HR Case Study

Group Name: Scooby Data Doo

- **Members:**

1. ***Siva Subramaniam Padmanabhan***
2. ***Shruti Saxena Das***
3. ***Shailesh Kumar Jha***
4. ***Chinmaya Sahu***

OBJECTIVE:

- To derive the factors which is the main cause for the increase of Attrition in the organization and to provide solution to curb the same.

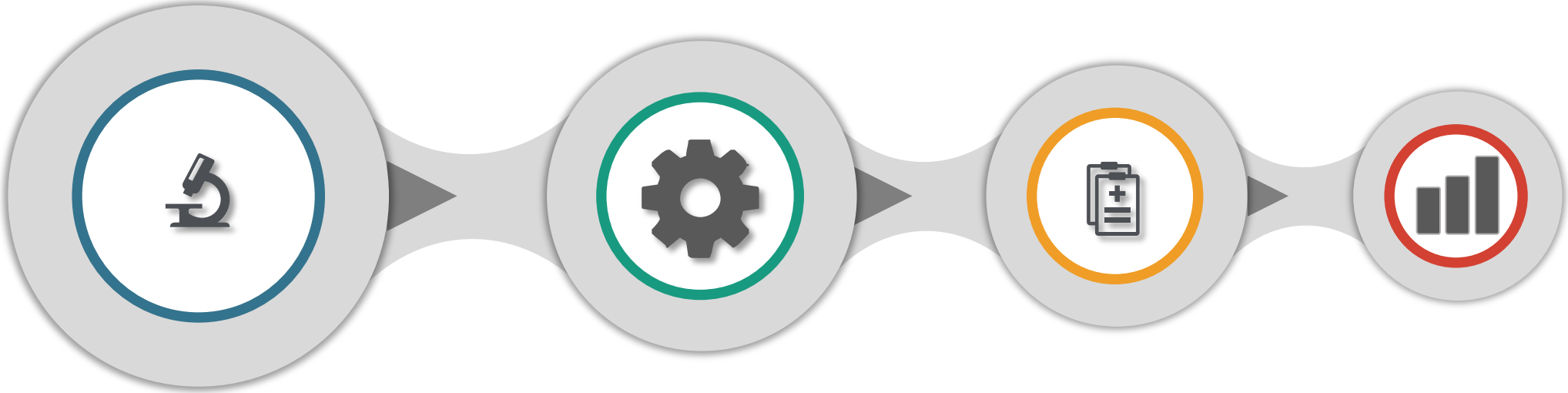
APPROACH STRATEGY:

- To analyze historical data and identify employees who might leave the organization.

•PROBLEM SOLVING METHODOLOGY:

- •Data Cleaning
  - •Eliminating column having NA's, single value & reducing the data file for further analysis.
  - •Standardizing Date format and removing special characters symbols like "%".
  - •Removing Outliers.
  - •Combining the employee data from different data sources
- •Applying logistic regression methodology to gain the insights

DATA FLOW



DATA UNDERSTANDING AND PREPARATION

(Identifying unique keys for given data frames, Collating files and checking for Data Quality issues)

DATA CLEANING

(DataCorrection, Data Conversion and Data Manipulation)

DATA COMPUTATION

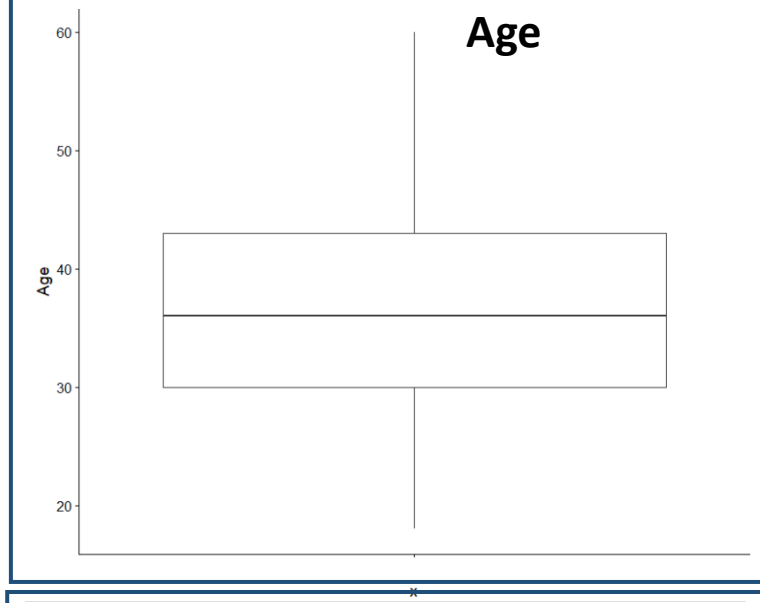
(Computing the attrition ratio)

DATA ANALYSIS

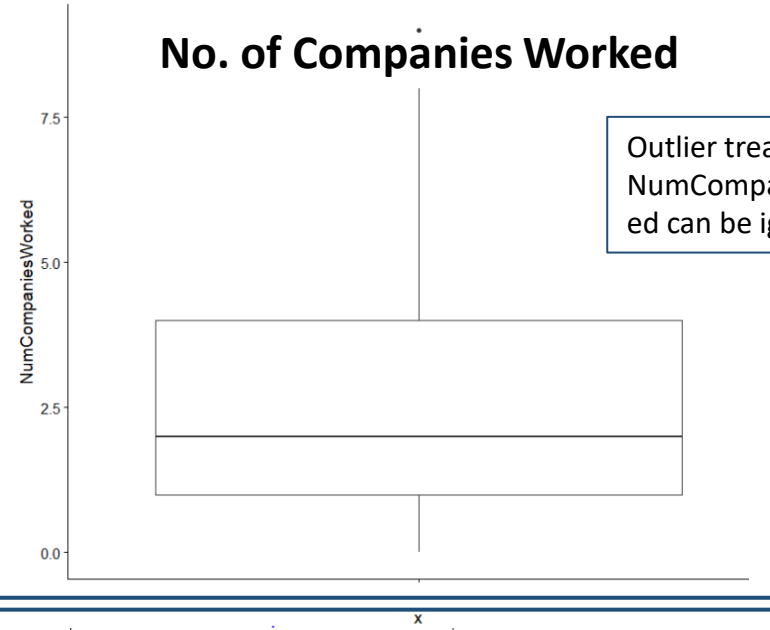
(Analyzing various factors related to employee leading to high attrition)

Outlier Analysis

Age



No. of Companies Worked

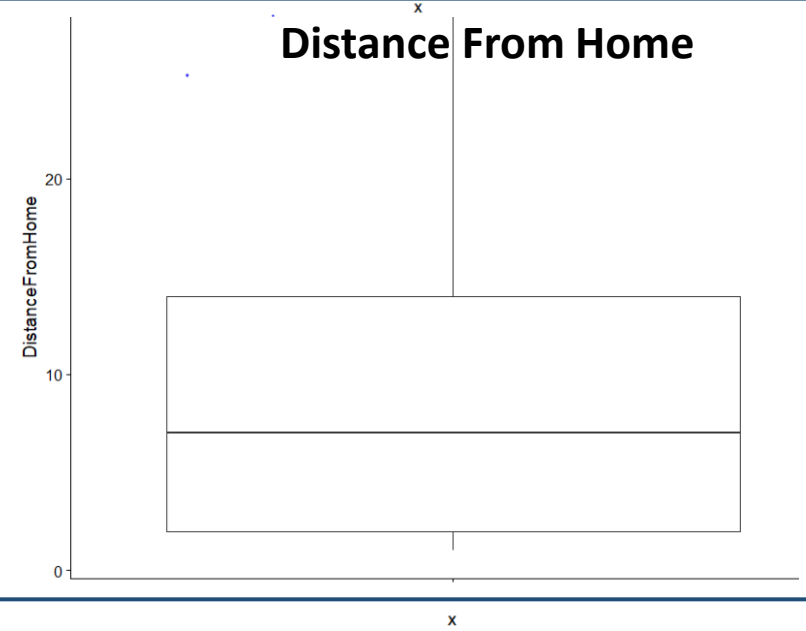


Outlier treatment for NumCompaniesWorked can be ignored

Monthly Income

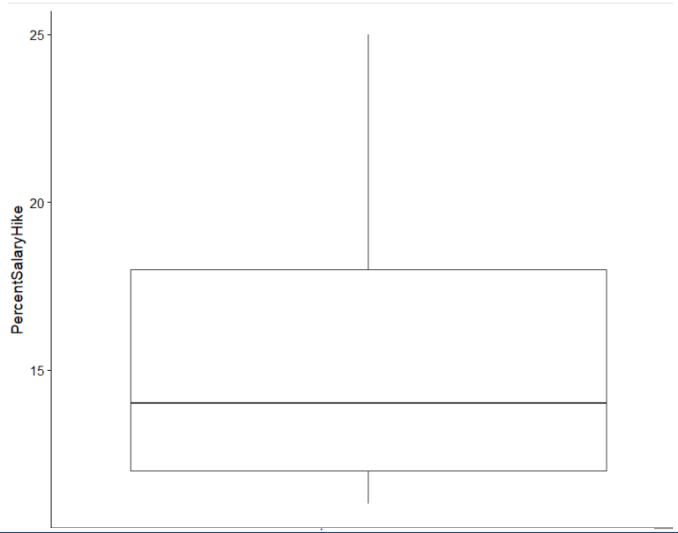
Outlier treatment required for Monthly Income

Distance From Home

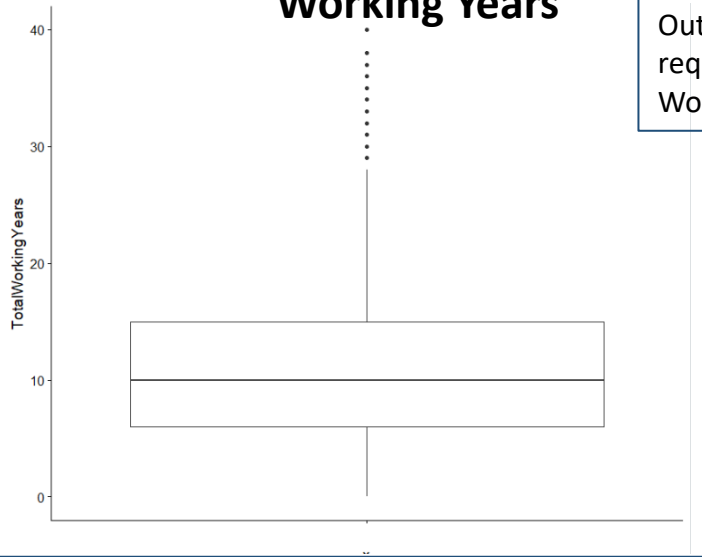


Outlier Analysis

Percentage Salary Hike

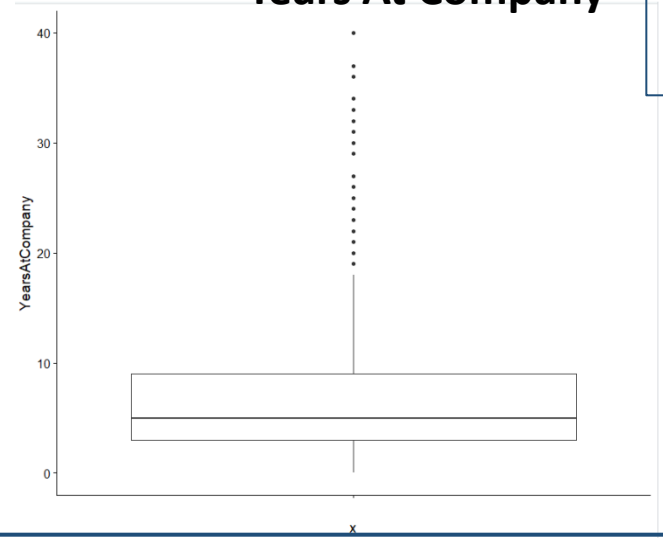


Working Years



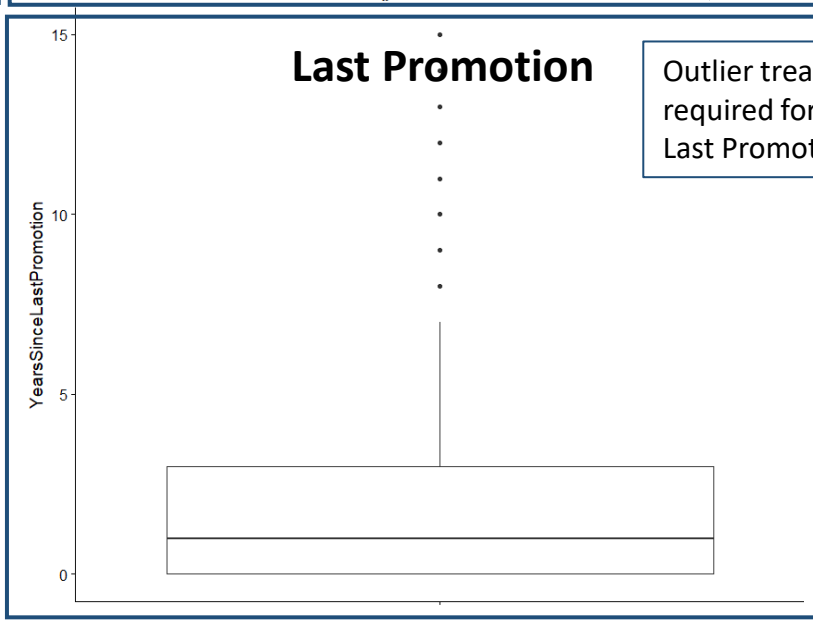
Outlier treatment required for Total Working Years

Years At Company



Outlier treatment required for Years At Company

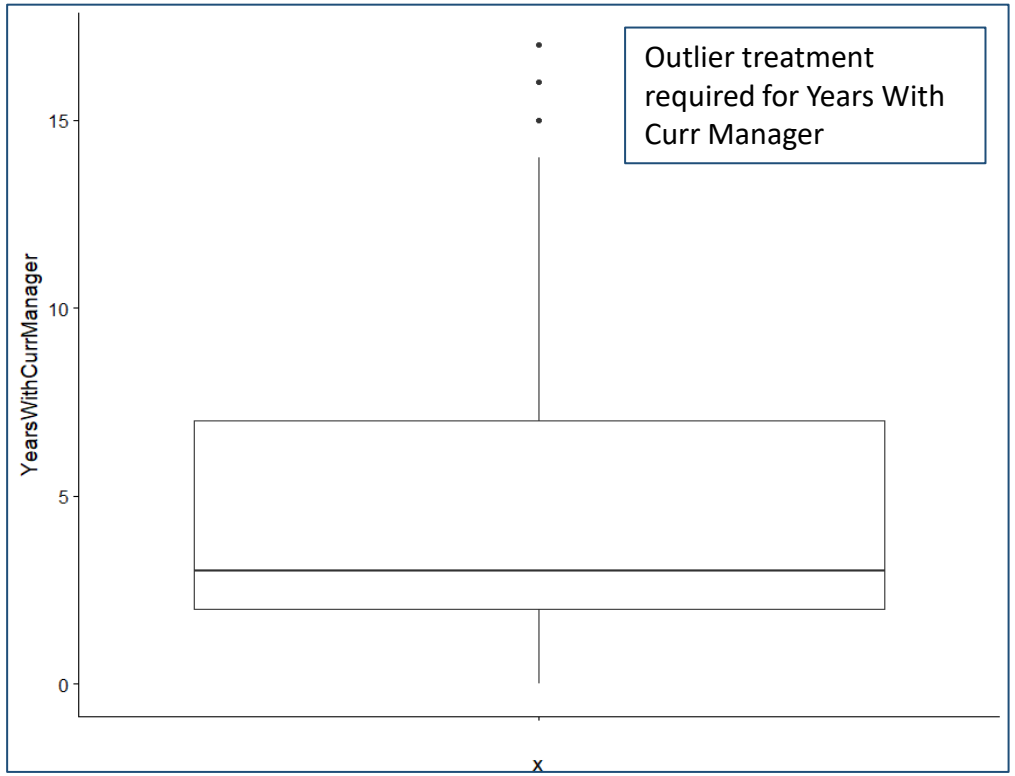
Last Promotion



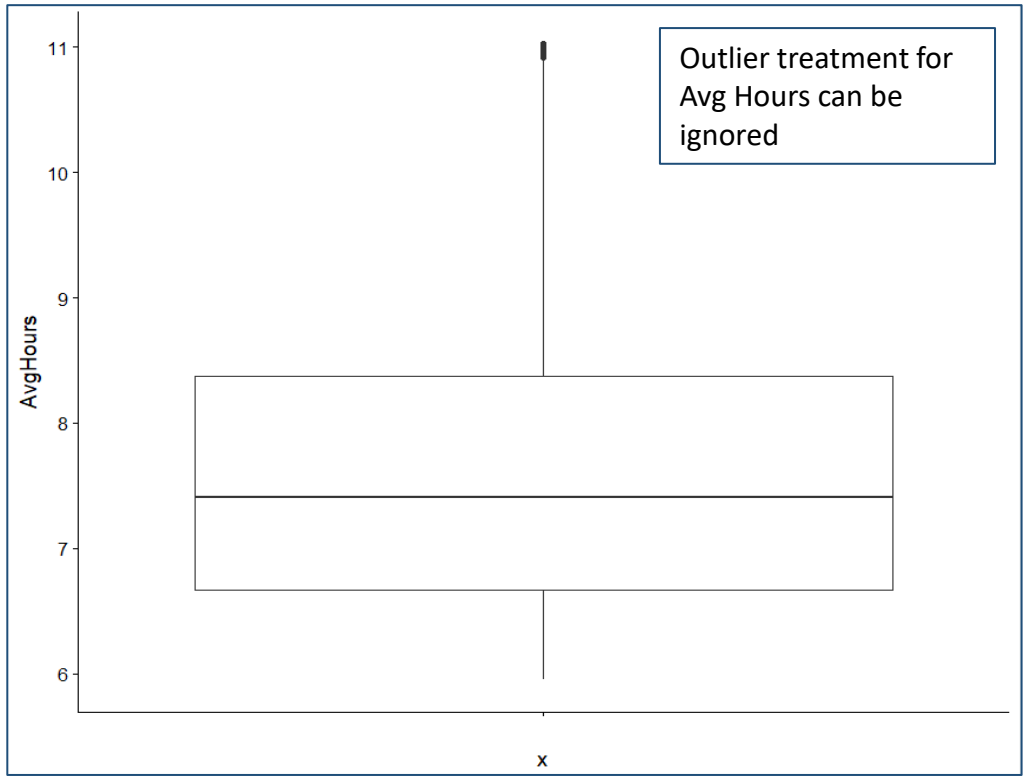
Outlier treatment required for Years Since Last Promotion

Outlier Analysis

Years with Current Manager



Average Hours



After the outlier treatment we have the final set of data on which we start applying our modelling.

Modeling

Building Model:

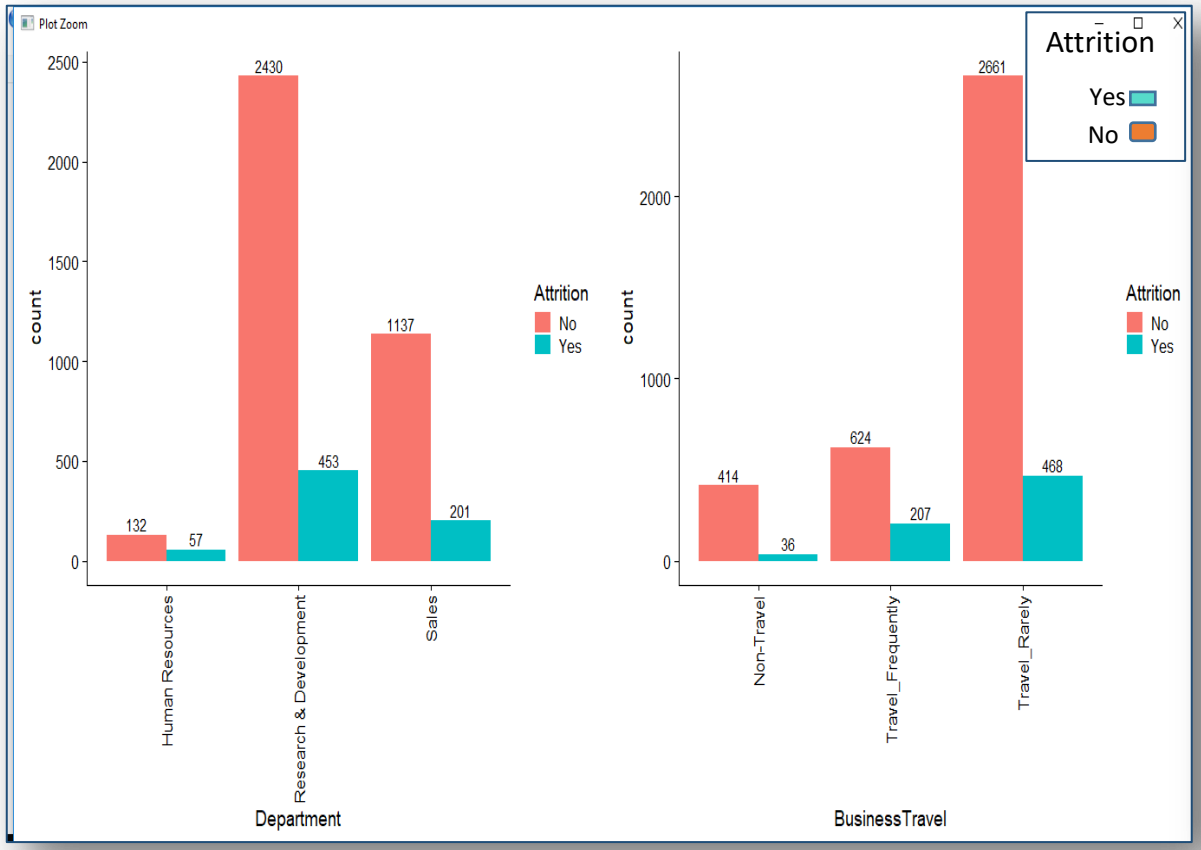
- Continuous variables are scaled for better analysis
- Categorical variables are converted to factors and then dummy levels are created for them
- Using the in_time & out_time data source's we are deriving a new metric AverageWorking hours.
- Data is divided into 70-30 ratio where 70% of the data is the training set and 30% of the data is Test set on which we predict our outcome.
- Applying Logistic Regression techniques for the outcome

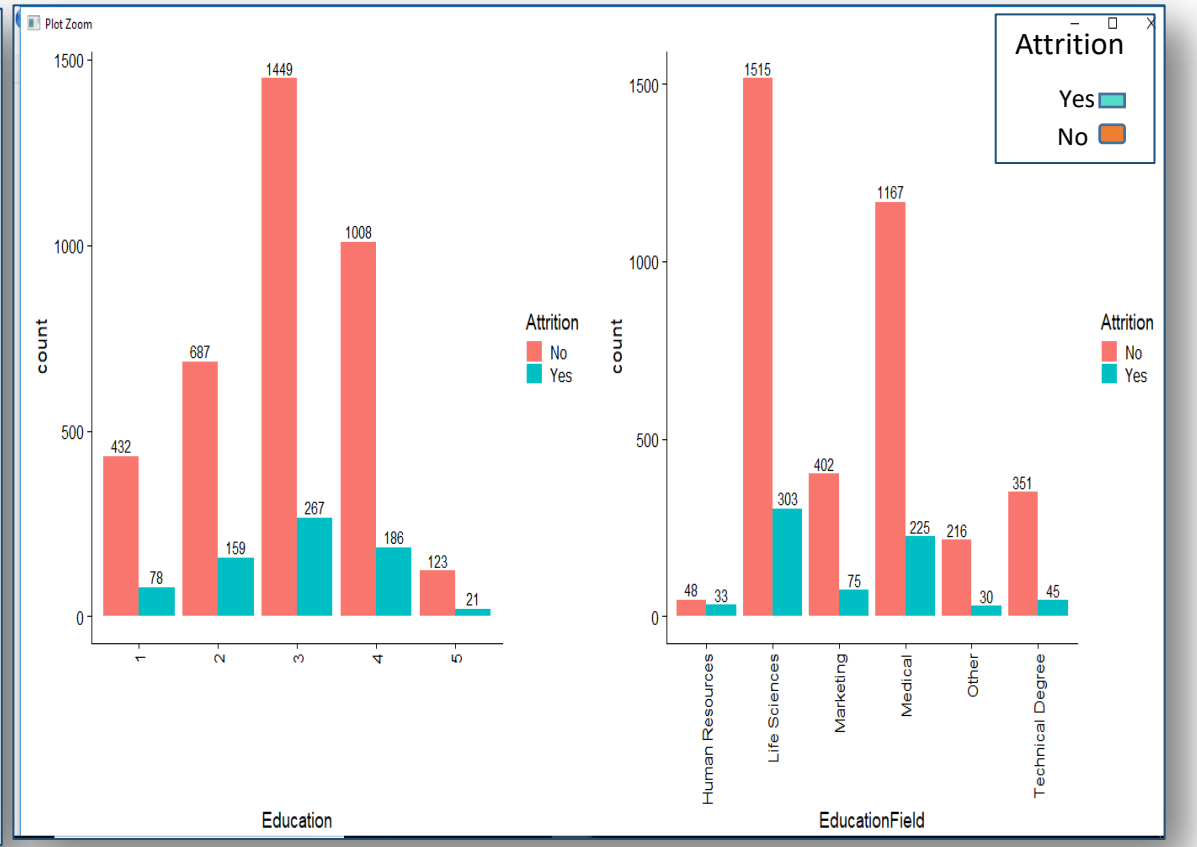
Evaluating Model:

- We fetch Accuracy, Sensitivity, Specificity for various cut-off's and predict the built model

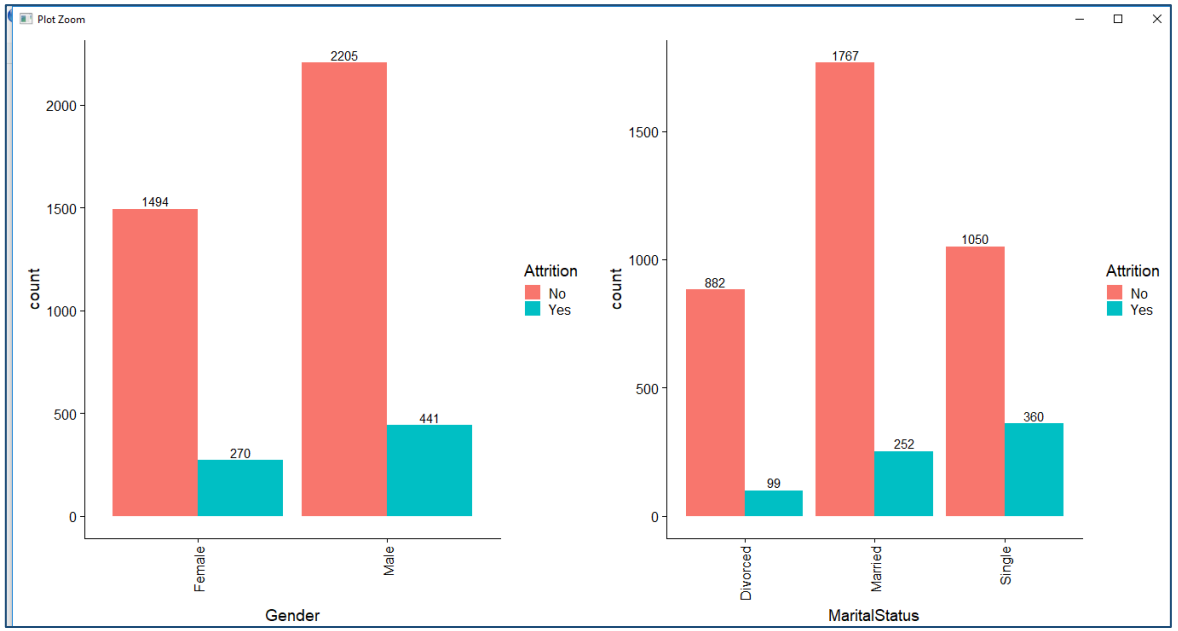
Univariate Analysis

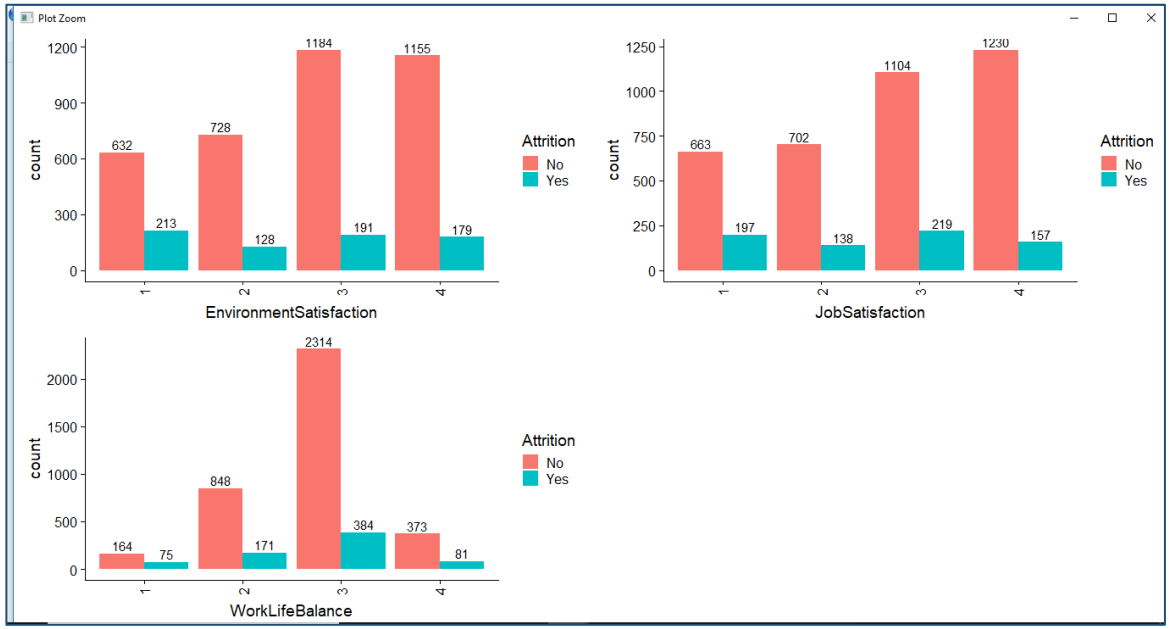
The below plots reveal the attrition rate with respect to various categorical variables

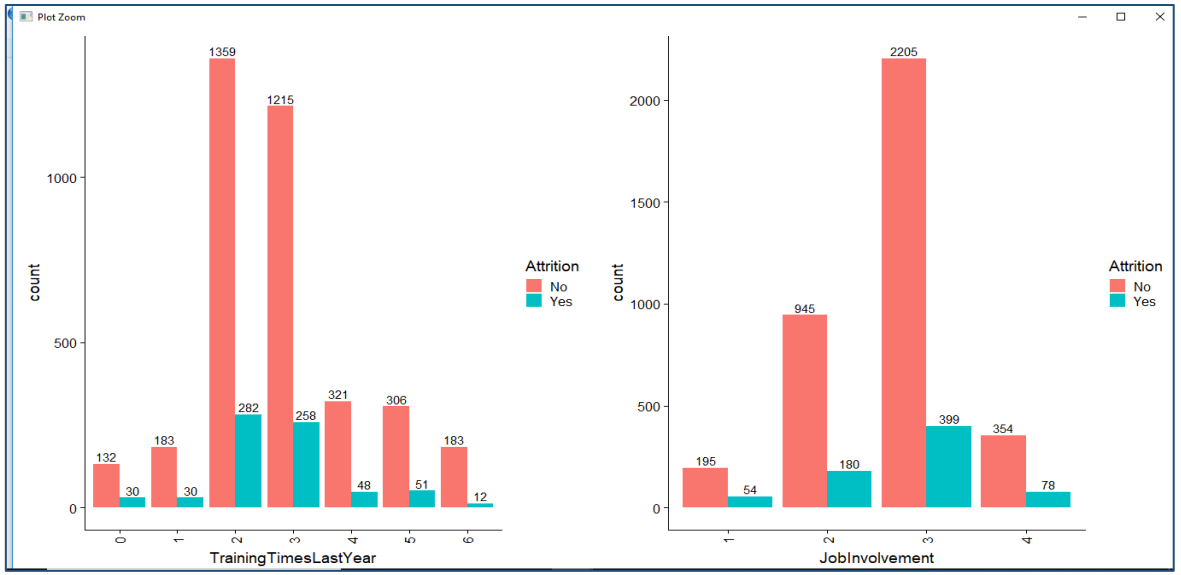


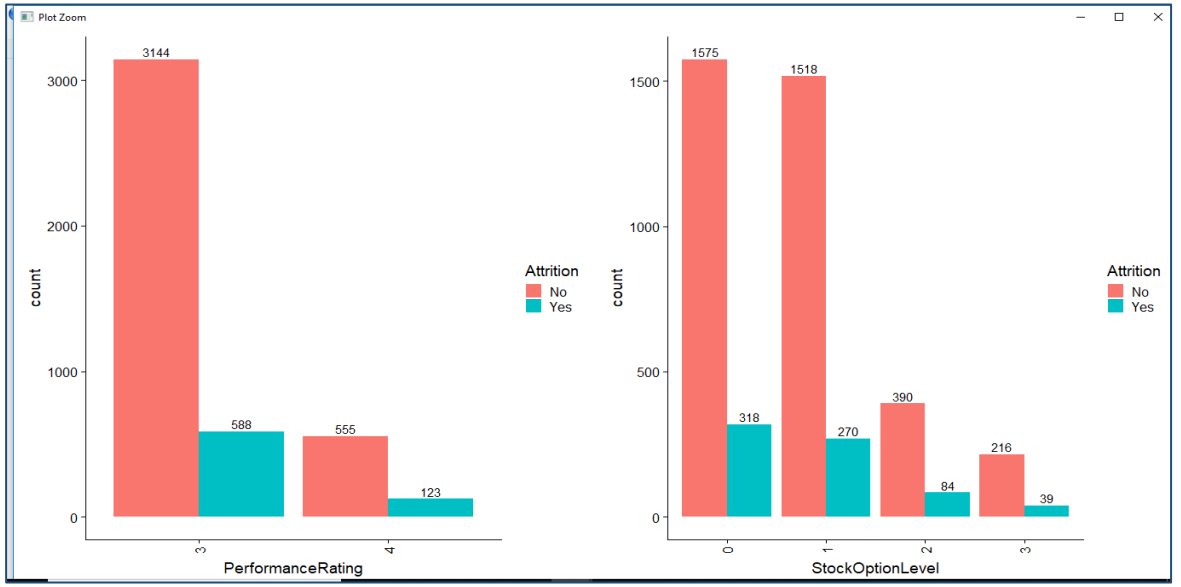


Univariate Analysis



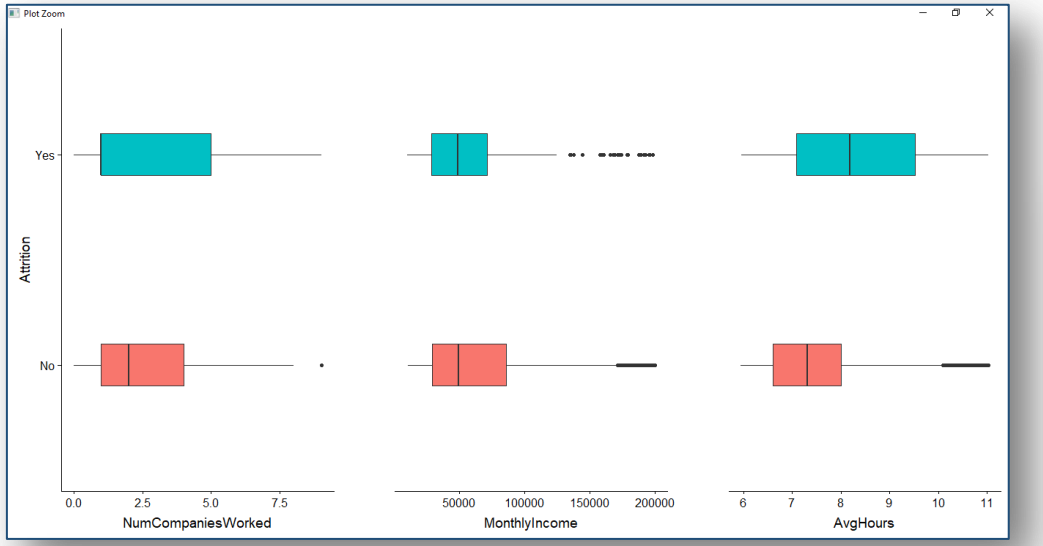


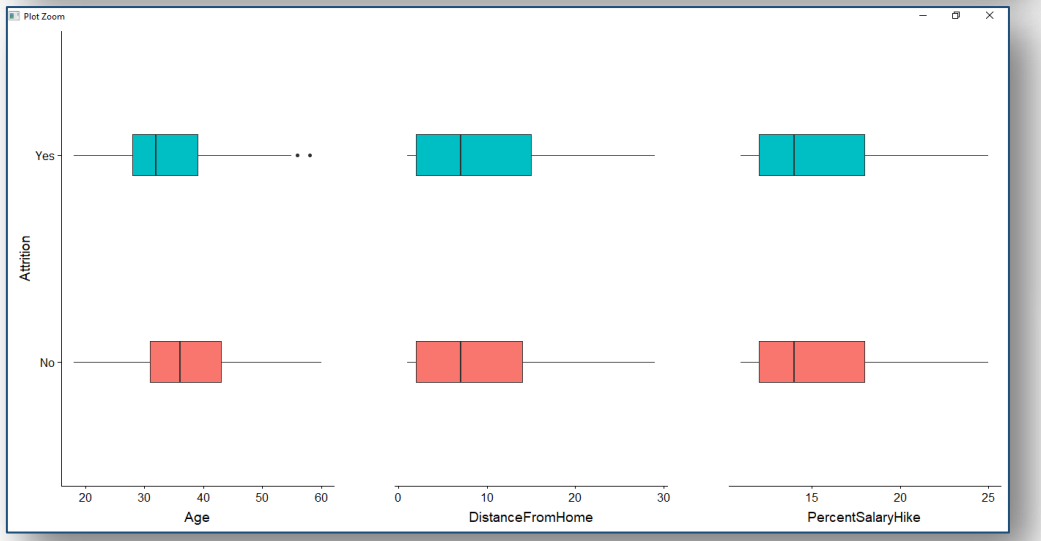


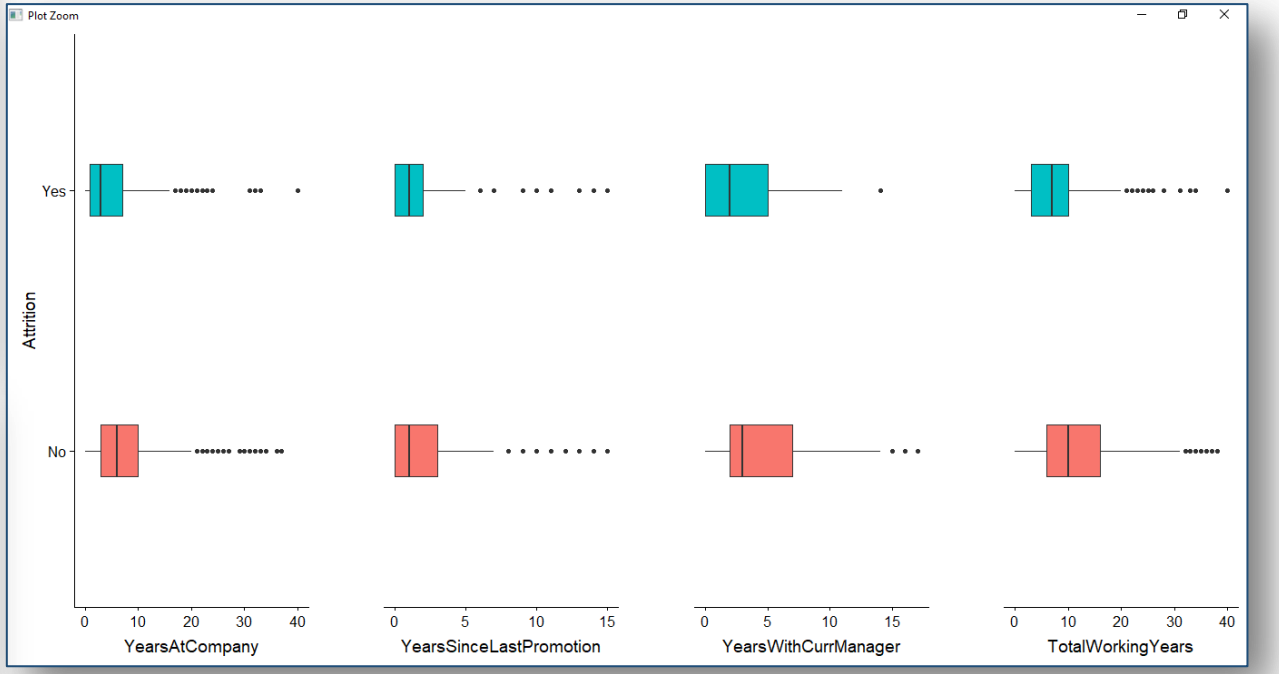


Bivariate Analysis

Attrition rate Vs Numeric variables





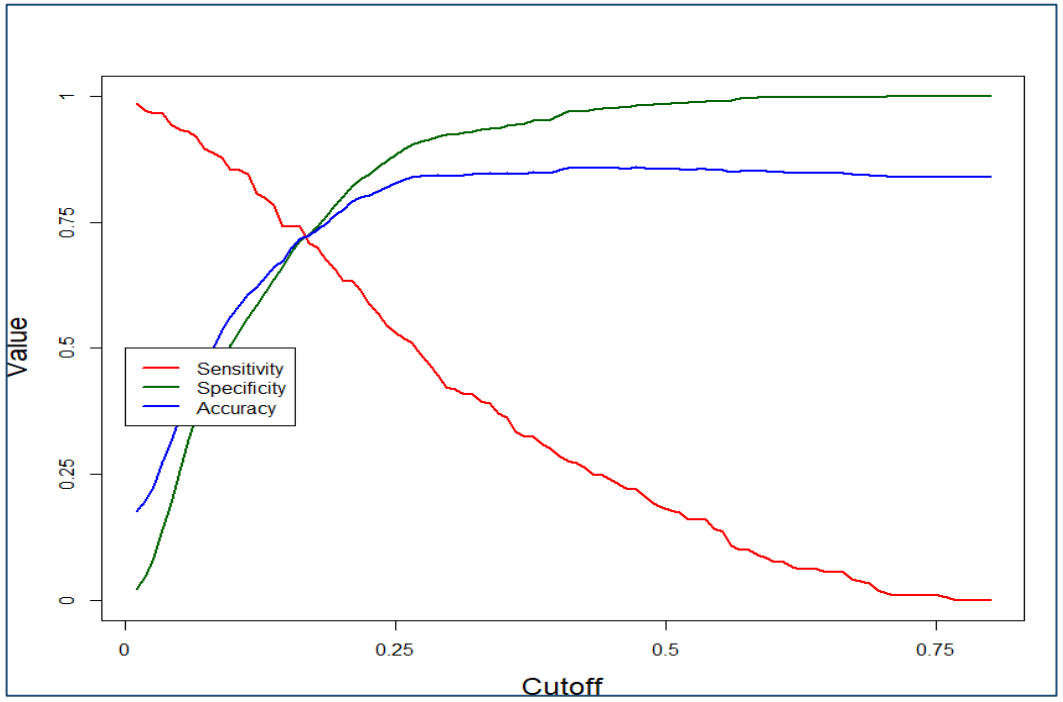


Correlation & Significance

- We apply the glm() function and create accurate models at every step by eliminating variables based on P-value & VIF().
- After 31 models we finally get an accurate model with 9 variables, which are as follows:
  - NumCompaniesWorked
  - TotalWorkingYears
  - AvgHours
  - BusinessTravel.xTravel_Frequently
  - MaritalStatus.xSingle
  - EnvironmentSatisfaction.x2
  - EnvironmentSatisfaction.x3
  - EnvironmentSatisfaction.x4

Accuracy , Specificity & Sensitivity of various cut-off's

Cut-off	Accuracy	Sensitivity	Specificity
0.5	0.8556	0.1784	0.98559
0.4	0.8534	0.28638	0.96216
Calculated by R: 0.1616162	0.7150416	0.7230047	0.7135135



Based on the intersection of the 3 components we R calculates the optimal cut-off which keeps the sensitivity and specificity at a controlled level. i.e. 0.1616162

CONCLUSION

- Company has to take care of the 9 factors mentioned in the previous slide to predict if their employee would stay or leave.
- As per the analysis, Performance rating to the 3rd level is the most contributing factor for employees to leave the company .Therefore the performance to every employee at third level must be scrutinized in future.