HR Analytics Case Study

Decode & Curb Attrition

Vijay Mudivedu
Subrat Samant
Abhinav Chandra
Apurv Gaurav

Business Objectives

Problem Statement:

- A 4000 employee Company XYZ that has 15% attrition wants identify the factors impacting attrition to reduce the costs as a result of it.

Business Goals:

- Find the factors affecting the 15% attrition rate
- Express the findings in terms of neat visualizations
- Find most important factors leading to attrition to be addressed on priority
- Suggest methods of minimizing the cause of attrition

Problem Solving Methodology

- **Data Sources**
  - General Data
  - Employee Survey
  - Manager Survey Data
  - In_time_dat, Outtime Data
- **Data Preparation**
  - Analysis of In_Out data frame
  - Calculate mean attendance, Workload, regularity, Stats
  - Missing Values, treatment
  - Outliers and Invalid record treatment
- **Exploratory Data Analyses**
  - Nominal Categories
  - Ordinal Categories
  - Interval Variables
  - Ratio Variables
  - Bivariate analysis
- **Variable Transformation**
  - Scaling of variables
  - Converting variables to factors
  - Creating Dummy Variables
- **Model Building**
  - StepAIC method to minimize insignificant variables
  - Iterating the model variables
  - Verifying the AIC and p-values of predictors
  - VIF values are iteratively measured
- **Model Evaluation**
  - Sens, Spec, Acc at Cutoff
  - Gain, Lift and KS-Static verification

Data Preparation & Data Cleaning

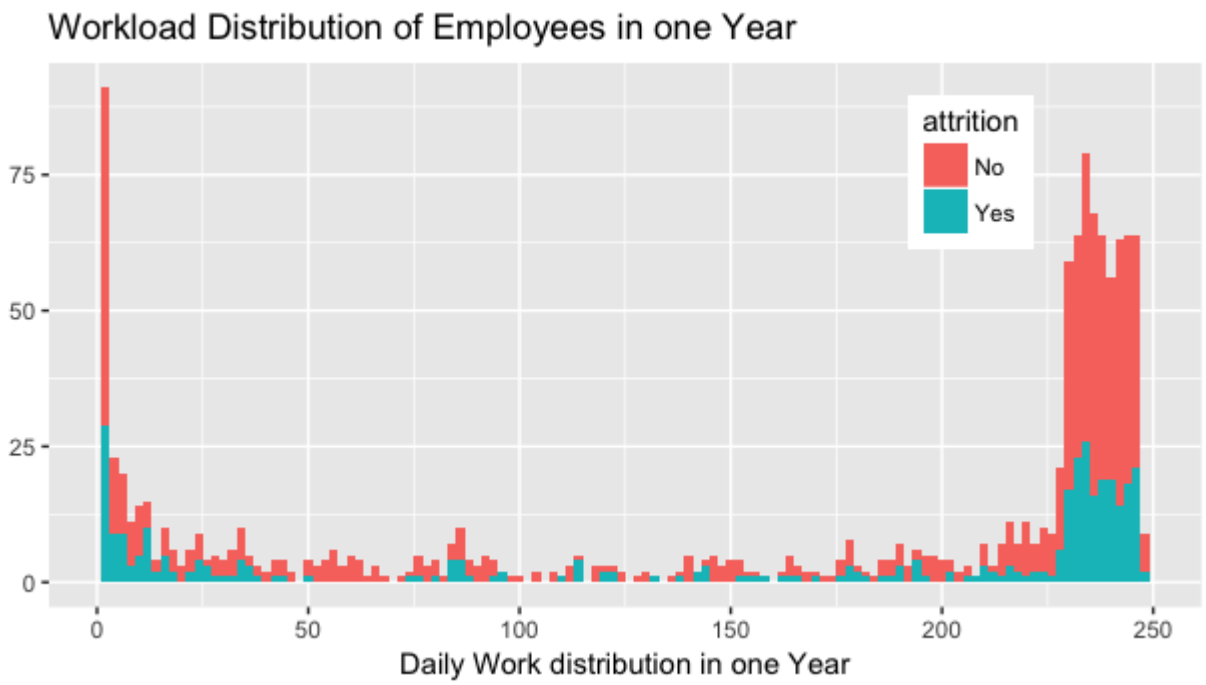
Data Preparation:

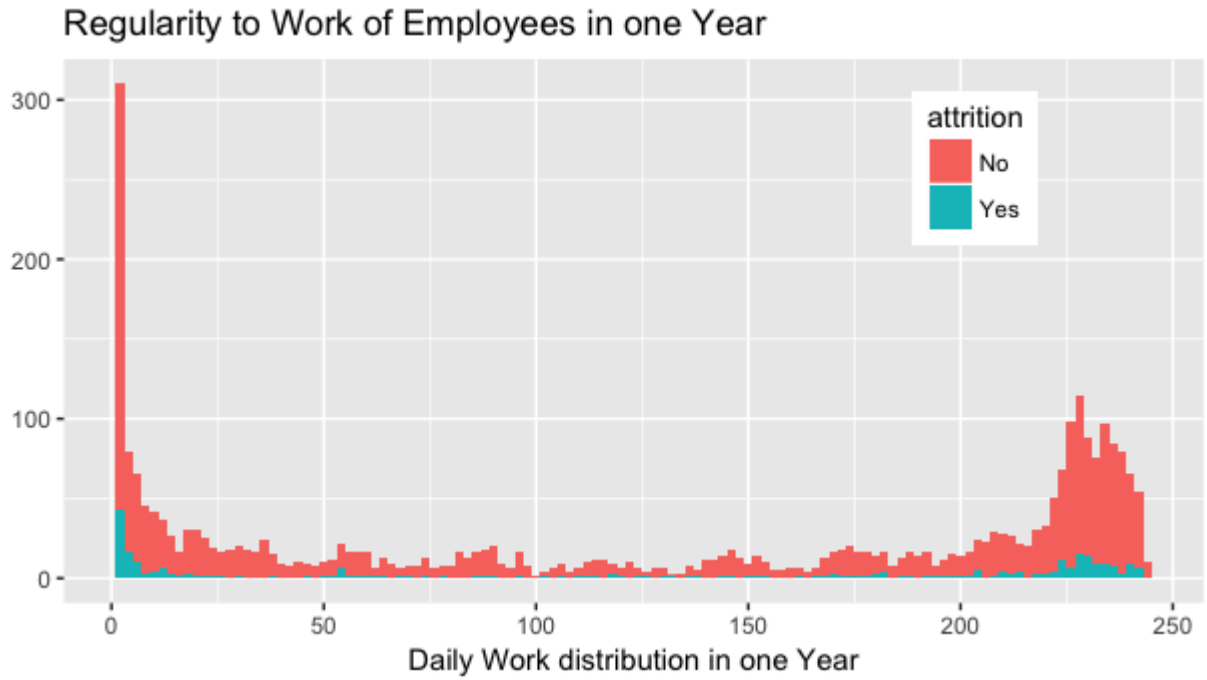
- *Employee_survey_data*, *Manager_survey* data, Derived metrics from *In_time_data* and *out_time* data, and general_data data frames were merged to collate into ***employee_master*** data frame using ***merge()*** function.

Data Cleaning:

- Preparation of *in_out_duration* dataframe from from *in_data* and *out_data* dataframes.
- Computation of derived metrics of each employees *mean_attendance*, *Workload*, and *Regularity*.
- Identified the NAs in each of the columns and removed the rows wherever NAs were found.
- Performed WOE analysis to impute the NAs "*TotalWorkingYears*"
- Finally, data quality checks of invalid records were followed to get to the *employee_master_cleaned* dataframe.
- Exploratory Data Analysis on the cleaned data to visualize the behavior of continuous variables, categorical variables, Nominal variables, Ordinals, Ratio Values and Interval variable.

Univariate Analysis – Derived Metrics

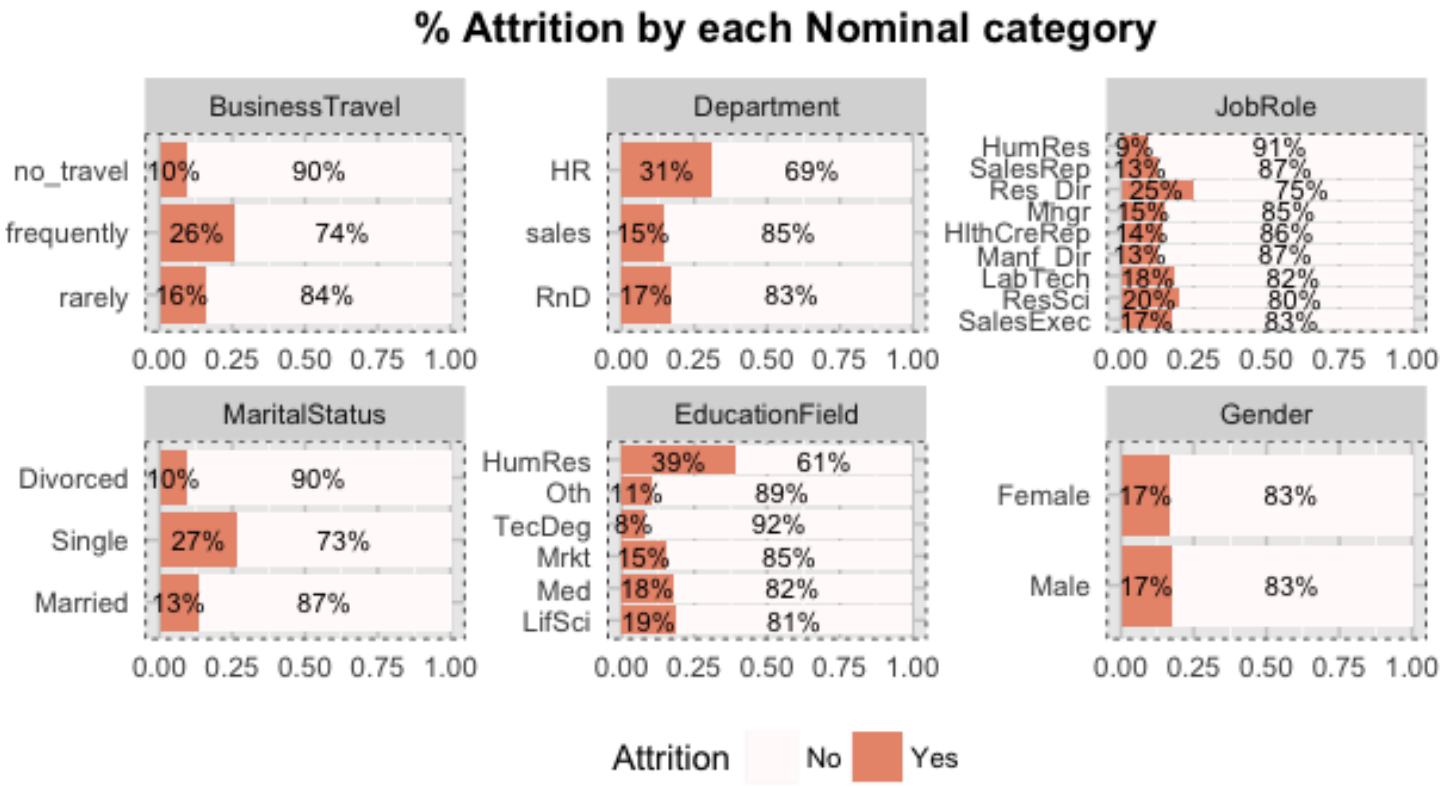




- 50 Percentile of data is spreads across 0 count indicating percentile of overworked
- 75 percentile is spread across 3 times overworked
- Rest fall under 100% overworked.

- 50% of data is < 20 times in 249 working days
- 75 % percentile of data is spread across 205 times,
- Rest 100% percentile is 249

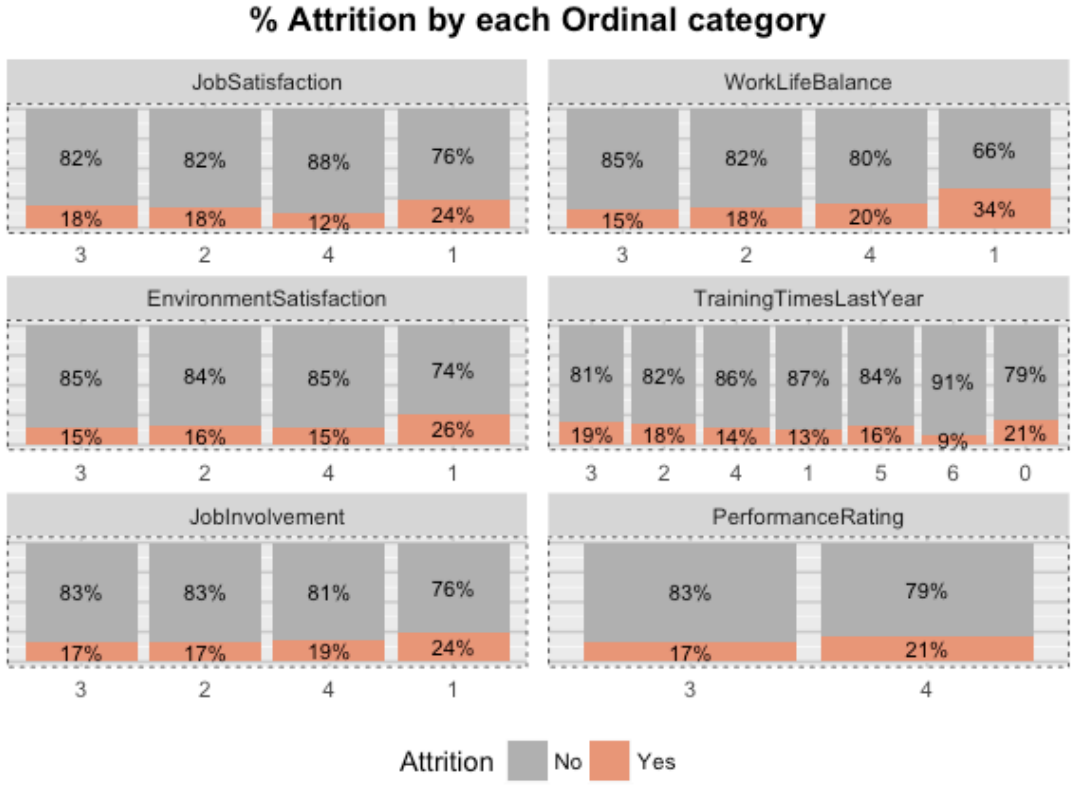
Univariate Analysis – Nominal Variables



Attrition rates:

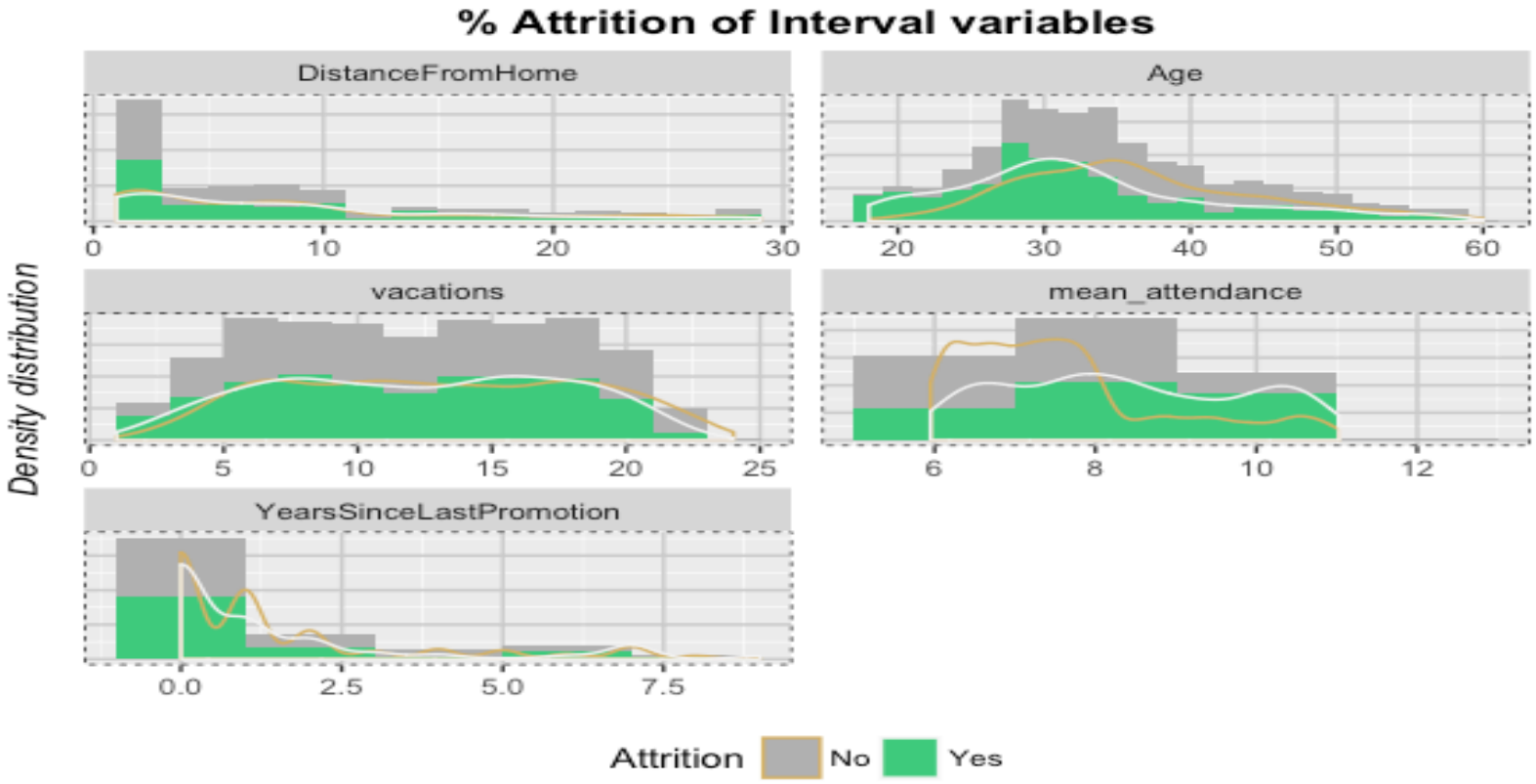
- Single & Married Employees
- Employees travelling Frequently and Rarely
- HR and R & D Departments
- Research Directors, Research Scientists, Lab Technicians
- HR degrees, Life Sciences, Medical degrees

Univariate Analysis – Ordinals



- Categories of employees with Low Performance Rating, Lesser Job Involvement, Low Environment Satisfaction, Job Satisifaction, No training received and lesser WorkLifeBalance are more likely part away
- Education levels, StockOptionLevels, JobLevels variables do not signify any attrition.
- Employes Heavily Worked, and under worked, irregular to work shows higher attrition over other factors indicators.

Univariate Analysis – Interval Variables



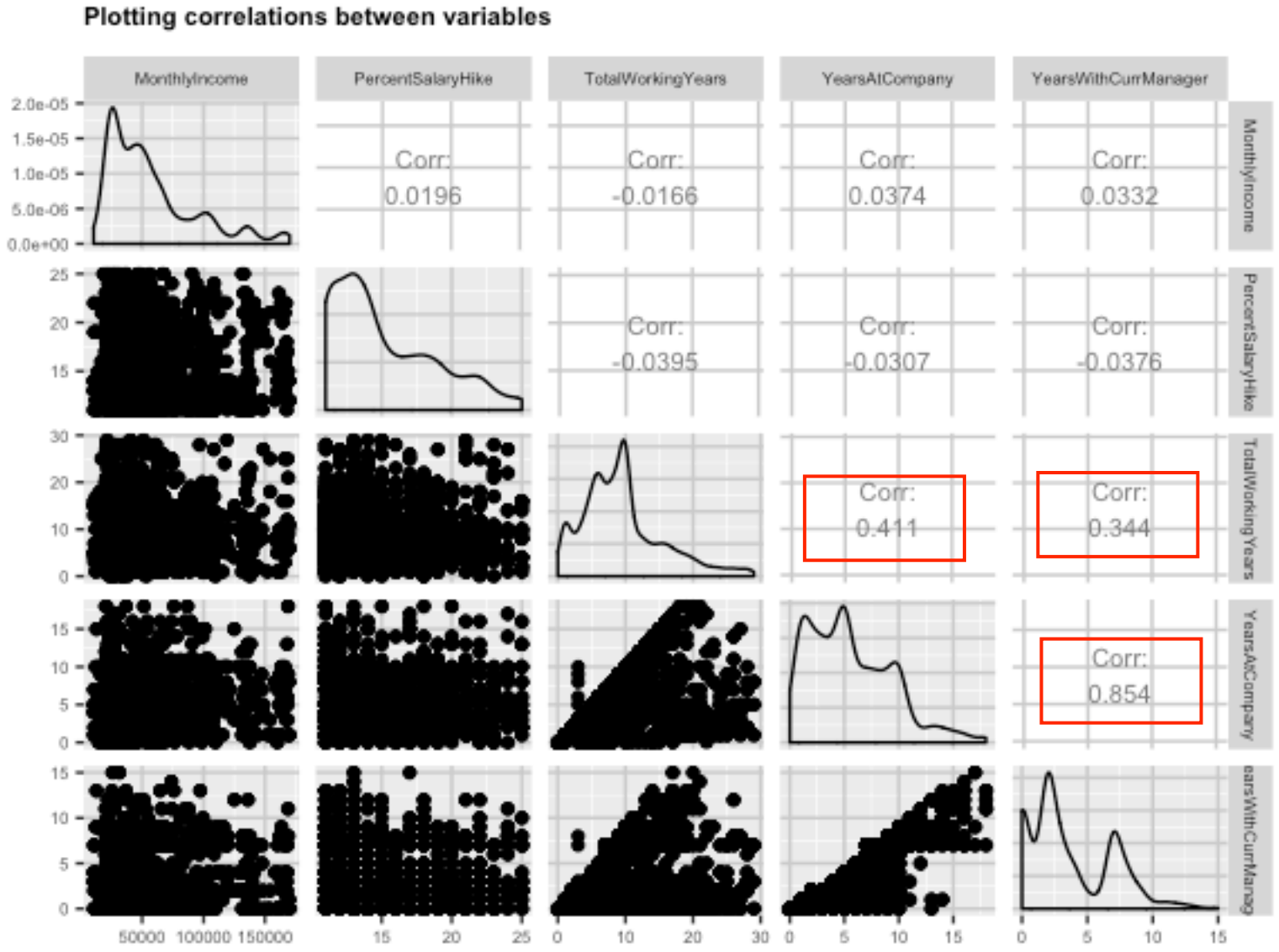
Attrition rates from Data:

- Employees in the Age group of 24-35 are highly like to part way
- Employees closest to the work location are also like to part away
- Mean attendance > 10 or < 7 are the ones that frequently have attrition.

Attrition rates from Data:

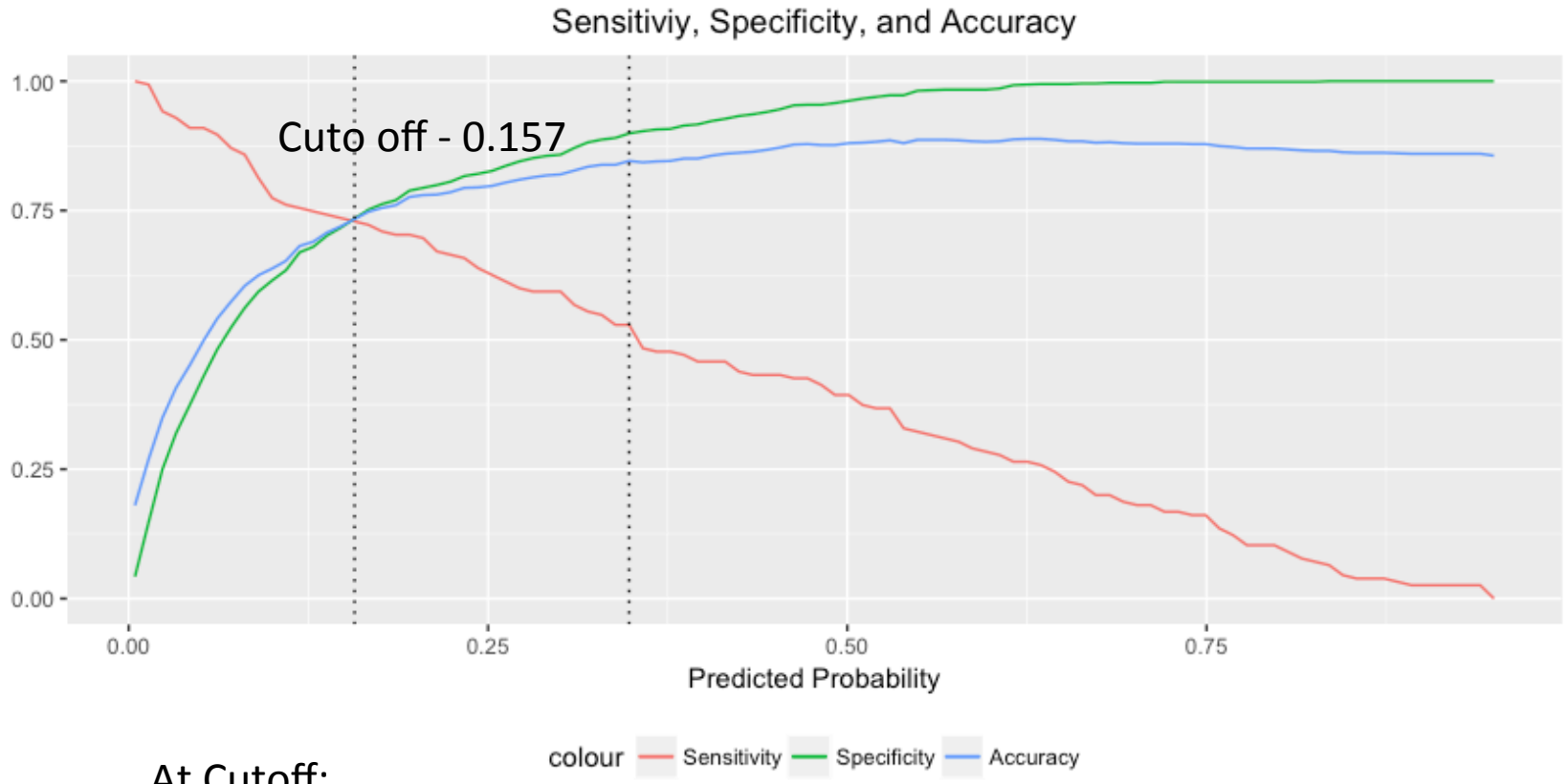
- Employees with no promotion for 7 years, 9 years and 15 years and very likely to part way
- Large portion of this concentrated between 0 to 2.5 years

Bivariate Analysis- Ratio Variables



- 85.4% employees with YearsatCompany work as associated with Currentmanager
- 41.4% YearsSinceLastPromotion & YearWithCurrManager (44.6%)
- Of the total work experience, 41.4% employees spent time in the company

Result – Model Performance

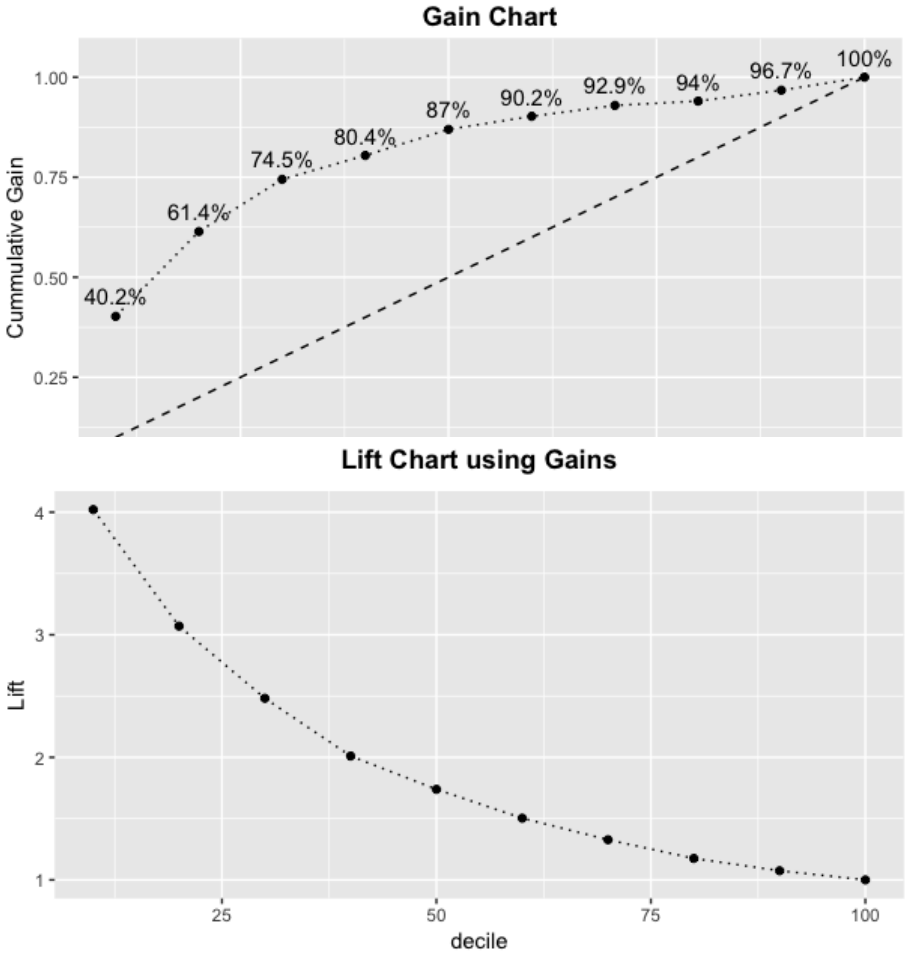


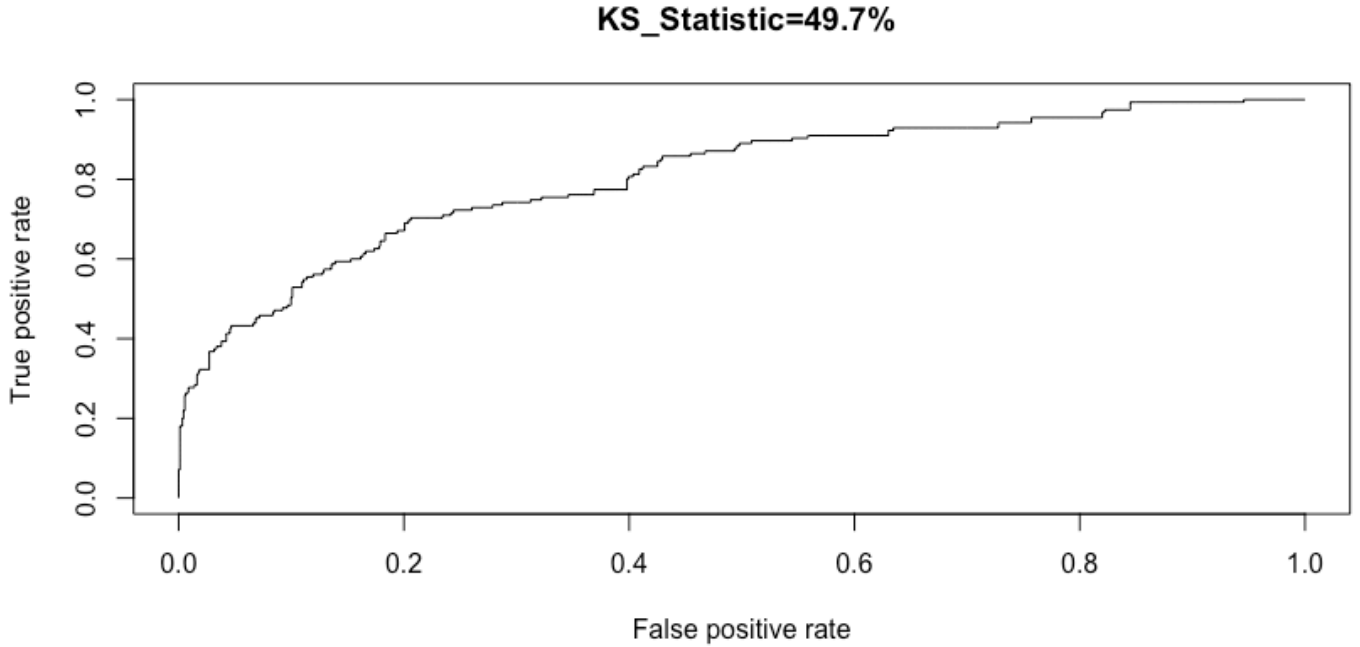
At Cutoff:

- Accuracy : 0.7335
- Sensitivity : 0.7290
- Specificity : 0.7343

- Attrition of True Positive or True Negative is 73.3% accurately Predicted by the model.
- Model predicts “Yes” Value 72.9% of times among available “Yes”
- Also, model predicts “Nos” in 73.43% times of all Nos.

Model Evaluation Gain and Lift Charts





- With KS- Statistic 54.7%, Area under the curve is 84%. model has greater ability to predict the Attrition
- With Progressive Lift predictive model scales very well against the random model,
- With each passing by decile, the gradual rise in cumulative decile scales very well with the model.

Conclusion

- **Age** - Higher Attrition for Lesser Age, It is higher fro the age groups beween 26-35.
- **Business Travel** – Frequently and Rarely,
- **R&D and Sales Department** - have negative relationship with Attrition. The JobRoles of the Department such as Lab Technicians and Research DIrector and Scientists are the one that are frequently looking for moving out.
- Employees with Marital_Status Singles are the ones who are frequently look out.
- Resources with NumberCompaniesWorked and Total Working Years are Correlated are directly. This impacts the attrition to some extent and the higher the number of companies a resource works, the higher is the attrition. Specifically, between 1 6,7,8
- Years with Current Manager and Years Since last promotion are signifcantly correlated with Each other. As an employee works with the CurrentManger attrition is apparently is showing the reduced trend.
- Employees with mean_attendance > 10 hours of workload are more prone for lookout and perhaps the reason be a stressful workenvironment and it is clear from the data variables where workload_1 and WorkLoad_2 that demonstrate higher positive values indictor to strong predictors.
- Further, employees with poor worklife balance, and poor Job Satisifaction, poor values of Environment Satisfaction and and lower Job_involvment_3 have Higher attrition. Thus there variables have a inverse relationship with attrition.