HR ANALYTICS CASE STUDY

HR Analytics for a company - XYZ

Group Name:

1. Atul Agarwal
2. Anurodh Kumar
3. Amit Kumar Verma
4. Debashish Bhattacharya

Contents

- Introduction & Problem Statement
- Data Cleaning
- Data preparation
- Building the model
- Evaluating the model
- Solutions/Inferences
- Recommendations

Problem Statement

- The Company is facing a problem of high attrition and analysis on probable causes is required
- Analysts are required to model the **probability of attrition** using a logistic regression
- The results obtained using the analytics is very helpful in driving insights and taking decisions in workplaces and management in order to get most of the employees to stay.

Data Cleaning

- NA Values Treatment: As the number of records with NA was less than 3 per cent, those records were eliminated

- Removing unwanted columns: following columns that have only a single value all through were removed from data sets
  - Employee Count
  - Over 18
  - Standard hours

- Factor Conversion: following columns were converted into factors
  - Education
  - Job level
  - Stock option level

Data Cleaning

- Outlier Treatment: Outlier treatment was done for the following variables:
  - Distance from home
  - Number of companies worked
  - Monthly income
  - Percentage salary hike
  - Total working years
  - Training times last year
  - Years at company

- Converting integer to categorical:
  - Education
  - Environment satisfaction
  - Job involvement
  - Job satisfaction
  - Performance rating
  - Work life balance

Data Preparation

- Manipulation in In-time and Out-time file:
  - Removing the holidays (dates with NA for all employees) in the In-time and Out-time files
  - Setting the correct date time formats
  - Calculating the daily office hours
  - Calculating the leaves per employee
  - Merging the mean office hours and leave information with the main dataframe
- Creating the dummy variables
- Scaling the numerical variables
- Creating final data frame: Merging the data frames created from the individual files based on employee id to create the final data frame for regression

Building the model

- Final dataset was split as – Train and Test (70-30 percent)
- Logistic regression was applied on the Training data
- StepAIC was used to remove the extremely insignificant variables
- Sequentially insignificant variables that were having high VIF and high p values were eliminated
- Iteration was stopped when all the variables had *** and had VIF between 1 to 2.5

Final Logit Model

The Final Model contains below significant attributes:

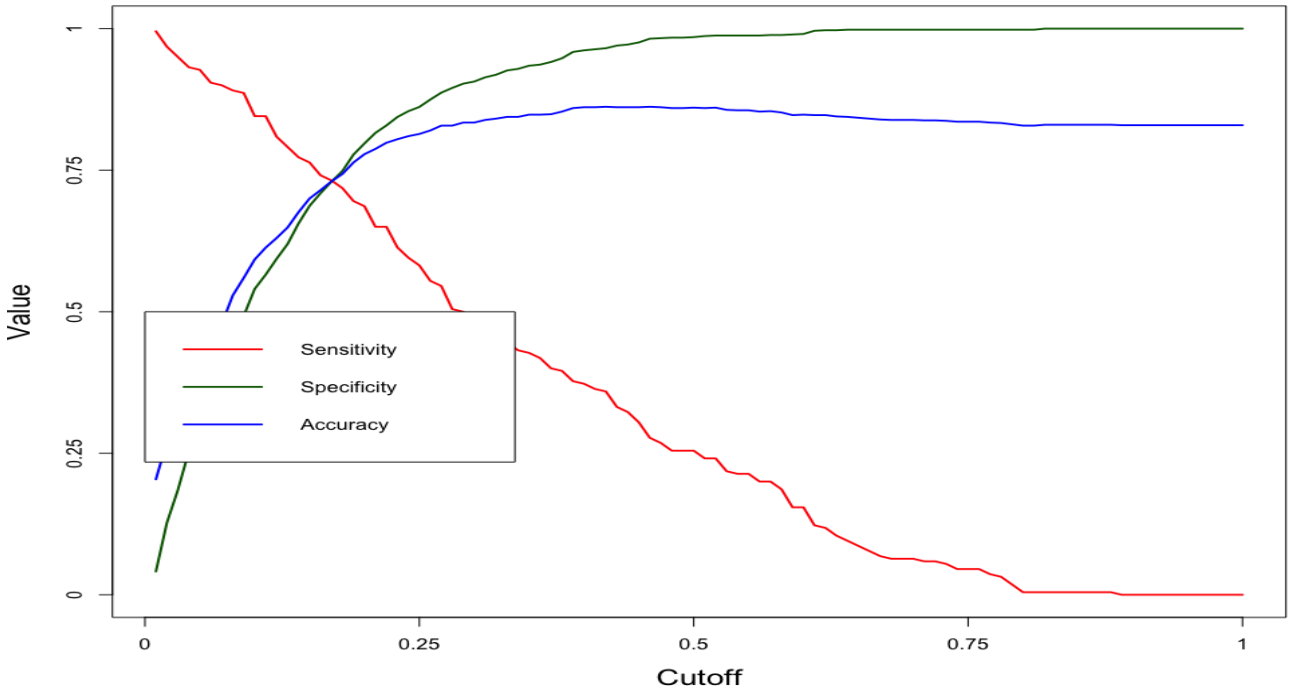
1. Age
2. NumCompaniesWorked
3. TotalWorkingYears
4. YearsSinceLastPromotion
5. YearsWithCurrManager
6. Average_office_hours
7. BusinessTravel.xTravel_Frequently
8. MaritalStatus.xSingle
9. EnvironmentSatisfaction.xMedium
10. EnvironmentSatisfaction.xHigh
11. EnvironmentSatisfaction.xVery.High
12. JobSatisfaction.xVery.High

Evaluating the model

- The final model was used on Test data and calculation of probability of the attrition was performed
- First we chose the probability cutoff as 50% and then we calculated optimal cut off, this classification generated insights for accuracy, specificity and sensitivity
- Cut off turned out to be 0.17 for our model, with accuracy, specificity and sensitivity around 0.73
- The model and its attributes are plotted in subsequent slides for better understanding using GGPLOT.

Evaluating the model – Cut off

- Cut off value=0.17



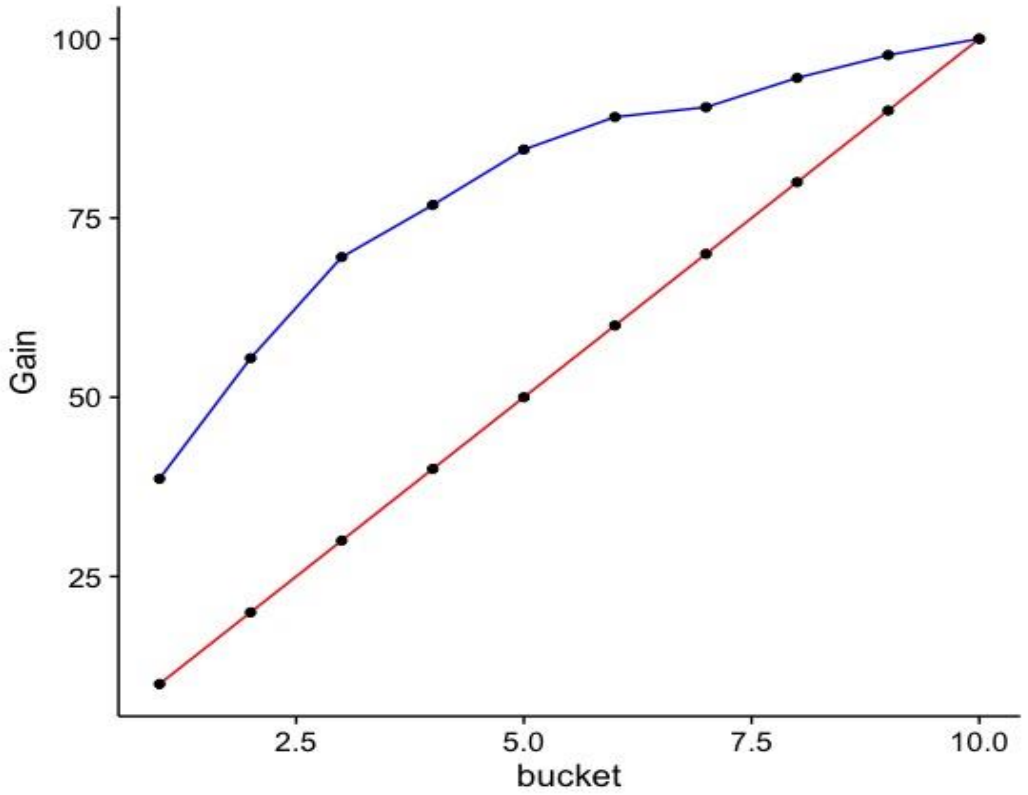
- For this cut off,
  - Sensitivity = 0.73
  - Specificity = 0.73
  - Accuracy = 0.73

Lift Gain Chart- Actual vs Ideal

Legends:

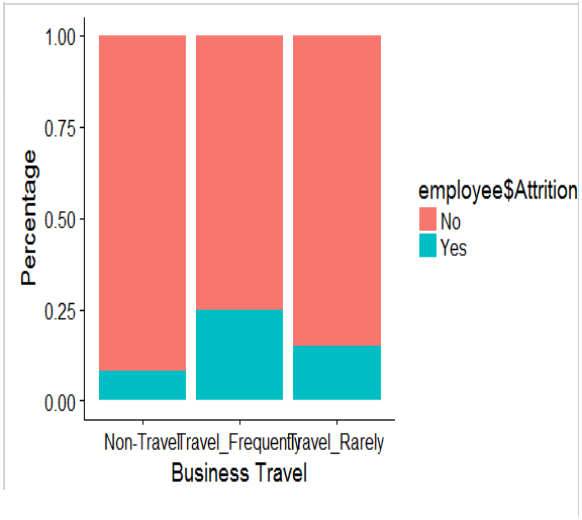
Actual

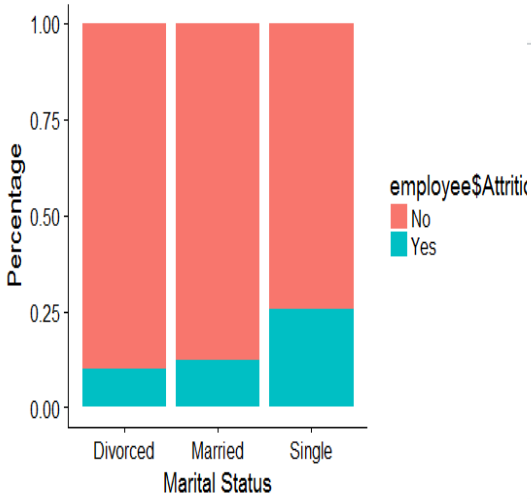
Ideal

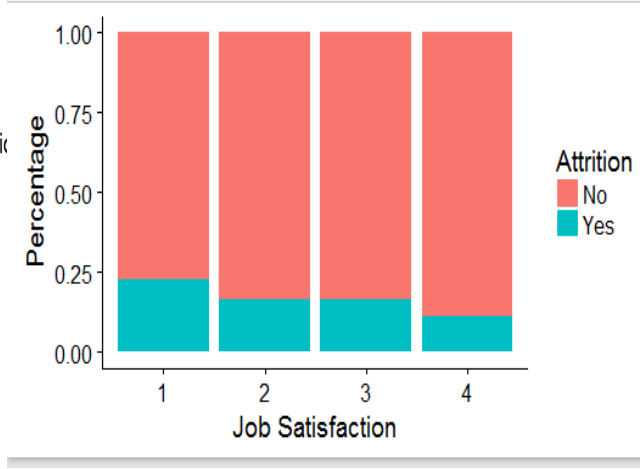


Inference: Model Attributes VS Attrition: Part-1

- Following class of employees are more likely to leave:
  - Employees who travel frequently
  - Employees who are single
  - Employees giving lower rating in Job satisfaction

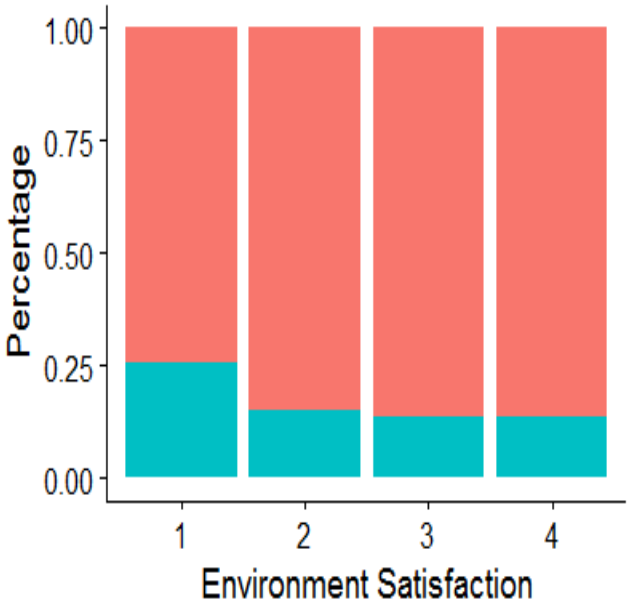


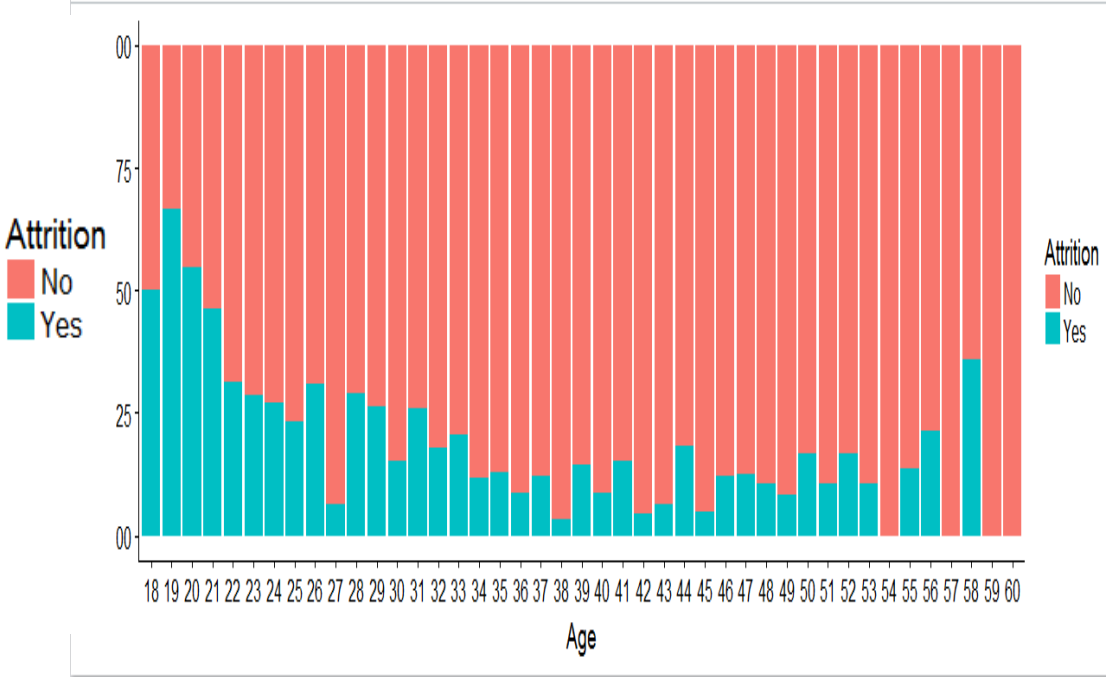




Inference: Model Attributes VS Attrition: Part-2

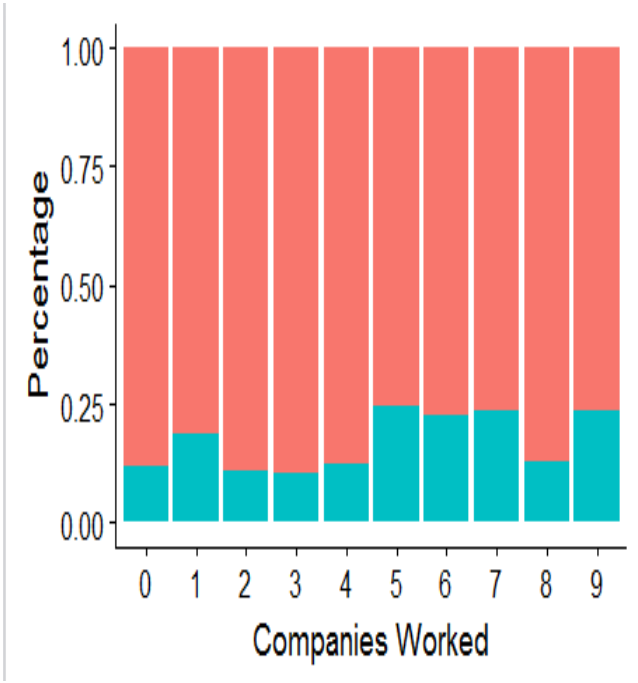
- Following class of employees are more likely to leave:
  - Employees who give lesser rating in environment satisfaction
  - Employees aged between 18 and 22

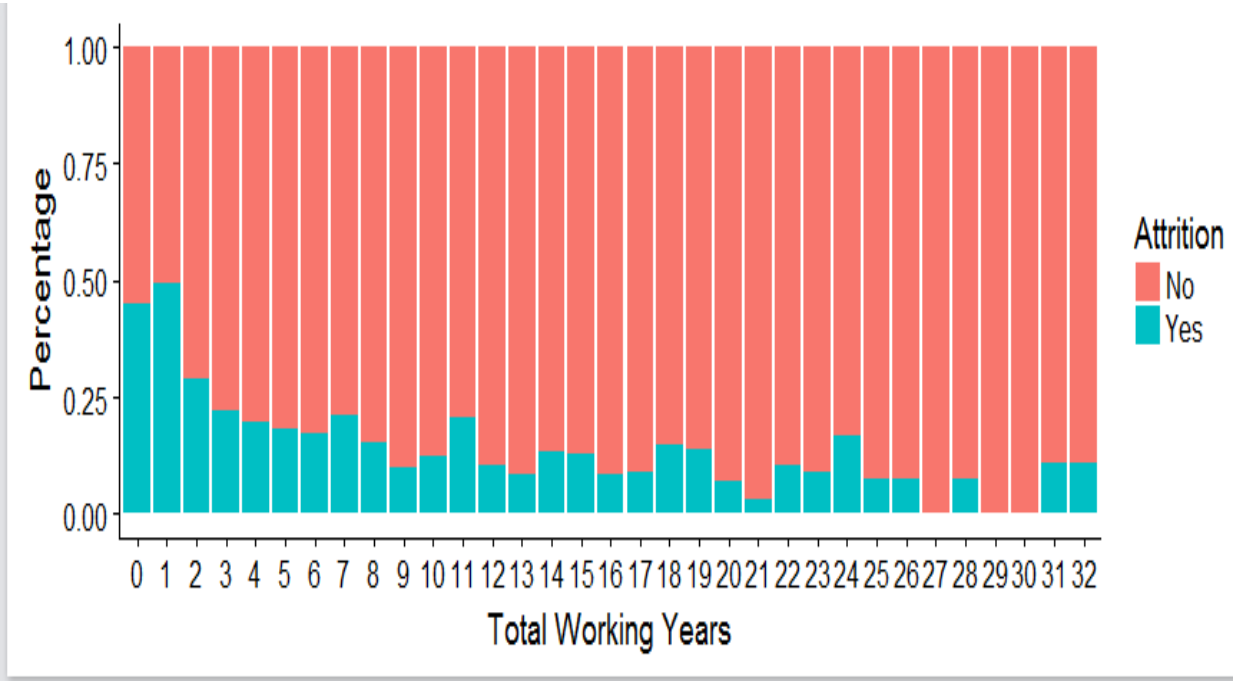




Inference: Model Attributes VS Attrition: Part-3

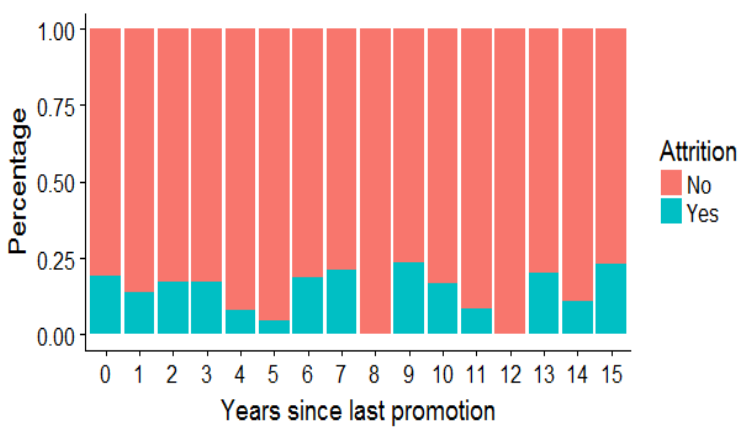
- Following class of employees are more likely to leave:
  - Employees who have worked in 5-7 companies earlier
  - Employees whose total working years is less than 3

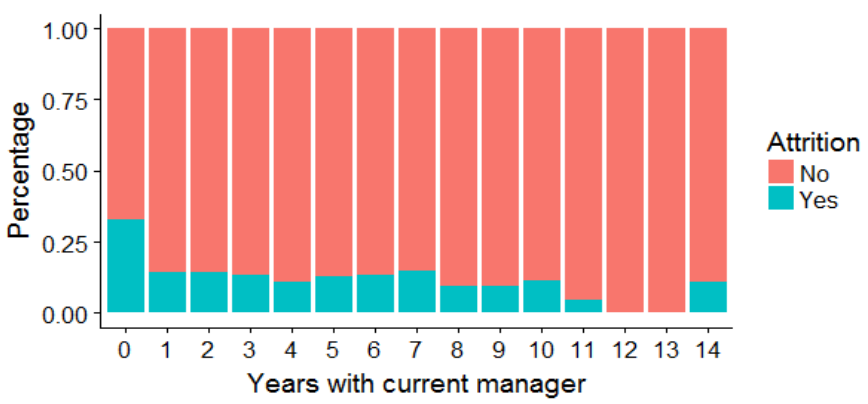




Inference: Model Attributes VS Attrition: Part-4

- Following class of employees are more likely to leave:
  - Employees who have spent 0-3 years since their last promotion
  - Employees who have worked under 3 years with current manager





Recommendations

- Management can take following steps to prevent attrition:
  - Promotion for the employees with experience between 0-3 years
  - Have a Growth Plan for Employees
  - Effectiveness of Reward & Recognition
  - Employee training