

# HR Analytics Case Study

## Group Members

Yogesh B S

Saishankar Palanisamy

Vartika Tewari

Ramakanth Joshi

# Problem Definition

A company with 4000+ employees has been facing a surprisingly high attrition rate of above 15% per year. The management believes that this level of **attrition** is bad for the company as, it leads to

- **Project delays** and eventually, loss of business
- **Loss of reputation** among accounts
- New **recruitment cost** and Training and Development expenses
- Negative motivation levels among current employees

The objective of the case study is to developing a model of the "Probability of Attrition" using Logistic Regression to zero down to the factors to focus on for curbing attrition rate from growing and hence,

- Understand the changes which are to be made in workplace environment to **minimise attrition**
- Prioritize among **factors attributing to high attrition** to attend them immediately.

## Assumptions:

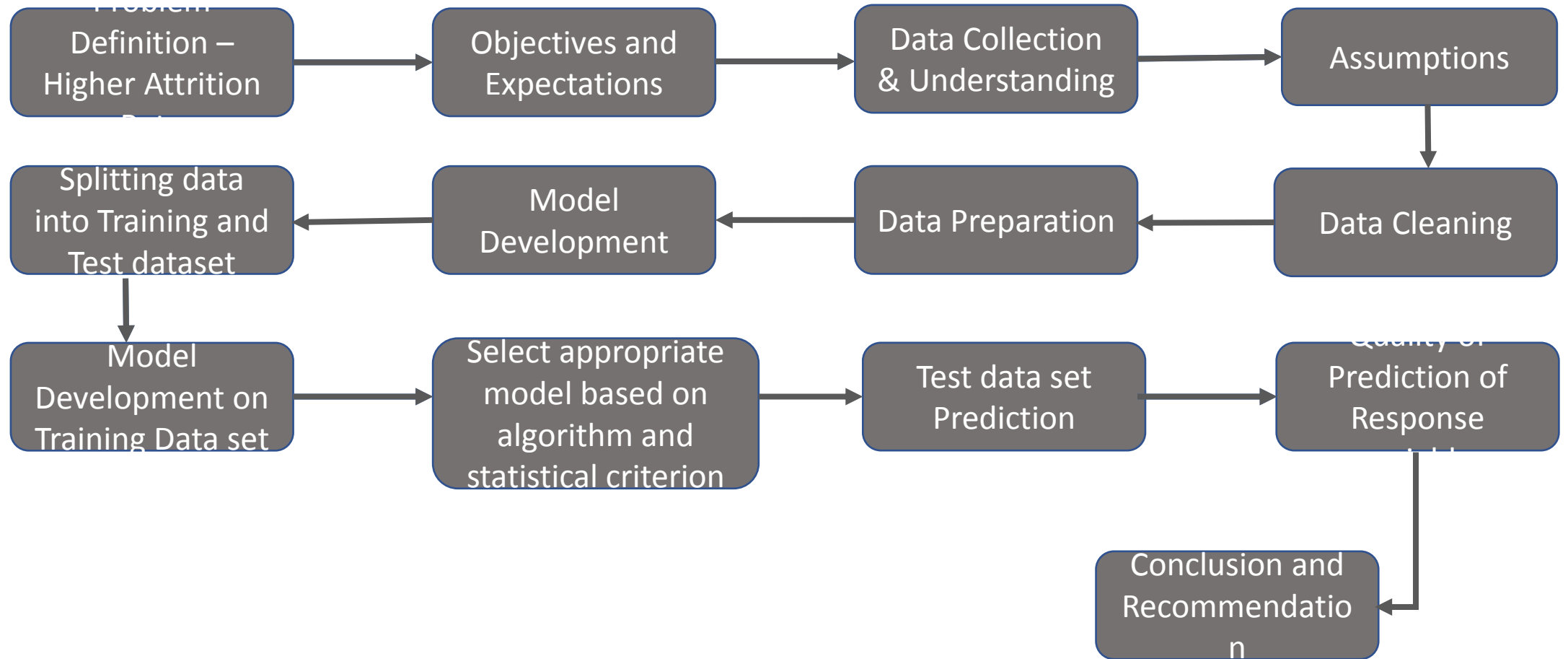
- The missing values for “No of companies worked” is filled by approximate values based on age as there is a high correlation among them.
- “Number of working years” is approximately assumed as actual age – 25 years.

## Deliverables:

The **Attrition Probability Model** based on logistic regression, which will help attaining results to

- Understand major drivers (factors) for high attrition rate
- Identify the fine tuning, which is required in workplace settings, to keep the employees from quitting.

# Problem solving approach



# Analysis Result

## Driver Variables :-

### Positive Impact Variables

(lowers attrition)

- Age
- Department
- Total Working Years
- Training Times Last Year
- Years With Current Manager
- Environment Satisfaction
- Job Satisfaction
- Work Life Balance

### Negative Impact Variables

(Increases attrition)

- Business Travel
- Marital Status
- Count of Companies Worked
- Years Since Last Promotion
- Work Load

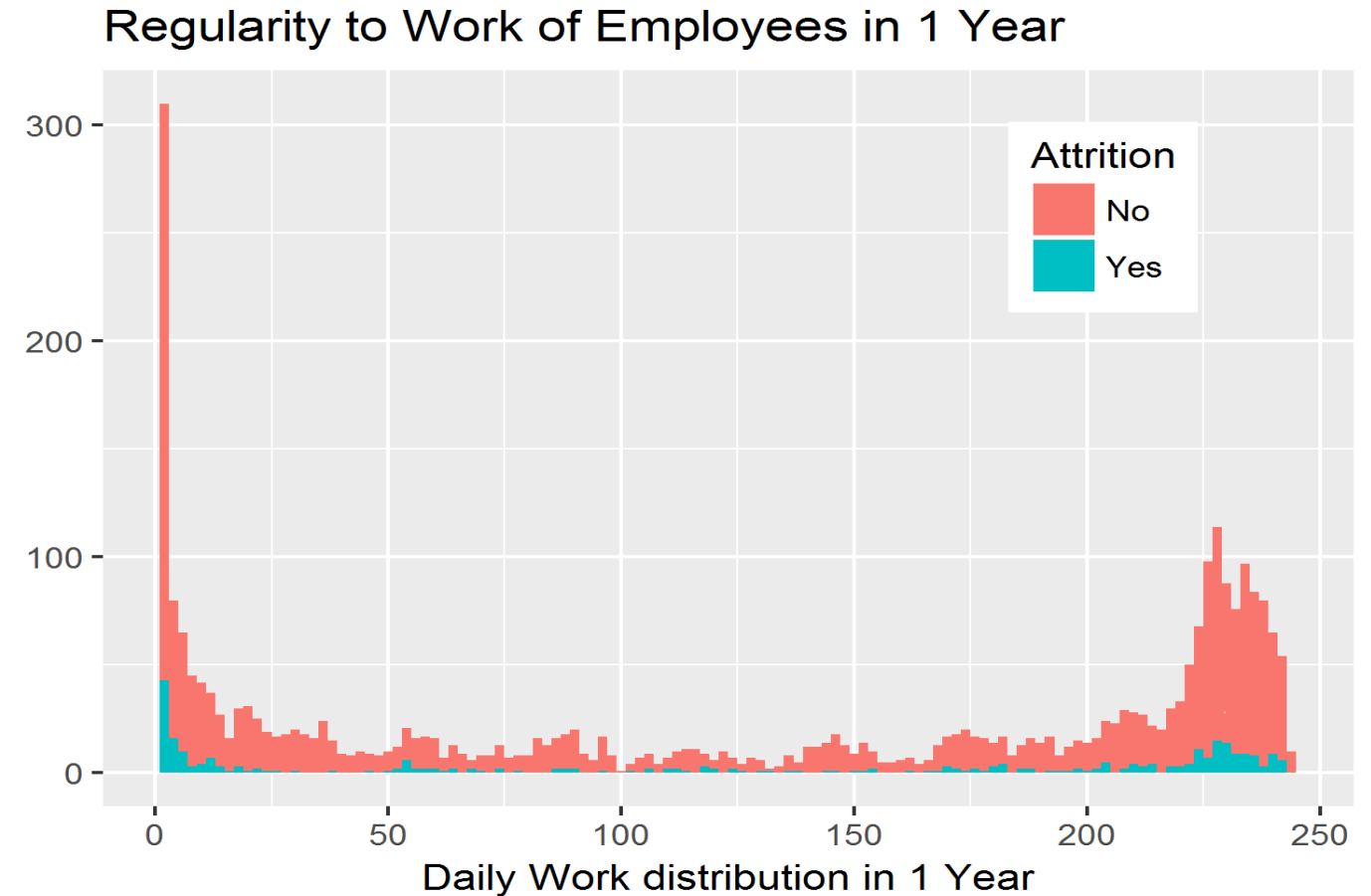
## Derived Variables:

- Vacation of each employee
- Mean attendance
- Work Load Check
- Work regularity check

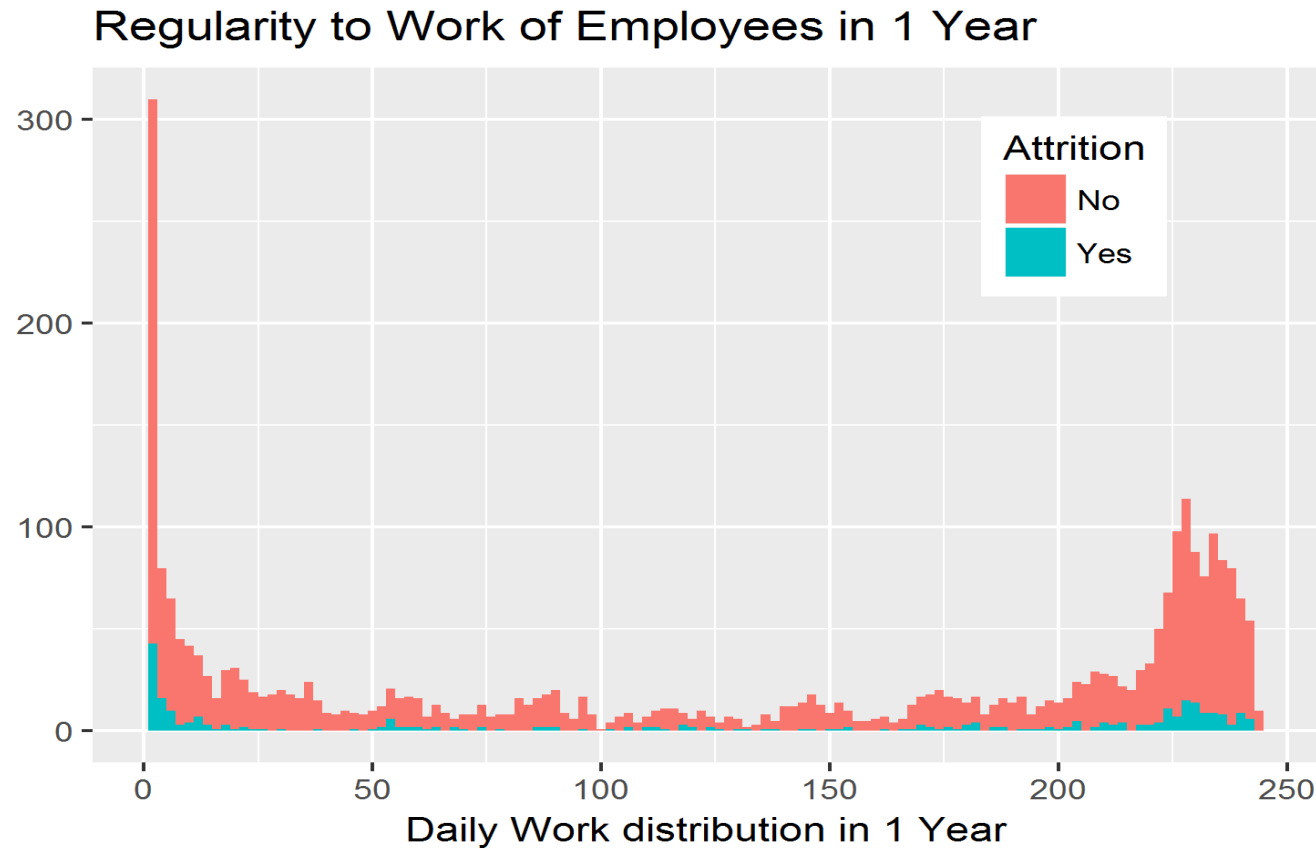
# Attrition by Daily Work Distribution

## Inference:

Ignoring the high spike at 0 which is due to null values in data, the **attrition shows high values for both very low and very high work distribution.**



# Attribute by the Regularity at Workplace



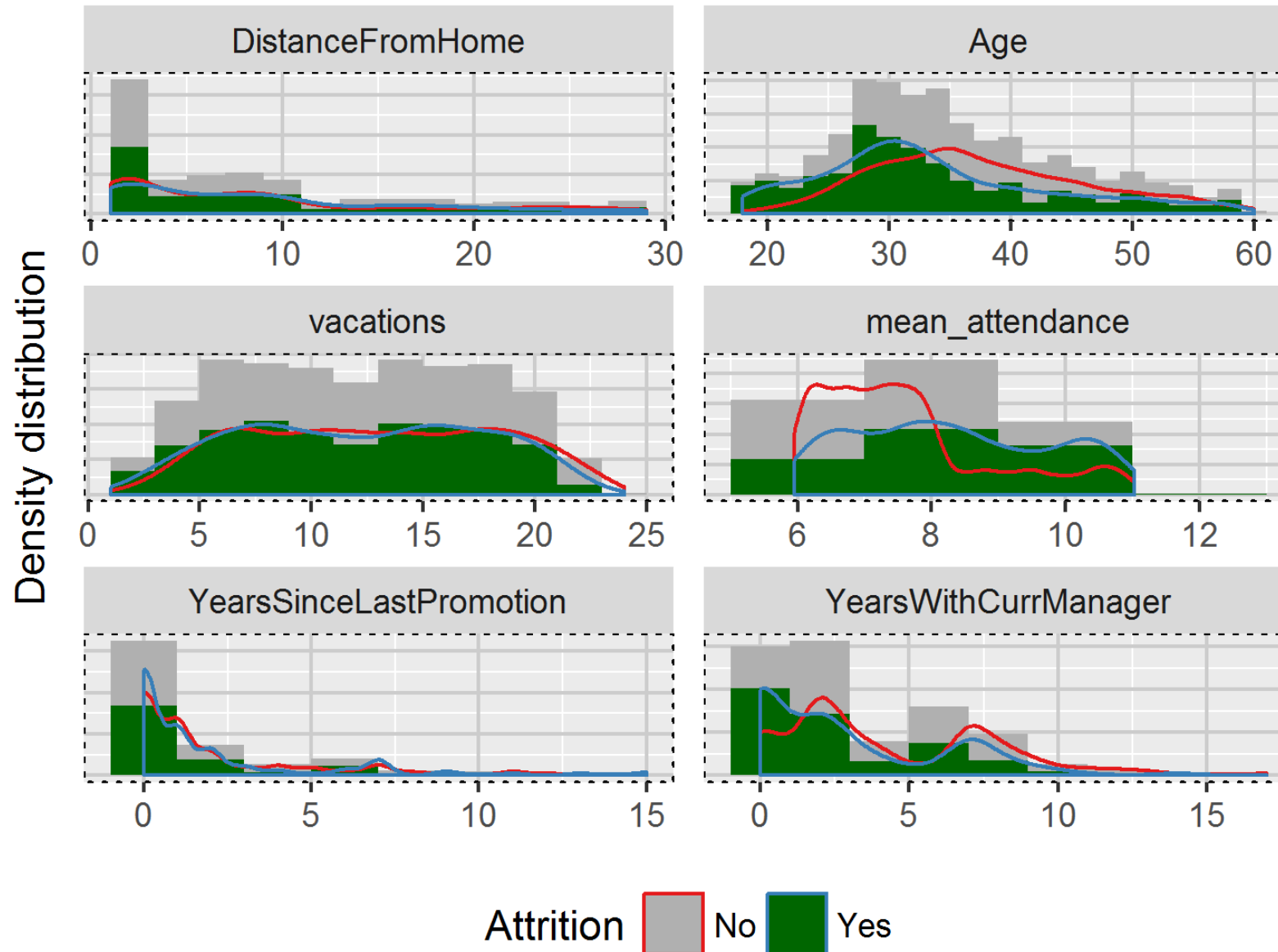
## Inference:

The trend shows that the employees who took most leaves and those who took least leaves are both prone to leave the employer.

The rest, with moderate amount of leaves taken are less likely to give up their employment with the concerned employer.

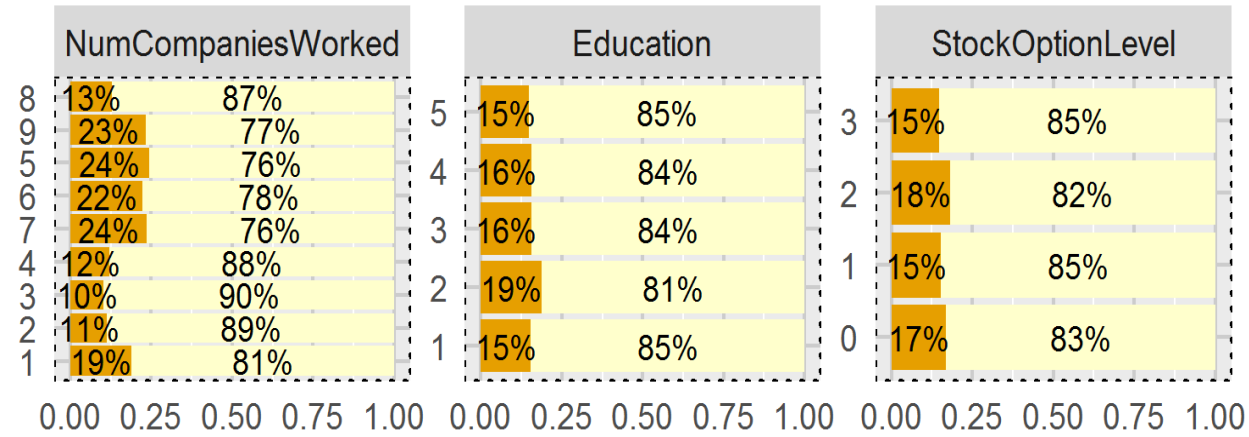
# Percentage attrition of continuous variables

## % Attrition of Continuous variables

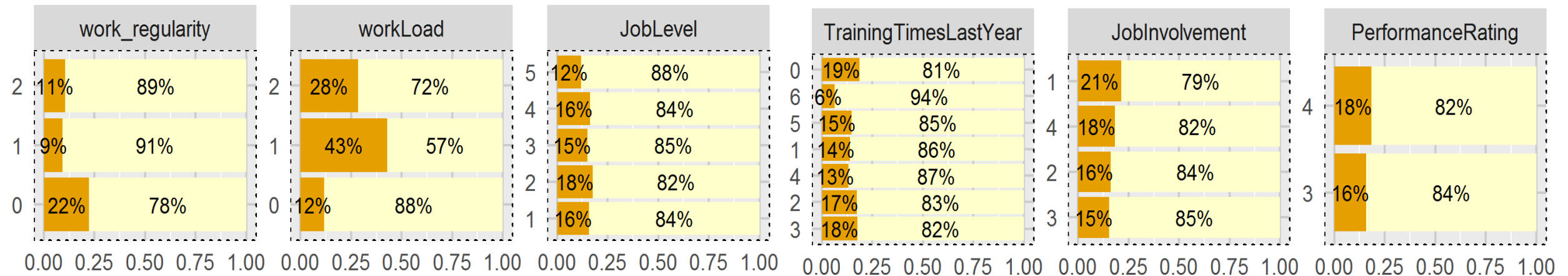
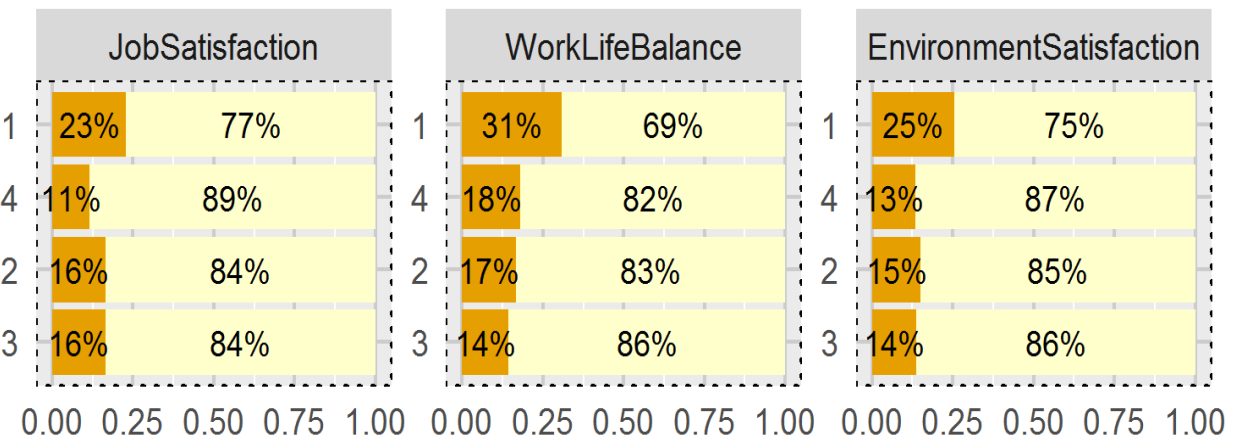




## Percentage Attrition by each Ordinal category 1



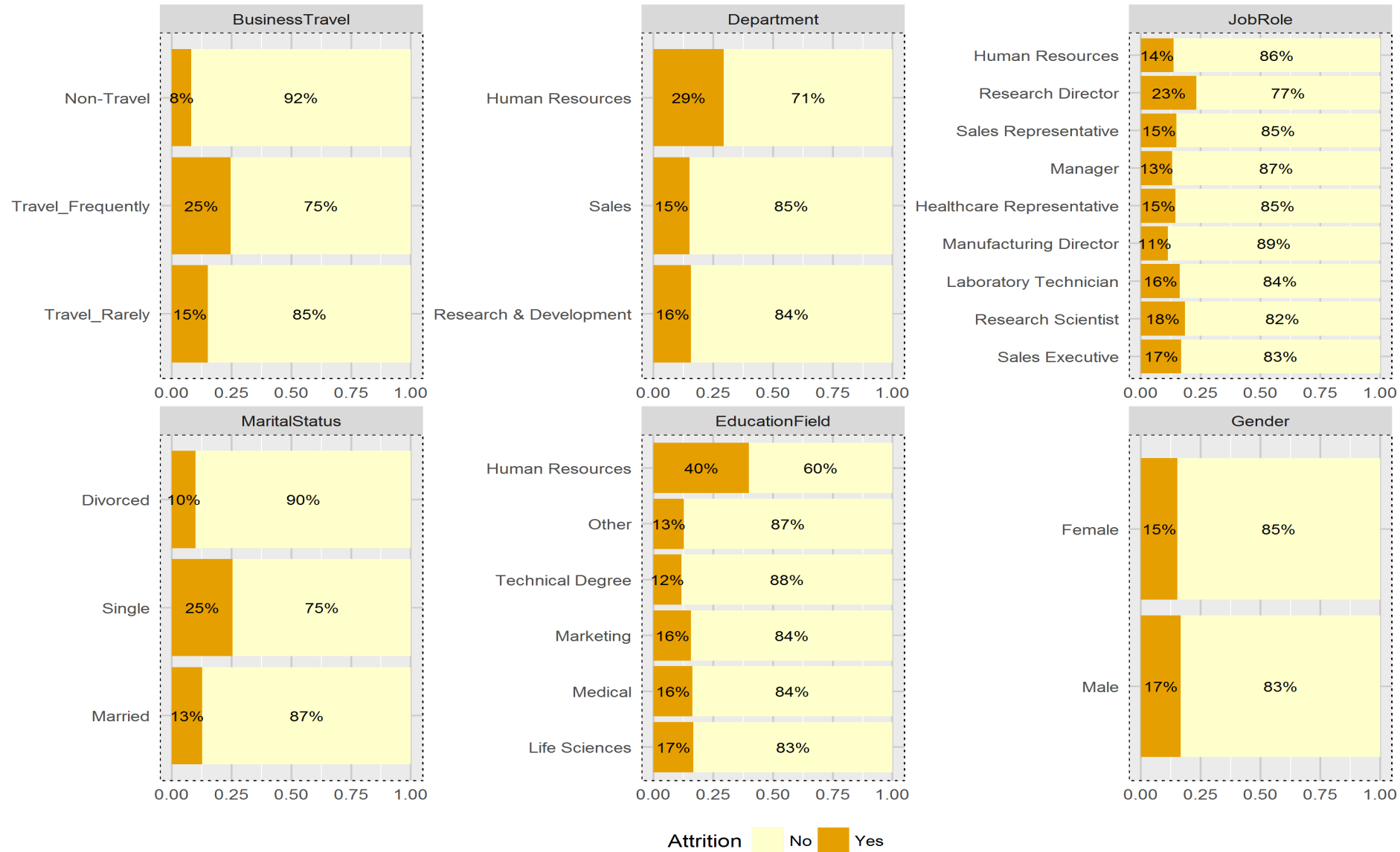
## Percentage Attrition by each Ordinal category 2



Attrition  No  Yes

Attrition  No  Yes

Percentage Attrition by each Nominal category

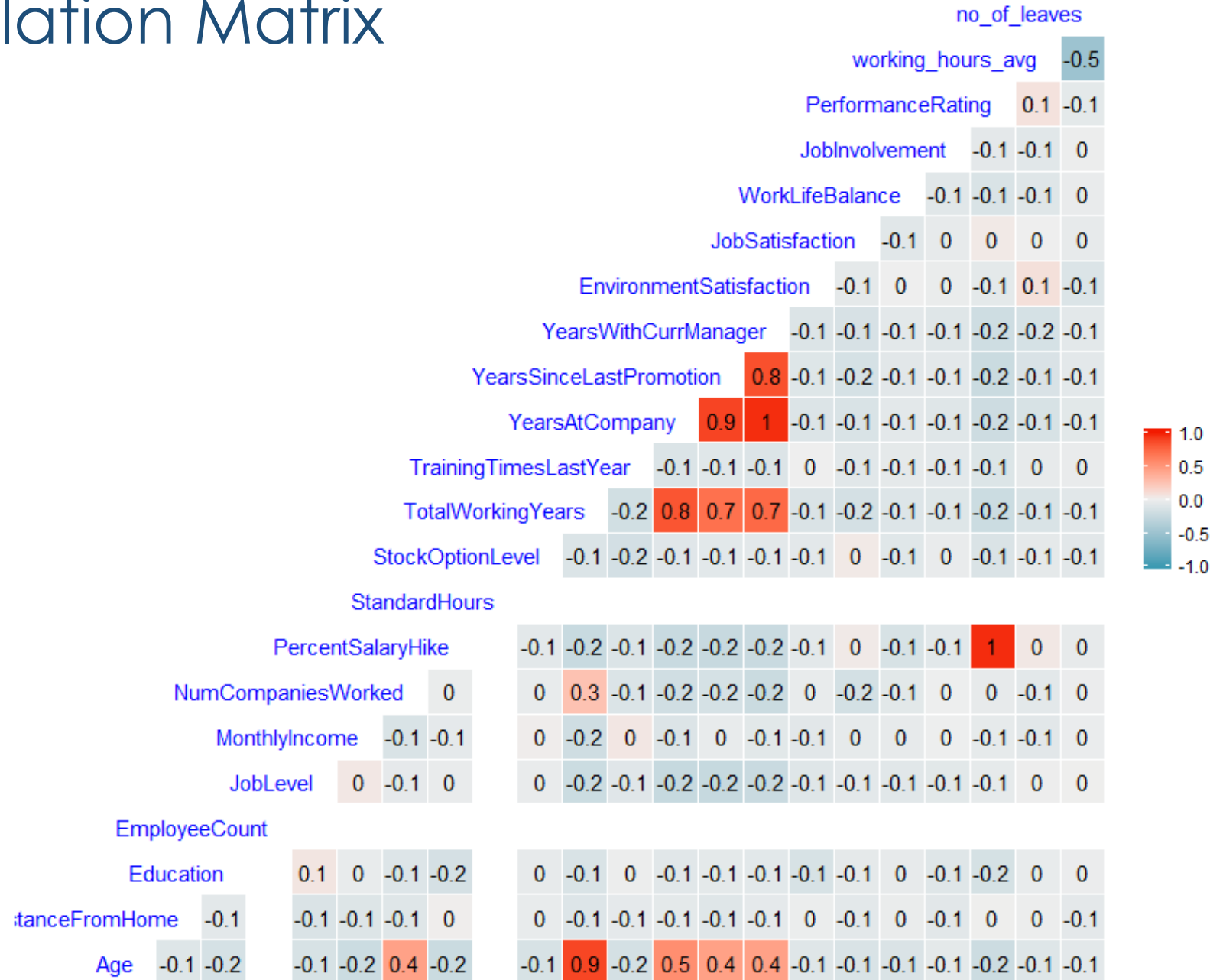


# Correlation Matrix

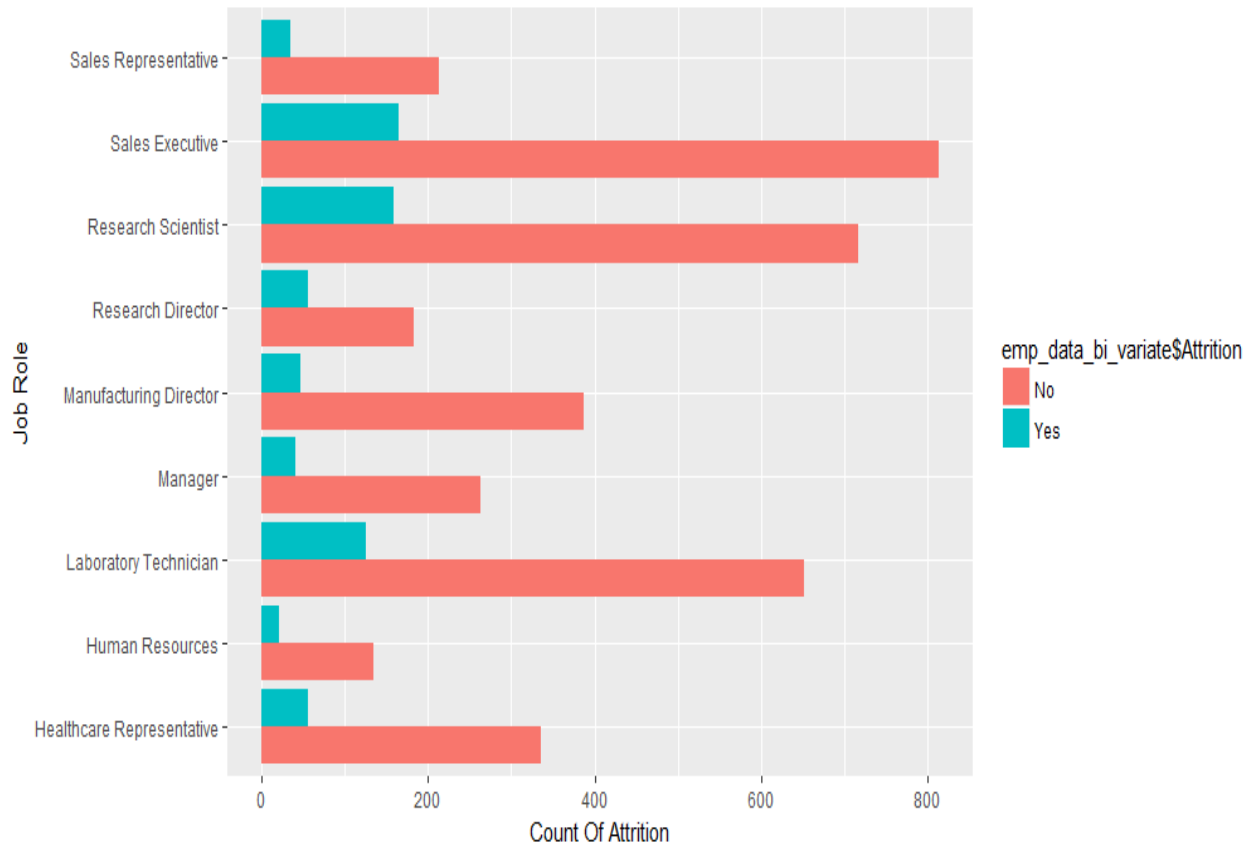
## Inference:

The correlation matrix, alongside, is built to find highly related variables.

The cubes in red show high correlation between the variables, namely, Performance rating and Percent Salary Hike and others.



# Attrition by the Job Role



Inference:

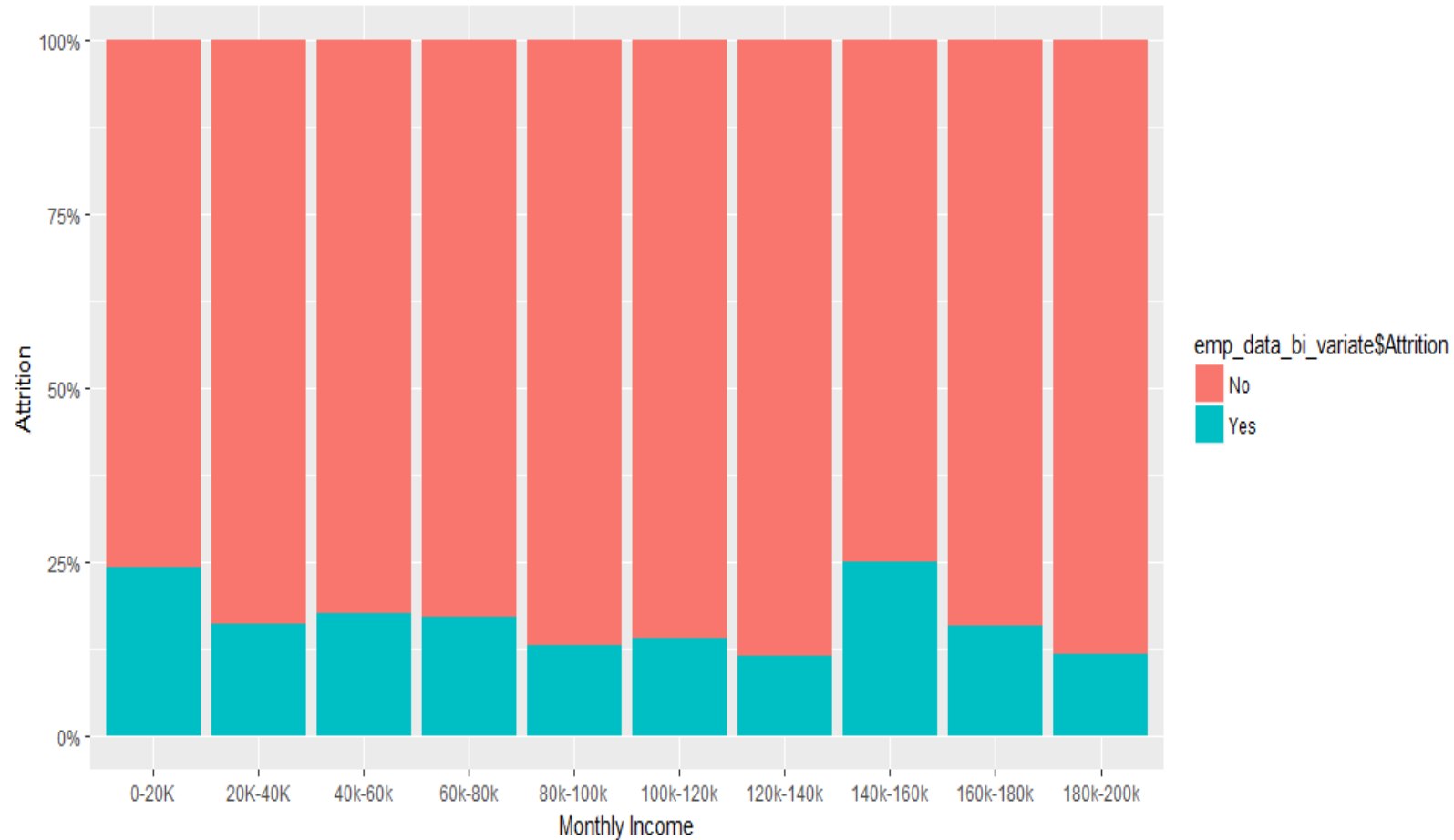
Of all the Job roles, the **sales executive** and the **research scientist** tend to show most records of attrition.

# Attrition by the Monthly Income

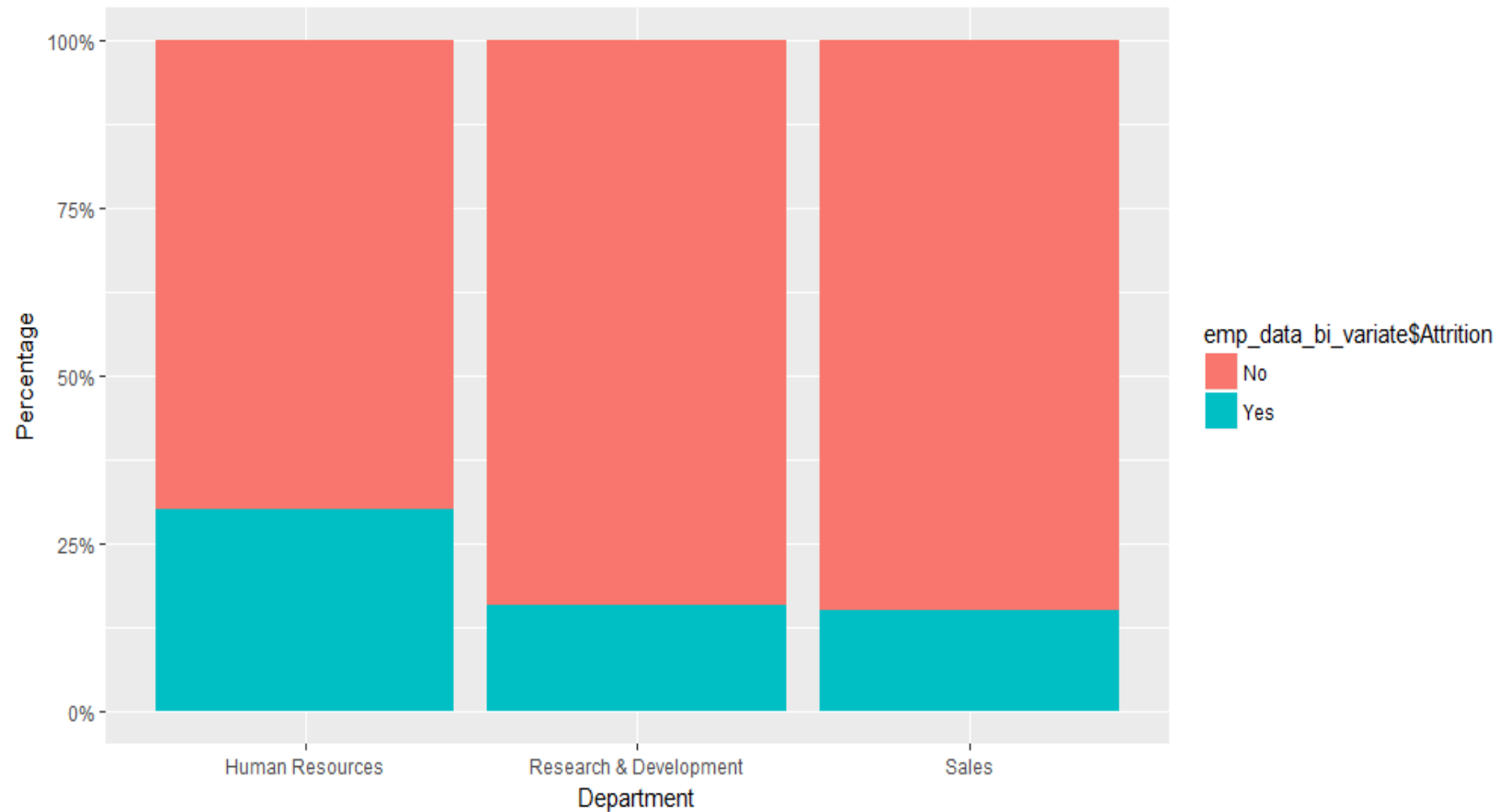
Inference:

With the **increasing income**, the **attrition rate appears to reduce**.

Also, the bin with salary range from 140k – 160k is an exception, and shows attrition even at a high salary.



# Attrition by the Department



Inference:

Ironically enough, the Human Resources department registers the maximum number of attrition.

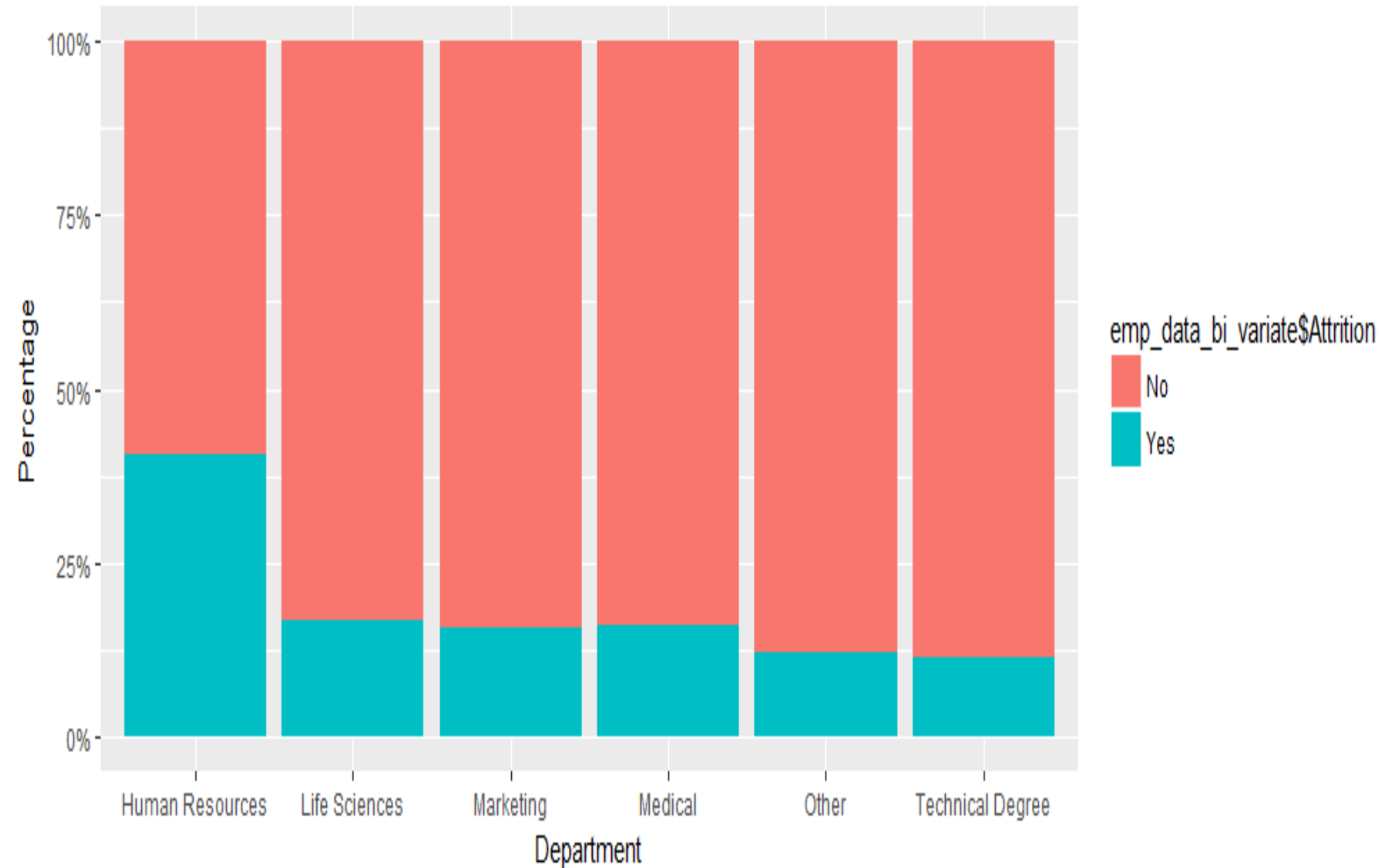
The Research and Development and the Sales departments are almost the same in the percentage of attritions registered.

# Attrition by the Education Field

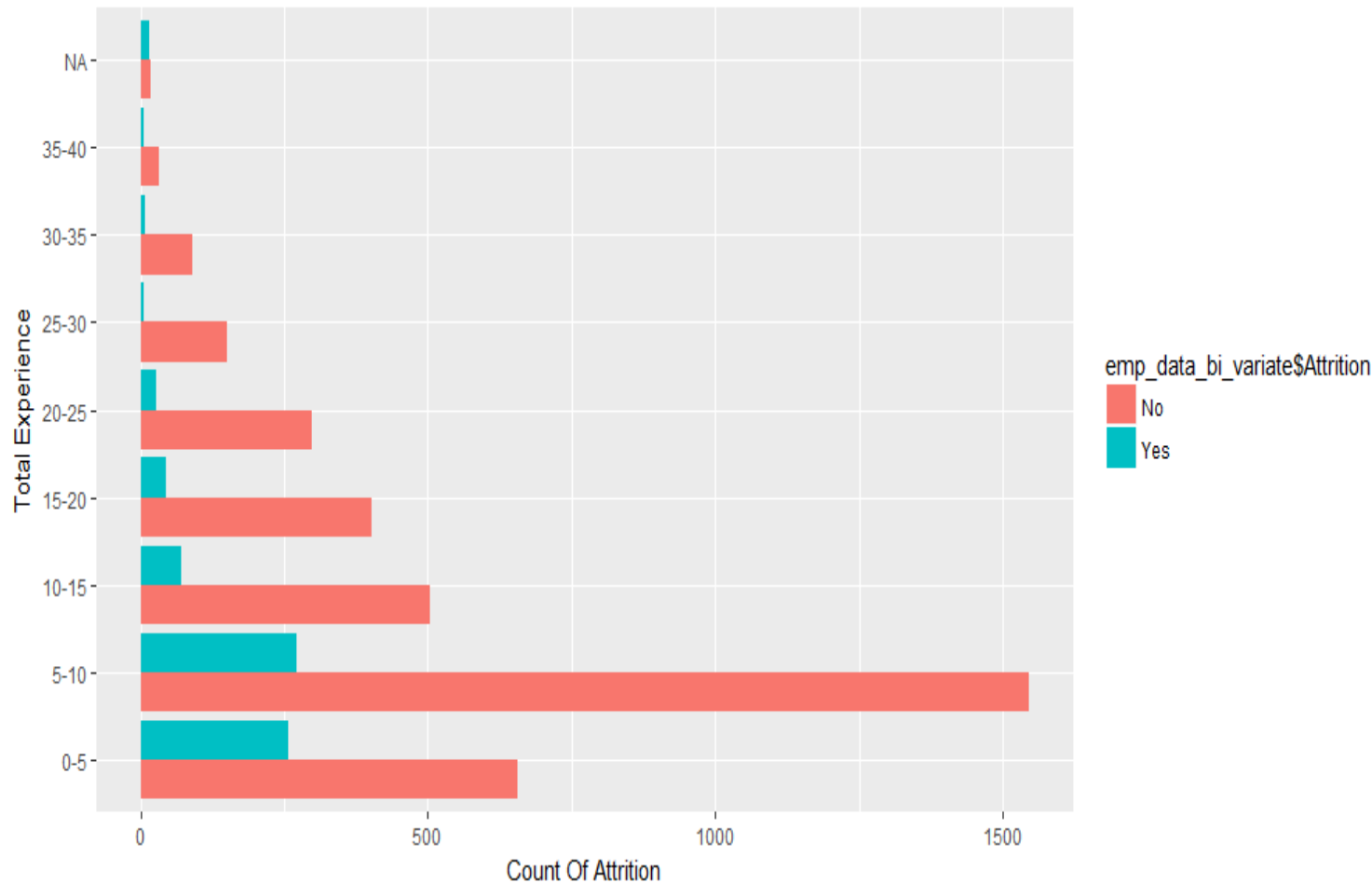
## Inference:

As expected, the Human Resources educated employees register the maximum number of attrition – almost twice of the average of all other departments.

The Medical students lie next as the people to leave jobs in the company; although not much difference exists for all the other departments.



# Attrition by the Years of Experience

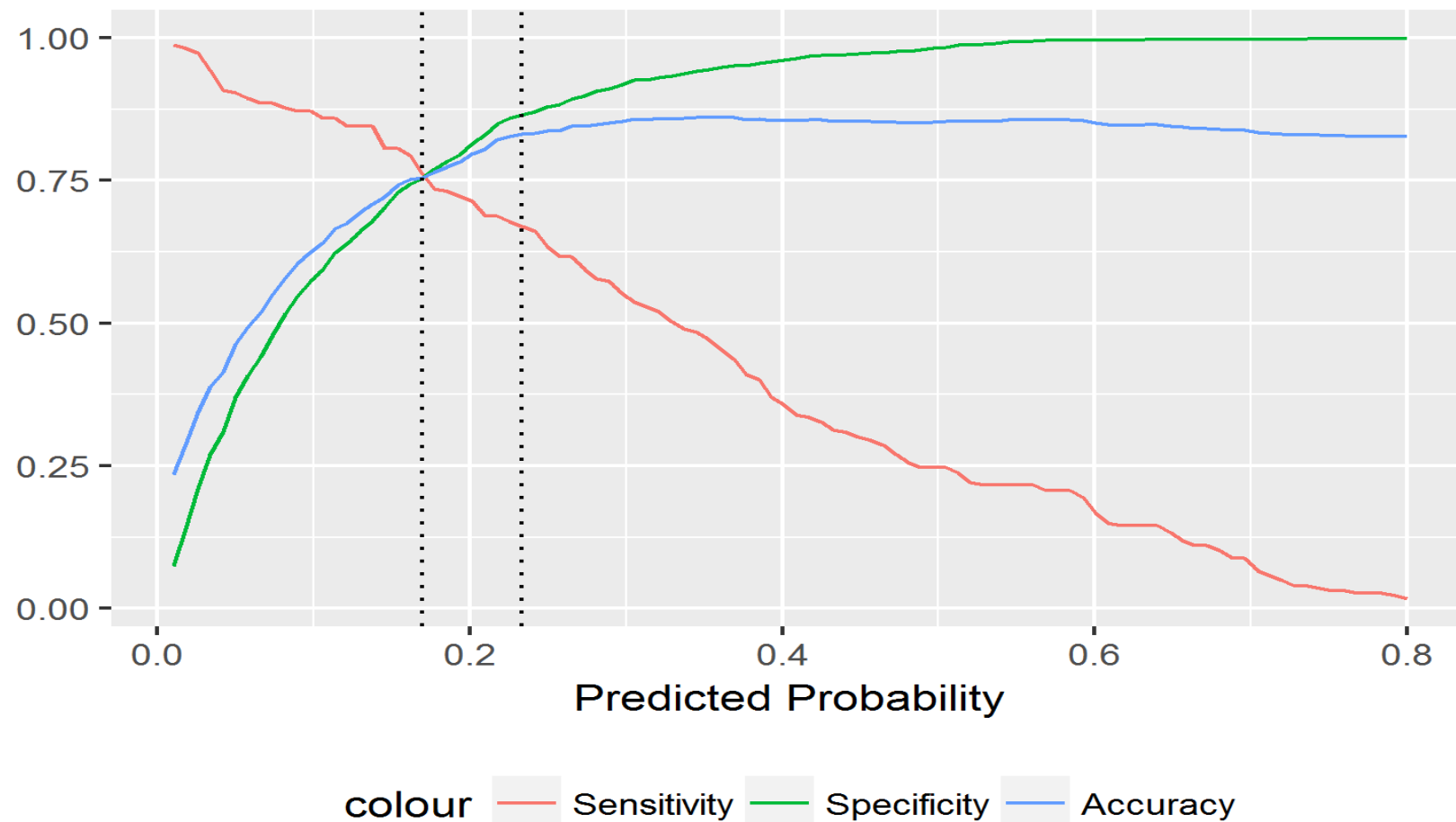


Inference:

Of all employees with varied years of experience, the ones with **lower years of experience prefer to leave jobs more**, probably to switch or due to unfavorable environments.



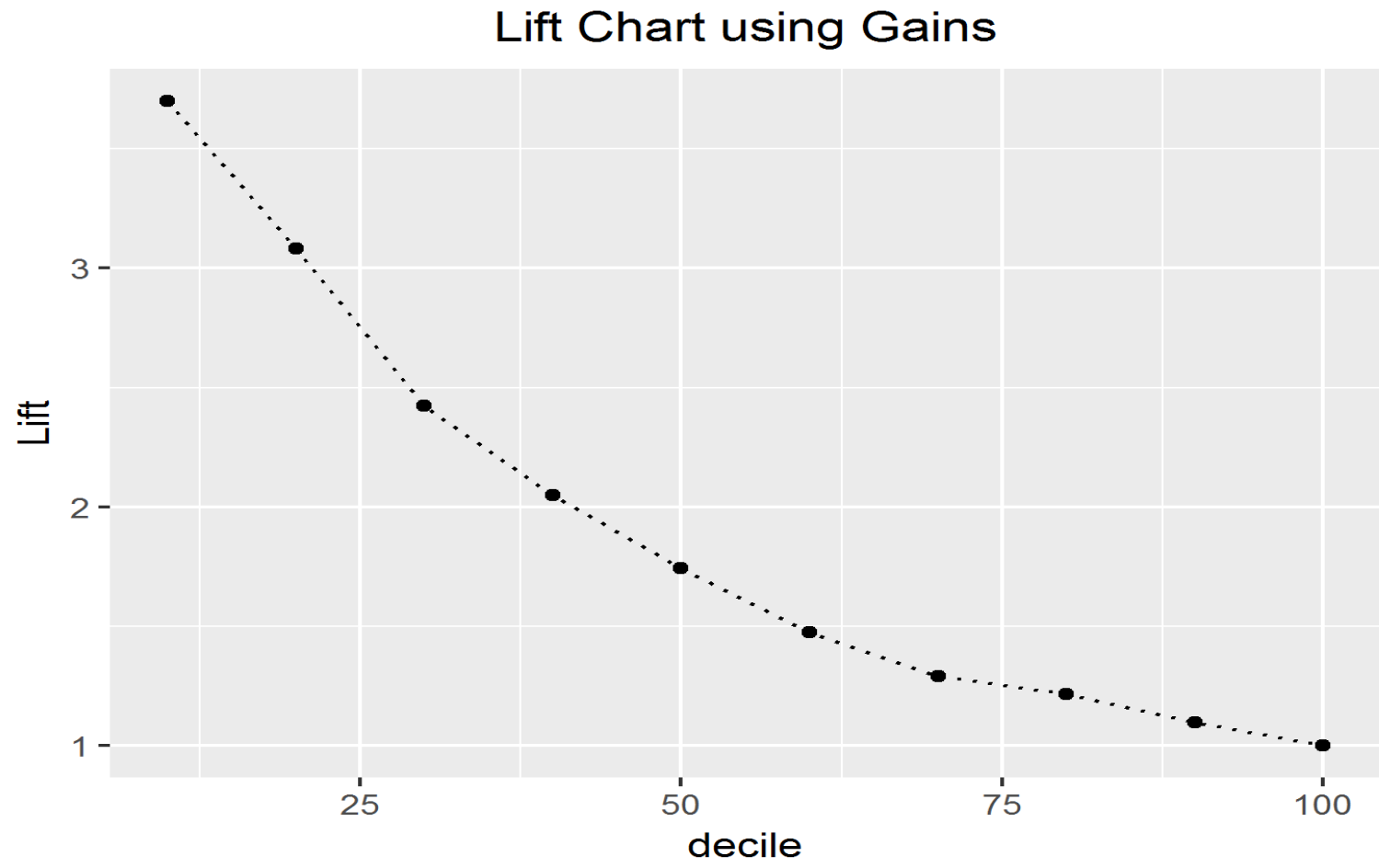
## Sensitivity, Specificity, and Accuracy



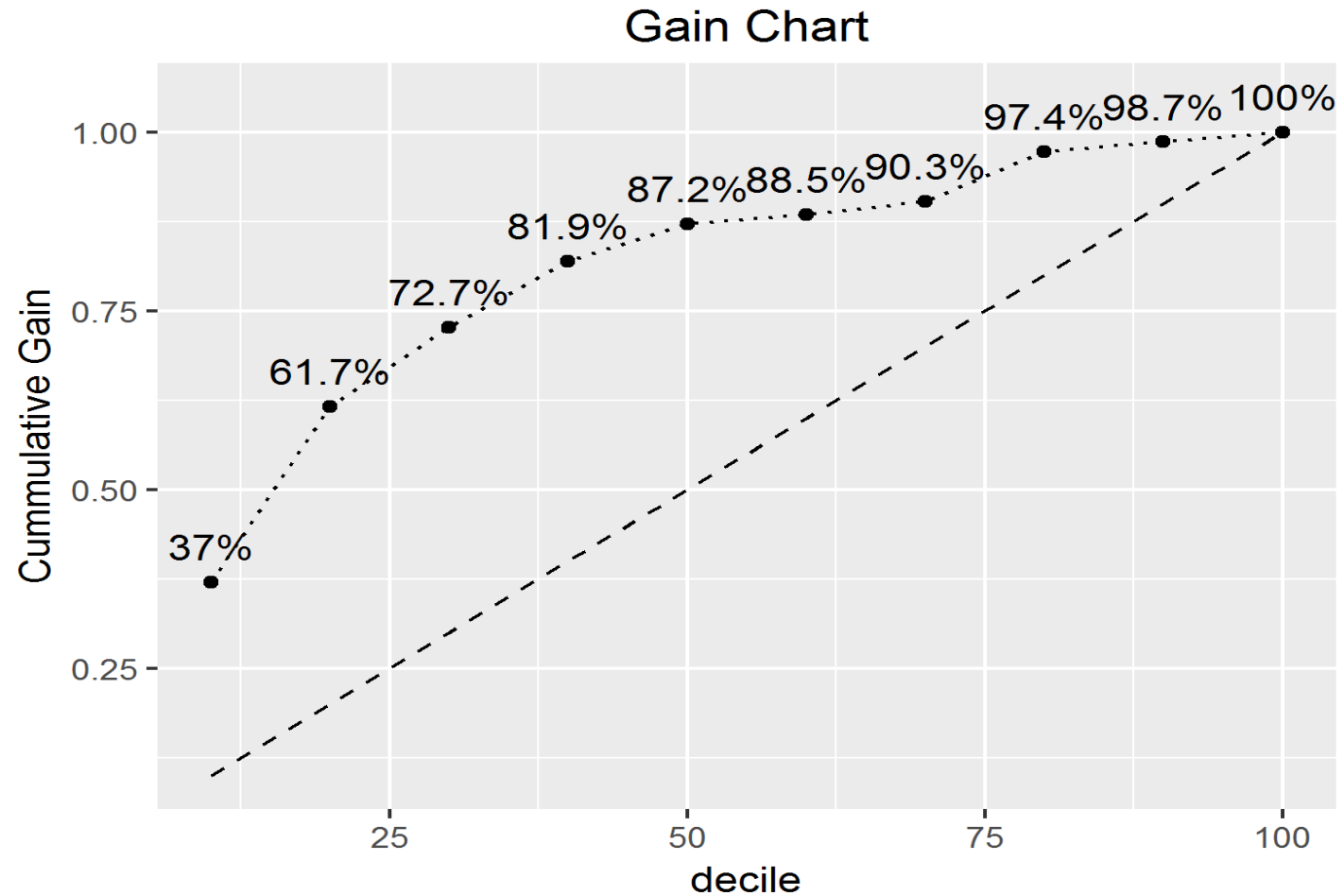
Outcome:

Optimum probability cutoff  
= 0.16, at 0.16, accuracy,  
sensitivity and specificity  
will be 75 percent

# Lift Chart using Gains



# Cumulative Gain per Decile



# Logistic Regression Model

$$\ln \frac{p}{1-p} = \beta_1 X_1 + \beta_2 X_2 \dots \dots + \beta_n X_n$$

The snippet shows the finally included, **decisive factors** in variables provided. These variables reportedly predict, with the most accuracy, the probability of attrition for each employee.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-0.7807	0.1610	-4.85	1.2e-06	***
Age	-0.2940	0.0785	-3.74	0.00018	***
NumCompaniesWorked	0.2741	0.0590	4.64	3.4e-06	***
TotalWorkingYears	-0.5185	0.1072	-4.84	1.3e-06	***
TrainingTimesLastYear	-0.1973	0.0563	-3.50	0.00046	***
YearsSinceLastPromotion	0.5495	0.0746	7.36	1.8e-13	***
YearsWithCurrManager	-0.5295	0.0845	-6.26	3.8e-10	***
working_hours_avg	0.5742	0.0529	10.86	< 2e-16	***
MaritalStatusSingle	1.0776	0.1130	9.53	< 2e-16	***
EnvironmentsSatisfaction2	-0.7961	0.1677	-4.75	2.1e-06	***
EnvironmentsSatisfaction3	-0.7886	0.1508	-5.23	1.7e-07	***
EnvironmentsSatisfaction4	-1.1240	0.1567	-7.17	7.3e-13	***
JobsSatisfaction2	-0.5534	0.1637	-3.38	0.00072	***
JobsSatisfaction3	-0.6675	0.1473	-4.53	5.8e-06	***
JobsSatisfaction4	-1.1707	0.1569	-7.46	8.5e-14	***
workLifeBalance3	-0.5111	0.1112	-4.60	4.3e-06	***



# Recommendations for management

## -conclusion



**Avg\_working\_hours** : The coefficient ( $\beta = 0.5495$ ) for working hours reiterates that as the working hours increases the rate of attrition also increases. This will help management to initiate necessary steps such as increasing the head count, making the work environment conducive.

**Years\_since\_last\_promotion** :  $\beta = 0.5495$ , Timely promotions of employee may reduce the attrition as indicated by the logistic regression model.

**Matril\_status\_single** :  $\beta = 1.07$ . Interestingly, unmarried people have low rate of attrition. Management cant do much about it!

**Job satisfaction/Environment satisfaction** : GLM confirms the obvious, higher the job/Environment satisfaction, lesser the attrition.

**Year\_with\_current\_manager** :  $\beta = -0.5295$ , as the number of years under same manager increases the attrition rate increases. Implies, the level of interest decreases as the tenure under the manager increases.