HR Analytics - CASE STUDY

SUBMISSION

By –

Bharat Saxena

Mohit Singh

Sunil Hegde

Shravani Shahapure

HR analytics case study - Outline

A large company XYZ, employs, around 4000 employees. Every year 15% employees leave the company and management needs to replace the talent pool available in the job market. Management of the company is worried about 15 % attrition as it can harm company reputation among stakeholders due to some of the following reasons:

- The project of ex-employee gets delayed , which makes it difficult to meet timelines.
- Need to train new employees for the job in short time.
- A sizable department has to be maintain for recruiting new talent

Based on above observations, problem statement is:

- Analyse employee data and identify the factors which will help to curb the employee attrition

- Highlight the important factors which needs to be addressed immediately

.
- This information will be generated by modelling the probability of attrition using logistic regression.

Problem Statement

- Analyze employee data and identify the factors which will help to curb the employee attrition.

- Highlight the important factors which need to be addressed right away.

- This information will be generated by modelling the **probability of attrition** using a logistic regression.

Business Objective

- Identify reasons which lead to employee attrition.
- Identify any patterns among multiple reasons leading to attrition.

Goal of Analysis

- Model the probability of attrition using a logistic regression.
- Find out which of the variables in the given data is most important and needs to be addressed right away.

Problem Solving Methodology

1. Data Sourcing and understanding

- There are 5 datasets provided for the analysis.

- Employee ID is unique among all the files.

2. Data Cleaning

- Missing value treatment

- NA value treatment

- Outlier treatment

- Scaling of continuous variables

- Dummy variable creation for categorical variables.
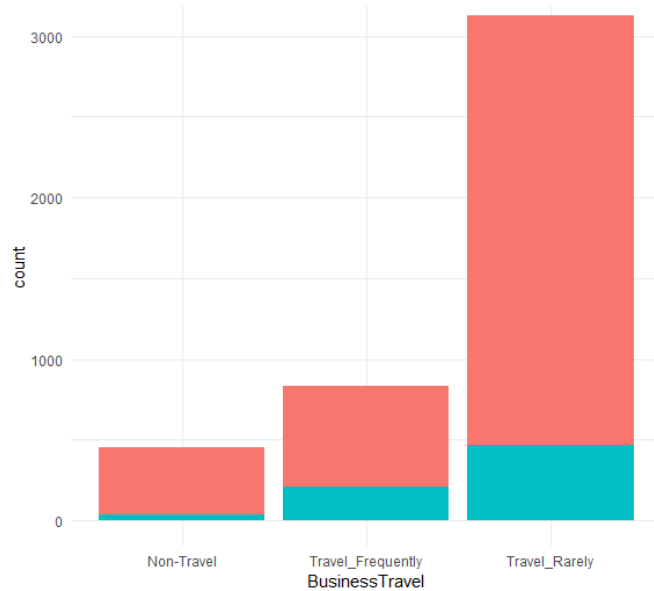
3. Derived Data creation

- Generate new data columns using the existing data. E.g work hours using in and out time.

## 4. Model Creation & optimization

- Create initial model.
- Using P values and VIF optimize the model

5. Evaluate the model

Exploratory Data Analysis

- Even though employee's who TRAVEL_RARELY have the highest attrition count.

- But % wise into individual category who TRAVEL_FREQUENTLY are **most likely to leave** the company.


- Employees who are in RESEARCH & DEVELOPMENT have the highest attrition count.
- But employees who are in HUMAN RESOURCES have the **highest attrition rate** 57 out of 189 left.
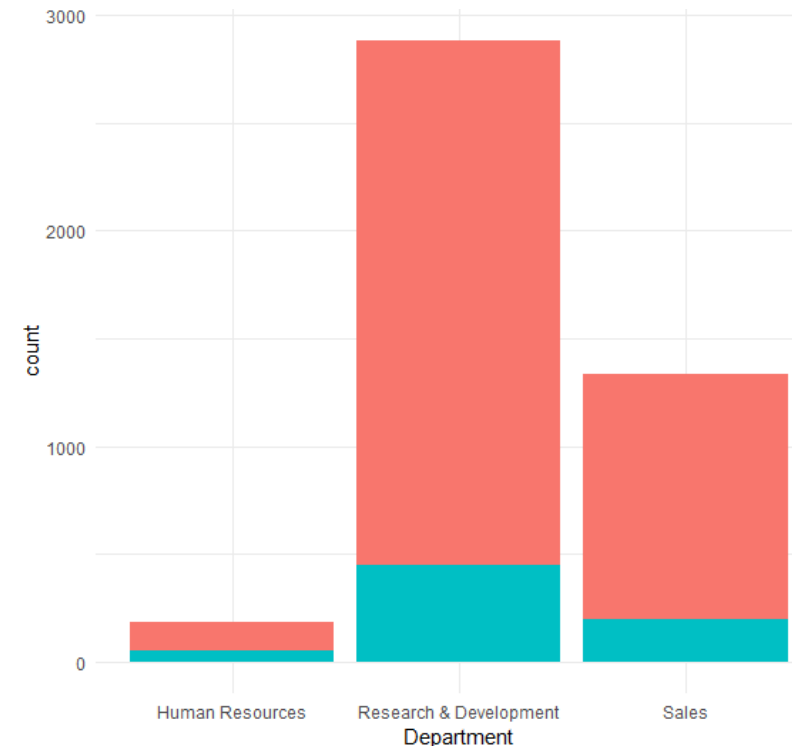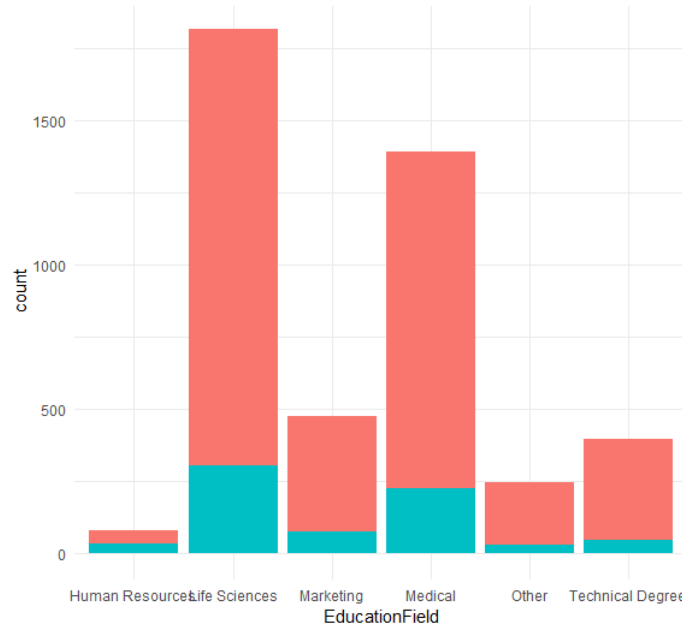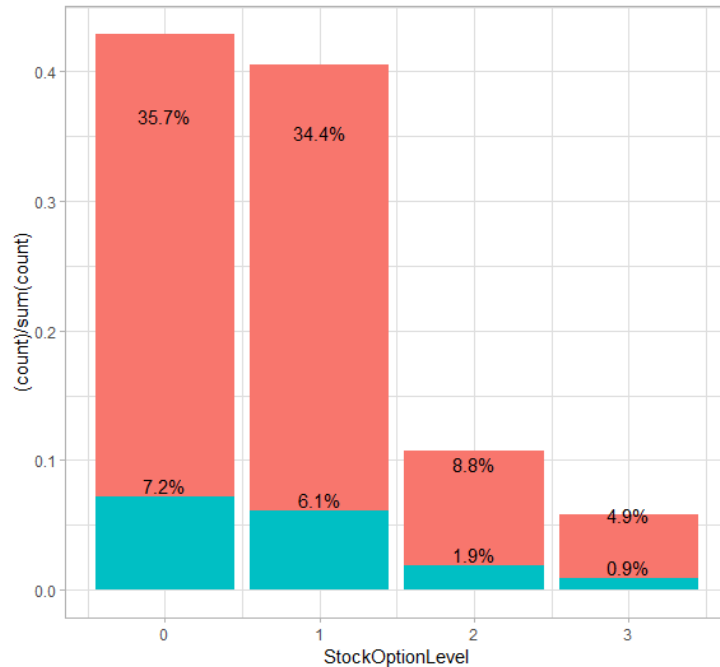
- Employees having Education Field of life sciences have the highest attrition count.
- But employees from HUMAN RESOURCES have highest attrition rate 40% (33 out of 81 left).


- SALES EXECUTIVE job role has the highest attrition count.
- But the RESEARCH Director job role has highest attrition rate (~24%) where 57 out of 240 left.
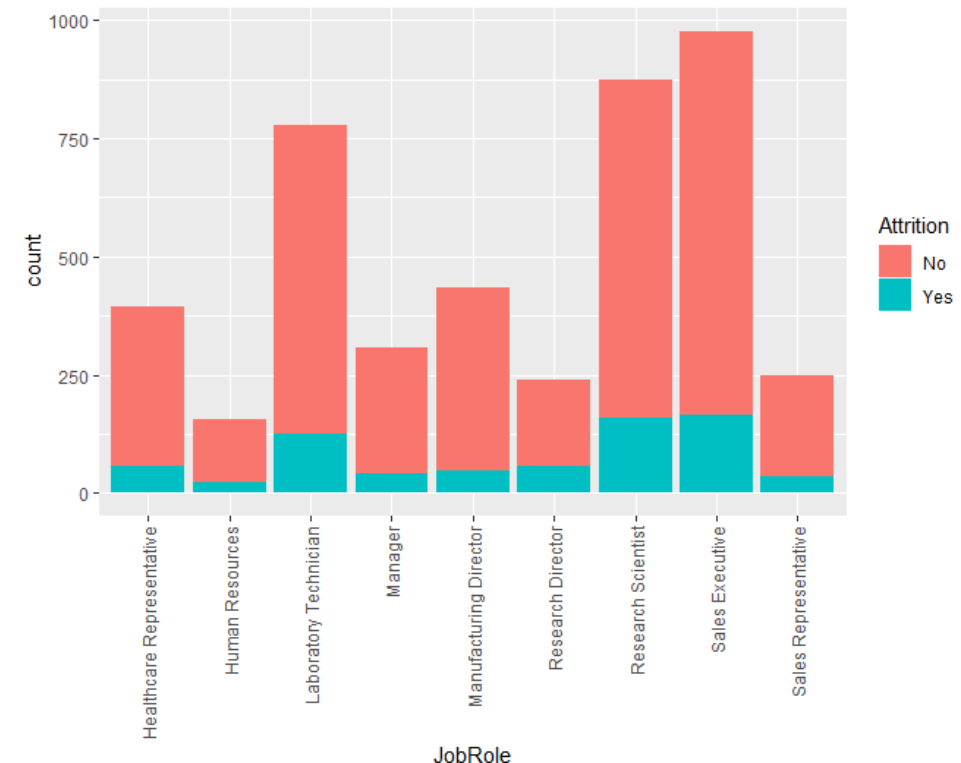
- Employees with 0 have highest attrition count (318) and highest attrition rate ~17%.

- Employees with NumCompaniesWorked=1 have the highest attrition count.
- But employees with NumCompaniesWorked=5 have highest rate at ~26% with 46 leaving out of 187.

Exploratory Data Analysis   contd...
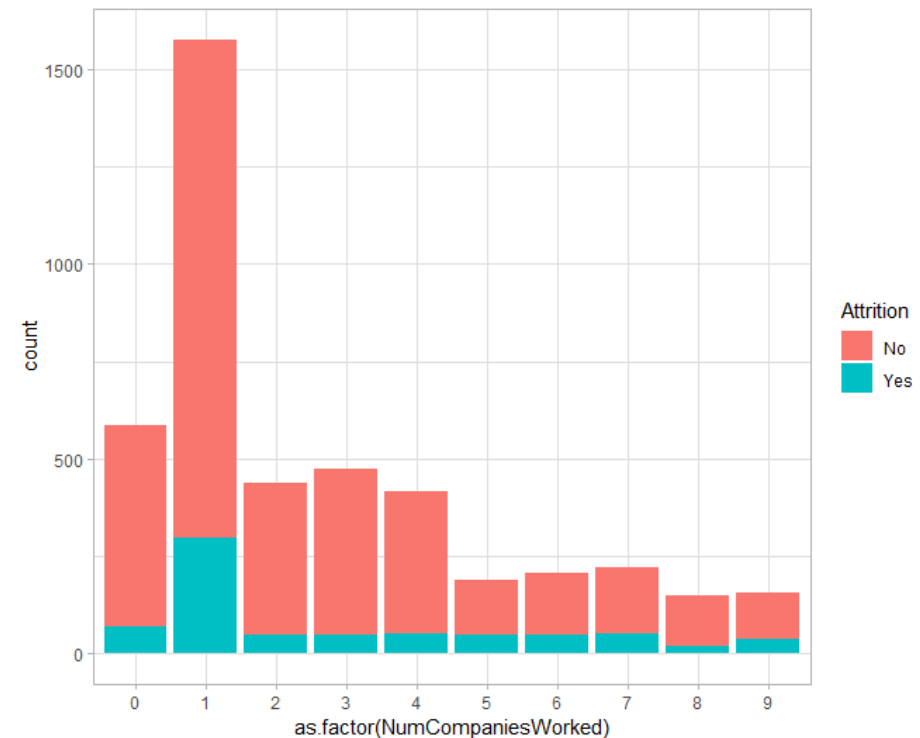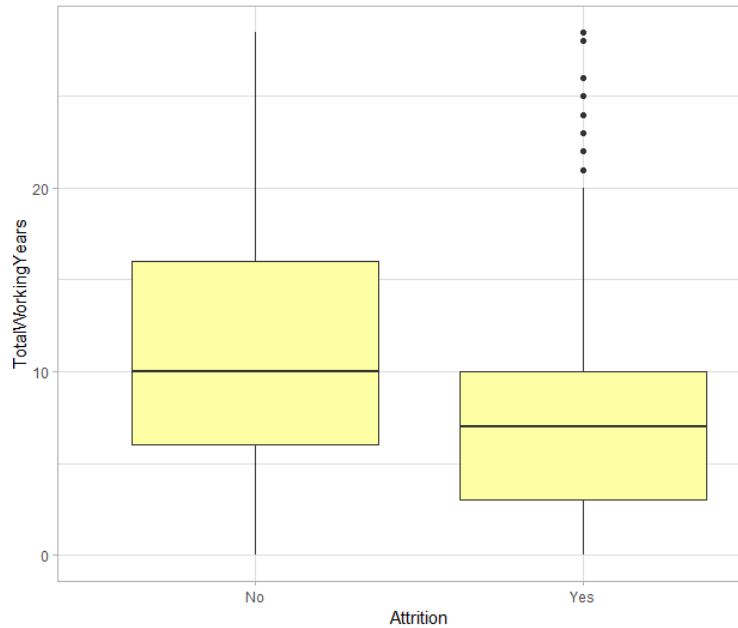
- Employees with less than 7 years of Total Working Years are more likely to leave.

- Employees with high job satisfaction level "HIGH" or have the highest attrition count 219,
- But employees with low Job Satisfaction levels "LOW or 1 have the highest attrition rate ~23%.
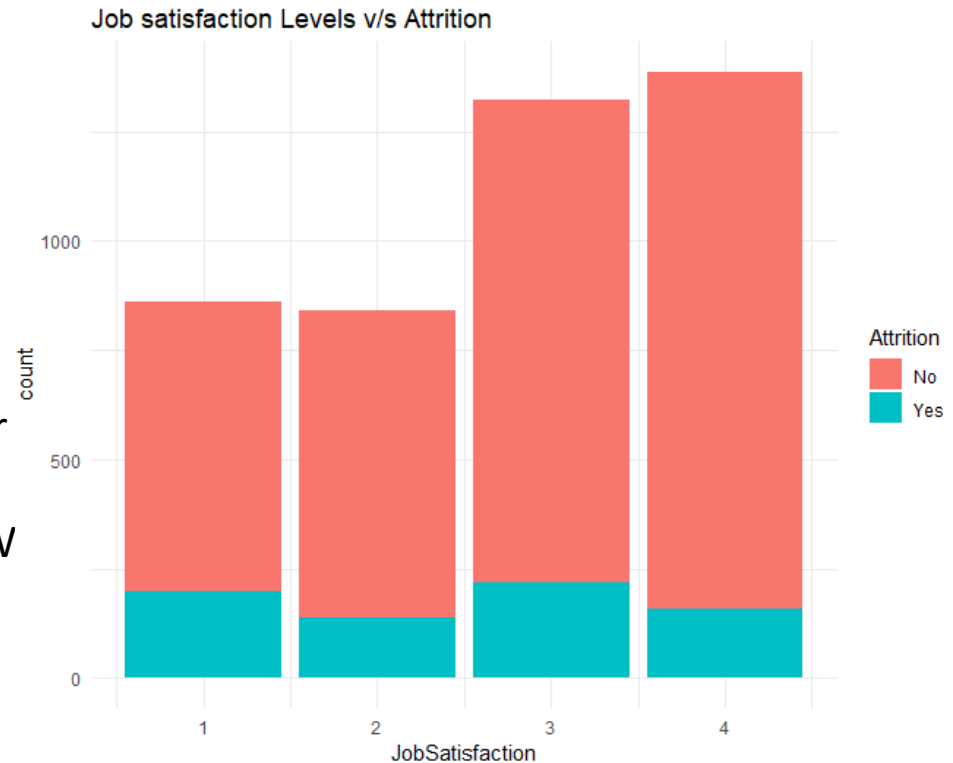
- Employees working more than average 8 hours are more likely to leave company.
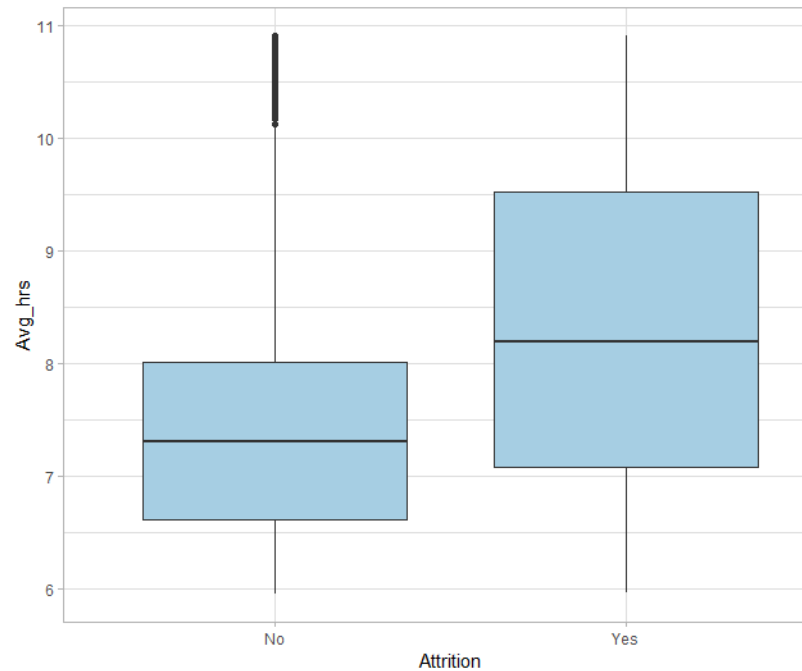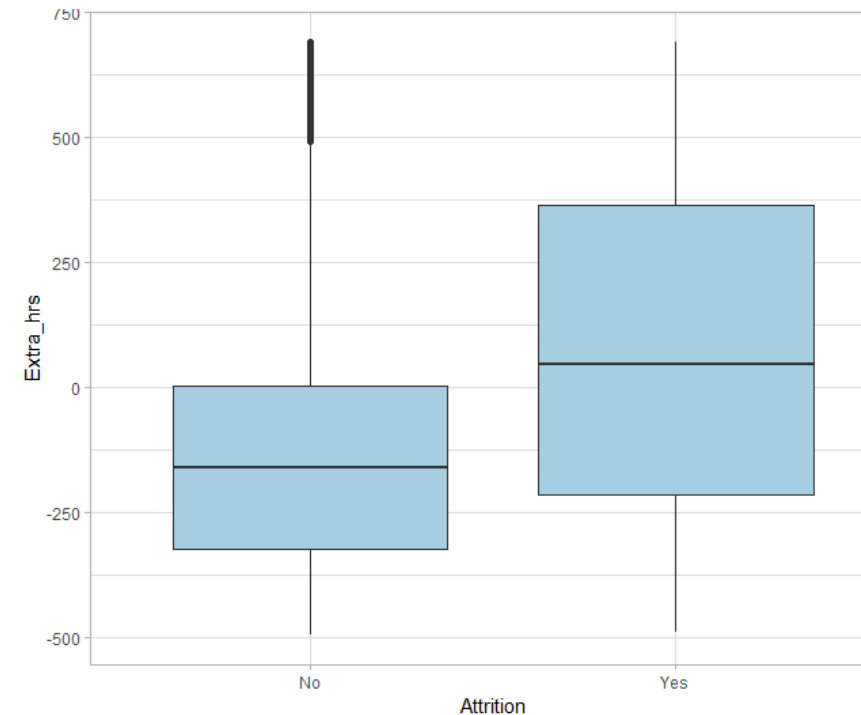
- Employees working more extra hours are more likely to leave company.

Logistic Model – Final Model

The final version of the model is based on the below 12 factors:

- `TotalWorkingYears`   `< 2e-16 ***`
- `Total_hrs`   `< 2e-16 ***`
- `MaritalStatusSingle`   `< 2e-16 ***`
- `YearsSinceLastPromotion`   `1.80e-15 ***`
- `EnvironmentSatisfaction`   `1.08e-14 ***`
- `JobSatisfaction`   `1.97e-12 ***`
- `BusinessTravelTravel_Frequently` `1.21e-10 ***`
- `NumCompaniesWorked`   `2.37e-09 ***`
- `YearsWithCurrManager`   `5.39e-09 ***`
- `JobRoleManufacturing.Director`   `2.28e-05 ***`
- `WorkLifeBalance`   `0.000210 ***`
- `TrainingTimesLastYear`   `0.000329 ***`

- `(arranged from most to least significance based on their P values)`

Logistic Model – Final Model

Model Evaluation statistics 1:

```
Confusion Matrix and Statistics
```

```
              Reference
Prediction   No   Yes
       No   1093  151
       Yes    23   56
```

```
              Accuracy : 0.8685
                95% CI : (0.8491, 0.8862)
   No Information Rate : 0.8435
   P-Value [Acc > NIR] : 0.006089
```

```
           Sensitivity : 0.27053
           Specificity : 0.97939
        Pos Pred Value : 0.70886
        Neg Pred Value : 0.87862
            Prevalence : 0.15646
        Detection Rate : 0.04233
  Detection Prevalence : 0.05971
     Balanced Accuracy : 0.62496

      'Positive' Class : Yes
```

```
                 Kappa : 0.334
 Mcnemar's Test P-Value : < 2.2e-16
```

Logistic Model – Final Model

Model Evaluation statistics 2:

```
   McFadden
  0.2039475
```

Variable Importance [t-statistics]

```
                                   Overall
NumCompaniesWorked                5.945444
TotalWorkingYears                 8.363161
TrainingTimesLastYear             3.648413
YearsSinceLastPromotion           6.855929
YearsWithCurrManager              5.567291
EnvironmentSatisfaction           7.681116
JobSatisfaction                   7.171940
WorkLifeBalance                   3.816703
Total_hrs                        12.101837
BusinessTravelTravel_Frequently   6.437628
JobRoleManufacturing.Director     4.256072
MaritalStatusSingle               8.596049
```

Logistic Model – Final Model

ROCR Curve

Logistic Model – Final Model

CONCLUSION
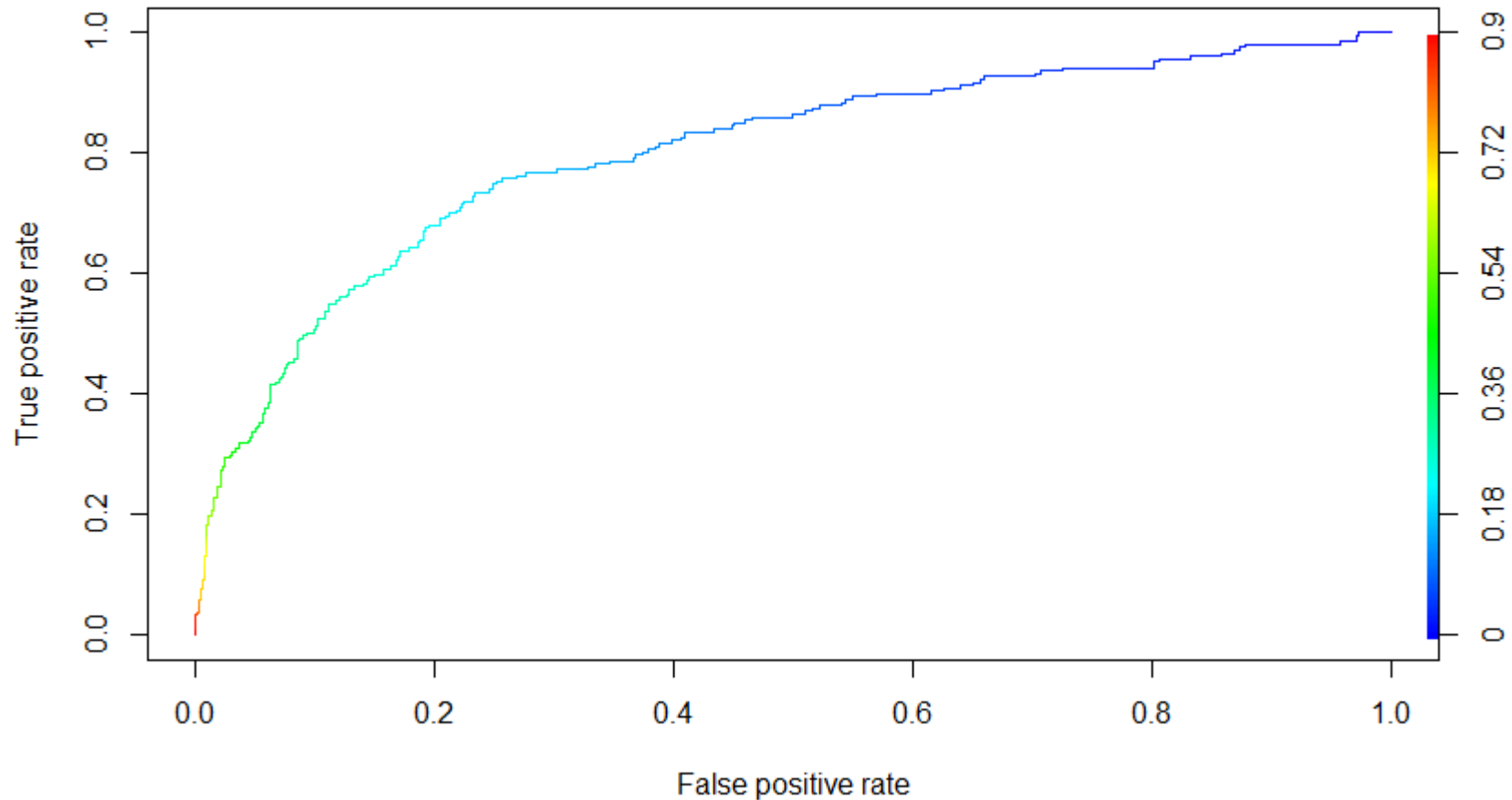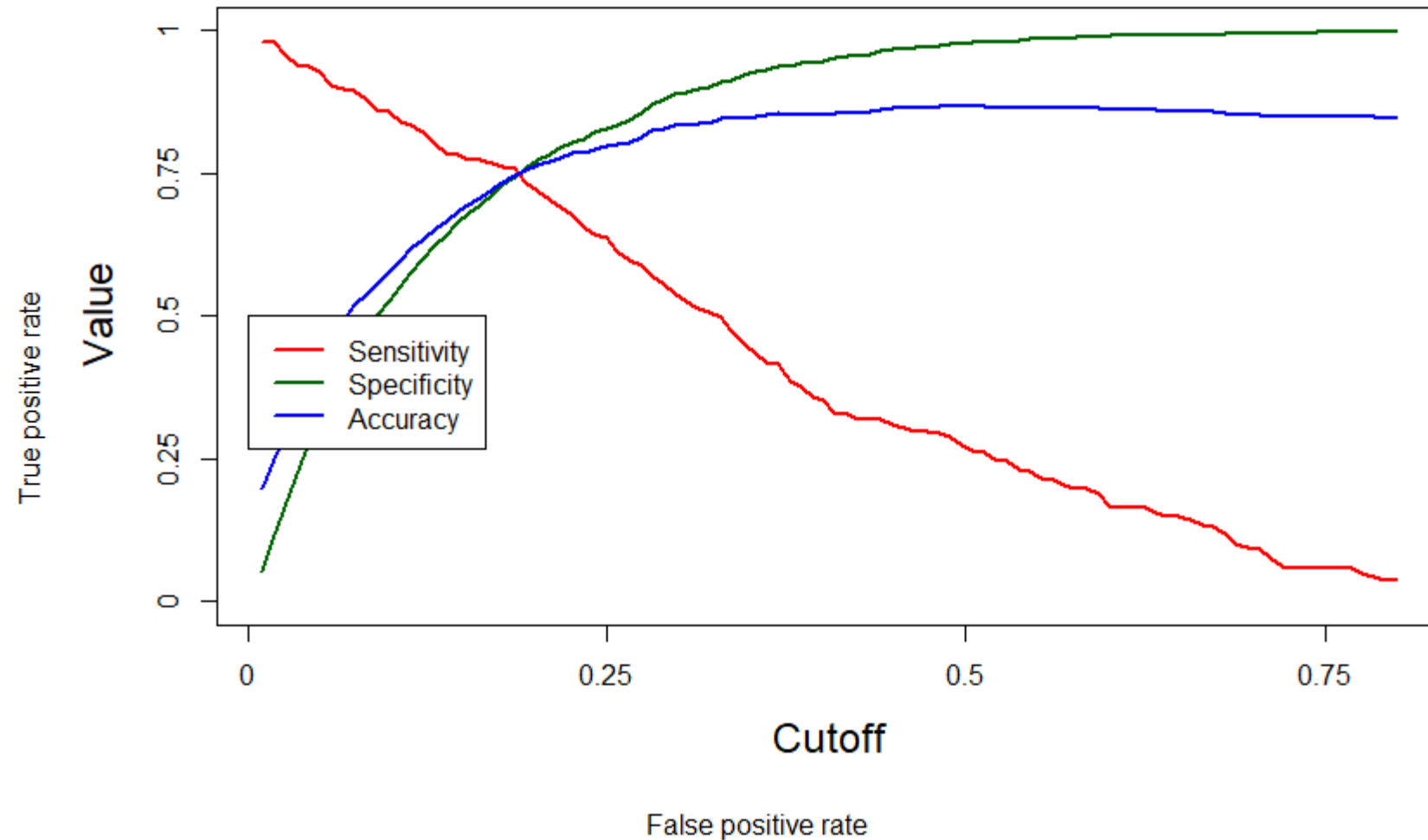
GOAL #1

- Based on the data provided below 12 factors have been identified to have a significant impact on employee attrition.

```
TotalWorkingYears                Total_hrs
MaritalStatusSingle              YearsSinceLastPromotion
EnvironmentSatisfaction          JobSatisfaction
BusinessTravelTravel_Frequently  NumCompaniesWorked
YearsWithCurrManager             JobRoleManufacturing.Director
WorkLifeBalance                  TrainingTimesLastYear
```

CONCLUSION

GOAL #2

- Below factors need to be addressed right away as they have the highest impact on employee attrition:
  - `TotalWorkingYears`
    Employees with less than 7 years of Total Working Years are more likely to leave. Special attention should be given to understand issues these set of employees are facing.

  - `Total_hrs`
    Employees work regularly above average total hours are more likely to leave. Policy around limiting Working hours might help alleviate this factor.

  - `MaritalStatusSingle`
    Employees who are single are more likely to leave the company. These set of employees should be engaged more in the project activities.