

HR ANALYTICS CASE STUDY

Presented By:

Srikant Balaji

Ravi Prakash

Preksha Tripathi

Snehalakshmi Balasubramanian

Abstract

A large company is struggling with attrition rate of 15%.

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 project delay, new recruitment and training new recruits

Business Objective

Identify factors which are causing attrition

Changes they should make to their workplace to curb attrition

Identify variables which needs to be addressed right away

1. Business Understanding



2. Data Understanding

Data Collection

Overview of data, Variables types



3. Data Preparation and EDA

Data Cleaning

Missing Values handling and imputation

Formatting Date & Time

Derived Metric, Plots

Variable Treatment

Dummy creation, scaling



4. Model Building

Split data into train and test

Initial Model using glm()

StepAIC



5. Model Evaluation

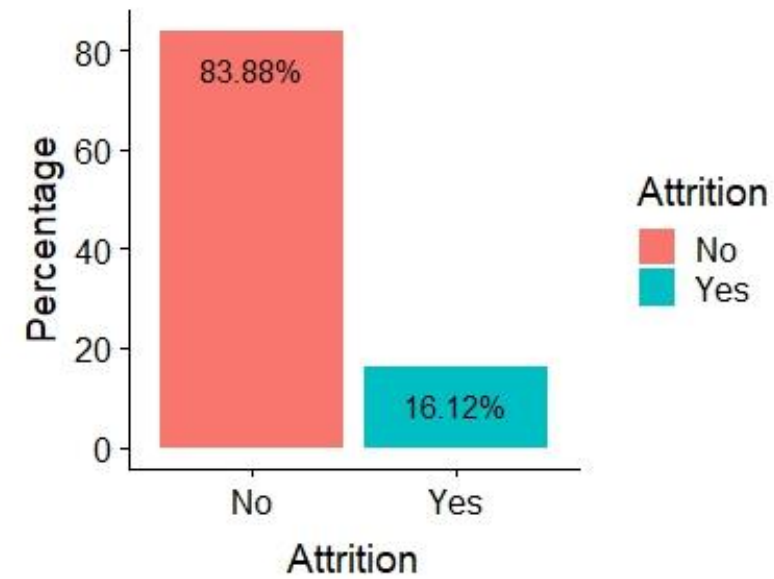
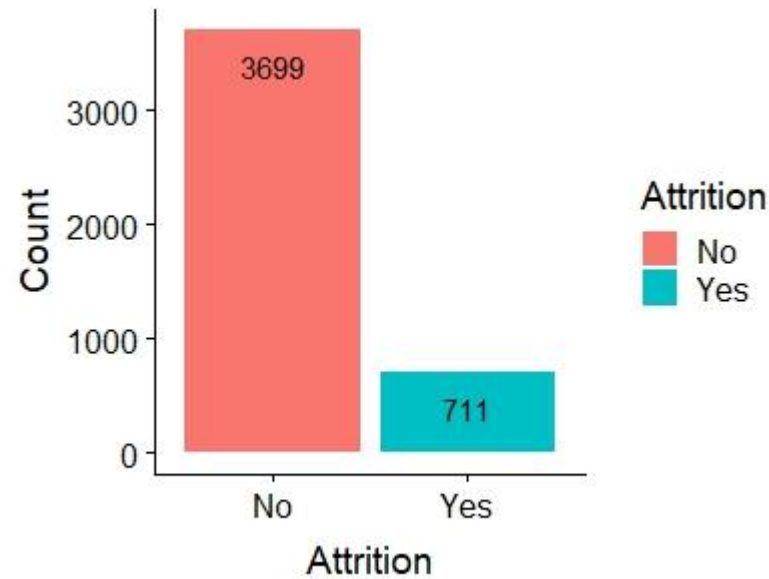
Confusion Matrix

KS Statistics

Gain and Lift



Results



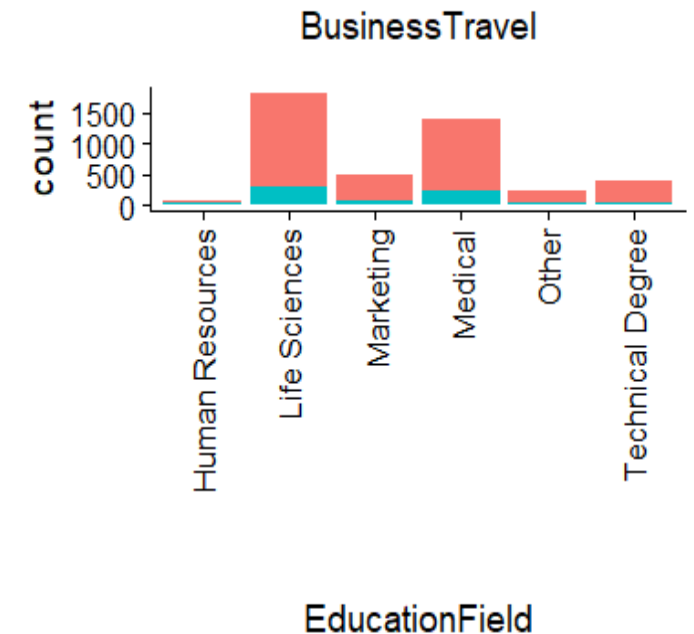
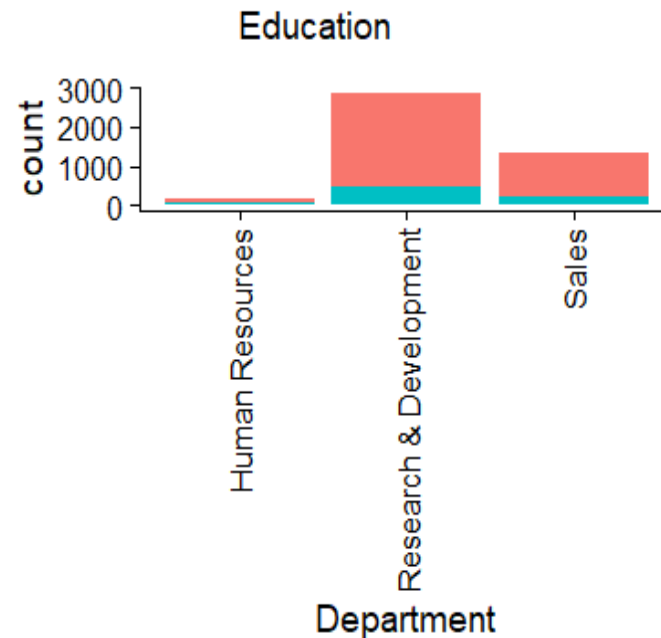
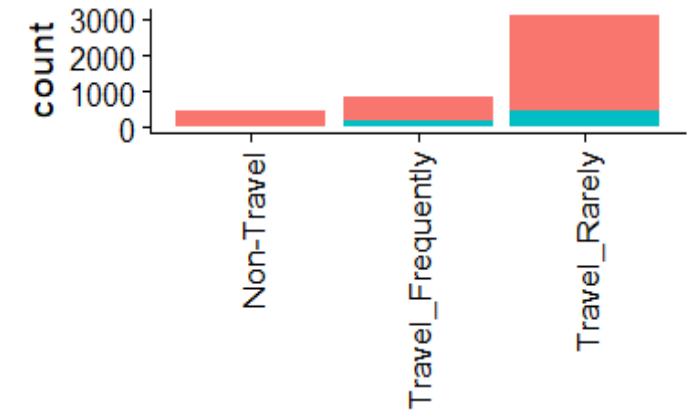
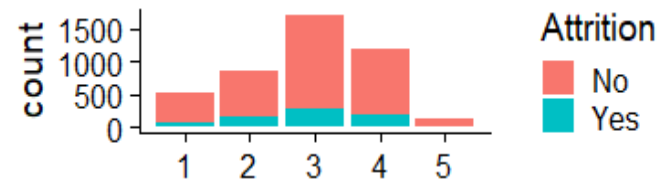
Preliminary analysis confirms Attrition rate is at 16%

Attrition level seems to be higher in

Bachelor Education ~ 3

Research & Development

LifeScience & Medical



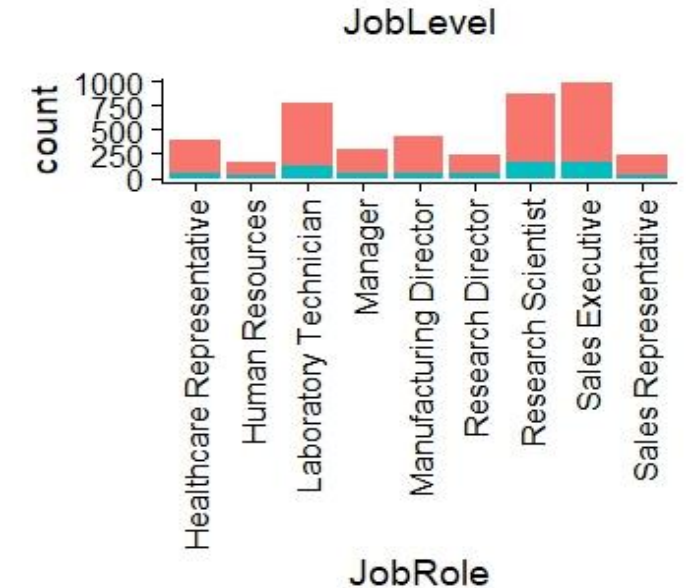
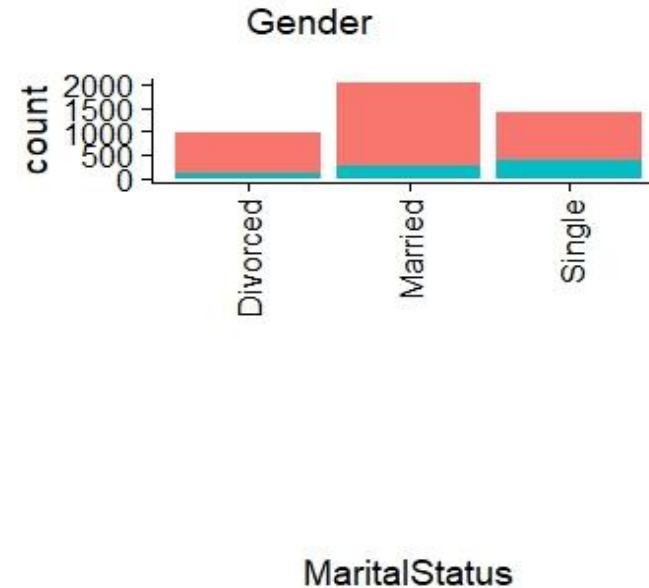
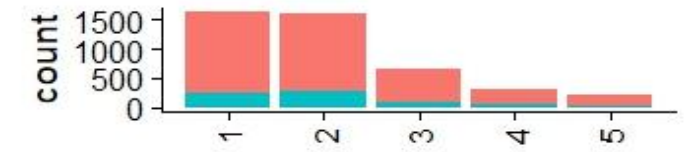
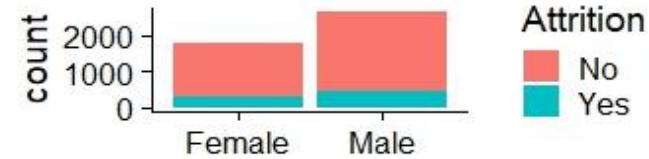
Attrition level seems to be higher in

Singles

Lower Job Levels

Research Scientist

Sales Executive

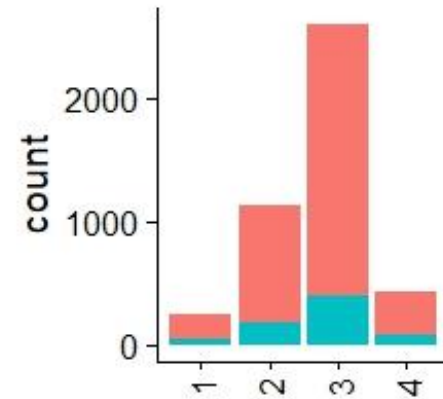


Attrition level seems to be higher in

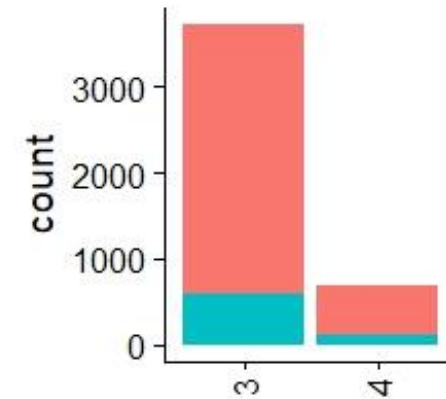
High Job Involvement

Lower Performance Rating

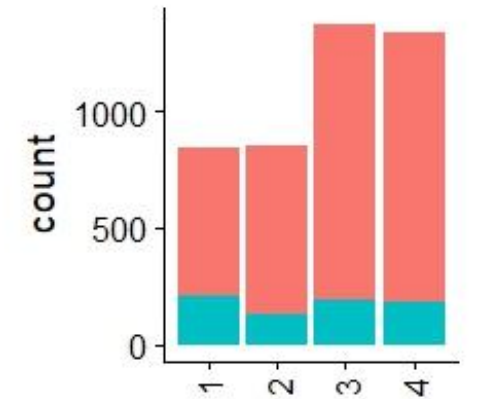
Lower Job Satisfaction



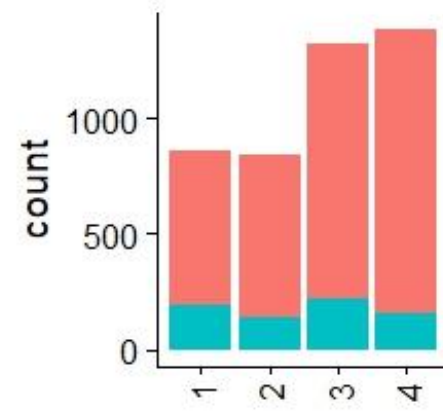
JobInvolvement



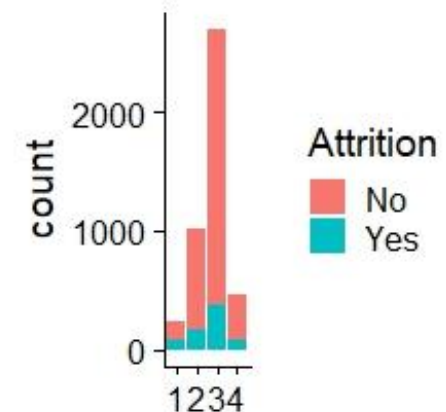
factor(PerformanceRating)



EnvironmentSatisfacti



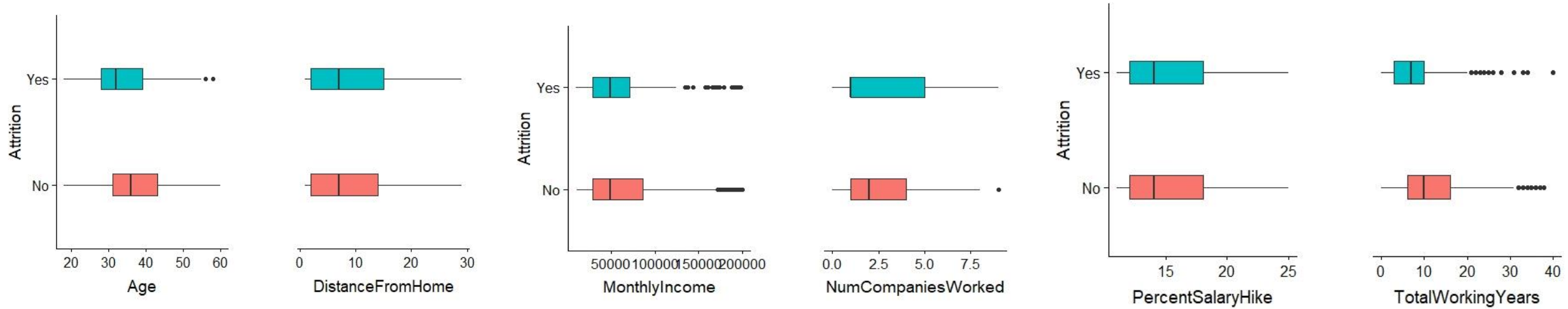
JobSatisfaction



WorkLifeBalance

Attrition
No
Yes

EDA on Continuous Variables.

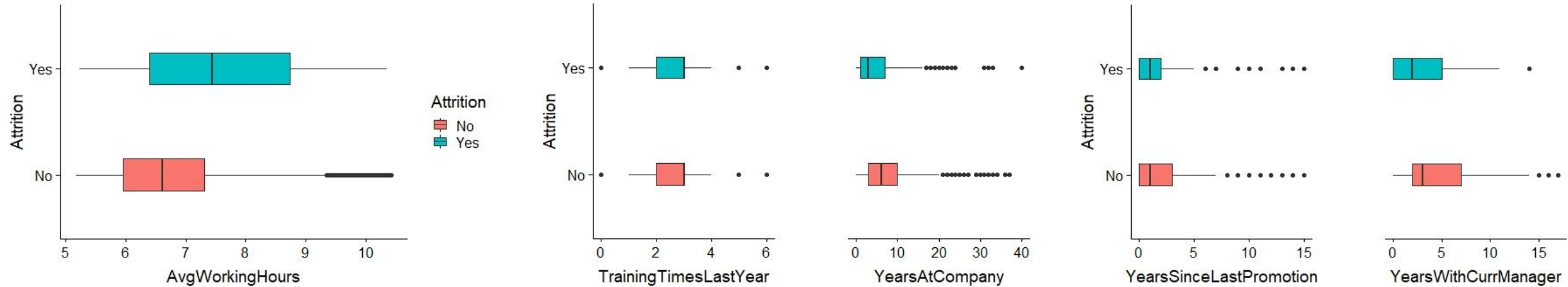


Attrition level seems to be higher in

Lower age group

Distance from home is high

Less Work Experience



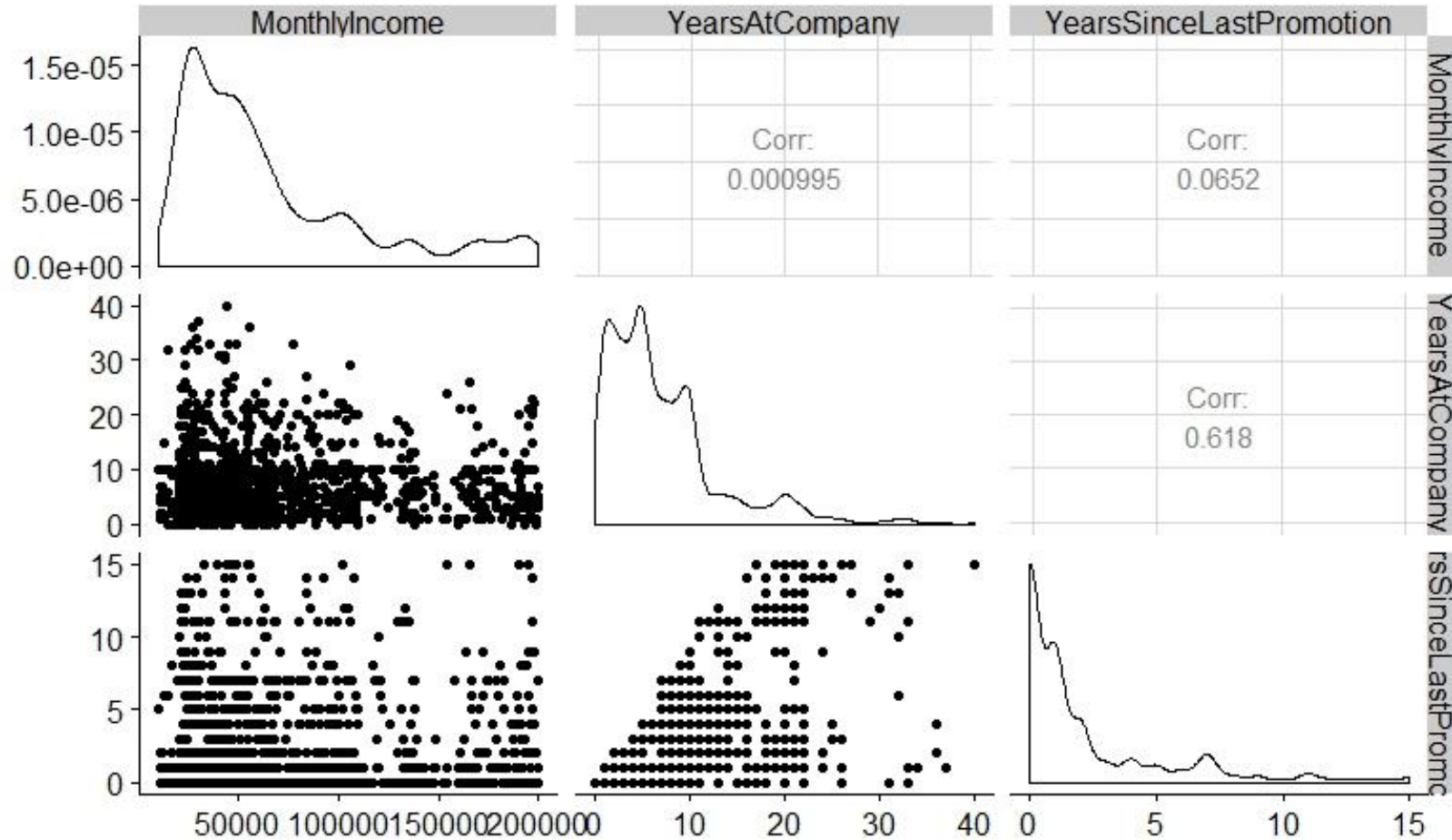
Attrition level seems to be higher for

Employees who spend more hours in office

Gets slow promotion

Relatively new at the company

EDA on Continuous Variables...



YearsAtCompany and YearsSinceLastPromotion has strong correlation (corr 0.618)

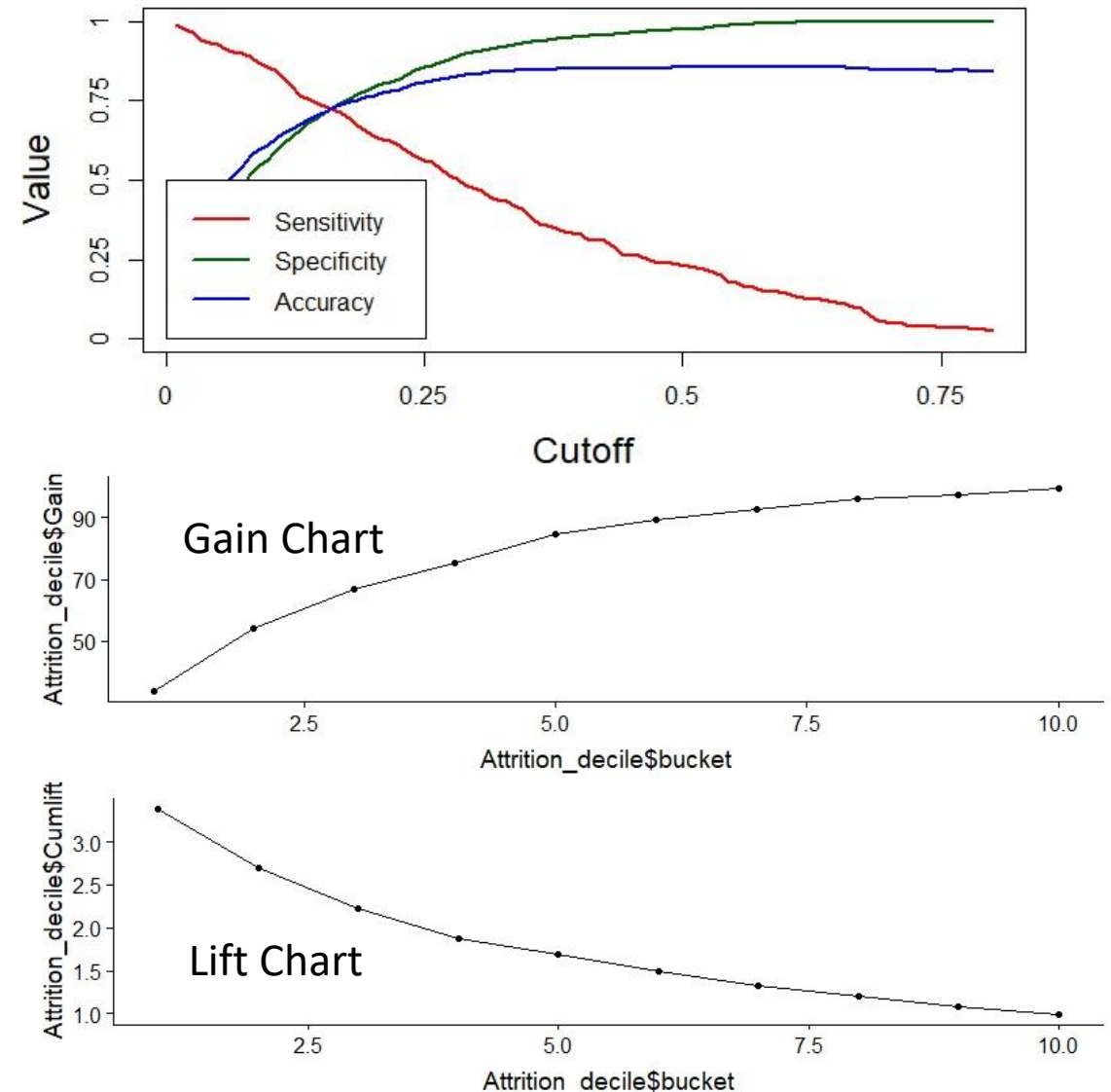
Steps

- ✓ Scaling of continuous variables
- ✓ Dummy creation for categorical variables
- ✓ Master Data frame creation
- ✓ Split master data into training and test data
- ✓ Perform first regression on training data
- ✓ StepAIC
- ✓ Remove variables on high VIF and low p-values
- ✓ Final model with all significant predictor variables

Ran model against test data
 Checked between 1% to 80% probability values
 Gain and Lift chart is monotonous

Findings

KS Statistics		0.449
Optimal Probability Cut off		0.1616
Confusion Matrix		
Prediction	Reference	
	No	Yes
No	802	59
Yes	301	153
Accuracy		0.726
Sensitivity		0.722
Specificity		0.727



1. Company should check the workload of an employee as working hour, environment satisfaction & Job satisfaction are interlinked.
2. Employee working in extended hours tends to leave
3. Training calendar to be improved
4. Promotion policy to be implemented and performance rating to be included as a parameter
5. Young employee to be mentored & motivated continuously
6. Experienced employees should be given leadership roles
7. Those who are singles are more likely to leave the company
8. Company should ensure healthy relationship between employees and manager.