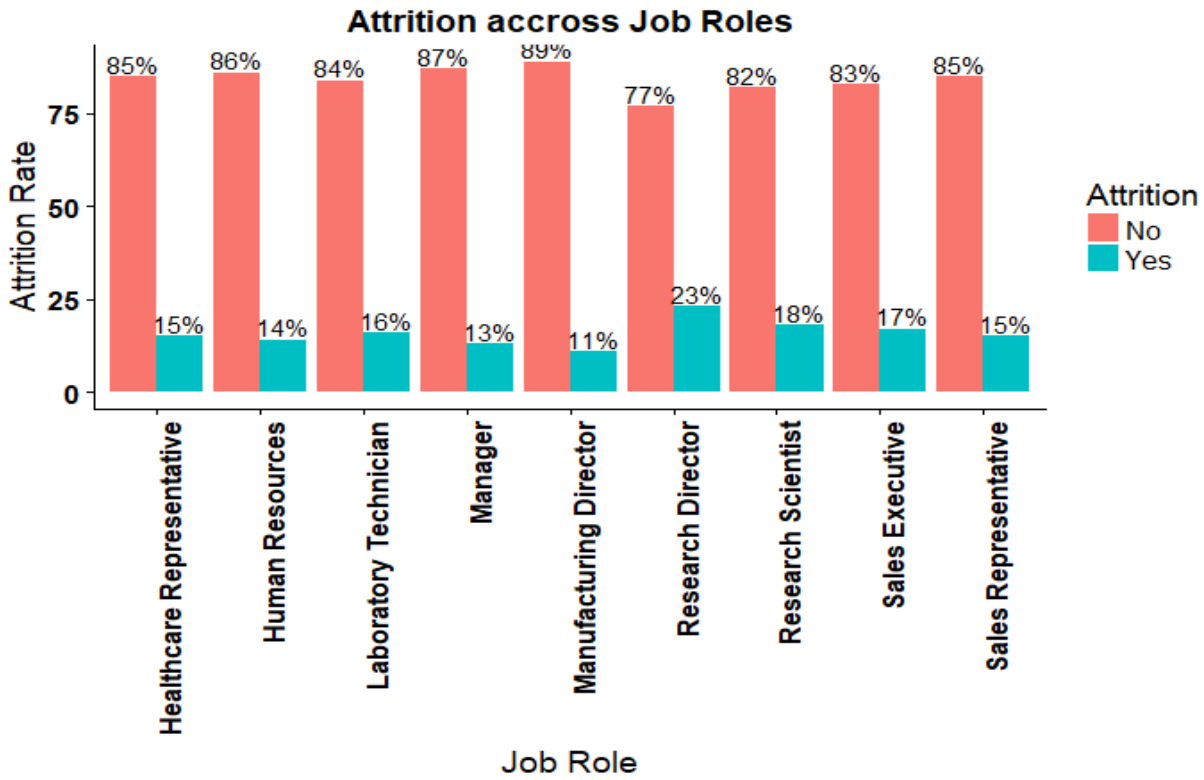




Univariate Analysis – Exhibit 3

- Employees who are single have a much higher attrition rate as compared to marriage and divorced.

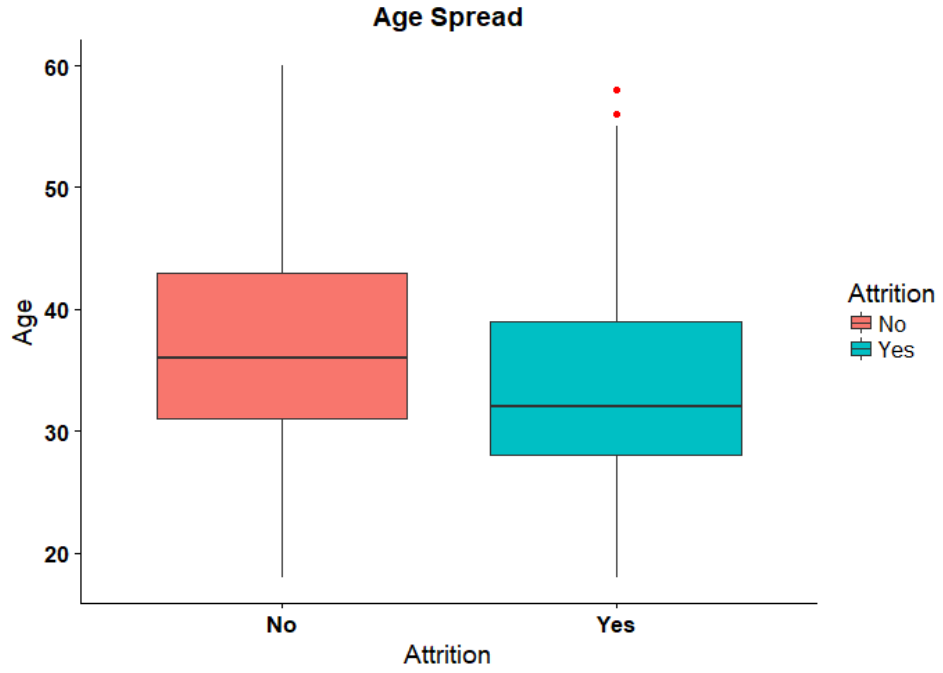


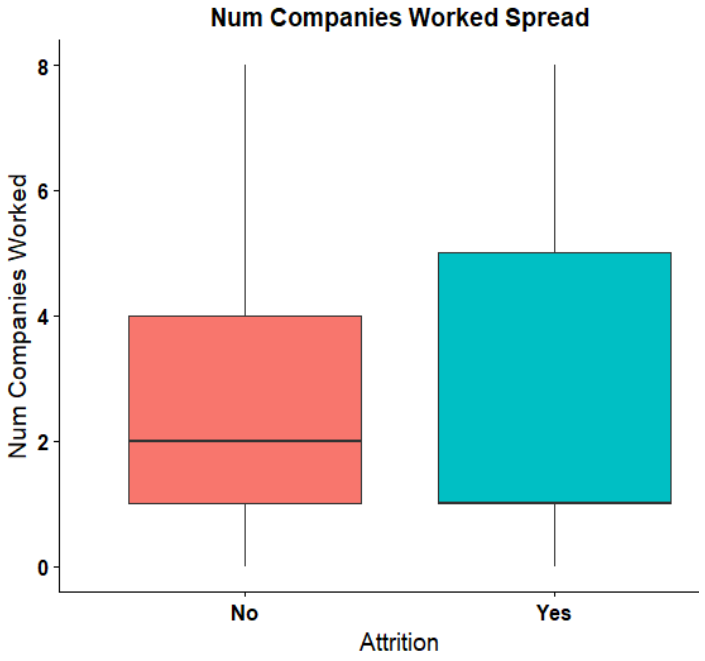


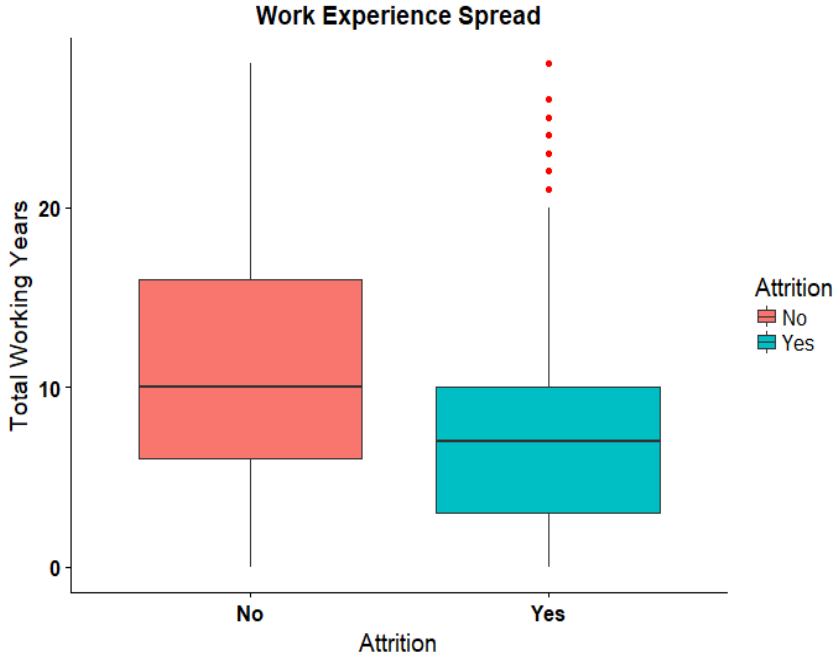
Job Role research director shows a little higher attrition rate.

Univariate Analysis – (Continuous)

- Age: Employees leaving the org definitely show a much lower median of Age.
- Fresher's and employees with less experience clearly have a higher tendency to leave.

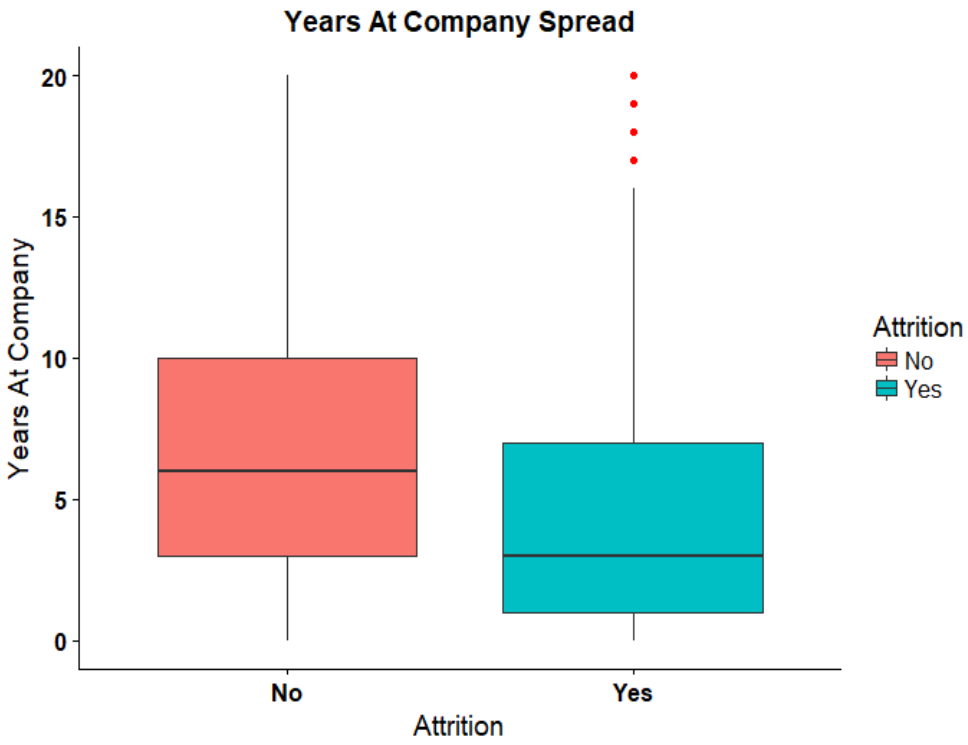


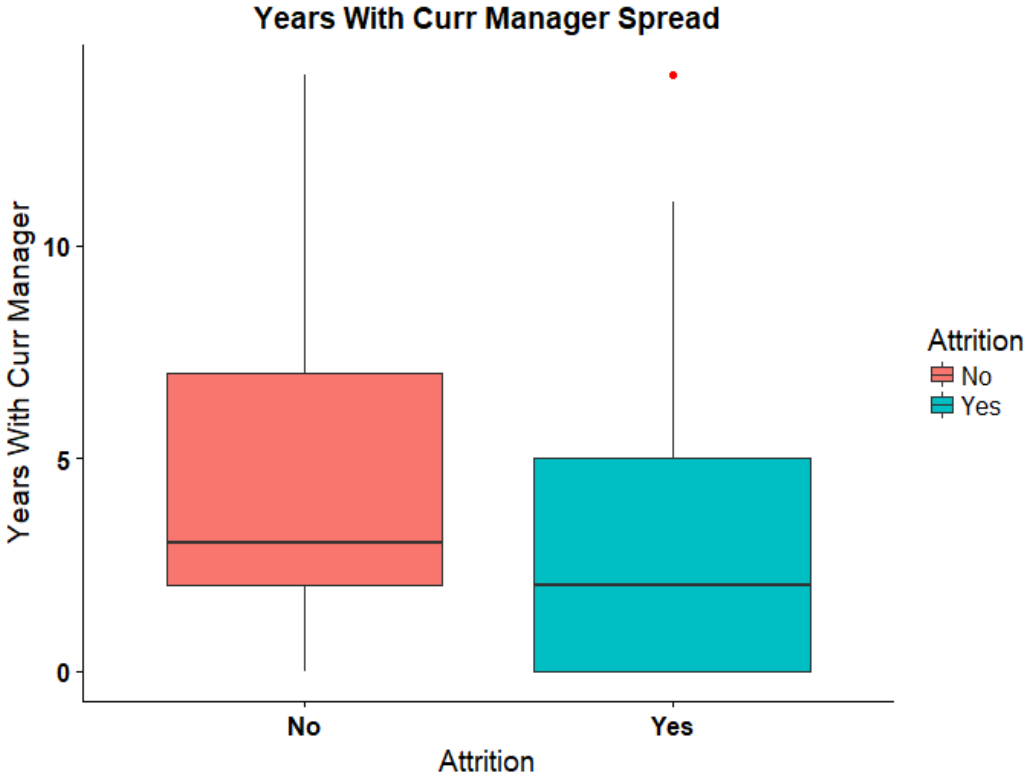




Univariate Analysis - (Continuous)

- Employees under the same manager for longer time tend to stay more in the company.



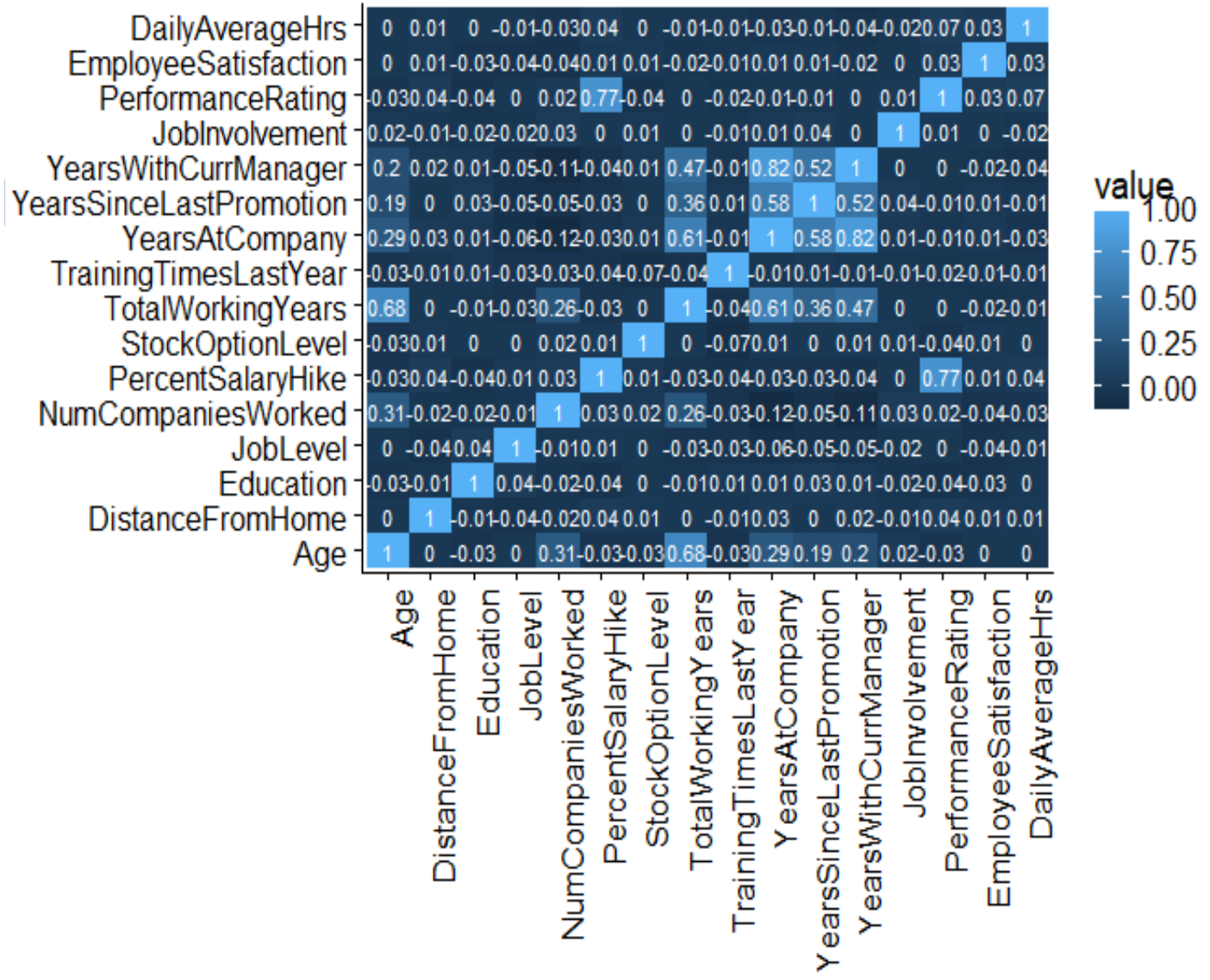


Employees leaving the company spend lesser time in the org as compared to ones that are not leaving.

Bivariate Analysis

- Conclusions:

1. Performance rating is highly correlated to salary hike.
2. Years with current manager and years at company are highly correlated.
3. Age, Total working years and years at company are highly correlated.



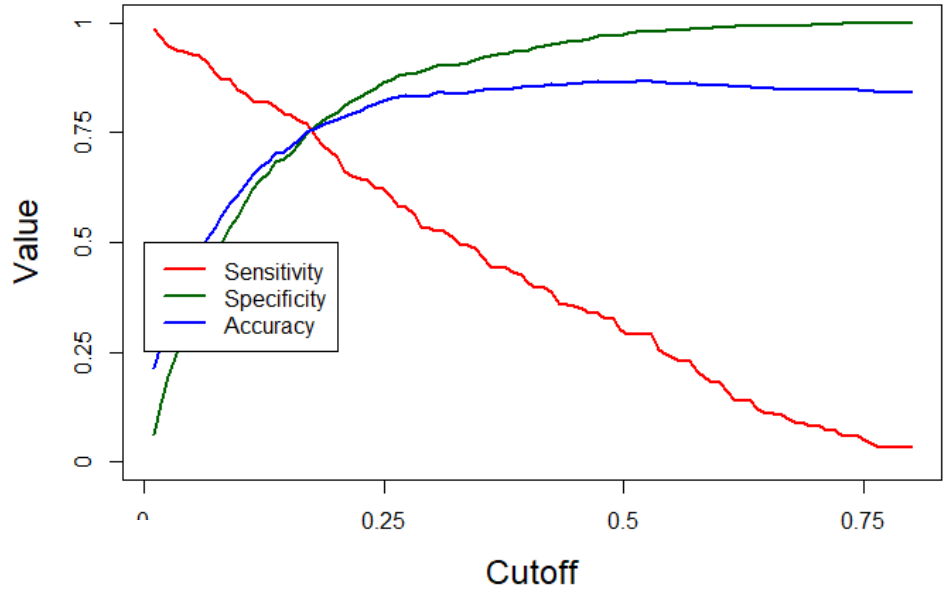
Logistic regression

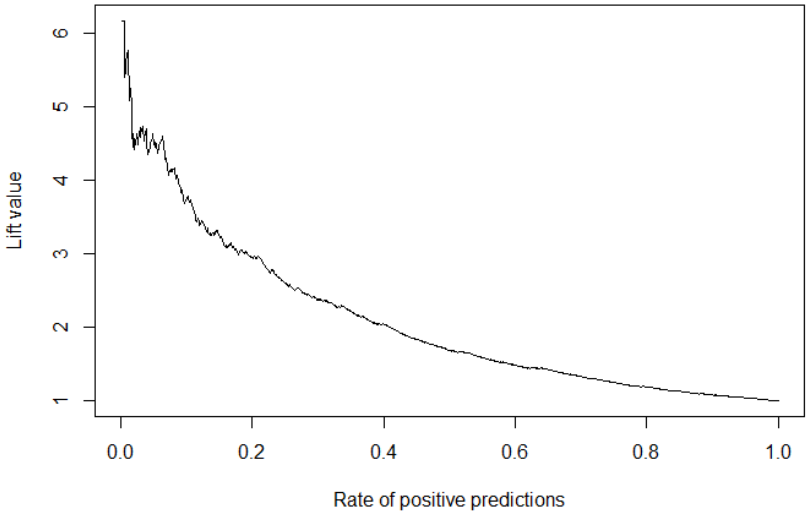
Steps:

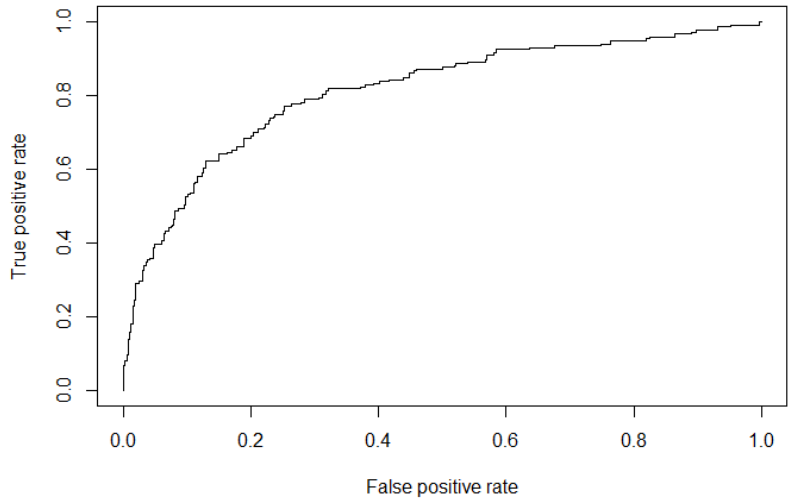
Gain And Lift Plots

Summary:

- The ideal cut off comes at 17.75%.
- ROC comes at 0.81 which is a good value.







Model Evaluation- K-Stats

1. Summary:

Decile	Observations	Attritions	Cum-Attrition	% Cum-Attrition	Non-Attrition	Cum-Non-Attrition	%Cum-Non-Attrition	Gain	Lift	(%Cum-Attrition) - (%Cum-Non-Attrition)
1	129	78	78	37.32%	51	51	4.72%	37.32057416	3.732057416	32.60%
2	129	45	123	58.85%	84	135	12.49%	58.85167464	2.942583732	46.36%
3	129	26	149	71.29%	103	238	22.02%	71.29186603	2.376395534	49.28%
4	129	21	170	81.34%	108	346	32.01%	81.33971292	2.033492823	49.33%
5	129	6	176	84.21%	123	469	43.39%	84.21052632	1.684210526	40.82%
6	129	10	186	89.00%	119	588	54.39%	88.99521531	1.483253589	34.60%
7	129	8	194	92.82%	121	709	65.59%	92.82296651	1.326042379	27.24%
8	129	4	198	94.74%	125	834	77.15%	94.73684211	1.184210526	17.59%
9	129	4	202	96.65%	125	959	88.71%	96.6507177	1.073896863	7.94%
10	129	7	209	100.00%	122	1081	100.00%	100	1	0.00%
Total	2110	209			1081					

Conclusions Recommendations

Variables affecting Employee Attrition