

# HR Analytics Case Study

Presenters:-

Anindya Chatterjee

Debabrata Saha

Ramya Mulukutla

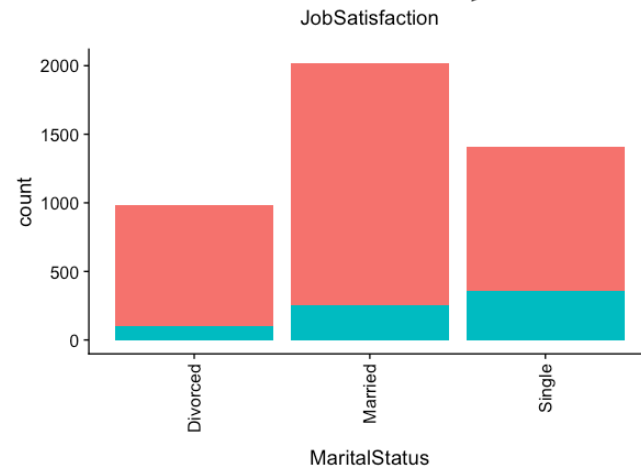
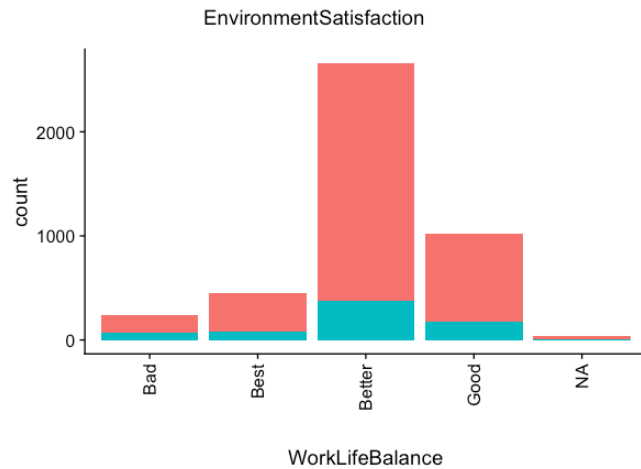
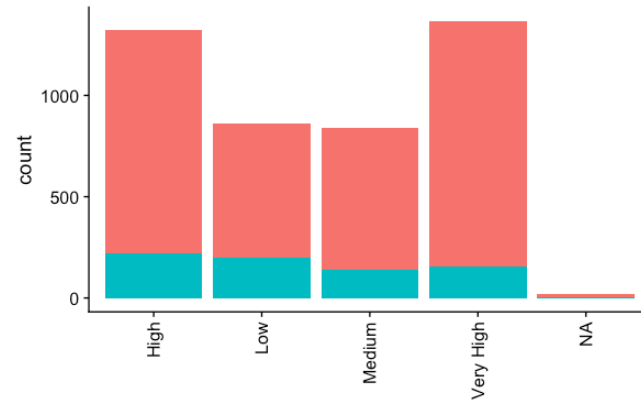
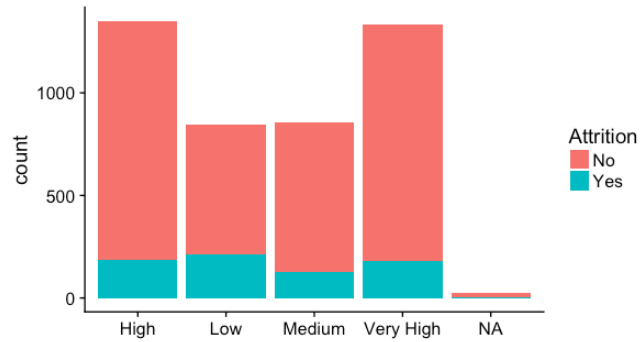
# Business Requirements

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, because of the following reasons -

- The former employees' projects get delayed, which makes it difficult to meet **timelines**, resulting in a reputation loss among consumers and partners
- A sizeable department has to be maintained, for the purposes of **recruiting** new talent
- More often than not, the new employees have to be **trained** for the job and/or given time to acclimatise themselves to the company

# Data Understanding

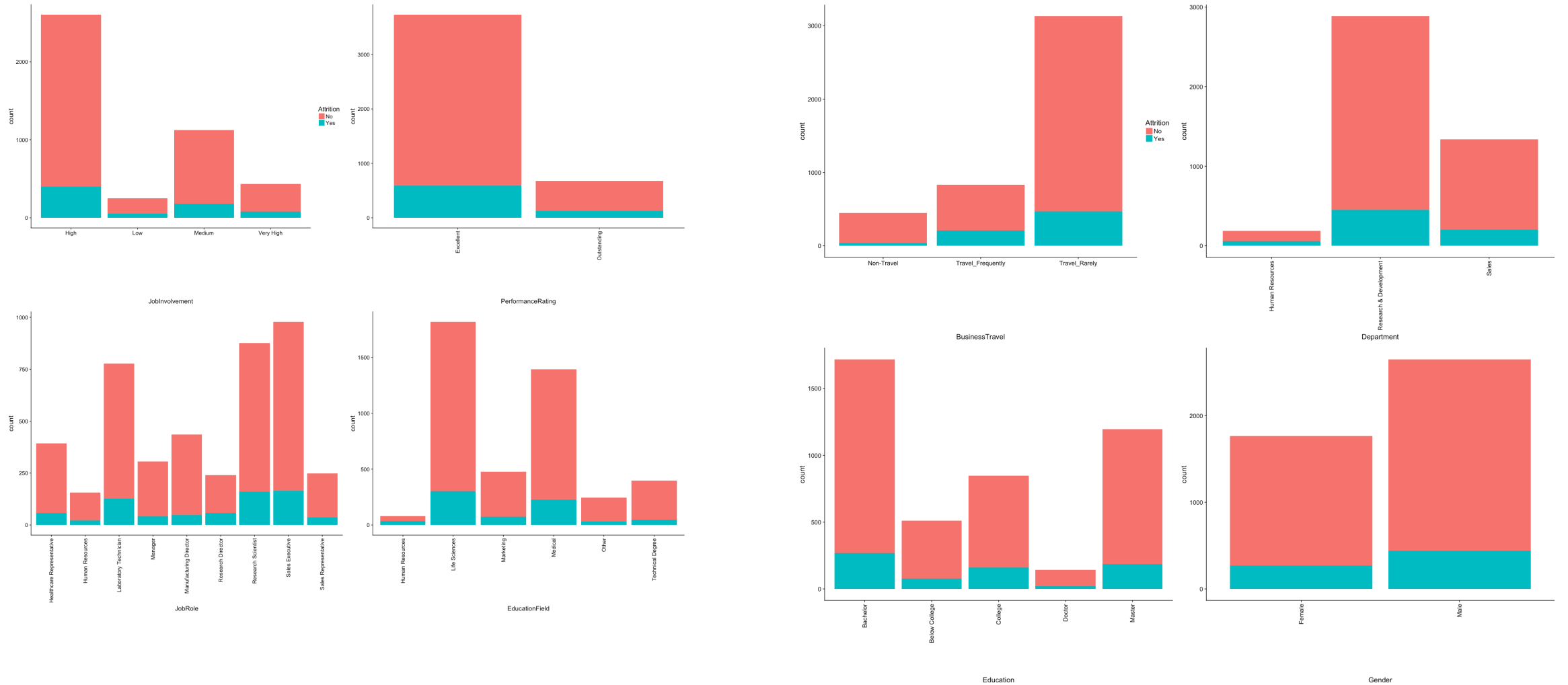
- Employee data is provided from 5 different sources.
- The files were merged, and checked for quality issues and duplicate records
- Employee average working time is calculated for using it further in the model building stage
- Converting numerical attributes to categorical values from data dictionary provided



# Exploratory Data Analysis

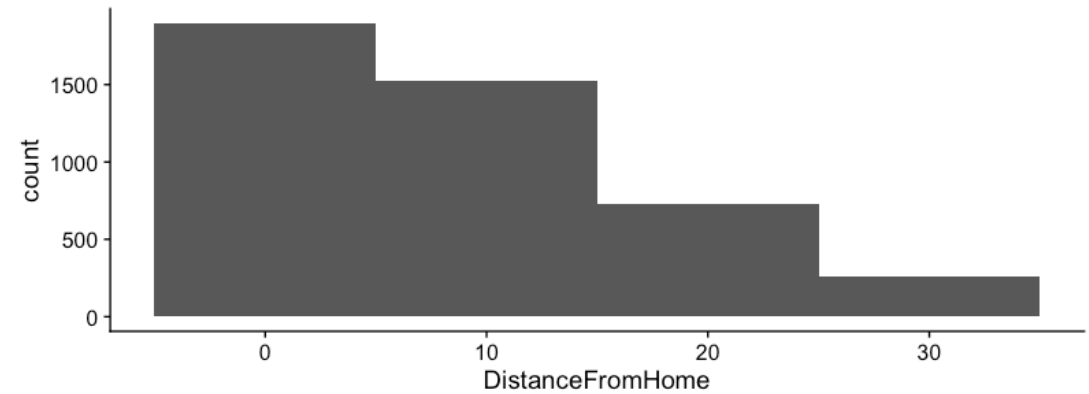
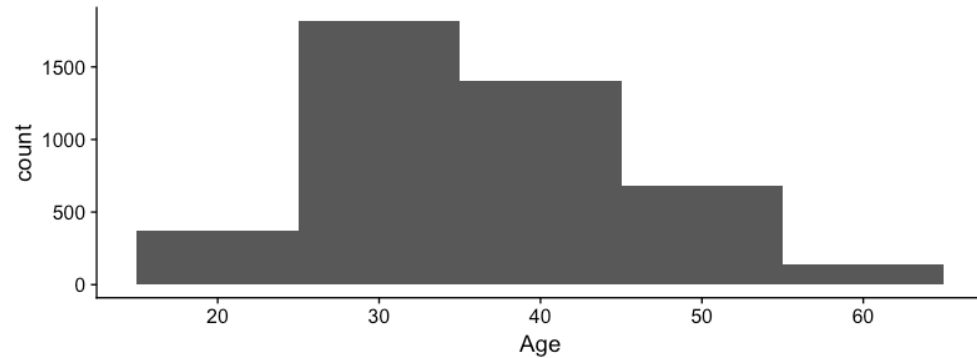
Bar graphs are plotted for categorical variables

# Exploratory Data Analysis



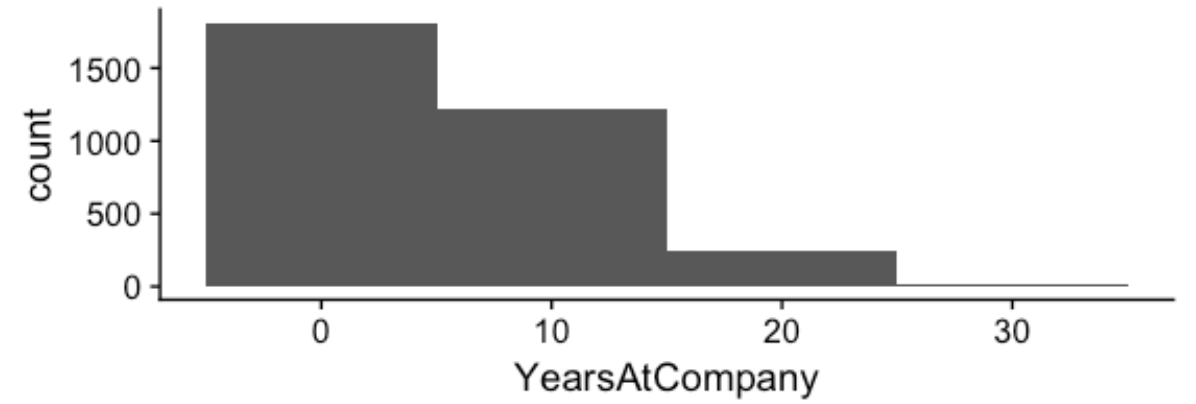
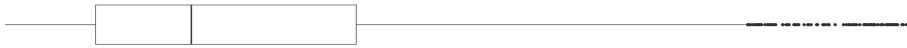
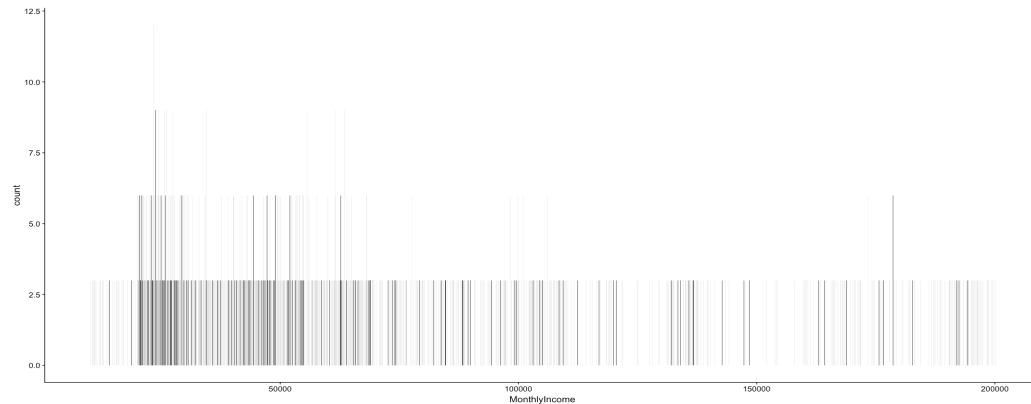
# Exploratory Data Analysis

For numerical variables box plots and histograms are created



# Exploratory Data Analysis

Also outliers were identified and were removed as a part of this process



## EDA - Observations

- Employees tend to leave the company when they are Single
- Employees leave when performance rating is excellent, job involvement is High, belong to the R&D department, and rarely travel
- Employees leave at higher percent when they are Research Scientist and sales executives



# Modelling

- After data cleaning is done and data is ready to be modelled , the attrition % is calculated and it came out to be 16.54 %
- Data is split into training and test data and logistic regression is used.
- Used Step AIC and VIF for the multicollinearity of the variables
- Kept on removing variables with high VIF and low significance.

## Variable selection and Final model

- After 23 rounds of regression, we found the below 9 variables which contributes to the final model
- Total Working Years ,Years With Current Manager, Average Hours , Environment Satisfaction – Low, Job Satisfaction – Low, Job Satisfaction– Medium, Business Travel – Frequently, Marital Status – Divorced, Marital Status - Married

# Model Evaluation – Confusion Matrix

	Reference	
Prediction	No	Yes
No	614	80
Yes	29	47

Accuracy: 0.8584

95% CI : (0.8318, 0.8823)

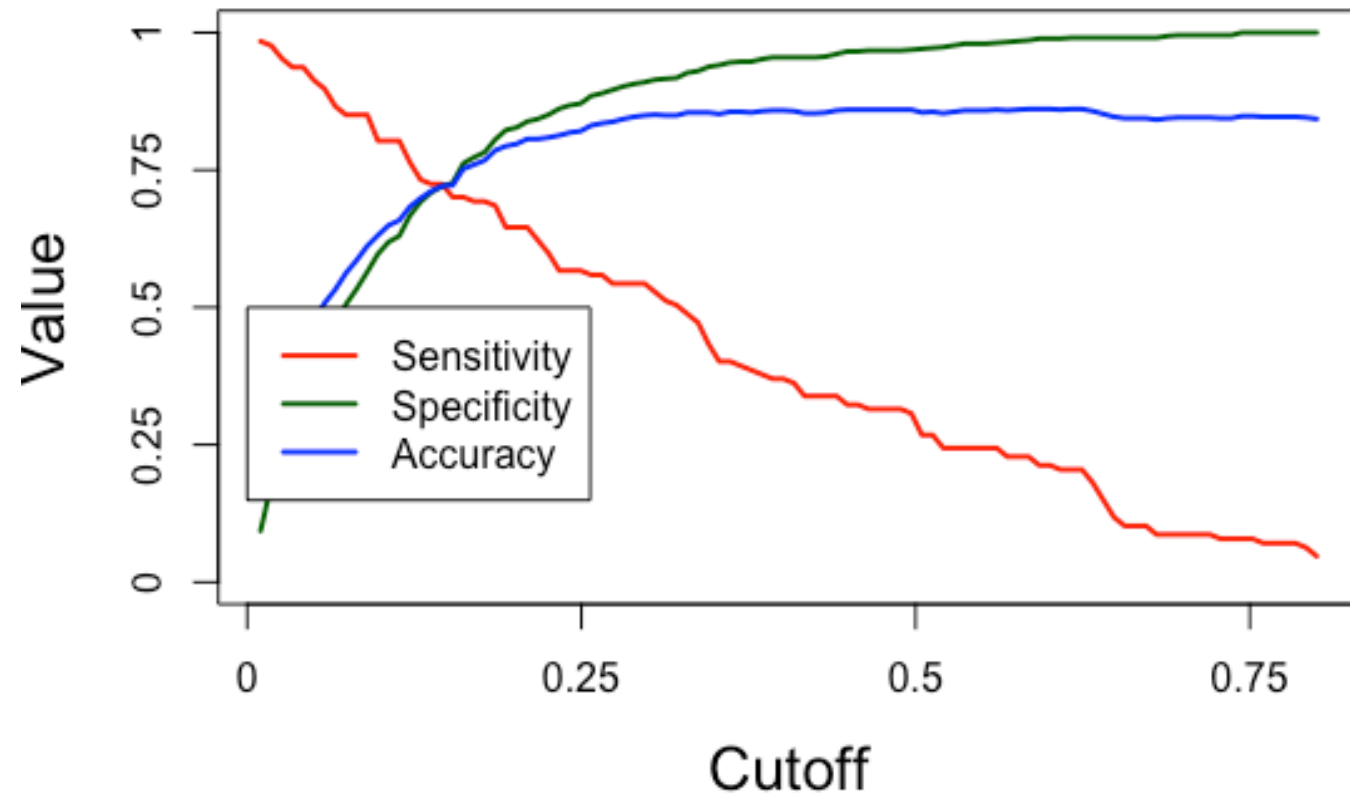
No Information Rate : 0.8351

P-Value [Acc > NIR] : 0.04258

## Accuracy , Sensitivity, Specificity

- Based on the cutoff levels, the optimal values of the following metrics
  - Accuracy 0.7207792
  - Sensitivity 0.7244094
  - Specificity 0.7200622

Accuracy,  
Sensitivity,  
Specificity –  
Optimal  
cutoff



## Business Recommendations

- Based on EDA and the model below are a few suggestions:
  - Tend to hire minimum resources when they have multiple previous companies as they will tend to leave the organization
  - Employees who have a high work experience and have worked under the same manager for a long time is more likely not to leave the organization; so the organization should not transfer the employee to a different unit where the resource would be reporting to a different manager
  - The organization should hire people as a non-travelling contract or else if they do not travel, they tend to leave the organization more often
  - People who are single are more likely to leave the organization
  - Employees who have to work more i.e. have high average hours tend to leave the company. The company should look into such cases and see why an employee has to extend the working hours and probably distribute the work among the team equally